2020-9

Stacked Convolutional Recurrent Auto-encoder for Noise Reduction in EEG

Eoghan Keegan
Technological University Dublin

Stacked convolutional recurrent auto-encoder for noise reduction in EEG

Eoghan Keegan

A dissertation submitted in partial fulfilment of the requirements of
Dublin Institute of Technology for the degree of
M.Sc. in Computing (Data Analytics)

June, 2020

Declaration

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Data Analytics), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: 

Date: September 30, 2020

Abstract

Electroencephalogram (EEG) can be used to record electrical potentials in the brain by attaching electrodes to the scalp. However, these low amplitude recordings are susceptible to noise which originates from several sources including ocular, pulse and muscle artefacts. Their presence has a severe impact on analysis and diagnoses of brain abnormalities. This research assessed the effectiveness of a stacked convolutional-recurrent auto-encoder (CR-AE) for noise reduction of EEG signal. Performance was evaluated using the signal-to-noise ratio (SNR) and peak signal-to-noise ratio (PSNR) in comparison to principal component analysis (PCA), independent component analysis (ICA) and a simple auto-encoder (AE). The Harrell-Davis quantile estimator was used to compare SNR and PSNR distributions of reconstructed and raw signals. It was found that the proposed CR-AE achieved a mean SNR of 5.53 db and significantly increased the SNR across all quantiles for each channel compared to the state-of-the-art methods. However, though SNR increased PSNR did not and the proposed CR-AE was outperformed by each baseline across the majority of quantiles for all channels. In addition, though reconstruction error was very low none of the proposed CR-AE architectures could generalize to the second dataset.

Keywords: *electroencephalography, event related potentials, noise reduction, convolutional-recurrent auto-encoder, signal-to-noise ratio*

Acknowledgements

I would like to express my deepest appreciation to Dr. Luca Longo for his invaluable insight, support and guidance throughout the dissertation process.

I would also like to acknowledge and convey my thanks to the Technological University Dublin staff that have supported me over the past two years. In particular, Dr. Sarah Jane Delaney, Andrea Curley and Deirdre Lawless, who gave me this opportunity.

I would also like to thank Conor Hanrahan and Alexander Suvorov, who kindly provided their dissertations and related work which I was able to leverage to deepen my understanding of the topic.

I would like to express my sincerest gratitude to Brendan O'Dowd and Kevin McTiernan, without whom I would not be where I am today. They have encouraged and supported me in my work and personal development, and are the reason I have completed this MSc.

Lastly, my family, who have an unwavering belief in my ability and who have never stopped supporting me no matter what. To you I am eternally grateful. This is for you.

Contents

Declaration **I**

Abstract **II**

Acknowledgements **III**

Contents **IV**

List of Figures **VII**

List of Tables **XI**

1 Introduction **1**

1.1 Background 1

1.2 Research problem 3

1.3 Research objectives 3

1.4 Research methodologies 4

1.5 Scope and limitations 4

1.6 Document outline 6

2 Related work **7**

2.1 Electroencephalogram 7

2.2 EEG artefacts 8

2.2.1 Ocular artefact 8

2.2.2 Muscle artefact 9

2.2.3 Pulse artefact 9

2.3 Artefact identification and reduction 10

2.3.1 Principal component analysis . . . 11
2.3.2 Independent component analysis . . . 13
2.4 Artificial neural networks . . . 15
2.5 Auto-encoders . . . 16
2.5.1 Classification . . . 17
2.5.2 Feature learning . . . 18
2.5.3 Noise reduction . . . 20
2.6 Summary . . . 22
2.6.1 Overview . . . 22
2.6.2 Gaps in the literature . . . 23
2.6.3 Research question . . . 25

3 Design and methodology **26**
3.1 Hypothesis . . . 26
3.2 Data collection . . . 27
3.3 Data preparation . . . 30
3.4 Auto-encoder design . . . 34
3.5 Evaluation of design . . . 39
3.5.1 Signal-to-noise ratio . . . 39
3.5.2 Hypothesis testing . . . 39
3.6 Summary . . . 41
3.6.1 Strengths . . . 42
3.6.2 Limitations . . . 43

4 Results, evaluation and discussion **45**
4.1 Results . . . 45
4.1.1 Signal reconstruction . . . 46
4.1.2 Signal-to-noise ratio . . . 54
4.2 Evaluation . . . 56
4.2.1 Proposed CR-AE architecture . . . 57
4.2.2 Principal components analysis . . . 58

4.2.3 Independent components analysis 62
4.2.4 Basic auto-encoder 66
4.3 Summary of Findings 69
4.4 Discussion 71
4.4.1 Strengths 72
4.4.2 Limitations 73

5 Conclusion **75**
5.1 Research Overview 75
5.2 Problem Definition 75
5.3 Design, Evaluation & Results 77
5.4 Contributions and impact 80
5.5 Future work & recommendations 81

Bibliography **83**

A Additional content **96**

List of Figures

2.1 Example of 2D data cloud where PCs are principal components (PC1 explains the maximum amount of variance, PC1 and PC2 are orthogonal). Reprinted from Zinovyev et al., 2013 11

2.2 Example of 2D data cloud where ICs are independent components (give maximally non-gaussian distribution of the projections). Reprinted from Zinovyev et al., 2013 13

2.3 Single hidden layer ANN network with each node representing an artificial neuron and arrows the connections from the output of one node to the input of another 16

3.1 High level overview of how each of the design components combines to create the experiment 27

3.2 10-10 electrode placement system with black circles indicating positions of the original 10-20 system. Reprinted from Oostenveld and Praamstra, 2001. 28

3.3 Left: Stimulus presentation sequence for each trial. Right: Finger-numeral configurations (right-hand only). Reprinted from Soylu et al., 2019 29

3.4 Single channel preprocessed EEG signal taken from primary dataset . 31

3.5 Single channel preprocessed EEG signal taken from secondary dataset 32

3.6 Top: Single trial split into windowed segments. Bottom: Two windowed segments recombined 33

3.7 A simplified representation of an auto-encoder architecture 34

3.8 Top: Structure of the encoder for the single layer architecture. Bottom: Structure of the decoder for same 35

4.1 Original signal overlaid with the corresponding reconstructed signal for architecture one (primary dataset) 47
4.2 Original signal overlaid with the corresponding reconstructed signal for architecture two (primary dataset) 47
4.3 Original signal overlaid with the corresponding reconstructed signal for architecture three (primary dataset) 48
4.4 Original signal used as input for each of the reconstructions below . . 49
4.5 Original signal overlaid with the corresponding reconstructed signal for architecture one (secondary dataset) 49
4.6 Original signal overlaid with the corresponding reconstructed signal for architecture two (secondary dataset) 50
4.7 Original signal overlaid with the corresponding reconstructed signal for architecture three (secondary dataset) 50
4.8 Single 300ms window of a signal used for the reconstruction below . . 52
4.9 Output from architecture one for the 300ms input window above . . . 53
4.10 Output from architecture one for the next 300ms window of the same signal 53
4.11 Combined outputs of the first and second windows with overlaps averaged 54
4.12 Original signal overlaid with the corresponding reconstructed signal for PCA 59
4.13 Original signal overlaid with the corresponding reconstructed signal for the CR-AE 59
4.14 Original signal overlaid with the corresponding reconstructed signal for ICA — electrode T7 63
4.15 Original signal overlaid with the corresponding reconstructed signal for ICA — electrode Fp1 64

4.16 Original signal overlaid with the corresponding reconstructed signal for basic AE 67
4.17 Original signal overlaid with the corresponding reconstructed signal for CR-AE 68

A.1 Technical model of proposed CR-AE 97
A.2 Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 1 107
A.3 Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 2 107
A.4 Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 3 108
A.5 Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 1 108
A.6 Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 2 109
A.7 Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 3 109
A.8 Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 1 110
A.9 Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 2 110
A.10 Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 3 111

A.11 Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 1111
A.12 Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 2112
A.13 Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 3112

List of Tables

3.1 Summary of datasets used in experiment 30
3.2 Summary of available hyper-parameters for each layer 38
3.3 Summary of training parameters used to fit each model 39

4.1 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 1 trial sampling 61
4.2 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 1 subject sampling 61
4.3 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 1 trial sampling 65
4.4 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 1 subject sampling 65
4.5 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 1 trial sampling 68
4.6 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 1 subject sampling 69

A.1 Common electrode placement sites for primary and secondary datasets 96
A.2 Chosen hyper-parameters for architecture 1 98
A.3 Chosen hyper-parameters for architecture 2 99
A.4 Chosen hyper-parameters for architecture 3 100
A.5 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 2 trial sampling 101
A.6 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 2 subject sampling 101

A.7 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 3 trial sampling 102
A.8 SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 3 subject sampling 102
A.9 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 2 trial sampling 103
A.10 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 2 subject sampling 103
A.11 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 3 trial sampling 104
A.12 SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 3 subject sampling 104
A.13 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 2 trial sampling 105
A.14 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 2 subject sampling 105
A.15 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 3 trial sampling 106
A.16 SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 3 subject sampling 106

Chapter 1

Introduction

Electroencephalogram is a method of recording electrical potentials produced by the brain (Binnie & Prior, 1994), commonly achieved by attaching non-invasive electrodes to the scalp which measure voltage fluctuations. In some cases, electrodes are placed directly against the brain. This method of invasive electroencephalogram (iEEG) is used when scalp EEG is not sufficient and high spatial and temporal accuracy is required (Ball et al., 2009). EEG is utilized in many fields including the study of event related potentials (ERP). ERP are small voltage changes resulting from the onset of a stimulus used to assess motor or sensory function. They are commonly analysed to detect the presence of abnormalities, such as for patients suffering from Alzheimers, Schizophrenia and Dementia (Himani et al., 1999).

1.1 Background

In the medical domain accuracy is paramount as patients' well-being is dependent on accurate diagnosis. For EEG readings this is no different. It is important to accurately record voltage fluctuations in the human brain to identify or detect the presence of cognitive abnormalities. As mentioned, in some cases EEG is not sufficient and iEEG has to be used. This is primarily in cases of epilepsy, where patients are resistant to pharmacological treatment (Engel et al., 2005) but is also due to limitations associated with EEG recording. As noted by Nair et al., 2008, given the varying shape and depth

of the skull at certain locations, scalp electrodes can lie at different distances from the brain. This can cause discharge to not be detected at the scalp or be obscured by background activity. iEEG is used because the electrodes are closer to the brain than scalp electrodes and produce higher amplitude readings with high spatial resolution (Ball et al., 2009), however, being an invasive procedure it is not utilized in the majority of cases.

Another reason for the use of iEEG is related to EEG artefacts. These are contaminating signals such as eye movement, blinking (Vigário, 1997) and muscle contraction (Crespo-Garcia et al., 2008) which interfere with the capture of other signals (Vaseghi & V., 2001). Since the implanted electrodes produce higher amplitude readings, they are generally less susceptible to artefact contamination (Marinković, 2004). Both artefacts and background activity are generally referred to as noise because they are unwanted signals present in the EEG recordings. Given that iEEG is only used in exceptional circumstances, an approach to removing these artefacts or distinguishing meaningful signal from background activity is required.

Several approaches have been applied to noise removal, including the use of filters (Roy & Shukla, 2015), mathematical transformations like the wavelet transform (Heydari & Shahbakhti, 2015) and blind source separation algorithms like ICA (Hyvarinen, 1999). In some cases, additional electrodes specifically placed to capture artefacts have been used, allowing for their subsequent identification and removal, however this can be uncomfortable for the patient. In other cases, machine learning techniques such as auto-encoders (B. Yang et al., 2016) have been utilized for their ability to reconstruct inputs from latent space representations. The aim of this paper is to propose a method of noise reduction based on convolutional and recurrent neural network layers. A stacked auto-encoder incorporating the spatial feature representations of convolutional neural networks and the temporal patterns extracted by long short-term memory recurrent neural networks, combined with the deep learning capabilities of stacked auto-encoders is explored as a potential solution.

1.2 Research problem

Two metrics are generally used to measure the level of noise present in a signal. Those being the signal-to-noise ratio which measures the power of a signal relative to the power of noise and the peak signal-to-noise ratio which measures the maximum amplitude of a signal relative to the power of noise corrupting it. Both are measured on the decibel scale and usually calculated using a clean signal and its noisy counterpart. Positive SNR indicates the presence of more signal than noise.

In order to maximize the SNR and PSNR, the level of noise present in a signal needs to be reduced. However, it is important to ensure that critical signal information is not lost during the process. This project investigates the use of a stacked convolutional-recurrent auto-encoder to increase the SNR and PSNR by reducing the level of noise present in a signal while retaining as much meaningful information as possible.

If this can be achieved, what is the magnitude of the increase, how effective is the solution compared to previously used techniques such as PCA and ICA and can it generalize to other research?

1.3 Research objectives

There are several objectives of this research. The first is to perform a literature review of noise reduction techniques and their application to EEG signals. In addition, the use of auto-encoders for both EEG noise reduction and other problems, will also be investigated. The focus of this review will be to discover any gaps or limitations that exist and to examine whether unsupervised machine learning algorithms can be used to address these limitations. Secondly, an empirical experiment will be designed to enable the hypothesis to be tested. This should allow the research to be replicated and validated by other researchers while also ensuring the rationale behind the chosen solution is clearly stated. Furthermore, a description of the data used and all pre-processing steps will be given. The third objective is to evaluate the experiment using

the chosen metrics and evaluation criteria. For this the proposed method will be compared to each baseline to determine whether an improvement has been observed. In addition, observations and key findings of the results will be highlighted. Finally, the overall purpose of this research is to enhance understanding of convolutional-recurrent auto-encoders and their effectiveness, as it pertains to noise reduction of EEG signals.

1.4 Research methodologies

The type of research being carried out is secondary, quantitative, empirical research using deductive reasoning. It involves a systematic review, summary, and extension of previous research. Two EEG datasets are used; One generated by Ford et al., 2014, the other by Soylu et al., 2019 both containing numeric voltage amplitudes recorded in microvolts (μV). These datasets were sourced from Kaggle and the Harvard dataverse respectively and each involved a set of trials in which subjects were presented with a stimulus. The original purpose of their collection was to measure the subjects' response by conducting an ERP analysis. In this case, mathematical models are applied to reconstruct these EEG signals with the aim of reducing the level of noise in the output compared to the input. This relationship is measured using the SNR and PSNR and compared using the Harrell-Davis quantile estimator. To determine whether the proposed method can reduce the level of noise, a hypothesis has been defined with a suitable experiment designed to test it. For this, empirical evidence is gathered from the data which tests the feasibility of the solution. This hypothesis has been derived from a theory, is tested through experiment and concluded upon to determine its validity.

1.5 Scope and limitations

The scope of this research is the use of convolutional and recurrent neural network layers in a stacked AE for noise reduction in EEG signals. Two datasets are used in this experiment. The first relates to a study on efference copy and corollary discharge

of schizophrenia patients in response to a stimulus. Data for that study was collected from 81 subjects (49 with diagnosed schizophrenia, 32 control) as they underwent 100 trials in each of three conditions. The second relates to a study on early perceptual and later semantic processing of canonical and non-canonical finger-numeral configuration in adults. Data for that study was collected for 46 right-handed undergraduate students as they underwent 960 trials comprised of randomized sets of finger-numeral counting configurations. In each case, data is epoched around the presentation of the stimulus and subset between 100 ms pre-stimulus and 500 ms post-stimulus. Pre-stimulus recordings are used to represent baseline noise, while the remainder represents meaningful signal.

This research is limited to ERP analysis, where a stimulus is used to illicit a response in the brain. It is assumed, that when a stimulus is used, the response contains both meaningful signal and noise, while pre-stimulus activity is only noise. This allows the SNR to be calculated using pre and post-stimulus activity as opposed to using a noise free reference signal which is not available. This assumption is required due to an inability to record perfectly clean EEG, and while the experiments are set up in a way to limit intra-trial neural activity, it is not guaranteed. It is also assumed that the output from the proposed and baseline methods are noise free signals and that no neural activity has been removed in the process. This limitation is important, because in the absence of clean reference EEG the actual information loss cannot be quantified.

Though artefact reference electrodes are captured in the datasets, no comparative regression-based methods are used to determine whether the neural activity from EEG artefacts present in other electrodes are reduced by the proposed method. This limitation means that evaluation on a specific type of EEG noise is not conducted and therefore it cannot be concluded which type of noise the proposed method is removing from the signals. Finally, due to differences in the electrodes present in both datasets, the research is limited to the 28 described in table A.1. Additionally, due to sample rate differences it is also limited to EEG recordings at 500 samples per second.

1.6 Document outline

The remaining chapters of this research structured as follows:

Chapter 2 - Related work

In this chapter, existing literature in the field of EEG noise reduction is reviewed and discussed with respect to the proposed solution. It is aimed at comparing and contrasting previous noise reduction approaches to convey the gaps identified that led to this research.

Chapter 3 - Design and methodology

This chapter focusses on experiment design; covering data collection, preparation and the proposed solution. Evaluation is discussed in detail to describe how the experiment will be conducted and the methods employed to test the hypothesis.

Chapter 4 - Results, evaluation and discussion

This chapter focusses on summarizing the results of the experiment in a clear and concise manner to evaluate the proposed method with respect to each of the baseline methods. In addition, the strengths and limitations of the proposed solution are discussed to highlight any areas for improvement.

Chapter 5 - Conclusion

In this final chapter, the research is summarized; presenting key findings, conclusions and areas for future research.

Chapter 2

Related work

Noise reduction in EEG signals is fundamentally grounded in the field of signal processing. Many of the techniques used for this purpose, which have their origins in speech and audio processing, have been implemented across the area. This research focuses on the application of unsupervised machine learning techniques to solve the problem. In particular, the use of convolutional and recurrent neural network layers in a stacked auto-encoder.

In this chapter, EEG noise reduction techniques are reviewed, from classical methods covering frequency filtering and transformation to decomposition using PCA and ICA. In addition, auto-encoders and both convolutional and recurrent neural networks are discussed in relation to their applications in noise reduction and other complex problems.

2.1 Electroencephalogram

As mentioned above, EEG is the recording of electrical potentials produced by the brain. They are made up of brief localized action potentials and slower widespread postsynaptic potentials (Binnie & Prior, 1994). EEG is recorded by placing electrodes on the scalp that measure the voltage of electrical potentials. Placement of the scalp electrodes was standardized by Klem et al., 1999 with the introduction of the interna-

tional 10—20 system, though this has been extended to the 10—10 and 10—5 system (Oostenveld & Praamstra, 2001) in recent years. They are so-called because placement is defined using proportional distances of 20, 10, and 5% of the total length along contours between skull landmarks respectively (Oostenveld & Praamstra, 2001). The amplitude of recordings at these electrodes usually lies between 10 and 100 μV and can be separated into a number of frequency components: Delta (0.1 - 4 Hz), Theta (4 - 8 Hz), Alpha (8 - 13 Hz), Beta (13 - 30 Hz) and Gamma (above 30 Hz) (Kaushal et al., 2016). Their low magnitude means that though the activity of a single neuron can be recorded at adjacent electrodes, it is not detected at a distance (Binnie & Prior, 1994). EEG is used to diagnose brain abnormalities such as sleep disorders, epilepsy, stroke, tumors, brain death and coma (Kaushal et al., 2016).

2.2 EEG artefacts

As EEG signals are very low amplitude, it makes them susceptible to various types of noise (Harender & Sharma, 2017). This noise originates from a number of sources such as eye blink and movement, muscle contraction, cardiac signals and line interference (Jung et al., 2000) commonly referred to as EEG artefacts.

2.2.1 Ocular artefact

Ocular artefacts which are the result of eye blinks and movement produce large electrical potentials at amplitudes more than ten times that of EEG (Peng et al., 2013). As discussed in Croft et al., 2005, there have been a number of methods used to remove these artefacts including the fixation/rejection technique and EOG correction. The former involves instructing the subject not to blink or including another task which involves the subject focussing on a fixed point, while the latter involves additional channels strategically placed to record ocular artefacts that are then subtracted, by means of regression estimation, from EEG channels. Though these are popular methods, the fixation/rejection technique has been shown to affect several components of the ERP (Verleger, 1991) while EOG correction can reduce important signal informa-

tion captured in the EOG channel. It was concluded in Croft et al., 2005 that, though each method resulted in cleaner data than no removal, a specific correction was needed for each type of ocular artefact.

2.2.2 Muscle artefact

Muscle artefacts, usually caused by movement, swallowing, or twitches, tend to be high frequency and normally affect electrodes located at the frontal and temporal regions of the brain, but can have an impact on any scalp electrode (van de Velde et al., 1998). Due to the high frequency nature of these artefacts filtering has been used as a method for their removal, however this can remove important underlying EEG information or sometimes obscure the muscle artefact. As noted in Crespo-Garcia et al., 2008, equivalent regression methods used for EOG correction are not possible for muscle artefacts since no regression channel exists for these sources.

2.2.3 Pulse artefact

Another of the more prominent artefacts is known as the pulse artefact. It is characterized by amplitudes and frequencies in the range of normal EEG synchronized with cardiac rhythm but with a delay of approximately 200 ms. In addition, it tends to exhibit more complete coverage across electrodes (Debener et al., 2009). Several factors make identification and removal of these artefacts quite difficult. In particular, because it lasts for approximately 500 ms and subjects can have heart rate differences, pulse artefacts associated with higher heart rates can overlap, which complicates their removal (J. L. Vincent et al., 2007). As discussed in Debener et al., 2009, one of the most frequently used algorithms for their removal is average artefact subtraction (AAS). It is constrained by the assumption that the EEG is not correlated to the electrocardiogram (ECG) information and that the artefact is stable across successive heartbeats. A simultaneous ECG is used to identify the exact onset of the heartbeat cycle, then a template is created for the artefact which is subtracted from the EEG. However, Debener et al., 2009 noted that due to the assumptions of AAS, correla-

tion between cardiac activity and the neuronal activity can cause EEG data quality issues. In addition, if the stability assumption is not met, the artefact can be wrongly estimated, resulting in greater residual artefact after subtraction.

2.3 Artefact identification and reduction

As mentioned, there have been a number of methods employed to identify and reduce EEG artefacts. The most common methods include filtering, regression, empirical mode decomposition (EMD), wavelet transformations (WT) and blind source separation (BSS) (Islam et al., 2016; Sheoran et al., 2015). Each of these have certain advantages and disadvantages. In particular, regression and adaptive filtering based methods require a reference channel for their implementation and it has been noted that regression-based methods can remove neural potentials contained in the reference channel (Croft et al., 2005). The wavelet transform, which is a time-frequency analysis, decomposes a signal into a set of functions that are translated versions of a mother wavelet, resulting in a set of coefficients. A threshold is then applied to the coefficients to de-noise the signal before being inverse transformed (Islam et al., 2016). It has been widely used (Harender & Sharma, 2017; Heydari & Shahbakhti, 2015; Kiamini et al., 2009) however, the issue with the WT is that the choice of decomposition method, mother wavelet, level of decomposition and threshold are all user defined with no appropriate means of selection (Sheoran et al., 2015).

The most frequently used method is BSS. In particular principal component analysis (PCA) and independent component analysis (ICA), both of which are decomposition methods. Primarily their use was motivated by the fact that an additional EOG or ECG channel was not required for their implementation (Jung et al., 1998). However, once successfully applied to one artefact, they were also implemented across a number of them. For example in Srivastava et al., 2005, ICA was extended to the identification and removal of pulse artefacts.

2.3.1 Principal component analysis

PCA is one of the most widely used multivariate statistical techniques, originally proposed by Pearson, 1901. The goal is to decompose a data table into a set of orthogonal variables called principal components. PCA is dependent on two things, an eigenvalue decomposition of positive semi-definite matrices and a single value decomposition (SVD) of rectangular matrices (Abdi & Williams, 2010). The extracted components are linear combinations of the original variables with their importance given by the proportion of explained variance. PCA differs to ICA in that it finds the orthogonal direction of greatest variance, whereas with ICA, components may not be orthogonal (Jung et al., 1998). Figure 2.1 shows PCA components from a simple example taken from Zinovyev et al., 2013. In comparison, the ICA components from the same example can be seen in figure 2.2.

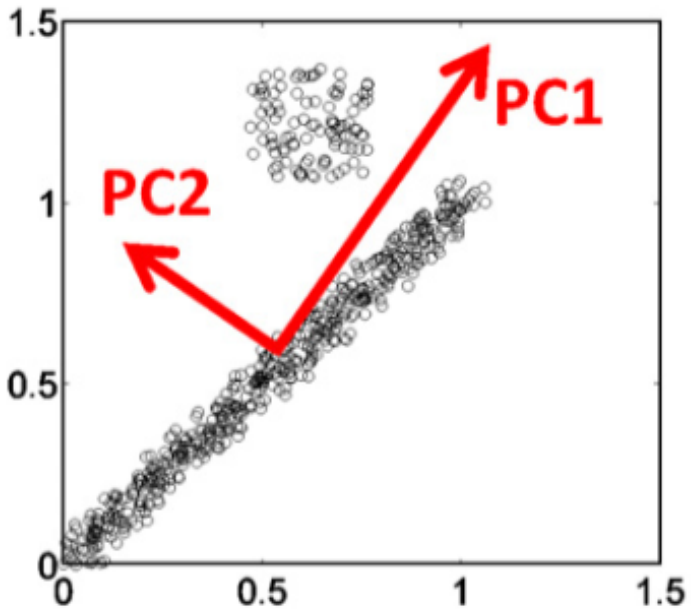


Figure 2.1: Example of 2D data cloud where PCs are principal components (PC1 explains the maximum amount of variance, PC1 and PC2 are orthogonal). Reprinted from Zinovyev et al., 2013

The method has been used in a number of cases for EEG de-noising. In particular, Casarotto et al., 2004 used PCA for the removal of ocular artefacts in ERP. Their method involved computing the correlation between each of the principal components and the EOG reference channel. The first or second component was subtracted if the correlation was above 0.9 or 0.95 respectively. Performance was measured by the number of useful trials and it was found that a significant increase was obtained. However, this requires a reference channel for its implementation which is not always available. Furthermore, as noted in Jung et al., 1998, PCA cannot completely separate ocular artefacts from brain signals when they have comparable amplitudes.

Though applied to magneto-encephalography (MEG), in de Cheveigné and Simon, 2007 time shift PCA was used to remove environmental noise. This method uses reference signals that are time-shifted by a set of positive and negative periods. PCA is then applied to these to obtain orthogonal signals. Finally, each sensor is projected onto the components and the projection is removed resulting in clean data. Performance was measured using the SNR and the results showed that a difference of about 20 db was observed between the clean and noisy data. Similarly to Casarotto et al., 2004, a reference signal is also required and the method is only applied to one EEG artefact.

In Kang and Zhizeng, 2012, PCA was combined with a density estimation blind source separation (DEBSS) algorithm. First wavelet decomposition was used to remove high frequency noise before using PCA for dimensionality reduction, keeping only those components whose cumulative explained variance was above 85%. These components were then seperated using the DEBSS algorithm and compared to the reference channels using cross-correlation. The highly correlated components were removed by setting them to zero before the EEG signals were reconstructed.

2.3.2 Independent component analysis

Introduced by Bell and Sejnowski, 1995 and optimized by Hyvarinen, 1999, independent component analysis is a decomposition method commonly used for feature extraction and BSS. The goal of ICA is to decompose an input into a set of statistically independent sources. As noted in Jung et al., 2000, the BSS problem is to recover independent source signals, $s = s_1(t), ..., s_N(t)$ after they are mixed, by an unknown mixing matrix A, into a set of N mixtures $x = x_1(t), ..., x_N(t)$, where $x = As$. ICA estimates $u = Ws$ by finding a square matrix W that acts as the pseudo inverse of the estimation of A. ICA has two assumptions, first that the data is a linear mixture of the underlying source signals and second that they are linearly independent (Nakamura et al., 2006). Figure 2.2 below, shows the first two ICA components from an example given by Zinovyev et al., 2013.

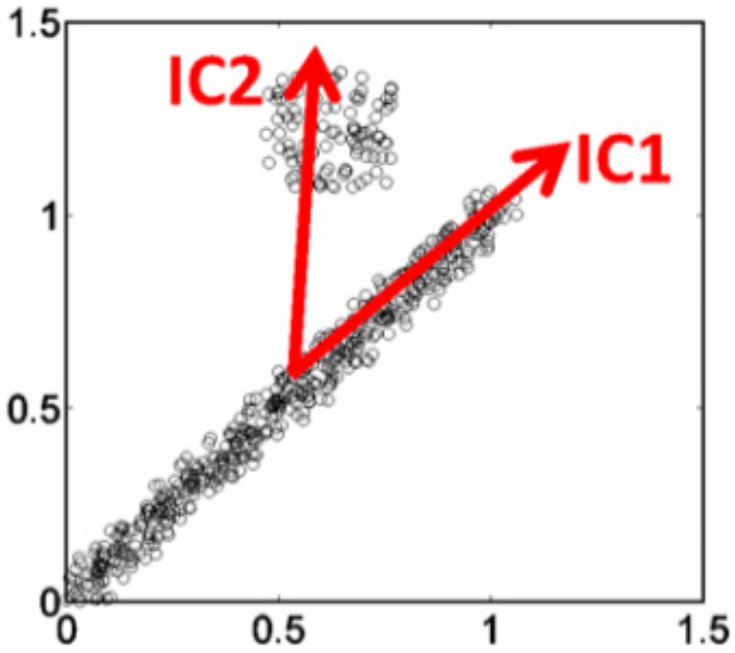


Figure 2.2: Example of 2D data cloud where ICs are independent components (give maximally non-gaussian distribution of the projections). Reprinted from Zinovyev et al., 2013

For EEG, the idea is to separate the signals from the projection of the artefact in each channel (Crespo-Garcia et al., 2008). It can be shown that EEG meets the criteria required for implementing ICA, however an assumption discussed in Jung et al., 2000 is that the number of sources is the same as the number of sensors. They note that this is questionable as the number of statistically independent signals contributing to scalp EEG is not known. Once the signal has been decomposed, EEG artefacts are removed by setting the components containing artefacts to zero before inverse transforming the decomposition.

ICA has been successfully implemented for a number of artefacts (Crespo-Garcia et al., 2008; Nakamura et al., 2006; Vigário, 1997). Makeig et al., 1997 examined its application to ERP analysis, where ICA was used to decompose the ERP into its components. Their results highlighted the use of ICA for EEG signal analysis and led to an extension of their work in which ocular, muscle, and line noise artefacts were removed (Jung et al., 2000). They noted however, that some limitations did exist. Specifically that, like PCA, ICA can only decompose at most N components from N channels and that without enough data its results are not meaningful. The principal limitation, which has been noted in several instances, is that ICA requires manual visual inspection and identification of the components for their removal (Campos Viola et al., 2009; Jung et al., 2000).

As a result, Campos Viola et al., 2009 attempted to develop a semi-automated implementation based on correlation between the inverse weights. It uses a template based method, similar to regression methods, that calculates the correlation between component weights and the template weights. Highly correlated components are then removed accordingly. The method introduced is designed to speed up the process by acting as an aid for the user. They found that performance was best for ocular artefacts due a high degree of overlap between user only selection and user aided selection of those artefacts.

Similarly, Ghandeharion and Erfanian, 2010 attempted to develop a fully automated implementation based on mutual information (MI) and wavelets. They suggest the use of four measures for identification, kurtosis of the coarse and detail component waveforms, the correlation coefficient between the components and the reference signals, the relative strength of each component at the vertical and horizontal EOG, and mutual information. Components with at least four maximal values are selected. When evaluated, the accuracy of ocular artefact identification was 97.8% using 4 second EEG epochs, however this still requires EOG reference channels and isn't applicable to artefacts that cannot be captured by reference electrodes.

2.4 Artificial neural networks

In the recent past, with the advances in deep learning and their expressive power, neural networks have been used for many complex problems (Bengio & Delalleau, 2011). An artificial neural network (ANN) is a system modelled after the human brain containing artificial neurons designed for complex non-linear tasks. Figure 2.3 shows a basic ANN network with an input, hidden and output layer. In an ANN, each connection is given a weight which is multiplied by the input before being passed to a node. At each node, a typically non-linear activation function is applied to the inputs to produce a single output. The purpose being to add non-linearity to the output enabling the model to learn complex relationships.

There are two central types of machine learning algorithm — supervised and unsupervised. In the former, a target variable is given and the algorithm learns a function that maps the attributes to each target. In the latter, the data is unlabelled which means the algorithm has to learn patterns in the data and add structure to meaningfully group instances. Two common types of supervised learning task are regression and classification. The difference between them is that in a regression task the target values are continuous while for classification they are categorical. In Miller et al., 1995 a single layer ANN was used to classify remote-sensing image data achieving within-class

discrimination comparable to humans; Highlighting their classification effectiveness. In Cigizoglu and Alp, 2006 a two hidden layer ANN was used to model river sediment yield and was found to be significantly superior to conventional methods; Highlighting the strength of ANN's for regression tasks.

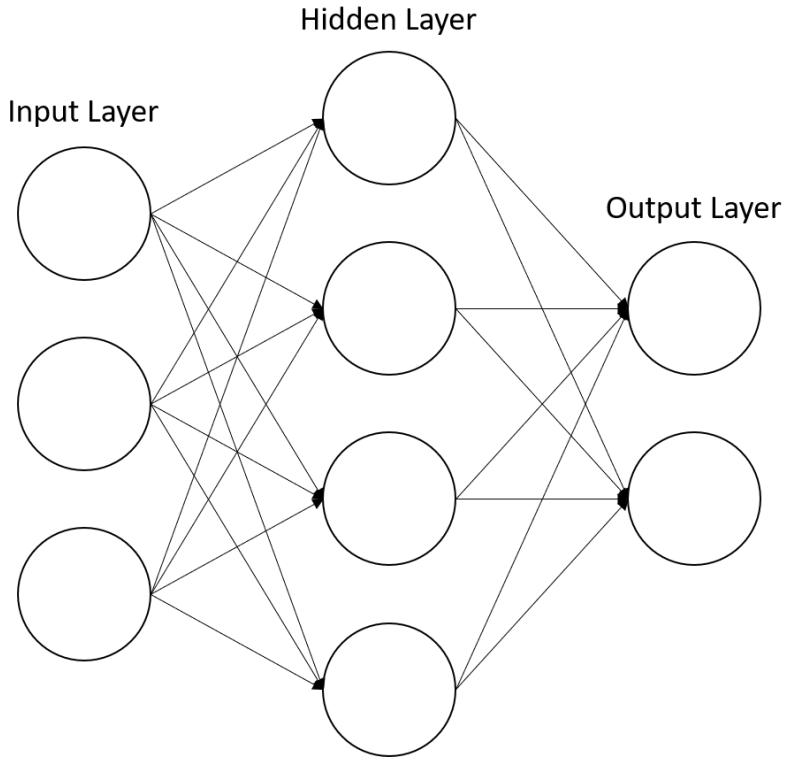


Figure 2.3: Single hidden layer ANN network with each node representing an artificial neuron and arrows the connections from the output of one node to the input of another

2.5 Auto-encoders

A specific class of ANN is the auto-encoder. This is an unsupervised learning algorithm proposed by Kramer, 1991 that is used to reconstruct a given input from a lower dimension representation called a bottleneck. As described by Bengio et al., 2013, the AE consists of a feature-extracting function called an encoder which creates a feature vector from the set of input vectors $h^{(t)} = f_\theta(x)$ and a decoder which maps from the feature space to the input space $r = g_\theta(h)$. In an AE the number of inputs is the same as the number of outputs and the model learns the set of encoder and decoder param-

eters simultaneously to minimize the reconstruction error $L(x, r)$ between the input values and their reconstruction. There are many types of AE including the de-noising auto-encoder (DAE) which corrupts an input signal before learning to reconstruct the original input (P. Vincent et al., 2008), sparse auto-encoder (SAE) which adds a sparsity constraint to the hidden units to control to the number of neurons that are active at once (B. Yang et al., 2016), contractive auto-encoder (CAE) which adds a penalty to the cost function to make the model more robust to slight variations in input data (Rifai et al., 2011), and the variational auto-encoder (VAE). In some cases these have been combined with convolutional neural networks (CNN) or recurrent neural networks (RNN) particularly due to their known strengths in image classification (Bengio et al., 2013) and time-series analysis (Bengio et al., 1994) respectively. As a result auto-encoders have been used for a number of EEG based problems including feature learning, classification and noise reduction.

2.5.1 Classification

A stacked de-noising auto-encoder was used by Yin and Zhang, 2016 for the classification of cognitive task load (CTL) into binary levels of low and high. One of their motivations for using a DAE was as a replacement for using a band-pass filter to remove potential noise in the signals. They found that based on classification error, the proposed method achieved 74% subject specific accuracy on average. A sparse DAE was used by Qiu et al., 2018 for classifying seizures in ictal EEG. They used the sparsity constraint for efficiency and similarly to Yin and Zhang, 2016, used the de-noising corrupting operation for robustness. In this case, they used a logistic regression classifier as an extension of the method for classification purposes. Results, measured using sensitivity and specificity, were very strong only falling to 92% for the five class classification problem. Despite these results, neither included a comparative method to determine whether their proposed method was better than an established classification technique.

2.5.2 Feature learning

Auto-encoders have been widely used in EEG as a means of extracting or enhancing features. This use of auto-encoders is similar to those mentioned above for classification, though in these cases they are specifically used for feature extraction. In Supratak et al., 2014, a stacked auto-encoder was used to extract features for a logistic regression classifier to detect seizures. The method employed is similar to that of Yin and Zhang, 2016, however a DAE was not used in this case. Performance was also evaluated with respect to sensitivity and it was found that given a single channel or set of three the model could classify all seizures, though when using 5 channels sensitivity fell to 87.18%. This result is very like that of Yin and Zhang, 2016 and in neither case was a comparative method used. This was addressed in Helal et al., 2017 who used a single layer AE for dimensionality reduction and feature extraction in comparison to PCA. Each output was used as input to a linear discriminant analysis (LDA) classifier and performance was evaluated with respect to classification accuracy and Cohen's Kappa. It was found that the AE performed better than PCA achieving a kappa of 0.56 compared to 0.52 with classification accuracy of 67% compared to 64%.

The use of convolutional auto-encoders was exploited by Wen and Zhang, 2018 for seizure detection in their proposed AE-CDNN. The deep convolutional network stacked layers of CNN's for both the encoder and decoder using nine different classification methods to compare performance. Once again PCA was used as a comparative method along with sparse random projection (SRP). Accuracy was used for evaluation and the proposed method was found to outperform PCA and SRP when feature reduction was greater than 16 reaching on average 92% accuracy. In cases of lower feature dimension reduction PCA and SRP were found to perform better. This result is consistent with that of Helal et al., 2017 who also found that PCA performed better than the proposed AE when dimension reduction increased.

As mentioned before auto-encoders have been combined with both convolutional and recurrent neural network layers due to their ability to extract both spatial and tempo-

ral features. In Jia et al., 2017, a spatio-temporal auto-encoder was designed using a stacked architecture of AE layers wrapped around long short-term memory (LSTM) recurrent layers. The LSTM is used to extract temporal features while the AE extracted spatial features. The feature output was then used in several classifiers to determine whether a patient was under the influence of alcohol or not. This research was specifically looking to determine whether the model could perform well with missing data, as was also done in Li et al., 2015 using a DAE, however their results showed the DAE did not outperform a support vector machine (SVM). In this case, three classifiers were tested using the full EEG data, 30% continuous and linear imputed data, and both imputed datasets using the spatio-temporal AE feature extraction method. The proposed method was shown to improve accuracy greatly for both imputed datasets with performance close to that of the full EEG dataset. The CNN classifier used achieved the highest accuracy, however it was not discussed why CNN layers weren't used in place of the AE layers of the spatio-temporal AE, especially considering the CNN's ability to extract spatial features. This strength could have been exploited which may have improved performance even more for the other simpler classification methods used.

The opposite configuration was used by Abdelhameed et al., 2018, who designed a deep convolutional auto-encoder for feature extraction as input to three classifiers — a multi-layer perceptron (MLP), an LSTM and a bi-directional LSTM for seizure detection. Several convolutional and max-pooling layers were stacked for the encoder with several convolutional and upsampling layers for the decoder. Sensitivity, specificity and accuracy were again chosen as evaluation metrics however, in this case no data was imputed and three other feature extraction methods were used for comparison. The comparative methods were PCA, wavelet transform and an ANN, and in all cases the proposed method using a bi-directional LSTM achieved better results across all metrics. The methods used in Abdelhameed et al., 2018 and Jia et al., 2017 highlight the strength of combining auto-encoders with both convolutional and recurrent neural network architectures.

2.5.3 Noise reduction

In addition to being used for feature extraction, auto-encoders have also been used for noise reduction and artefact removal. Generally the focus seems to have been on the removal of EOG artefacts, though auto-encoders have been used for de-noising in ECG signals (Xiong et al., 2016) and medical images (Gondara, 2016). In the former, a DAE was used in combination with a wavelet transform (WT) to remove artificially added noise from ECG signals. The SNR was used to measure performance and a significant improvement was observed along with very low root-mean-square error (RMSE) for reconstruction loss. In the latter, a deep convolutional AE was used to remove noise, that similarly to Xiong et al., 2016 had been artificially added, to medical images. In this case, structural similarity index measure (SSIM) was used instead of PSNR for consistency and accuracy. Results showed that the proposed method was better than a simple median filter, however other more established image de-noising methods could have been used for comparison.

A common method used for EOG artefact removal is that of the SAE (Nguyen et al., 2019; B. Yang et al., 2018; B. Yang et al., 2016). In B. Yang et al., 2016 an SAE was combined with a least squares adaptive filter for this purpose. Though the focus of this research was on EOG artefact removal, classification accuracy and time consumption were used as evaluation metrics. Therefore, despite an increase in classification accuracy the ability to remove EOG artefacts was not evaluated. This was addressed in B. Yang et al., 2018, where an SAE was also used though evaluation of artefact removal was specifically assessed using the power spectral density (PSD) along with RMSE and classification accuracy as additional metrics. In both cases comparisons were made to other state-of-the-art methods including ICA in the first instance and ICA, k-ICA and second-order blind identification (SOBI) in the second. In both cases the proposed methods were shown to outperform the state-of-the-art. PSD was also used by Nguyen et al., 2019 along with frequency correlation (FC) to evaluate the performance of their proposed deep wavelet SAE which uses the wavelet coefficients as input to an SAE for EOG removal. Similarly to B. Yang et al., 2018, SOBI is also used for comparison

along with a wavelet neural network (WNN). They showed that the proposed method addressed some limitations of the other methods, specifically single channel online use.

The method used by Leite et al., 2018, assessed the ability of a deep convolutional AE to remove EOG and jaw clench artefacts. As with Wen and Zhang, 2018, several stacked convolutional and max pooling layers were used for the encoder with upsampling used in the decoder. In this case, PSNR was used to measure the peak amplitude of signals compared to the noise affecting them. Unlike in both Xiong et al., 2016 and Gondara, 2016 artificial noise was not added to the signals. As an alternative, specific tasks were developed to evoke a response which would create the desired noise. In this way, though still artificially generated, the noise would be more indicative of real artefacts. As a result, the proposed method could be assessed for both types of noise individually. Evaluation was done using the PSNR and it was found that overall the mean difference was positive across all channels with specific channels performing best for each type of noise. Channel Cz had the highest overall mean PSNR for eye blink noise while Fz had the highest for jaw clench noise. In addition, F4 and Fz exhibited the highest PSNR difference for eye blink and jaw clench noise respectively.

An interesting concept used by Ghosh et al., 2019, is that of combining an SVM to classify windowed segments as noisy or not with an AE to remove noise from noisy segments. The proposed method uses a 0.45 second sliding window with a 50% overlap. This window covers the typical eye-blink duration and the average eye-blink duration so none are missed. Training of the AE involved 1000 pre-classified segments of noisy data as input with clean data as the target. Similarly, for the SVM, training involved classification of pre-labelled noisy and clean EEG data using three features, namely, variance, kurtosis and peak-to-peak amplitude. In total, five metrics were used to evaluate the artefact removal method, including RMSE, signal-to-artefact ratio (SAR), mean absolute error (MAE), correlation coefficient (CC) and MI with accuracy used to evaluate the SVM. Results showed that SVM identification of corrupted segments was consistent across the entire EEG, and that the proposed de-noising method achieved

better results across each metric than both adaptive noise cancellation (ADC) and the wavelet transform. In addition, this method does not alter the entire EEG; Only specific windowed segments are changed by the auto-encoder.

2.6 Summary

2.6.1 Overview

EEG is a low amplitude recording of electrical potentials in the brain. As such they are susceptible to noise being present in the signal which could impact analysis and consequently diagnosis. This noise can originate from several sources including eye and muscle movement, heartbeat pulse and line interference which are known as artefacts. Each has different characteristics which make their identification and reduction difficult. Many approaches have been used to identify and reduce each artefact, including fixation, subtraction, linear and non-linear filtering, wavelet transformations and blind source separation. The most widely used methods have been PCA and ICA though a lot of research has been done on wavelets. All approaches have been successfully implemented, however there have also been several limitations noted for each.

With advances in modern machine learning techniques, many supervised and unsupervised algorithms have been used for both classification and regression tasks. In particular, auto-encoders, which are used to reconstruct an input from a dimensionally reduced representation with several variations, have become increasingly useful across a spectrum of tasks. From an EEG perspective they have been used primarily for classification and feature extraction, however recently they have also been implemented for noise reduction and artefact removal. In addition to the various types of auto-encoder, other neural network architectures including CNN's and RNN's have been embedded in them. These utilize the spatial and temporal feature maps that have shown CNN's and RNN's to be so successful for image classification and time series analysis. Primarily they have been used in stacked auto-encoders to take further advantage of deep architectures for highly non-linear and complex tasks.

2.6.2 Gaps in the literature

Classic noise reduction approaches, including regression-based techniques, adaptive filters and blind source separation, have often required a reference signal for artefact removal. This means that additional electrodes, placed specifically to record these artefacts need to be used. For pulse and ocular artefacts this is possible, however, it cannot be implemented for all artefacts. Furthermore, it has been noted, that specifically for ocular artefacts, these reference signals can also include neural information which is subtracted in the process. Therefore, an automated method is required that can be applied for the identification and removal of all noise sources and will limit the amount of information loss.

Decomposition methods like PCA and ICA have been used to address the limitation of regression-based methods. However, the effectiveness of PCA has primarily been for oculur artefacts and it has been noted to struggle when the amplitudes of artefacts and signal are comparable. In addition, ICA has been noted as requiring extensive knowledge and time to implement. This is due to the fact that careful consideration is needed for the identification of those components that contain artefacts. Therefore, automated and semi-automated approaches have been developed to address this limitation. The latter method uses components identified for one subject to identify similar components for other subjects. This still requires time for identification but significantly reduces the time required for other subjects. The former approach uses a combination of ICA, MI and WT's along with a number of metrics at maximal value to identify components. Though this has been successfully implemented there is still an opportunity to utilize the power of deep architectures and modern machine learning methods to improve the effectiveness. In particular, because of information loss, since the N channel EEG is decomposed into N components. Each component likely contains some important neural information which is then removed during the process and unlike PCA cannot be quantified. Therefore, further validation is given to the need for an automated method that is applicable to more noise sources and can limit information loss.

Auto-encoders, which are specifically designed to minimize construction loss between the input and output, are particularly useful for tackling the issue of information loss. Since they are designed to minimize reconstruction loss between the input and output they are by default trying to reduce the amount of information lost as a result of the dimensionality reduction. Not only this but their unsupervised algorithm enables them to learn important features of the latent representation which are retained in the reconstruction. The assumption here is that only important neural activity will be retained and noise will be reduced. In addition, given that no reference signal is required, they therefore also address the limitation of previous regression-based methods and become applicable to a broader spectrum of artefacts.

Their use to date has primarily been limited to classification tasks and feature extraction though they have been implemented recently for noise reduction. In many of these cases they have been used because of their supposed de-noising ability, in particular the DAE, and used to pre-train neural networks. The assumption is that the auto-encoder architecture would retain only the important information, however, in the majority of classification and feature extraction cases they were evaluated based on measures of accuracy or loss and have rarely been evaluated for the purpose they were implemented. In fact even when used for artefact reduction they have not always been evaluated with measures of signal purity. Despite this fact, they have been very successful and have achieved higher accuracy than other classic methods. This highlights their effectiveness, though it has not been thoroughly quantified from a noise reduction perspective. This indicates that there is a need to evaluate these methods based on metrics such as SNR and PSNR which determine the level of noise compared to the signal.

Recurrent neural networks have been shown to perform very well for time-series analysis (Wan et al., 2019). In particular, gated architectures such as the LSTM and GRU allow RNN's to learn long-term dependencies which are particularly powerful in

problems for which these are essential (Bengio et al., 1994). Furthermore, they have been used in stacked AE architectures for EEG feature learning and shown to greatly improve accuracy. Convolutional neural networks on the other hand, are a proven state-of-the-art image classification method (Szegedy et al., 2015) and, similarly to RNN's, have been shown to perform well in stacked AE's for EEG feature learning. Their respective ability to extract temporal and spatial features has been one the key motivations for their use. In some cases they have been used in combination, where one is used for feature extraction and the other as a classifier or for other complex problems such as in Cakir et al., 2017, Marchi et al., 2015, and Trigeorgis et al., 2016. However, there has been limited exploration of their combined use in a single auto-encoder architecture for noise reduction despite it being noted as a potential extension (Leite et al., 2018). As can be found in the literature, they are often used in deep architectures and have been used in stacked auto-encoders for EEG feature learning. In fact, the most commonly used auto-encoder architecture for artefact removal has been the stacked SAE because of its depth and sparsity constraints making it a robust learning algorithm. This combination of deep learning and advanced neural network architectures has proven to be very powerful for a number of complex tasks. For these reasons, a logical extension of the CNN or RNN based AE is to combine both into a single convolutional-recurrent auto-encoder. Furthermore, stacking layers to add depth seems like a reasonable approach, given the already successful use of stacked AE architectures for EEG artefact removal.

2.6.3 Research question

This literature review led to the following research question:

"Can the signal-to-noise ratio of EEG signals produced by a stacked auto-encoder be improved when compared to PCA, ICA and a traditional auto-encoder?"

Chapter 3

Design and methodology

The purpose of this chapter is to introduce the research methodology for this quantitative empirical study on whether a convolutional recurrent auto-encoder can be used to improve the signal-to-noise ratio of EEG signals. In this chapter the null and alternative hypothesis are stated, data collection and preparation methods are discussed and the experiment design used to test the hypothesis is described. Additionally, a summary of the method including its strengths and limitations are outlined.

3.1 Hypothesis

H_1: If a stacked auto-encoder, built with convolutional and recurrent neural network layers, is applied to EEG signals, the signal-to-noise ratio will be increased when compared to PCA, ICA and a basic auto-encoder.

H_0: If a stacked auto-encoder, built with convolutional and recurrent neural network layers, is applied to EEG signals, the signal-to-noise ratio won't be increased when compared to PCA, ICA and a basic auto-encoder.

Figure 3.1 below shows a high level overview of the experiment and its components.

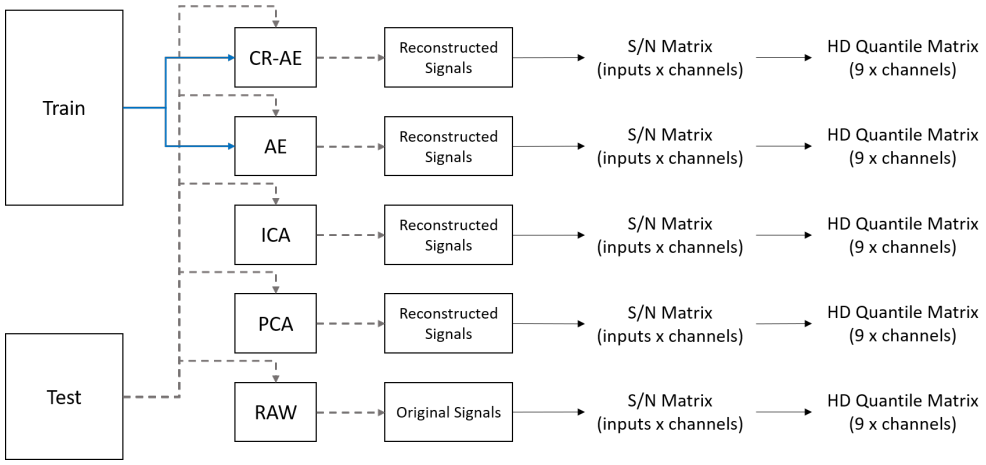


Figure 3.1: High level overview of how each of the design components combines to create the experiment

3.2 Data collection

Two datasets were used in this research. The primary dataset, sourced from Kaggle [1] [2] was recorded by Ford et al., 2014 during research into efference copy and corollary discharge of schizophrenia patients in response to a stimulus. The 64 channel EEG data, with an additional 8 external sites, was collected for 81 subjects as they underwent 100 trials in each of 3 conditions. It was recorded at 1024 Hz using a BioSemi ActiveTwo system with the 64 channels placed as per the international 10-10 system described by Oostenveld and Praamstra, 2001 (see figure 3.2). Of the 81 subjects, 49 had been diagnosed with DSM-IV schizophrenia while the remaining 32 were healthy controls.

Following recording, the data was re-referenced to averaged earlobe electrodes and bandpass filtered between 0.5 and 15 Hz. It was then divided into 3000 ms epochs, 1500 ms before the onset of the stimulus and 1500 ms after, and baseline corrected at -600 to -500ms.

[1] https://www.kaggle.com/broach/button-tone-sz
[2] https://www.kaggle.com/broach/buttontonesz2

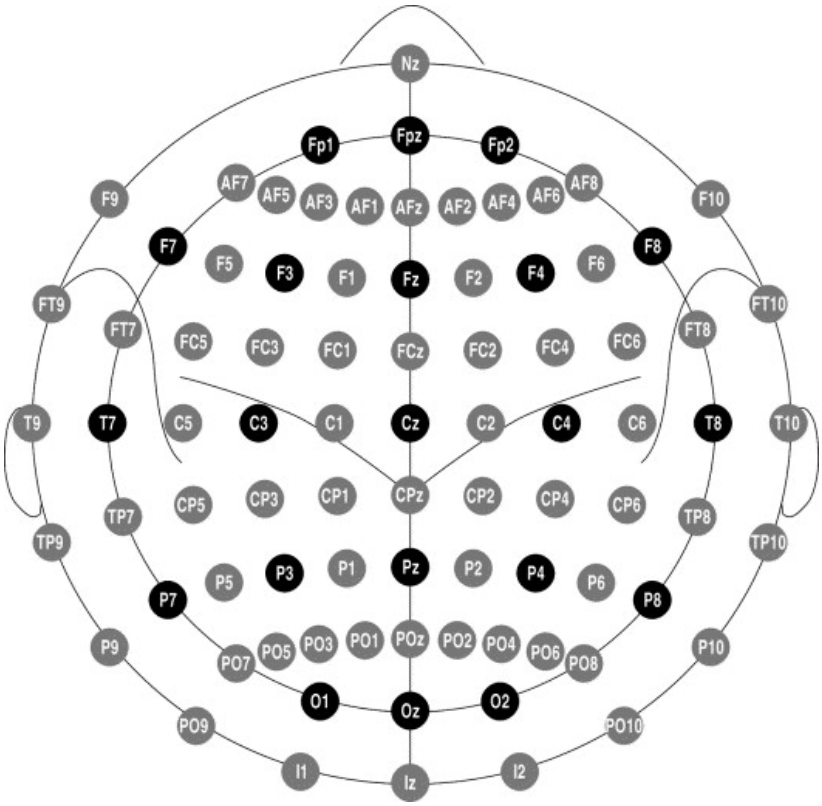


Figure 3.2: 10-10 electrode placement system with black circles indicating positions of the original 10-20 system. Reprinted from Oostenveld and Praamstra, 2001.

Each condition involved a simple set of actions: for condition 1, each subject pressed a button every 1 to 2 seconds which generated, without delay, a 1000 Hz, 80 dB tone (Button Tone), condition 2 involved the listening back to the tones generated in the first condition (Play Tone), finally, for condition 3, subjects once again pressed a button but no tone was generated (Button Alone). In total, the primary dataset contained 71,273,391 samples with each trial containing approximately 3000.

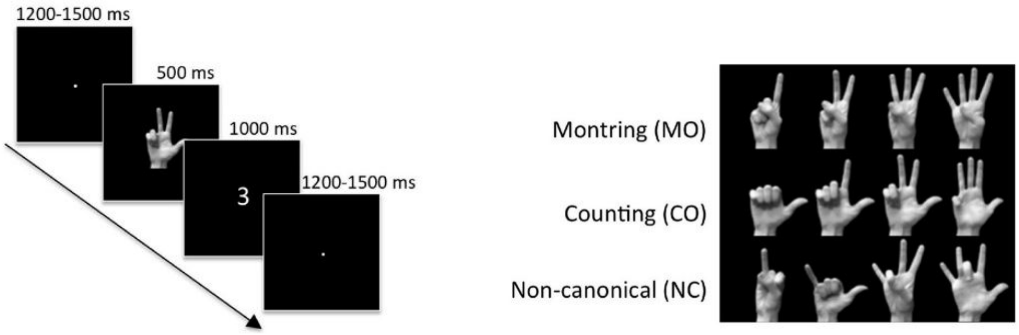
The secondary dataset, sourced from Harvard Dataverse [3] was recorded by Soylu et al., 2019 for research into early perceptual and later semantic processing of canonical and non-canonical finger-numeral configurations in adults. The 32 channel EEG data was collected for 46 right-handed undergraduate students as they underwent a total of 960 trials but was excluded for 8 of those who counted on their left-hands. It was

[3]https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BNNSRG

recorded at 500 Hz using a BrainVisio ActiChamp system with electrodes placed as per the international 10-20 system.

Once recorded, the data was re-referenced to the average reference and filtered between 0.5 and 15 Hz using a Butterworth filter. Each recording was segmented into epochs representing 200 ms before the onset of the gesture presentation to 500 ms after and baseline corrected using the 200 ms pre-stimulus period.

In each trial a finger-numeral configuration was presented for 500 ms followed by a validation step in which a single-digit Arabic numeral was presented (see figure 3.3). To reduce predictability and evenly distribute the stimuli, each 96 trial block was made up of 4 randomized sets of 24 finger-numeral configuration images; 4 montring (MO), 4 counting (CO), and 4 non-canonical (NC) separately for left and right hands (see figure 3.3). Participants pressed one of two buttons using their right or left index finger to indicate whether the Arabic numeral corresponded to the number represented by the preceding finger-numeral configuration. In total, the secondary dataset contained 47,569,380 samples with each trial containing approximately 1500.



(a) (b)

Figure 3.3: Left: Stimulus presentation sequence for each trial. Right: Finger-numeral configurations (right-hand only). Reprinted from Soylu et al., 2019

3.3 Data preparation

For the primary dataset, each subject's data is divided per condition into three separate datasets. Separately, all trials for every subject are combined into a single dataset per condition. This is done to enable both subject and trial wise sampling. Full details of both datasets can be found in table 3.1 below.

Dataset	Source	Subjects	Trials	Channels	Sample Rate	Sample Size	Data Format
Primary	Kaggle	81	100	64 + 8	1024 Hz	71,273,391	Epoched
Secondary	Harvard Dataverse	46	960	32	500 Hz	47,569,380	Continuous

Table 3.1: Summary of datasets used in experiment

For preprocessing, both the baseline and meaningful signals are extracted from each sample. In Ford et al., 2014 a 100 ms baseline was used, whereas in each of Bijma et al., 2003; Hu et al., 2014; Min and Herrmann, 2007; Soylu et al., 2019 the baseline varied between 100 and 500 ms. For consistency with the original research, each sample is subset to included 100 ms pre-stimulus and 500 ms post-stimulus as the baseline and meaningful signal respectively. Though this is inconsistent with the 700 ms used in Ford et al., 2014, the meaningful signal from the second dataset is constrained to 500 ms due to the subsequent presentation of a single-digit Arabic numeral in each trial. Therefore, each trial is subset such that it contains 600 samples corresponding to 100 ms pre-stimulus and 500 ms post-stimulus at a rate of 1000 samples per second.

Additionally, the data is downsampled by a factor of 2 due to the sample rate difference between the datasets. Furthermore, since a smaller subset of electrodes was used in Soylu et al., 2019, only channels present in both datasets are preserved. With a difference of three channels and only 31 present in the secondary dataset, 28 channels are used for this research. Details of the channels used can be found in table A.1. Finally, each dataset is re-shaped such that its dimensions are $(trials, timesteps, channels)$. An example pre-processed signal can be seen in figure 3.4.

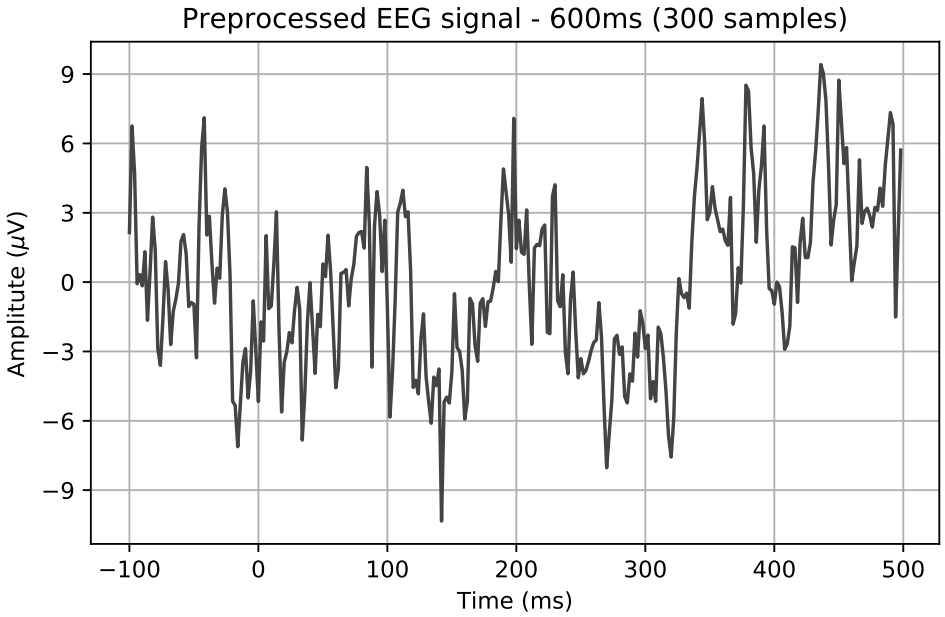


Figure 3.4: Single channel preprocessed EEG signal taken from primary dataset

For the secondary dataset, all 960 trials per subject were recorded continuously and contained in an approximately 1.3 million sample observation per channel. Stimulus introductions were annotated for each trial at the point of onset. Using these, each trial is subset such that it contains 300 samples, corresponding to 100 ms pre-stimulus and 500 ms post-stimulus at a rate of 500 samples per second.

For every subject a separate dataset is created for each of the 24 stimuli. Since each subject was presented with 4 randomized sets of the 24 in each block and there were 10 blocks altogether per subject, a total of 40 trials for each stimulus is present in a given dataset. As before, all trials for each stimulus are then combined into a single dataset containing data across all subjects for the corresponding stimulus. Finally, the channels are reordered to be consistent with the primary dataset and each dataset is re-shaped such that its dimensions are $(trials, timesteps, channels)$. An example pre-processed signal can be seen in figure 3.5.

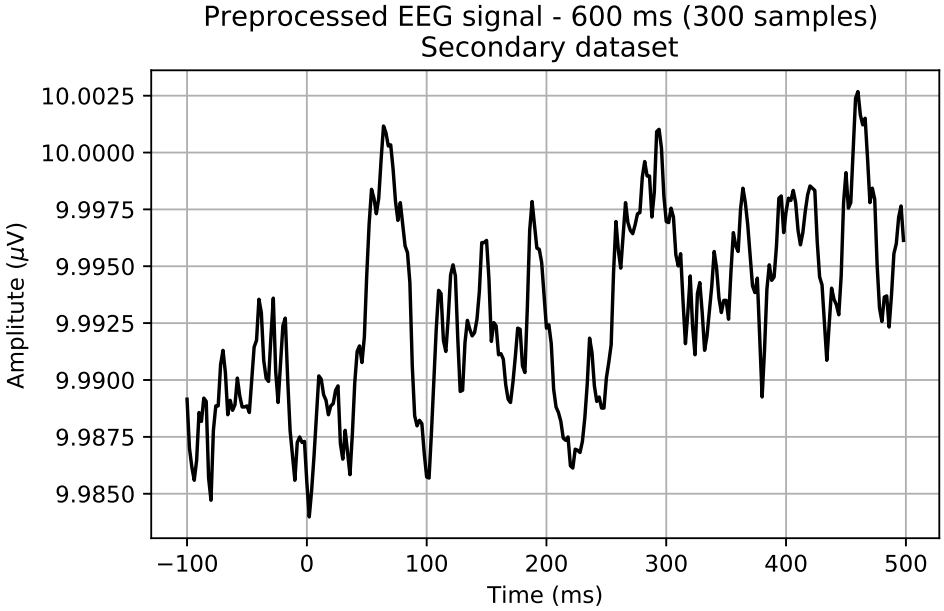
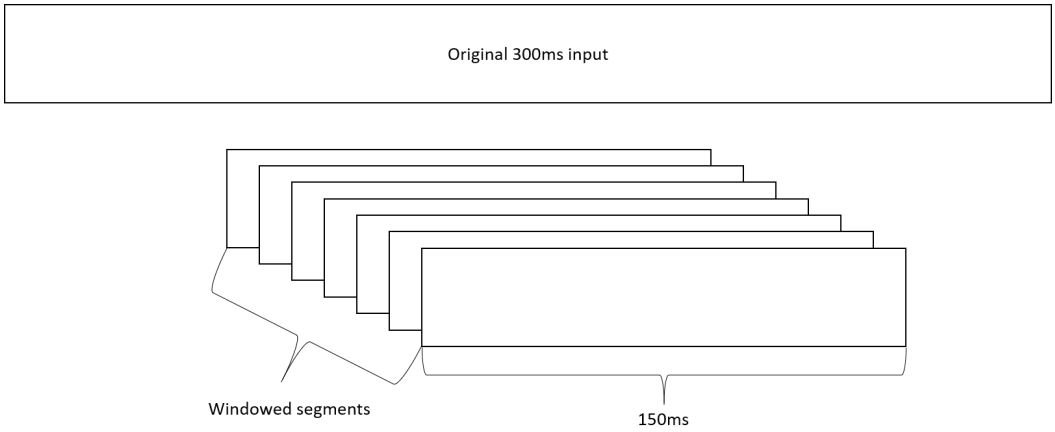


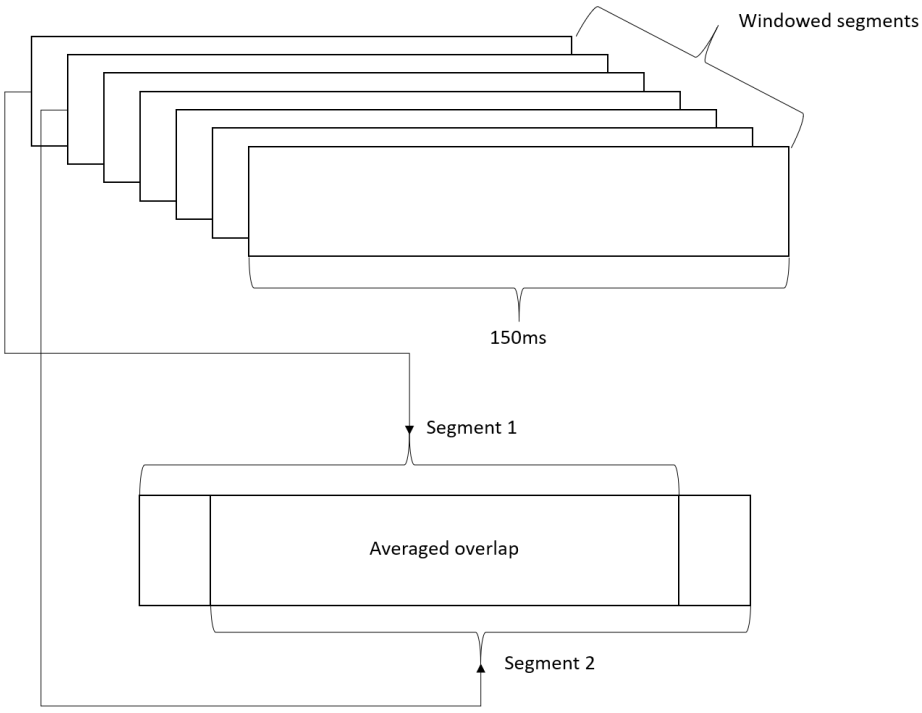
Figure 3.5: Single channel preprocessed EEG signal taken from secondary dataset

Two methods of sampling are used to generate train and test datasets for the primary data — subject-wise and trial-wise. For the former, each condition dataset is randomly shuffled and then split in the ratio 70:30; For the latter, the set of subjects is randomly shuffled and split using the same ratio. Then train and test data is created using each set of subjects. This allows the model to be tested on both unseen subjects and unseen trials.

Windowing is used to augment both the train and test datasets by extracting additional information from each input. A sliding window of 300 ms with a 25 ms shift is used to divide each input into seven overlapping slices. In doing so, the number of inputs for training increases by a factor of 6. Order is preserved by applying the sliding window, after sampling, to the train and test data. This is done to ensure that all predicted outputs can be combined correctly into the corresponding original signal for evaluation. Predicted output signals are recombined by averaging the values in every overlap and taking the actual values to the left and right of each overlap. This process is shown in the graphic below 3.6.



(a)



(b)

Figure 3.6: Top: Single trial split into windowed segments. Bottom: Two windowed segments recombined

3.4 Auto-encoder design

The proposed stacked auto-encoder is based on a combination of one-dimensional convolutional neural network (CNN) layers and long short-term memory (LSTM) recurrent neural network layers. In total three architectures are evaluated. The first combines a single layer of each for both the encoder and decoder, the second utilizes a single CNN layer with multiple LSTM layers, while the third combines multiple CNN layers with a single LSTM layer. Figure 3.7 shows a simplified example of the auto-encoder architecture.

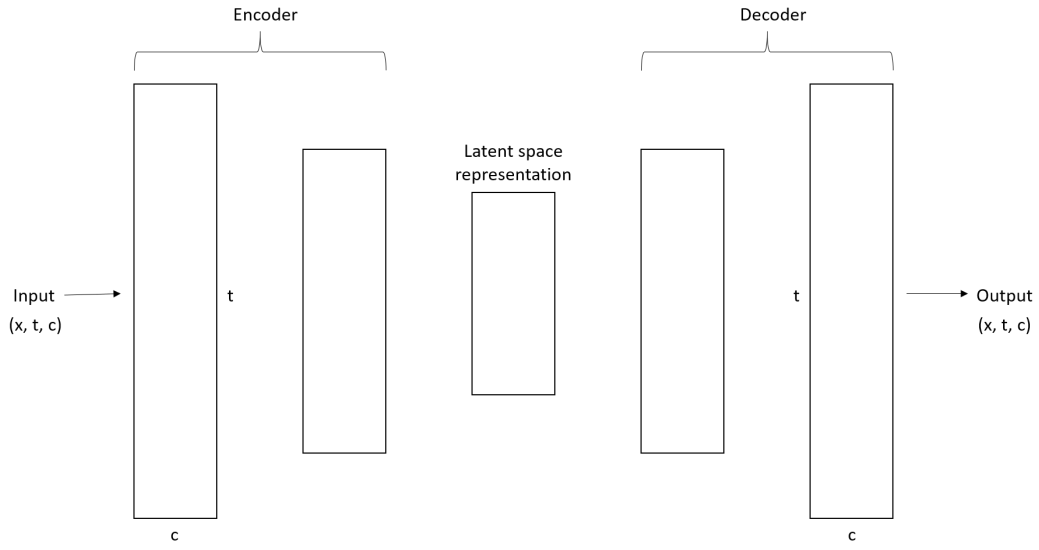
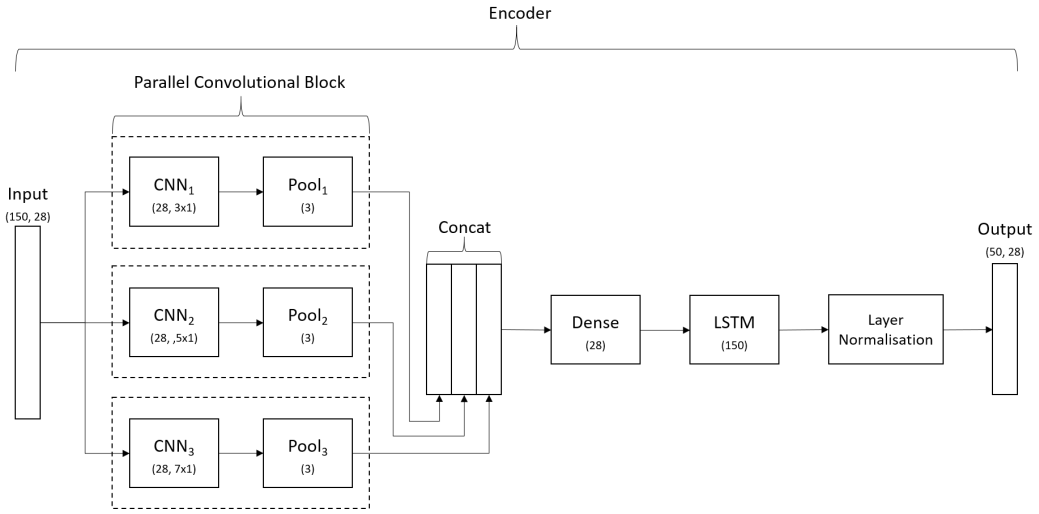


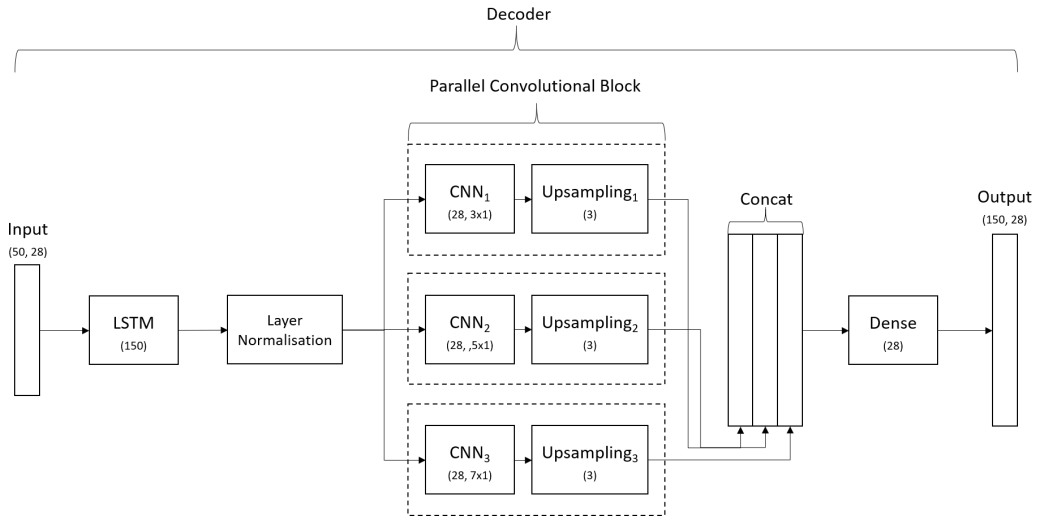
Figure 3.7: A simplified representation of an auto-encoder architecture

Each convolutional layer is constructed as a triple of parallel CNN's utilizing different size kernels to extract varied temporal information similar to the method used for inception modules (Szegedy et al., 2015). This enables each CNN to extract information for the same window but over a larger number of time-steps. For the encoder, pooling is applied to each CNN individually and the outputs are concatenated. Before being input to a recurrent layer, comprising an LSTM and layer normalization, the concatenated outputs are passed to a dense layer to extract the most meaningful information. For the decoder, this sequence is reversed and pooling is replaced by upsampling. Figure 3.8, shows the structure of the single layer architecture for both the encoder and

decoder. When multiple convolutional layers are used, each parallel block is made up of sequential CNN and pooling layers.



(a)



(b)

Figure 3.8: Top: Structure of the encoder for the single layer architecture. Bottom: Structure of the decoder for same

Pooling is used to reduce the input resolution, making it more robust to small variations from previous learned features (Zheng et al., 2014). There are several pooling strategies, though the most commonly used are maximum and average pooling. The former emphasizes prominent features while the latter smooths them by taking their

average. Though max pooling has been shown to produce excellent results across a variety of tasks in both signal and image processing (Abdelhameed et al., 2018; Gondara, 2016; Leite et al., 2018; Wen & Zhang, 2018), average pooling is chosen for its smoothing property which should aid signal de-noising. As noted in Bengio et al., 1994, training recurrent neural networks is complicated due to long-term dependencies and vanishing gradient. In Glorot and Bengio, 2010, normalization was proposed to speed up convergence — an approach that was further developed by Ioffe and Szegedy, 2015 who introduced batch normalization. This strategy standardizes each summed activation of the previous layer at each batch. One of the issues when training recurrent neural networks using this strategy is that the activations tend to vary with the length of the sequence which indicated a requirement to have different statistics for different time-steps (Ba et al., 2016). This led to the development of layer normalization which normalizes the activations of the previous layer for each given example in a batch independently. For that reason layer normalization was chosen to help improve model convergence.

Further optimization is achieved through the use of the Adam algorithm — a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments. This method uses the exponential moving average of the gradients to scale the learning rate which is similar to the RMSProp algorithm, however it incorporates momentum which helps to accelerate gradient descent. The benefit of the Adam algorithm is its computational efficiency and limited memory requirement (Kingma & Ba, 2014). Another consideration was the NAdam algorithm which incorporates Nesterov momentum into Adam. This calculates the gradient updates with respect to the future steps as opposed to the current step. It is useful for noisy or high curvature gradients but was noted to perform best when dropout was included but not work well in all cases (Dozat, 2016). As discussed, dropout is not used in this model so for that reason Adam was chosen over NAdam.

A key choice for the model is the loss function. Since this is a regression based model, potential options include mean square error (MSE), MAE and Huber loss. Both MSE and MAE are commonly used loss functions though MAE is more robust to outliers, while MSE is more stable and converges even with a fixed learning rate. In this case, the EEG signals have high variance and contain a significant number of outliers. These outliers are important to model as they could be related to meaningful signals. However, since there are a significant number and MSE is sensitive to outliers other predictions may be skewed which could impact signal de-noising. Huber loss incorporates both by adding a threshold value which determines whether the residuals are minimized using MSE or MAE. This means it is less sensitive to outliers but is still capable of accurate reconstruction. Therefore, Huber loss is chosen for this research.

Several hyper-parameters can be tuned for the CNN and LSTM layers. Specifically, for a CNN layer those include the number of filters, kernel size, strides, padding and activation. Since signals can take any real value r $\in \mathbb{R}$, linear activation is used for the convolutional layers. Given the computational requirements for training recurrent neural networks, the cuDNN backed LSTM implementation is used which takes advantage of GPU optimization. Since this is only compatible with hyperbolic tangent activation, sigmoid recurrent activation and zero recurrent dropout, hyper-parameter choices for the LSTM layers are limited. A summary of the available hyper-parameters can be found in table 3.2 while full details of the hyper-parameters chosen for each architecture can be found in tables A.2, A.3 and A.4 in appendix A.

A concept discussed in van den Oord et al., 2016 is that of using causal convolutions for modelling time series data. This ensures that a prediction at time t does not depend on any future time-step — thus preserving temporal order. Additionally, dilated convolutions are used which maintain input shape but allow the model to operate on a coarser scale. Stacking layers of dilated convolutions thus increases the receptive field of the network. For this research, causal padding is used for all CNN layers, while dilated convolutions are used for sequential CNN layers only.

Layer	Hyper-parameters	Description
CNN	Filters	The dimensionality of the output space
	Kernel size	Length of the convolution window
	Strides	The stride length of the convolution
	Padding	The padding to use, one of valid, same or causal
	Dilation rate	The dilation rate to use for dilated convolution
	Activation	Activation function to use
Average Pooling	Pool size	Factor by which to downscale
Up-sampling	Size	Upsampling factor
LSTM	Units	Dimensionality of the output space
	Unit forget bias	Whether to add 1 to the bias of the forget gate
	Return sequences	Whether to return the last output in the output sequence, or the full sequence
Dense	Units	Dimensionality of the output space

Table 3.2: Summary of available hyper-parameters for each layer

When fitting the model, mini-batch sizes over 10 are recommended by Bengio, 2012 due to the computational advantage of matrix-matrix products over matrix-vector products. Given its use in (Glorot & Bengio, 2010; Liu & Yang, 2019; Supratak et al., 2014) a batch size of 10 is chosen here. A validation split of 20% is used to track the out-of-sample error. Using the validation loss, early stopping and model checkpoints can be implemented which monitor this value. Early stopping stops training when a given criteria is met while a checkpoint saves the model after each epoch if the validation loss has improved. For this research, patience is used to stop training if there has been no improvement after 5 epochs. The total number of epochs used is 100. All neural networks are implemented using Keras with a TensorFlow back-end and trained on an Nvidia GeForce GTX 1050 GPU with 4GB dedicated and 16GB shared RAM. Details of the chosen training parameters can be found in table 3.3

ML Library	API	Batch Size	Epochs	Early stopping	Validation split	Loss	Optimizer	Model checkpoint
TensorFlow	Keras	10	100	5 epochs	20%	Huber loss	Adam	Best

Table 3.3: Summary of training parameters used to fit each model

3.5 Evaluation of design

3.5.1 Signal-to-noise ratio

Both the peak signal-to-noise ratio (PSNR) and signal-to-noise ratio (SNR) were used for evaluation. The PSNR was calculated using the expression in (3.1), while the SNR was calculated using the expression in (3.2).

$$PSNR = 10 \cdot log_{10} \left(\frac{MAX_S}{N} \right)^2 \tag{3.1}$$

$$SNR = 20 \cdot log_{10} \left(\frac{S}{N} \right) \tag{3.2}$$

$$S = \sqrt{\frac{\sum (signal)^2}{len(signal)}} \tag{3.3}$$

$$N = \sqrt{\frac{\sum (noise)^2}{len(noise)}} \tag{3.4}$$

Where *signal* is the meaningful input (0ms - 500ms), MAX_S is the maximum amplitude of the meaningful input and *noise* is the unwanted baseline ($-$100ms - 0ms).

3.5.2 Hypothesis testing

Two state-of-the-art signal de-noising techniques are used to examine the effectiveness of the proposed method — PCA and ICA. In addition, a basic auto-encoder is included for comparative purposes to ensure the proposed method outperforms a simplified variant.

For PCA, it is generally considered that the main information from a signal is retained in the principal components when the cumulative explained variance is $\geq 85\%$ (Kang & Zhizeng, 2012). For the primary dataset, that is captured by the first six principal components (87.35%). For the second dataset, 87.2% is captured by the first nine principal components. Therefore, the number of components chosen in each case is 6 and 9 respectively.

ICA is discussed in Jung et al., 1998 with regard to its use for EEG artefact removal. They note that one of the assumptions of ICA is, given a set of N sensors there exists exactly N sources. However, for EEG the effective number of statistically-independent contributing signals is not known. When using ICA for artefact removal, components are usually manually omitted based on observation and visual inspection of source scalp locations (Berkovsky & Freyne, 2010; Jung et al., 2000; Scott Makeig et al., 1995). In other circumstances EEG artefacts are removed semi-automatically by identifying components that contain EEG artefacts for one subject and using correlation to identify similar components for other subjects (Campos Viola et al., 2009). In all cases, continuous EEG is used to identify artefacts because those like eye blink and heartbeat tend to exhibit an element of regularity.

For this research, EEG signals were epoched and subset to 100ms before the onset of a stimulus to 500ms after. As a consequence, the signals used are no longer continuous, and manual artefact removal techniques based on visual inspection of the ICA components may not be successful. To account for this and for evaluation purposes only, ICA is applied to the full continuous EEG before it is subset. Artefacts are removed for a single subject and correlation is used to identify similar artefacts for other subjects as per Campos Viola et al., 2009. This method of implementation is consistent with the literature and provides a meaningful baseline estimation of the SNR and PSNR for comparison.

Using test data as input, each state-of-the-art method and proposed model is used to re-construct the input signals. While the test data used for evaluation is the same in all cases, ICA is applied to the full continuous signals, PCA to each 500ms subset, and the proposed models to each 300 ms window. Before the PSNR and SNR are calculated for both the raw and re-constructed test data, all outputs are returned to the original test data shape. In total, fourteen distributions are produced — seven SNR and seven PSNR. Each resulting distribution for a given metric and proposed model is then compared individually using the Harrell-Davis quantile estimator (Harrell & Davis, 1982) to the PCA, ICA, basic auto-encoder and raw distributions for the corresponding metric.

All calculations are implemented channel-wise to reduce the influence of channel outliers. Individual decile differences are calculated and also averaged across the deciles. A positive increase for a given metric is indicted by either a positive mean difference or by five or more positive decile differences.

3.6 Summary

Three different architectures are proposed to test that, when built using convolutional and recurrent neural network layers, a stacked auto-encoder applied to EEG signals will increase the signal-to-noise ratio compared to ICA, PCA and a basic auto-encoder.

Performance and generalizability is tested using two data sources. The respective 64 and 32 channel primary and secondary datasets were captured for ERP research. The former comprises 81 subjects in each of 3 conditions for approximately 100 trials, while the latter is made up of 38 right-handed subjects in each of 4 randomized sets of 24 finger-numeral counting configurations for 10 sets of 96 trials. In both cases 100ms pre-stimulus is used as the baseline and 500 ms post-stimulus as the meaningful signal. The primary dataset is downsampled by a factor of 2 due to a sample rate difference between the recordings before both are reshaped to $(trials, timesteps, channels)$. Sampling is

applied at subject and trial level to generate train and test data. A ratio of 70:30 is used in combination with random shuffling to divide the subjects and trials respectively. In addition, windowing is implemented using a window length of 300 ms with a shift of 25 ms to augment the training data. Predicted outputs are combined before evaluation by reversing the windowing process; averaging across overlapping segments.

The PSNR and SNR, which are calculated channel-wise for the reconstructed signals produced when each proposed architecture and baseline method are applied to the test data, are used for evaluation. In addition, both are calculated for the raw signals. The Harrell-Davis quantile estimator is used to compare the resulting distributions for the proposed model to those of each baseline method and the raw signals. A positive increase for a given metric is indicated by a positive mean difference across all deciles or by 5 or more positive decile differences.

3.6.1 Strengths

- **Computational efficiency**: All models are trained using Tensorflow — a framework which utilizes the GPU-accelerated NVIDIA CUDA® Deep Neural Network library (cuDNN) [4]. The resulting computational speed-up which enables in the range of 6000 to 7000 tokens per second compared to ~400 on a standard CPU is thus approximately 14 times faster.

- **Stacking neural network architectures**: As noted in Bengio and Lecun, 2007 deep network architectures can generalize in non-local ways and model complex relationships between variables for the purpose of AI. In addition, combining both recurrent layers, which have been noted as being very powerful for modelling sequences (Bengio et al., 2013), and convolutional layers, which can utilize local perception to extract signal features (Wen & Zhang, 2018), should improve the overall performance.

[4] https://developer.nvidia.com/cudnn

- **Limited domain knowledge required**: Unlike ICA which requires manual intervention to remove noise related components and thus specific domain knowledge and expertise to identify EEG artefacts, the proposed CR-AE is an unsupervised method that can be implemented without prior domain knowledge to reduce noise in EEG signals.

- **Second dataset to test generalizability**: With the addition of a second dataset, a more robust estimation of the generalizability of the proposed models to unseen data can be given.

- **Robust statistical method for distribution comparison**: The Harrell-Davis quantile estimator is a weighted average of all the order statistics and provides a robust decile based statistical method to calculate the difference between deciles of two groups.

3.6.2 Limitations

- **No baseline EEG signals**: The SNR is usually calculated between a clean signal and its noisy equivalent. However, it is not possible to record perfectly clean EEG signals. Therefore, a 100 ms baseline is used to represent the noisy input, under the assumption that, before a stimulus is presented, the recorded signal represents noise and not neural activity. This means it is not possible to accurately determine whether noise is actually removed from the reconstructed signals. An increase in SNR or PSNR does not necessarily indicate that noise has been removed, important signal information could also have been removed during the process. In some cases synthetic EEG data is created and noise is artificially added which allows the researcher to quantify what information is removed through the process (Ahmadi & Quian Quiroga, 2013).

- **Unquantifiable information loss**: As a consequence of the above there is no way to determine how much information is lost through the reconstruction process of the proposed convolutional recurrent auto-encoder. This is unlike ICA

and PCA where information loss can be quantified by the components removed during analysis and in PCA specifically by the total explained variance of the components removed.

- **Limited hyper-parameter tuning**: The cuDNN backed implementation of LSTM is optimized only for use with hyperbolic tangent activation, sigmoid recurrent activation and zero recurrent dropout, therefore the available hyper-parameters are limited and consequently the potential for tuning.

- **Non-use of spatial information**: Though the scalp position of electrodes is known, this information is not utilized by the proposed method. A 2D convolutional layer applied to temporal spatial projections of the electrode placement in 2D space could make use of this information.

In the next chapter, implementation of the experimental process and results are discussed along with evaluation of same.

Chapter 4

Results, evaluation and discussion

The presence of noise in EEG is a significant problem as it interferes with the capture of other signals. The identification and removal of noise is a key area of research as it improves the quality of EEG recordings and could reduce the number of trials required for ERP analysis. This research is focussed on improvement of the PSNR and SNR of EEG signals in comparison to the state-of-the-art methods — ICA and PCA. A basic auto-encoder is also included as a baseline to measure the improvement of the proposed CR-AE over a simple auto-encoder.

In this chapter, the experimental results are discussed and evaluated with respect to those obtained from each of the baseline methods and in relation to previously conducted research. Limitations of the study and findings are highlighted and a summary of the results is presented.

4.1 Results

The results for each proposed architecture are discussed in relation to reconstruction error, SNR and PSNR. Results for the primary dataset were generated using test data as input across each condition for both trial and subject wise sampling. In contrast, results for the secondary dataset were generated using all data points and used to assess model generalizability.

4.1.1 Signal reconstruction

The purpose of these auto-encoders is to reconstruct an input signal from a latent space representation and in doing so reduce the level of noise present. As such, an important aspect of this research is the accuracy to which the signals are reconstructed. During training, Huber loss was used to measure this. As mentioned previously, Huber loss utilizes a combination of the mean squared error and mean absolute error depending on a given threshold. In this way it is robust to outliers but still provides a balanced prediction error across all instances.

Primary dataset

The mean reconstruction error for architectures one, two and three is 1.32, 1.65 and 1.61 respectively. From the condition and sampling method point of view, mean reconstruction error is lowest for condition 3 (1.46, sd=0.21) and trial wise sampling (1.40, sd=0.17). Architecture 1 has the lowest reconstruction error, however, the overall mean of 1.52 (sd=0.21) indicates the performance does not vary greatly.

Lower error rates associated with condition 3 could be due to the fact that, in that task, each subject pressed a button but no tone was generated. From the analysis conducted by Ford et al., 2014, in healthy controls this condition was associated with a smaller response than pressing a button to deliver a tone. Therefore, condition 3 may contain fewer outliers and have lower variance than other conditions which would influence reconstruction error.

Visual inspection of the reconstructed signals shows, that for each architecture, the output closely resembles the original input. This can be seen in figures 4.1, 4.2, and 4.3 below which highlight examples of reconstructed signals from condition 1 for each architecture overlaid with the original input.

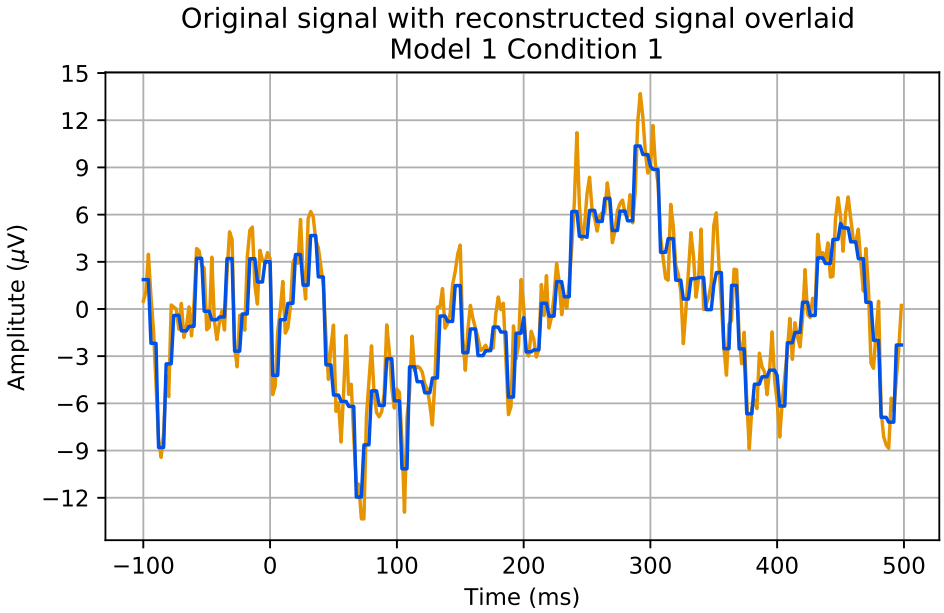


Figure 4.1: Original signal overlaid with the corresponding reconstructed signal for architecture one (primary dataset)

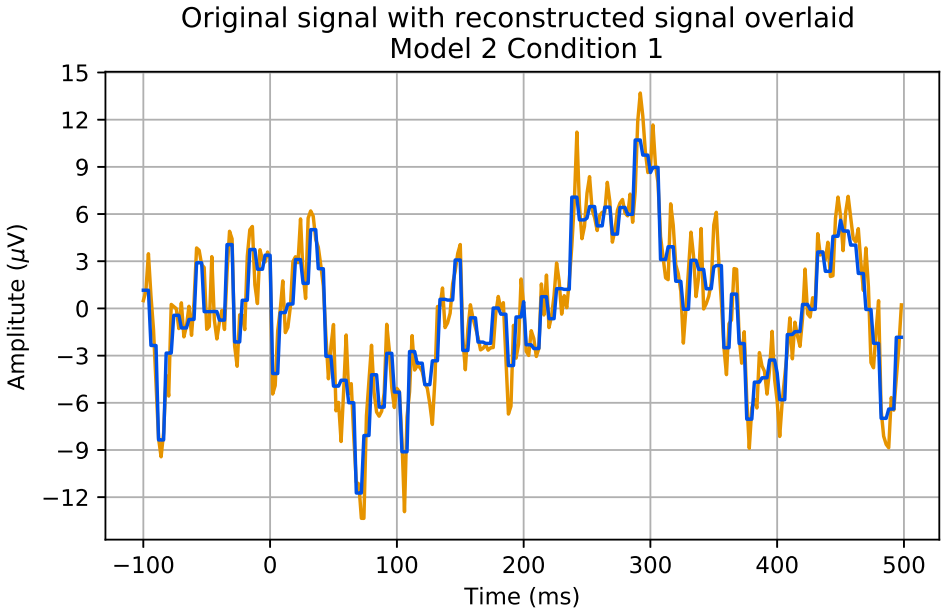


Figure 4.2: Original signal overlaid with the corresponding reconstructed signal for architecture two (primary dataset)

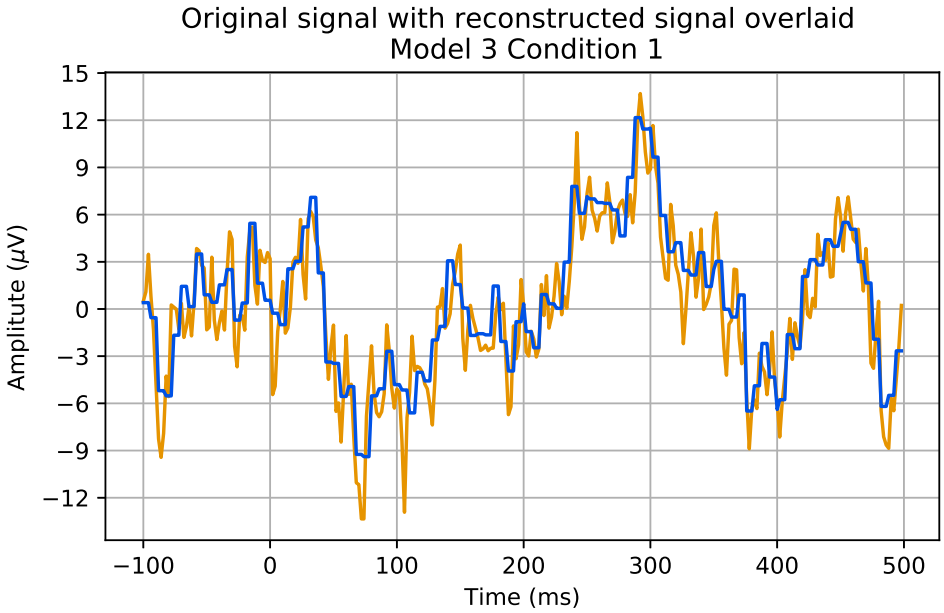


Figure 4.3: Original signal overlaid with the corresponding reconstructed signal for architecture three (primary dataset)

Secondary dataset

The purpose of including a second dataset, is to determine how well the proposed CR-AE can generalize to other datasets without needing to be retrained. If the model can perform well on another dataset it shows that there is the potential it could be used as a noise reduction technique in other scenarios. Mean reconstruction error for each architecture shows that architecture one (0.14) performs better than architecture two (1.23) and three (0.38) though there is little between architecture one and three. These results are in fact also better than the reconstruction error on the primary dataset. However, visual inspection of reconstructed signals across the three architectures shows that none are able to accurately reproduce the input signal. Figures 4.4, 4.5, 4.6, and 4.7 highlight this across each architecture.

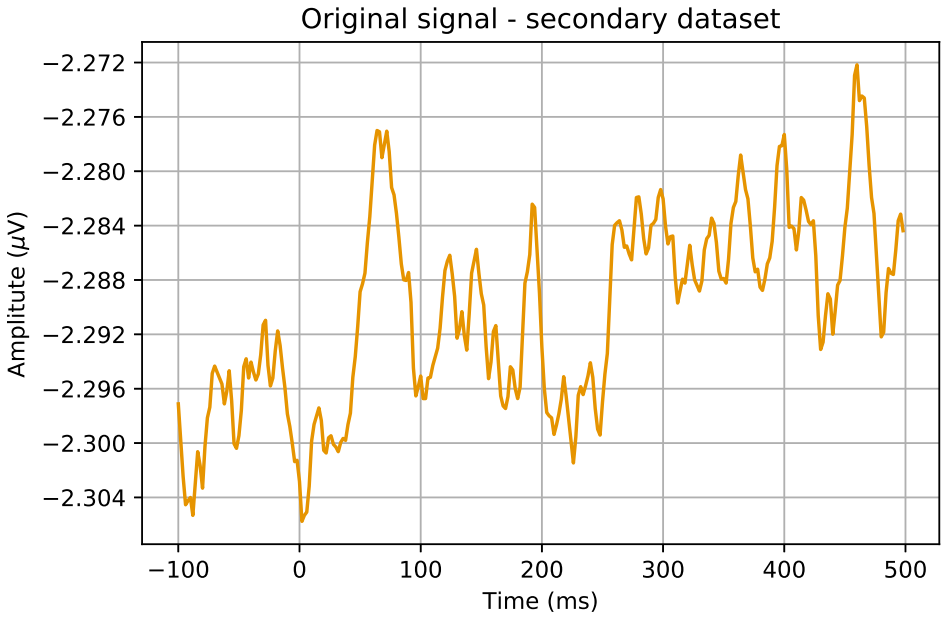


Figure 4.4: Original signal used as input for each of the reconstructions below

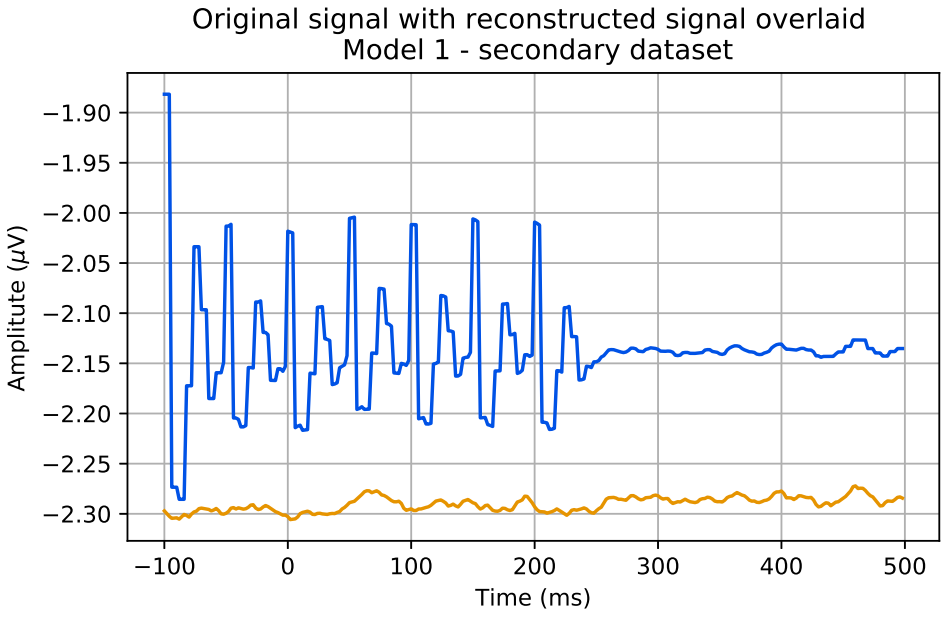


Figure 4.5: Original signal overlaid with the corresponding reconstructed signal for architecture one (secondary dataset)

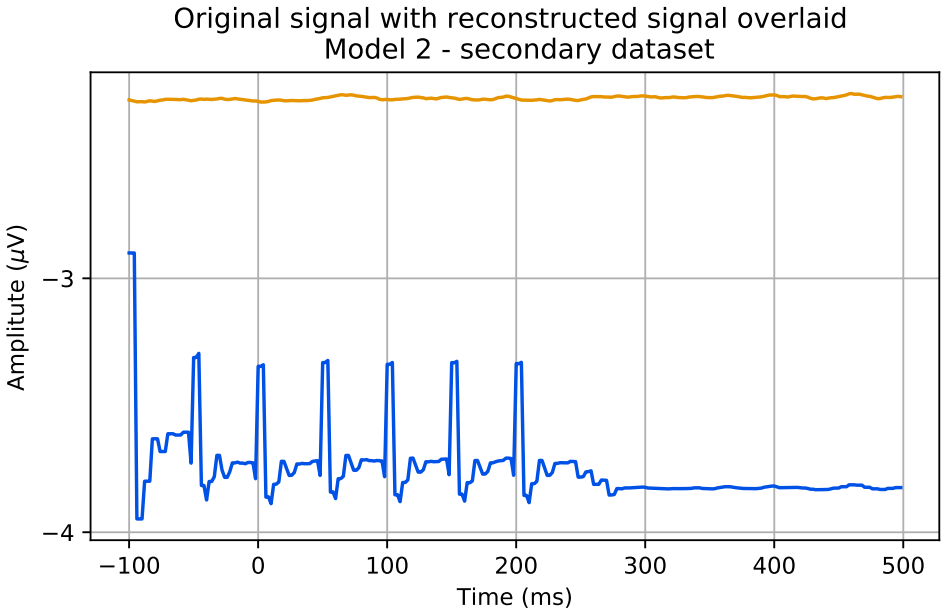


Figure 4.6: Original signal overlaid with the corresponding reconstructed signal for architecture two (secondary dataset)

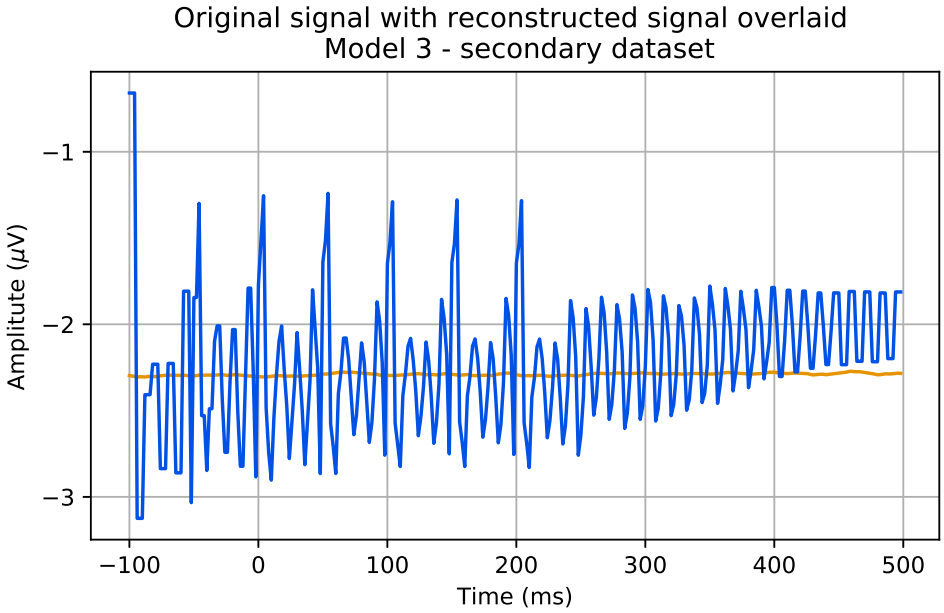


Figure 4.7: Original signal overlaid with the corresponding reconstructed signal for architecture three (secondary dataset)

It can be clearly seen that there are significant issues with each reconstructed signal. One of those is partly caused by the method employed to recombine the windowed signals and partly due to inaccurate reconstruction for early samples. As mentioned previously each 600 ms input is windowed using a 300 ms window length with a 50 ms stride, to augment the training data. This results in seven inputs representing each single input from the raw data. Each consecutive set of seven inputs is then recombined post prediction into the original 600 ms signal.

This is achieved by averaging the predicted values in each overlapping window. This method of combining the windowed segments would in most cases be fine, however, in this case all three architectures fail to correctly predict early samples; an issue which is then exaggerated across much of the reconstructed signal due to how the outputs are combined. Reconstruction error for early samples is likely due to a lack of historic information. Each model was trained on data from the primary dataset which means the weights are adjusted during training to predict values from that dataset. The convolutional layers in this model use causal padding which means they only take into account previous time steps. The weights associated with early time-steps would have been updated during training using only initial values from all instances and therefore are optimized to predict those. Failure to correctly predict early values in the secondary dataset could be due to a difference in the range of values. Certainly the variance is much less which could also influence predictions. This reconstruction error highlights the fact that the proposed architectures do not generalize well to other datasets, particularly for early samples. Using padding that can violate temporal order could solve this issue.

For architecture one and two, later samples are more accurately predicted, preserving at least signal shape despite not being in the same range. This can be seen in figure 4.5 where, after 250 ms, the signal resolves and the shape begins to resemble that of the input. Since the stride length is 25 ms and there are seven windows in total, at 175 ms the last window is combined with that of all previous windows to create the final

output. Shortly after that point, the signals resolve. Since early samples from each prior window are poorly predicted, when these are combined the error is repeated throughout the reconstruction. Figures 4.8, 4.9, 4.10, and 4.11 show reconstructed signals from consecutive windows before and after being combined. It highlights the issue with poor prediction and the method used to recombine the windows. Another way to combine these would be to only use early samples from the first window and use later samples from all others in place of the early samples of subsequent windows. This would solve the problem, however, if the models could generalize it would not have to be solved and therefore should not be implemented.

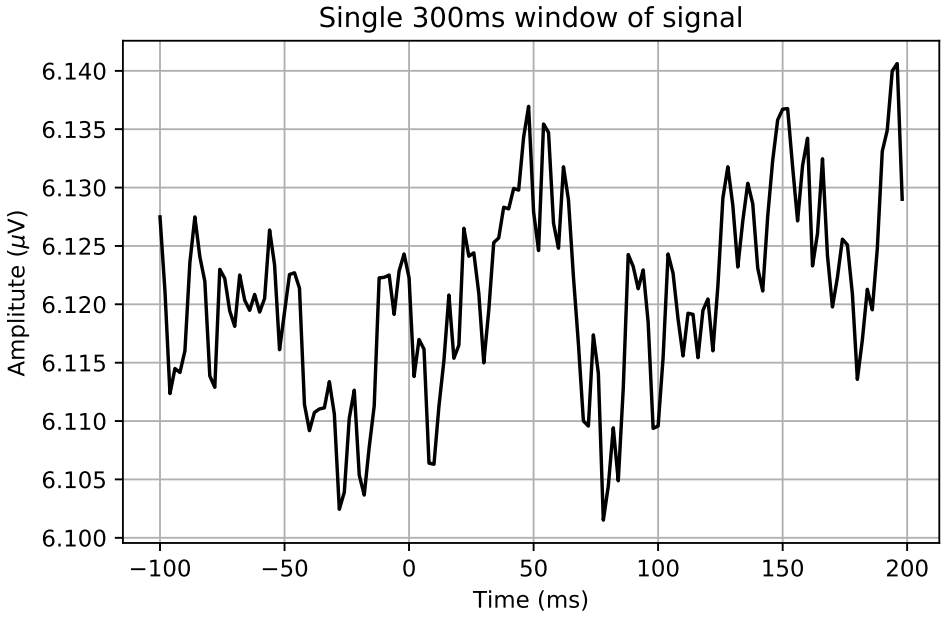


Figure 4.8: Single 300ms window of a signal used for the reconstruction below

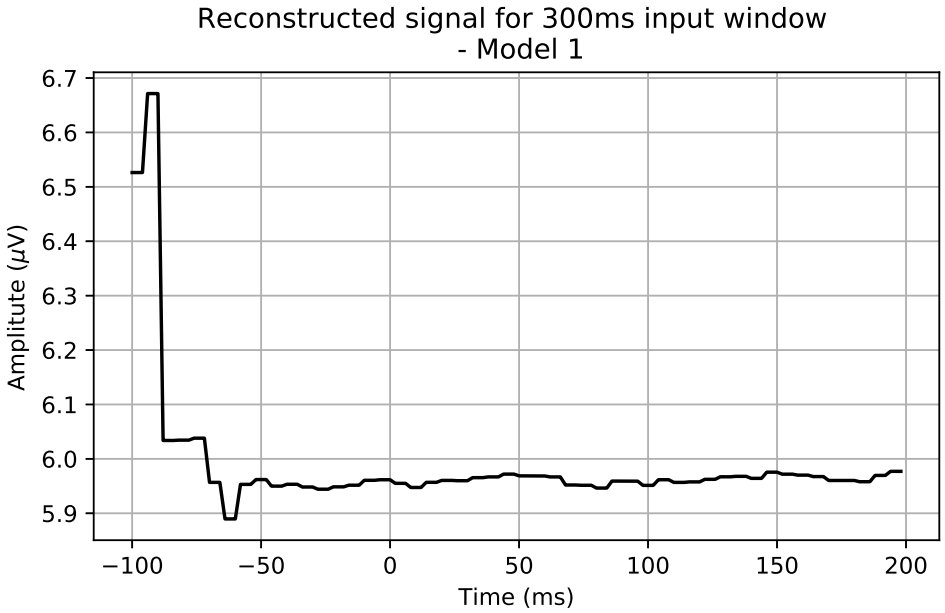


Figure 4.9: Output from architecture one for the 300ms input window above

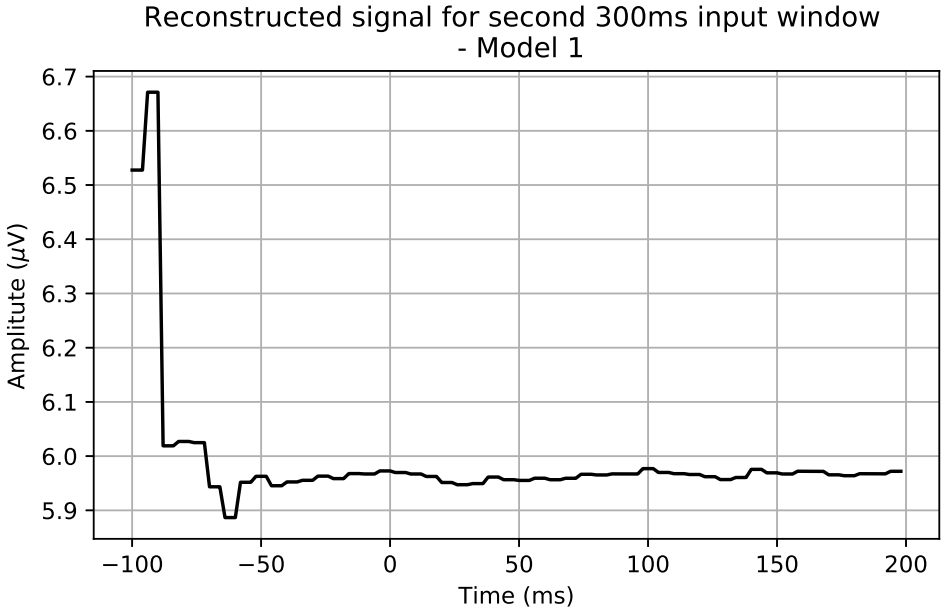


Figure 4.10: Output from architecture one for the next 300ms window of the same signal

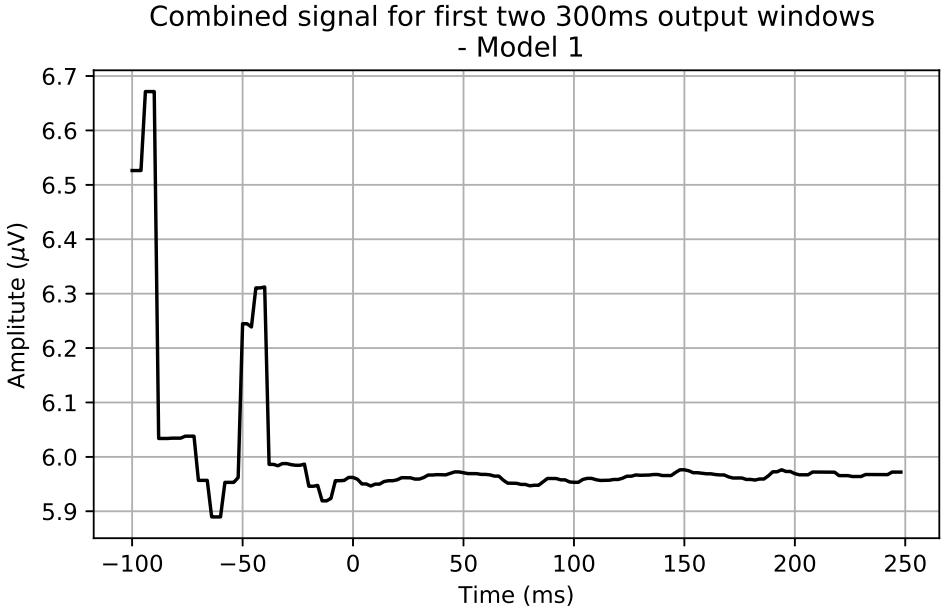


Figure 4.11: Combined outputs of the first and second windows with overlaps averaged

Given that the reconstructed signals are not accurate representations of the input, reconstruction accuracy is not taken into consideration for architecture selection or model evaluation. Additionally, SNR and PSNR are not reported for the second dataset as they do not provide any meaningful information for solution evaluation. These findings show that the proposed method does not generalize to all other EEG datasets.

4.1.2 Signal-to-noise ratio

To compare the level of signal to the level of noise in each reconstructed input, the SNR is used. Additionally, PSNR is used to compare the maximum possible power of a signal to the level of noise affecting it. Each of these are measured on the decibel scale. Since an uncorrupted EEG signal is not possible to record, the 100 ms baseline is used to calculate the level of noise in each signal. The purpose of this research is to increase these ratios for reconstructed signals compared to each baseline and the raw signals. Both are calculated channel wise to reduce the influence of outliers and provide more detailed results.

Architecture three achieved the highest mean SNR across all conditions and sampling methods with 5.53 db (sd=4.49) compared to 5.33 db (sd=4.36) and 5.30 db (sd=4.36) for architectures one and two respectively. Similarly, it achieved the highest PSNR with −24.44 db (sd=6.67) compared to −24.46 db (sd=6.46) and −24.48 db (sd=6.47) though the difference between each architecture in both cases is minimal. This result could highlight the benefit of using dilated convolutions in sequential convolutional layers to increase the receptive field of the network. By doing so, the model learns coarser representations of its input which would result in less detailed signals and by consequence less noise. As can be seen from figure 4.3, this results in the loss of signal information, though importantly, critical information appears to be retained. However, as a result of not having clean EEG signals the information loss cannot be quantified.

Condition 1 was consistently highest in terms of both metrics with mean values of 5.38 (sd=4.31) and −23.81 db (sd=6.24) for SNR and PSNR respectively across the three proposed architectures. In comparison, condition 2 and 3 achieved very similar results for both, with mean SNR values of 5.01 db (sd=4.13) and 4.99 db (sd=4.23) and mean PSNR values of −24.33 db (sd=6.06) and −24.24 db (sd=6.17) respectively. As mentioned before, this could be due to the nature of condition 1, wherein subjects pressed a button to deliver a tone. As greater responses was observed for that condition, higher SNR and PSNR values would be expected. Indeed, for the raw signals, both metrics were higher for condition 1. This indicates the values are not necessarily the result of improved noise reduction for that condition, though this will be explored later to quantify the difference.

The sampling method chosen did not have a significant influence on the results with mean SNR of 5.09 db (sd=0.008) and 5.16 db (sd=0.008), and mean PSNR of −24.22 db (sd=0.011) and −24.03 db (sd=0.011), for trial and subject sampling respectively. This result shows that the models can be trained on a subset of subjects and perform

equally well for unseen subjects. The benefit of this is that the model does not need to be trained with data for a given subject before being able to accurately reconstruct signals with performance equivalent to a model trained on a portion of the subjects' data.

Certain channels reached higher mean SNR than others. Specifically, for electrode TP10, each proposed architecture reached above 6 db at a maximum of 6.93 db (sd=5.07) for architecture three. In fact, mean SNR reached over 6 db for five electrodes for that architecture, those being, TP10, F7, Fp1, F8 and Fp2; in that order from highest to lowest. Furthermore, the highest mean SNR values were observed for each of those electrodes across all proposed architectures. Given that Fp1 and Fp2 are located above the left and right eye respectively, they are usually associated with strong eye activity (Vigário, 1997) and therefore could contain more noise than other signals. Additionally, both F7 and F8 are located in proximity to Fp1 and Fp2 which could also indicate the presence of noise in those signals. Since the research conducted by Ford et al., 2014 involved an auditory experiment, activity around electrode TP10, located close to the ear, may have been more prevalent and hence noisier. Early auditory activity would likely have been present in these signals and given the focus of ERP analysis on electrodes Fz, FCz and Cz, the overall indication is that TP10 did not contain relevant ERP information. These findings would suggest that noise reduction is most effective when signals are noisier, rather than for those that contain meaningful ERP information. Once again, this will be explored later in comparison to the raw signals and other baseline methods.

4.2 Evaluation

SNR and PSNR distributions are compared using the Harrell-Davis quantile estimator. This robust statistical method provides better estimation of the difference between two distributions than using a standard comparative test. As a weighted sum of sorted values it is used in place of point estimates like mean or median to show how much one group must be shifted to be comparable to another group at each quantile (Rous-

selet, 2010; G. A. Rousselet et al., 2009). The purpose of this research is to determine whether a statistically significant difference exists between the SNR and PSNR produced by applying the proposed CR-AE to EEG signals, compared to those produced by ICA, PCA and a basic auto-encoder. To achieve this, HD quantiles are calculated for the reconstructed signals and the raw signals and their difference computed. For each method the differences are directly compared to determine whether there is a positive shift.

4.2.1 Proposed CR-AE architecture

To choose the best CR-AE architecture, each model was evaluated, by condition and sampling method, individually to the raw signals at both an overall and channel level. The chosen architecture was then used to establish whether the null hypothesis could be rejected.

For condition 1, architecture three had the highest SNR HD quantile increase for both trial (mean=0.640) and subject (mean=0.633) sampling compared to both architecture one (trial mean=0.426, subject mean=0.447) and two (trial mean=0.406, subject mean=0.434). However, for trial sampling channel CP1 showed a decrease of (mean=-0.078) with only 3 positive quantile differences. For PSNR, all three proposed CR-AE performed worse than the raw signals; Each showing a mean HD quantile decrease across all channels. Furthermore, none had any channel showing more than 5 positive quantile differences. Architecture three, had the lowest decrease for trial sampling (mean=-0.457) but performed worst for subject sampling (mean=-0.640). Despite this, architecture three had the most positive decile differences overall in both cases, including 4 positive quantiles for channels Fp1, Fp2, F7 and TP10.

Similar results were found for condition 2 and 3 with respect to both metrics. However, in both cases architecture three performed better than one and two in terms of PSNR for both trial and subject sampling. Despite this, HD quantiles remained negative in all cases. Though all three architectures performed worse than the raw signals, architec-

ture three had the lowest decrease for both trial (condition 2 mean=−0.473, condition 3 mean=−0.549) and subject (condition 2 mean=−0.583, condition 3 mean=−0.604) sampling for both conditions. As regards SNR, architecture three performed best in condition 2 for both trial (mean=0.651) and subject (mean=0.602) sampling and similarly for condition 3 in respect of same (trial mean=0.626, subject mean=0.612).

Given its performance across all conditions and sampling methods, architecture three is chosen as the best CR-AE. The technical model can be seen in figure A.1. As highlighted above, it outperformed both alternatives for SNR in all cases and PSNR in most. It was only outperformed once for PSNR in condition 1 using subject sampling. Overall, PSNR performance was poor for all architectures which could be due to the use of average pooling. As can be seen in figure 4.3, the reconstructed signals do not accentuate the peaks or troughs of the original signal. Therefore, the max amplitude is reduced by the CR-AE. As can be seen by expression 3.1, if the max amplitude of the meaningful signal is reduced, and the mean square amplitude of the noise is not reduced by an equivalent amount, the PSNR will decrease. Utilizing max pooling to extract the most prevalent features, instead of average pooling, could improve this metric, though it could also increase the level of noise and thus lower the PSNR.

4.2.2 Principal components analysis

Six components were used for PCA with retained variance of 87.35% to reconstruct the EEG signals. The mean SNR of 4.88 db (sd=4.04) was lower than the proposed CR-AE (mean=5.53 db, sd=4.49), however, at −23.78 db (sd=5.76) the mean PSNR was higher. Huber loss calculated for the reconstructed signals shows that the proposed CR-AE (1.61) produced more accurate results than PCA (2.46). Figures 4.12 and 4.13 show an example of the PCA reconstructed signals compared to the raw signals and those of the proposed CR-AE compared to same. It can clearly be seen that the CR-AE produces a smoother signal that better models the general shape of the original input in comparison to PCA.

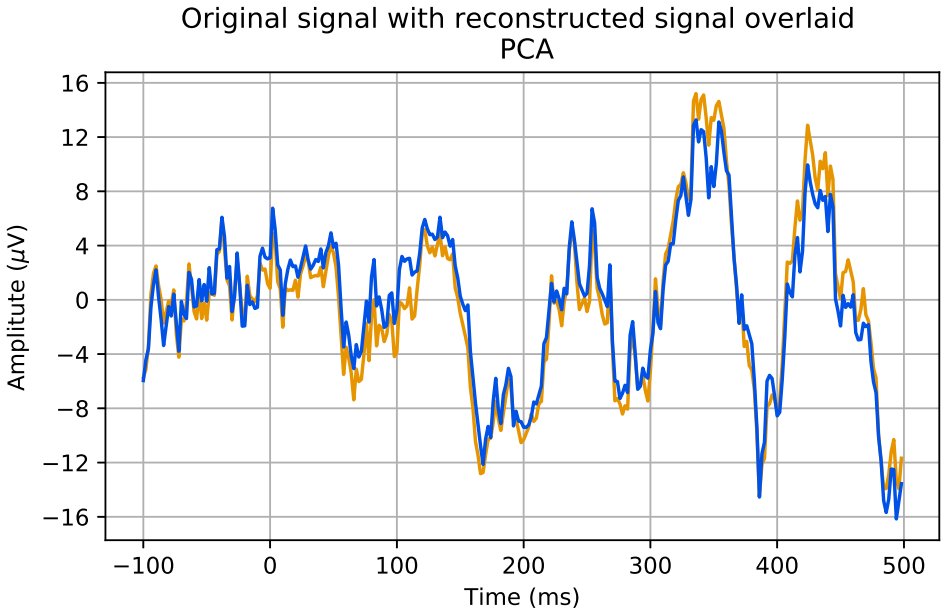


Figure 4.12: Original signal overlaid with the corresponding reconstructed signal for PCA

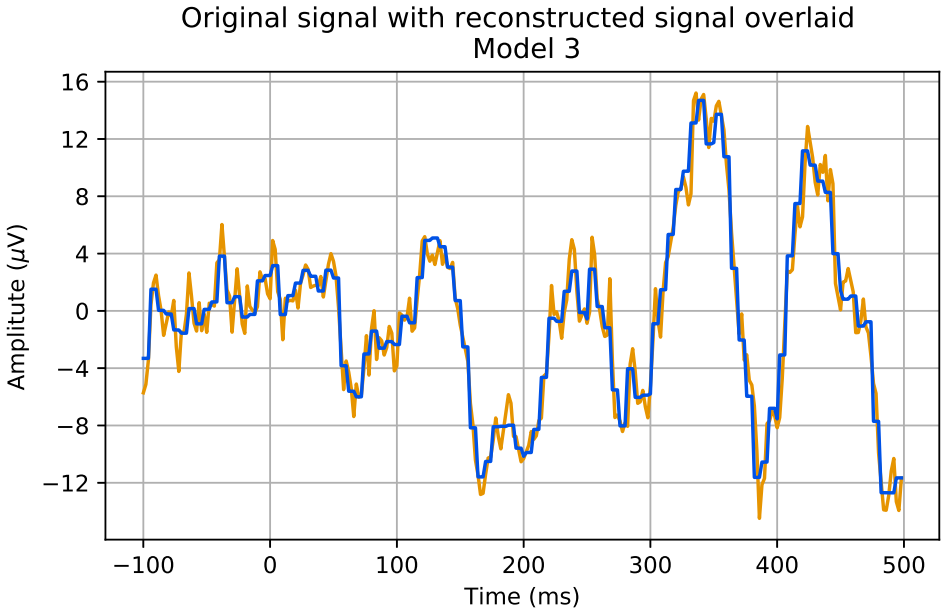


Figure 4.13: Original signal overlaid with the corresponding reconstructed signal for the CR-AE

For comparative purposes, both the SNR and PSNR of the PCA and CR-AE reconstructed signals were compared to those of the raw signals with each set of results then being compared to determine whether there was an improvement.

Results of the experiment, averaged across all channels, for condition 1 can be seen in tables 4.1 and 4.2. These results show that, for condition 1, PCA performs approximately equivalently to the raw signals across all quantiles for both SNR and PSNR. In each case, the overall mean difference is positive but only marginally with maximum shift of 0.108 and 0.121 in the 8^{th} and 9^{th} quantiles for PSNR with subject sampling. In contrast, for SNR, the proposed CR-AE shows significant shift across the majority of quantiles except for quantile 1, where the increase is marginal. For each consecutive quantile the shift gets gradually bigger, indicating that shift is more to the right of the distribution. However, performance for PSNR was greatly reduced across the majority of quantiles, with notable increases in only the 9^{th} quantile for both sampling methods. Since the improvement for PCA is marginal the overall difference remains similar to that of the proposed CR-AE.

From a channel perspective, PCA showed positive shift across the majority of quantiles for electrodes FC1, F3, F4, F8, P8, Pz and T7, while the proposed CR-AE showed positive shift across all quantiles for every electrode except CP1. Figures A.5 and A.2 highlight the HD decile differences for both PCA and the proposed CR-AE in condition 1 across all channels when compared to the original signals. The original research on this dataset analysed ERP's from reference nodes, FCz, Fz and Cz, for which information from FC1, F3, F4, F7 and F8 would be important. Therefore, it is essential to improve SNR on these channels. This has been achieved by both methods however, the positive shift is greater for the proposed CR-AE.

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.003	0.0830	0.0862	-0.0174	-1.4540	-1.4366
	2	-0.004	0.2388	0.2426	0.0160	-0.9659	-0.9819
	3	0.005	0.3834	0.3786	0.0197	-0.7547	-0.7744
	4	-0.001	0.5139	0.5151	0.0435	-0.5827	-0.6262
Trial	5	-0.025	0.6233	0.6482	0.0370	-0.4364	-0.4733
	6	-0.007	0.7420	0.7486	0.0480	-0.2675	-0.3155
	7	0.012	0.8926	0.8805	0.0442	-0.1125	-0.1566
	8	0.023	1.0418	1.0187	0.0414	0.0822	0.0408
	9	0.000	1.2444	1.2446	0.0436	0.3804	0.3367
	Mean	0.000	0.6403	0.6403	0.0307	-0.4568	-0.4874

Table 4.1: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 1 trial sampling

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.065	0.068	0.133	0.004	-1.781	-1.785
	2	-0.004	0.248	0.253	-0.011	-1.224	-1.213
	3	0.024	0.400	0.376	0.034	-0.954	-0.987
	4	0.024	0.516	0.492	0.028	-0.756	-0.785
Subject	5	0.022	0.624	0.603	0.051	-0.564	-0.614
	6	0.014	0.749	0.735	0.064	-0.399	-0.463
	7	0.051	0.871	0.820	0.063	-0.230	-0.293
	8	0.046	1.031	0.985	0.108	-0.024	-0.132
	9	0.026	1.197	1.171	0.121	0.170	0.048
	Mean	0.015	0.634	0.619	0.051	-0.640	-0.692

Table 4.2: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 1 subject sampling

Results for condition 2 and 3, which can be seen in tables A.5, A.6, A.7 and A.8, are very similar to those of condition 1. In each case, the proposed CR-AE outperforms PCA across all quantiles for SNR, but as expected, heavily underperforms in terms of PSNR with meaningful positive shift observed only in the 9^{th} quantile for both trial and subject sampling. As with condition 1, the proposed CR-AE showed minimal positive shift for SNR on one electrode (F7) across all quantiles in condition 2, though not for the same electrode. In contrast, condition 3 showed positive shift across all quantiles for every electrode. Though SNR performance improved from condition 1 to 3, PSNR declined with significant negative shift across quantile 2 and 3, indicating that as SNR performance improves, PSNR performance declines.

4.2.3 Independent components analysis

ICA was implemented in a semi-supervised manner using all components and PCA pre-whitening. Components containing artefacts were identified for one subject and correlation was used to find similar components for other subjects. In all cases the component was set to zero before the signal was reconstructed to remove that component from the original input. A correlation coefficient equal to or above 0.7 was used to identify other components. As ICA is usually applied to continuous EEG, ICA was implemented before the signals were epoched and subset. As this is a semi-supervised method, domain knowledge of EEG artefacts is required for identification. Due to a lack of expertise, the focus was to remove ocular artefacts which are easier to identify. These usually appear as large, sometimes periodic spikes with scalp activity focused around electrodes Fp1 and Fp2. The mixing matrix was used to determine how the components mapped to the electrodes. This was used in combination with visual inspection of component plots to identify potential ocular artefacts.

Overall, SNR (mean=4.75 db, sd=3.88) was lower for ICA than that of the proposed CR-AE. However, PSNR (mean=$-$23.95 db, sd=5.62) was higher. This is consistent with earlier findings for the CR-AE due to poor PSNR performance. Huber loss calculated for the reconstructed signals shows that ICA (0.35) performed better than the

proposed CR-AE (1.61). However, this is skewed by the fact that in some cases no components are removed which means the original input is reconstructed perfectly. This can be seen in figure 4.14 where the original input is completely obscured by the reconstruction. In this case the Huber loss is zero. This occurs when none of the components are correlated high enough with the example component.

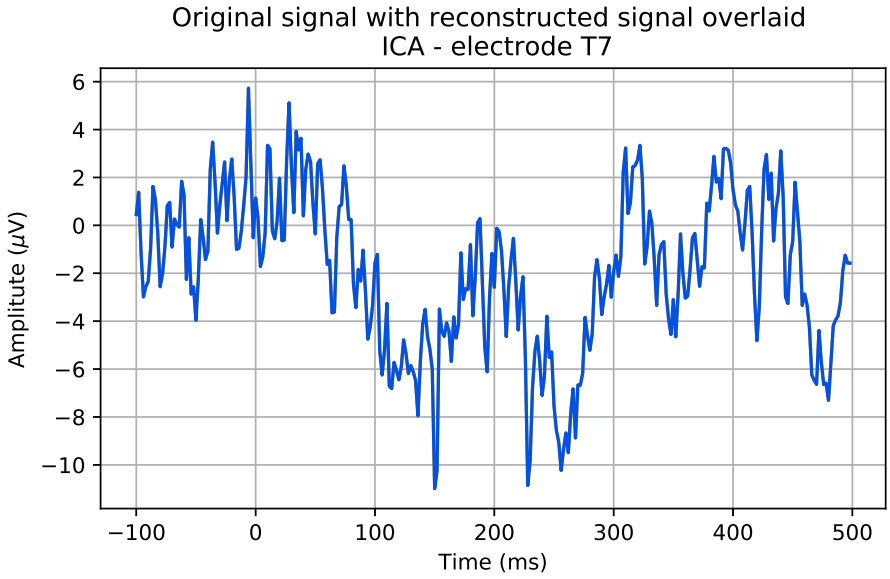


Figure 4.14: Original signal overlaid with the corresponding reconstructed signal for ICA — electrode T7

Results averaged across all channels, for condition 1 can be seen in tables 4.3 and 4.4 with all other results available in A.9, A.10, A.11, and A.12. These show that ICA performed worse than the raw signals across all quantiles in almost every case — except for condition 1 using subject sampling where it performed approximately equal which is likely due to its implementation. It is quite clear from the channel breakdowns for conditions 1, 2 and 3 (figures A.8, A.9 and A.10 respectively) that for both metrics, the majority of negative shift is focused on electrodes Fp1 and Fp2. However, nearby electrodes have also been heavily impacted in most cases. This would indicate that important information has been removed during the process impacting SNR and PSNR. This can be seen in 4.15 which shows a reconstructed

signal for electrode Fp1. Most peaks and troughs are removed leaving a relatively flat signal. This is consistent for most signals of affected electrodes. Though this may be appropriate in some circumstances, it appears too much of the variation has been removed. This may be the result of setting entire components to zero and could be modified by only setting localized spikes in the component to zero instead.

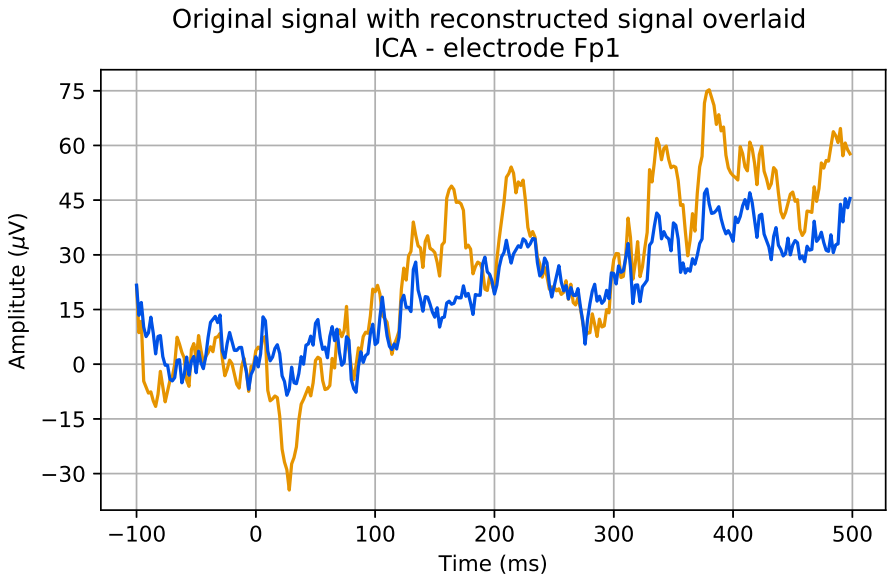


Figure 4.15: Original signal overlaid with the corresponding reconstructed signal for ICA — electrode Fp1

As a consequence, the results for ICA are similar to PCA. The proposed CR-AE outperformed ICA in terms of SNR across all conditions and sampling methods for all quantiles. In addition, channel performance shows significant positive shift across all quantiles between the CR-AE and ICA. Though SNR performance improved, once again the proposed CR-AE was outperformed in terms of PSNR across the majority of quantiles. All but the 8^{th} and 9^{th} quantiles showed positive shift. Given the potential issues with its implementation, it is very hard to consider these results as significant in the context. In contrast to PCA, ICA requires domain knowledge. As a result, it would be better to compare the proposed CR-AE to ICA using a dataset where the results are available and on which ICA has already been implemented by an expert.

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.088	0.083	0.171	-0.154	-1.454	-1.300
	2	-0.125	0.239	0.364	-0.160	-0.966	-0.806
	3	-0.148	0.383	0.532	-0.157	-0.755	-0.598
	4	-0.171	0.514	0.685	-0.192	-0.583	-0.390
Trial	5	-0.181	0.623	0.804	-0.228	-0.436	-0.209
	6	-0.195	0.742	0.937	-0.258	-0.267	-0.010
	7	-0.200	0.893	1.093	-0.283	-0.112	0.171
	8	-0.201	1.042	1.243	-0.332	0.082	0.414
	9	-0.198	1.244	1.442	-0.443	0.380	0.824
	Mean	-0.168	0.640	0.808	-0.245	-0.457	-0.212

Table 4.3: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 1 trial sampling

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.006	0.068	0.073	-0.017	-1.781	-1.764
	2	0.002	0.248	0.246	-0.022	-1.224	-1.202
	3	0.000	0.400	0.400	-0.011	-0.954	-0.943
	4	-0.005	0.516	0.521	-0.009	-0.756	-0.747
Subject	5	-0.009	0.624	0.633	-0.013	-0.564	-0.550
	6	-0.006	0.749	0.755	-0.015	-0.399	-0.384
	7	-0.001	0.871	0.873	-0.027	-0.230	-0.203
	8	0.005	1.031	1.026	-0.031	-0.024	0.006
	9	0.025	1.197	1.172	-0.019	0.170	0.188
	Mean	0.001	0.634	0.633	-0.018	-0.640	-0.622

Table 4.4: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 1 subject sampling

4.2.4 Basic auto-encoder

A basic auto-encoder was included as a simple baseline to ensure the proposed CR-AE could outperform the non convolutional-recurrent alternative. From an SNR perspective, the basic auto-encoder (mean=4.84 db, sd=4.02) underperformed compared to the proposed CR-AE (mean=5.53 db, sd=4.49), PCA (mean=4.88 db, sd=4.04) and the raw signals (mean=4.89 db, sd=4.02). However, for PSNR (mean=−23.80 db, sd=5.73), it outperformed the proposed CR-AE (mean=−24.44 db, sd=6.67) and performed on par with the raw signals (mean=−23.80 db, sd=5.75), with only PCA (mean=−23.78 db, sd=5.76) showing improved performance over it.

Huber loss calculated for the reconstructed signals shows that the proposed CR-AE (1.61) once again produces more accurate results than the baseline method (2.30). Figures 4.16 and 4.17 show an example of an AE reconstructed signal overlaid with the original signal alongside the equivalent for the proposed CR-AE. As mentioned previously, the CR-AE accurately models the shape of the input signal producing a smoother representation compared to the AE which only accurately reconstructs the initial 200 ms but begins to diverge thereafter. It can be seen that like PCA the AE accentuates the peaks far more than the CR-AE which is likely the reason for its performance with regard to PSNR.

As before, results are evaluated with respect to the performance of each method when compared to the raw signals as determined by the HD quantiles. These are then directly compared to evaluate performance. Results for the experiment can be seen in tables 4.5 and 4.6 for condition 1, and tables A.13, A.14, A.15, and A.16 for condition 2 and 3 respectively. As highlighted by its mean SNR, the AE showed negative shift across all quantiles for each condition in both trial and subject sampling, except for condition 2 with subject sampling where two quantiles showed marginal positive shift. As with PCA, the CR-AE outperformed the AE for SNR across all conditions and sampling methods with mean positive shift difference of approximately 0.674.

In terms of PSNR, despite parity between the mean for the AE and the raw signals, the AE showed negative mean shift for all conditions except condition 1. However, as before, since the CR-AE showed significant negative shift in all conditions across every quantile but the 9^{th}, the mean shift difference of -0.545 further reinforces the trade-off between increased SNR at the expense of PSNR.

Heat-maps illustrating channel performance, show that the majority of the negative shift was confined to electrode TP10 (A.11a, A.12a, A.13a). In comparison, the CR-AE showed significant positive shift for that channel. Across all conditions, for SNR, the majority of the positive shift was seen across electrodes C3, C4, CP1, CP2, though for condition 1 this shift was minimal. Differences across the quantiles for PSNR was much more sporadic, except for condition 1 which showed consistent improvement for all central and central parietal electrodes. Once again, the majority of the negative shift was isolated to electrode TP10.

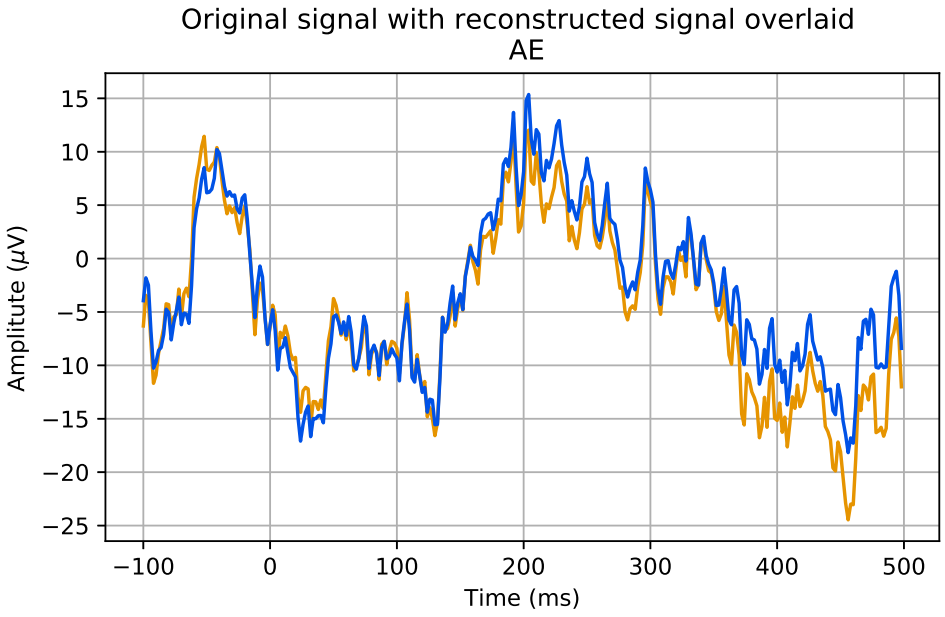


Figure 4.16: Original signal overlaid with the corresponding reconstructed signal for basic AE

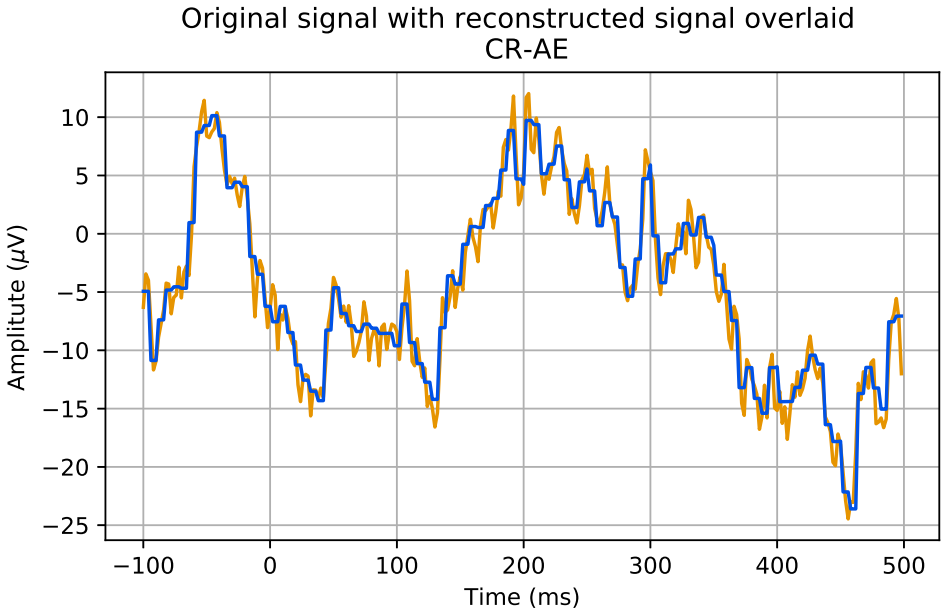


Figure 4.17: Original signal overlaid with the corresponding reconstructed signal for CR-AE

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
Trial	1	-0.052	0.083	0.135	0.019	-1.454	-1.473
	2	-0.028	0.239	0.267	0.001	-0.966	-0.967
	3	-0.026	0.383	0.410	0.024	-0.755	-0.778
	4	-0.032	0.514	0.545	0.036	-0.583	-0.618
	5	-0.052	0.623	0.676	0.017	-0.436	-0.453
	6	-0.088	0.742	0.830	0.003	-0.267	-0.270
	7	-0.062	0.893	0.954	-0.010	-0.112	-0.103
	8	-0.069	1.042	1.111	-0.027	0.082	0.109
	9	-0.112	1.244	1.357	0.005	0.380	0.376
	Mean	-0.058	0.640	0.698	0.007	-0.457	-0.464

Table 4.5: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 1 trial sampling

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
	1	-0.085	0.068	0.153	0.047	-1.781	-1.828
	2	-0.040	0.248	0.289	0.013	-1.224	-1.236
	3	-0.008	0.400	0.408	0.036	-0.954	-0.989
	4	-0.019	0.516	0.535	0.041	-0.756	-0.798
Subject	5	-0.014	0.624	0.638	0.046	-0.564	-0.610
	6	-0.044	0.749	0.793	0.030	-0.399	-0.430
	7	-0.027	0.871	0.899	0.031	-0.230	-0.261
	8	-0.051	1.031	1.083	0.057	-0.024	-0.081
	9	-0.064	1.197	1.261	-0.003	0.170	0.172
	Mean	-0.039	0.634	0.673	0.033	-0.640	-0.673

Table 4.6: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 1 subject sampling

4.3 Summary of Findings

Three CR-AE architectures were proposed for this experiment, each containing at least one convolutional and one recurrent neural network layer. Performance was evaluated based on reconstruction error, SNR and PSNR. Both SNR and PSNR were compared channel wise using the Harrell-Davis quantile estimator. In addition, two datasets were used to generate results with the second being included to test model generalizability.

Initially, the proposed architectures were evaluated with respect to each other to determine which would be chosen as the final CR-AE. Architecture one (1.32) had the best overall reconstruction error though differences were marginal. Visually, all three proposed architectures showed strong performance on the primary dataset, however for the secondary dataset this was not the case. Due to poor prediction error for early samples exaggerated by windowing, none of the proposed methods could generalize.

As a result, SNR and PSNR were not reported for the second dataset as they do not make sense in the context.

Performance with respect to SNR and PSNR was evaluated by first comparing each architecture to the raw signals with those results compared at quantile and channel level to each other. Architecture three performed the best overall for both SNR and PSNR. It had the highest mean SNR (5.53 db) and maintained minimum mean quantile difference of 0.187, across all conditions and sampling methods, to a maximum of 0.223 in comparison to architecture one and two. For PSNR, though the results were poor compared to the raw signals, architecture three (-24.44 db) had the highest mean PSNR. In addition, the mean quantile difference was positive in the majority of cases with a minimum of -0.060 and maximum of 0.139. Given overall performance, architecture three was chosen for the CR-AE.

To test the hypothesis, the proposed CR-AE was compared to each baseline methods using both SNR and PSNR. A positive increase was indicated by a positive mean quantile difference or when five or more deciles were positive. In all cases the proposed method outperformed each baseline in terms of SNR but exhibited worse performance for PSNR.

For PCA, the mean SNR (4.88 db) was lower than that of the CR-AE, while the PSNR (-23.78 db) was higher. All SNR quantile differences were positive for all conditions and sampling methods with a mean difference of 0.631. In contrast, all but the 8^{th} and 9^{th} quantiles in most cases were negative, with a mean difference of -0.564.

In terms of ICA, the semi-supervised nature of its implementation and the lack of domain knowledge available brings into question the validity of the results. Reconstructed signals and heat-maps of quantile differences by channel suggest that important signal information may have been lost during the process. In addition, comparison to the raw signals shows that ICA performed worse in the majority of quantiles for

both SNR and PSNR. With that in mind, the mean SNR (4.75 db) was lower than that of the CR-AE, while the PSNR (-23.95 db) was higher. All quantile differences were positive for SNR with a mean difference of 0.716. The mean quantile difference for PSNR (-0.427) was negative in all but some of the 7^{th}, 8^{th} and 9^{th} quantiles.

A basic auto-encoder was used to determine the effectiveness of the proposed CR-AE wth respect to a simpler implementation. Mean SNR for the AE (4.84 db) was also lower than that of the CR-AE, while again for PSNR the mean (-23.80) was higher. Similar to the results for both PCA and ICA, the SNR quantile difference was positive for all quantiles with mean difference of 0.674. In addition, PSNR results were also similar with negative differences across all but the 8^{th} and 9^{th} quantiles with a mean difference of -0.545.

These results show, that for PCA and a basic AE, there is evidence to support rejecting the null hypothesis from the point of view of SNR. However, for ICA, there is not enough evidence. In addition, for all three baseline methods, there is no evidence to support rejecting the null hypothesis in terms of PSNR. Finally, reconstruction error on the second dataset shows the proposed model does not generalize well to unseen data from other datasets. Therefore, it must be concluded that there is not enough evidence to support rejecting the null hypothesis that a stacked auto-encoder built with convolutional and recurrent neural network layers can increase the signal-to-noise ratio of EEG signals when compared to PCA, ICA and a simple auto-encoder.

4.4 Discussion

In this section the strengths and limitations of the results are discussed along with potential improvements that could be made to improve performance.

4.4.1 Strengths

The inclusion of a second dataset allowed model generalizability to be tested. Importantly, it highlighted an issue with the proposed CR-AE and showed that the CR-AE could not generalize well. The model's inability to predict early signal values for the second dataset, indicates that historic information may be required for accurate predictions. A potential solution could be to include an additional amount of data for each input that would be omitted from the final output. This would be used only to aid prediction of early values. As a solution this works though it is not optimum as it only shifts the error back a number of time-steps. From an architecture point of view this issue has a number of potential solutions. One of those is the use of non-causal padding. Causal padding is used to preserve temporal order which prevents the model from looking forward to future time-steps. As a consequence the model can only look back to previous time-steps which could be why it seems to require historic data to generalize. Using same or valid padding could help solve this issue as the CNN would use both prior and subsequent time-steps. This however violates temporal order which would be required if the model were to be used in a real-time application.

The use of windowing to augment training data was important as it increased the number of inputs seven-fold and enabled the model to extract more meaningful information from each overlapping slice of a signal. Order was preserved so that all windows from a single input were passed sequentially to the model, in that was they could be recombined accordingly post training. This worked well for the primary dataset but caused issue for the second due to the issue above. Re-combining the windows meant the issue was repeated throughout the reconstructed signal.

Incorporating both trial and subject sampling enabled the model to be tested on both unseen data and unseen subjects. By doing so, it showed whether the model could generalize to data it has never seen. Importantly, it also showed whether the model could generalize to subjects for whom it has seen no previous data.

Using dilated convolutions in combination with stacked parallel CNN's each utilizing different kernel sizes enabled the model to extract multiple feature maps over varied length temporal slices at a coarser scale. This increased the models' receptive field while maintaining input shape. These output feature maps were then concatenated and passed through a dense layer to obtain the most meaningful information.

4.4.2 Limitations

Though EEG inherently contains spatial information, given by the scalp location of each electrode, that information is not utilized by the proposed method. In this case only temporal information is used without explicitly taking into consideration proximity to nearby electrodes. Another approach to the problem would be to use a two-dimensional spatial representation of the electrodes at each time-step. Convolutional layers could then extract spatial information in addition to temporal information from each time-step. This, in combination with max pooling, could have the effect of enhancing the ERP information while reducing the level of noise originating from other electrodes.

The results obtained for ICA were poor due to the lack of domain specific knowledge required for its correct implementation. ICA is usually implemented in either a supervised or semi-supervised manner. The approach used in this case required at least one artefact to be identified for a single subject before being applied broadly across all subjects. The results suggest that this was done incorrectly as important information was lost across several electrodes. This meant that any evidence to support rejecting the null hypothesis could not be used as the results were not reflective. An alternative approach would be to use results from another piece of research where ICA was applied correctly and compare the results to those of the proposed CR-AE using the same dataset. As long as the design of the original research is mirrored, results should be comparable.

Though one of the motivations for conducting this research is to reduce the number of trials required to extract ERP, this is never explored. Increasing the signal-to-noise ratio is just one aspect of preparing the EEG for analysis and does not necessarily result in better quality signals. Being able to show that the number of trials required to extract ERP information has reduced, would further enhance the results and show that critical information has not been lost in the process. This would be especially powerful as a direct consequence of not being able to quantify information loss.

Chapter 5

Conclusion

5.1 Research Overview

The electroencephalogram is a method of recording electrical potentials in the brain used to diagnose different brain abnormalities including sleep disorder, epilepsy, stroke, and coma. These electrical potentials are measured using electrodes attached to the scalp which capture voltage fluctuations. Their placement is usually defined by one of the international standards depending on the number of electrodes and scalp coverage. This high temporal resolution recording is usually conducted during trials of a specific task where event related potentials can be extracted to analyse a subjects' reaction. Despite the availability of other recording methods, the electroencephalogram has remained in use since the first human EEG was recorded in 1924 by Hans Berger.

5.2 Problem Definition

The low amplitude nature of EEG makes it susceptible to interference. This can arise in the form of noise from eye and muscle movement, heart beat, line interference and underlying brain activity. These sources of noise are known as EEG artefacts. Given the proximity of electrodes on the scalp this noise tends to be present across a number of electrodes even if it is from ocular artefacts that originate at the frontal parietal area of the skull.

Since EEG is used to diagnose brain abnormalities it is imperative that accurate readings are taken so as a misdiagnosis is not given. The presence of artefacts is therefore a significant issue for EEG analysis. In the past, a number of approaches for removing these artefacts has been suggested. Early methods included selective rejection of EEG epochs based on visual inspection, verbal instruction not to blink given to the subjects and a fixation dot on a screen to reduce the likelihood of ocular artefacts. These proved useful, but were not effective in all cases and were very time-consuming. As an alternative, specific electrodes were placed to record artefacts and regression-based methods were used to remove the specific artefact from all other electrodes. This method was not applicable to all artefacts and any neural activity present in the reference channel would also be removed. Therefore, methods based on wavelet transformations and blind source separation were evaluated as they did not require a reference channel and addressed some issues with previous methods. However, wavelet transforms required a number of parameters to be chosen with little appropriate means for their selection and methods like PCA, which was primarily useful for ocular artefacts, failed to be fully effective when amplitudes were similar for the artefact and EEG. In addition, ICA involved similar domain knowledge and time to early rejection methods as components had to be inspected and removed manually. To address the limitations of ICA semi-automated and fully automated methods were developed though removing entire components could also remove important neural activity.

More recently, advanced machine learning algorithms have been used to solve very complex problems and their use for EEG analysis has been explored. Unsupervised learning algorithms, specifically auto-encoders, have the potential to address many of the issues identified for classic noise reduction techniques. Moreover, their design encourages accurate input reconstruction which could address the issue of information loss. If a method such as this could be used to reduce the amount of noise in an EEG signal in a robust and generalizable manner, without requiring manual intervention or significant time while minimizing information loss, then the number of trials required

for ERP analysis could be reduced and the accuracy of clinical diagnose could be improved.

For this research, a stacked auto-encoder designed using convolutional and recurrent neural network layers was proposed. The goal was to determine if the method could reduce the level of noise in a signal by improving the signal-to-noise ratio while retaining as much neural activity as possible. To evaluate its effectiveness, the method was compared to PCA and ICA — two state-of-the-art noise reduction techniques widely used for EEG signals. In addition, the method was compared to a simple auto-encoder variant to determine the level of improvement over a similar method. Finally, to ensure robustness and generalizably a second dataset was included in the research.

5.3 Design, Evaluation & Results

Two datasets were used in this research with both having previously been collected for ERP analysis. In each case a stimulus was introduced to elicit a cognitive response. Each trial was epoched around the onset of the stimulus and subset between 100 ms pre-stimulus and 500 ms post-stimulus. Pre-stimulus values were used to represent the noise while post-stimulus values represented the meaningful signal. The primary dataset was then divided into three seperate datasets; one for each condition. Differences in sample rate and channels mean that the primary dataset was downsampled by a factor of two and only common channels were retained. Two sampling methods were applied to each; trial and subject sampling. For the former, 30% of the instances were set aside as test data with the rest for training. In the latter 30% of all subjects were set aside for testing with the remaining 70% being used for training. Windowing using a 300 ms window length and 25 ms shift was then implemented on the train and test data to augment it. Augmentation was done at this point to preserve order so that the signals could be easily combined post prediction.

Three architectures were proposed to solve the problem. Each was a stacked auto-encoder created by combining CNN and RNN layers. The first was a combination of a single CNN and single RNN for both the encoder and decoder. The other two utilized multiple CNN or RNN layers as an extension of architecture one. Causal padding was used for all CNN layers with dilated convolutions being introduced for sequential CNN's. In addition, every layer was made up of three parallel CNN's each with a different kernel size followed by average pooling. LSTM was the chosen architecture for all RNN layers. A dense layer was used to combine the outputs of each parallel CNN and layer normalization was introduced to normalize the activations after every RNN layer. Huber loss was used to measure reconstruction error with early stopping implemented to prevent over-fitting. Finally, GPU accelerated TensorFlow was used to enable faster training and smaller batch size with more weight updates per epoch.

Each 300 ms window for all 28 channels was used as input to the proposed methods. Following prediction, these were re-combined with all other relevant windows to create the original 600 ms signal. The method used to re-combine the windows was to average the overlapping segments. Two metrics were used for evaluation, those being the SNR and PSNR. Both were implemented channel-wise across each input to generate a distribution of values. For comparative purposes, the Harrell-Davis quantile estimator was used to compare distributions. A positive shift was indicated by either a positive mean quantile difference or by five or more positive quantile differences. This was done to account for potential outliers in either the quantiles or channels. The best architecture was chosen by comparing reconstruction loss, SNR and PSNR of the reconstructed signals to those of the raw signals for both the primary and secondary datasets. Architecture three performed the best overall with higher mean SNR and PSNR than the other architectures and higher mean quantile difference across all conditions and sampling methods. However, architecture one had the best reconstruction error. In addition, none of the architectures generalized to the second dataset which resulted in the SNR and PSNR not being reported as they did not make sense in the context. As a result, architecture three was chosen for the CR-AE.

Three comparative methods were used to determine whether the CR-AE could increase the SNR and PSNR of EEG signals; PCA, ICA and a simple auto-encoder. PCA with retained variance of 87.35% was used to reconstruct the EEG signals. PCA was applied to the whole EEG signals before windowing. Results showed that, for SNR, the proposed CR-AE outperformed PCA in all conditions for both sampling methods. However, PCA outperformed the CR-AE in terms of PSNR for same. Mean quantile differences showed that for SNR there was a positive shift but for PSNR there was an equivalent negative shift.

ICA was implemented in a semi-supervised manner by identifying EEG artefacts for a single subject and using correlation to identify similar components across all subjects to remove. Components were removed by setting the values to zero before reconstructing the signals. This was done using the full continuous EEG as per the literature before being subset into trials. Results for ICA indicated that the method had not been implemented correctly. This was likely due to the lack of expertise required for identifying EEG artefacts and called into question the strength of the results. The CR-AE achieved higher mean SNR than ICA though lower mean PSNR. Similarly to PCA, a positive mean quantile difference was observed for SNR with a negative mean difference for PSNR.

Finally, a simple AE was used ensure the proposed method performed better than the basic implementation. This was also trained on the windowed signals and separately for each condition and sampling method. As with PCA and ICA, the proposed CR-AE achieved higher mean SNR but lower mean PSNR. The quantile differences showed that the CR-AE exhibited positive mean shift for SNR but negative mean shift for PSNR.

Given these results and the issues identified for ICA, it was concluded that there was not enough evidence to support rejecting the null hypothesis.

5.4 Contributions and impact

This research evaluated the applicability of a convolutional-recurrent auto-encoder to the task of noise reduction of EEG signals. A CR-AE, using stacked parallel CNN layers and an LSTM recurrent layer, was successfully implemented and evaluated using the SNR and PSNR showing this was possible.

Through evaluation, it was found that an increase in SNR was coupled with a corresponding decrease in PSNR. It was also found that the CR-AE could not generalize to a second dataset due to issues with early sample predictions. However, performance on subject sampling indicated that the CR-AE could generalize to data from the same research for unseen subjects.

By comparing performance on the raw signals to the performance of PCA, ICA and a simple AE on same, it was found that in all cases the CR-AE increased SNR across all Harrell-Davis quantiles for every condition and sampling method. However, it was also found that PSNR decreased in all cases across the majority of quantiles except for the 8^{th} and 9^{th}.

Issues with identifying EEG artefacts in ICA components highlighted both the need for domain knowledge when doing so, and the benefit of using an unsupervised method that achieves equivalent results.

The performance of architecture two suggests that, in this case, sequential LSTM layers do not improve performance when preceded by a CNN. In most cases the SNR, PSNR and reconstruction error were worse than both other architectures. In comparison, the performance of architecture three suggests that there is a benefit to using consecutive parallel CNN layers with dilated convolutions.

5.5 Future work & recommendations

This research focussed on the application of auto-encoders for noise reduction of EEG signals. Another topic of potential interest where this research could be expanded, is the area of mental workload (MWL) modelling. MWL is a highly complex concept, that explores the interaction of humans with technological devices. It is used to assess the cognitive effort involved in completing a task to ascertain performance (Longo, 2014, 2018a; Moustafa & Longo, 2019; Rizzo & Longo, 2017, 2019). In particular, future research could focus on the area of cognitive load theory, which has evolved within Education Psychology but is still grounded in the same concepts as MWL (Orru & Longo, 2019). With applications in medicine (Longo, 2015), human computer interaction (HCI) (Longo, 2011) and education (Longo, 2018b; Longo & Orru, 2019) there is plenty of scope within which to explore the application of auto-encoders to EEG signals for noise reduction in this field.

Prediction accuracy, on early samples from the secondary dataset inputs, highlighted the inability for the CR-AE to generalize to other datasets. As mentioned previously, this could be a result of using causal padding which limits the CNN to prior time-steps for predictions. An extension of this work could look to determine whether an alternative padding could be used to improve model generalizability. Another potential option would be to use a bi-directional wrapper for the LSTM layers, which learns from the original input and a reversed copy of it. In both cases this does not preserve temporal order so could not be used in a real-time scenario but it could enable the model to generalize better.

One of the limitations with using the GPU accelerated LSTM layers is the constraint imposed on hyper-parameter selection and consequently tuning. For this research, therefore there was little experimentation with LSTM hyper-parameters. In particular, recurrent dropout, which probabilistically excludes units from activation and weight updates could be used to improve model performance. For future work, I

would recommend focussing on how model performance is impacted by altering model hyper-parameters particularly in relation to performance on the secondary dataset if applicable.

As noted before, this research was motivated by the potential to reduce the required number of trials for ERP extraction. An extension of this research, which may also give context to the information loss incurred as a result of de-noising, would be to determine the number of trials required to extract ERP from the EEG signals. This could be achieved by averaging over an incremental number of trials before and after applying the CR-AE and generating a distribution of error values between the actual ERP, extracted from averaging all 100 trials per condition and subject, to the ERP's generated for the raw and reconstructed signals at each incremental number of averaged trials. In doing so, this would show the point at which error was lowest, the point at which it stopped improving and the information gain at each point. It would however be imperative to maintain the original order of the trials to ensure each ERP is generated using only information that had been captured to that point.

Given the spatial topography of EEG scalp electrodes, and the use of convolutional layers it would be interesting to extend this work to a two-dimensional implementation of the CNN by generating a spatial array of the values at each electrode on a given channel for each time-step. This could be achieved by creating a sparse $(n \times m)$ matrix M where each value $M(i, j)$ would represent an electrode location with the value being either zero, representing no electrode, or the voltage amplitude recorded for the electrode at time τ. This would then become an $(n \times m \times \tau)$ matrix which could be used as input to the two-dimensional CNN. By doing this, spatial and temporal information could be extracted thus enhancing the model. If used in combination with windowing, this could be very powerful as a single trial could be split into overlapping windows which would enhance the temporal information available to the model. As an alternative to simple data augmentation, this could actually then become an $(\omega \times n \times m \times \tau)$ input, where ω is the number of windows, which can be used in a

three-dimensional convolutional layer, commonly used for MRI or CT scan imagery.

As model performance improved when using stacked convolutional layers, a final suggestion would be to add more sequential CNN layers to take additional advantage of deep learning and its numerous benefits. With more layers even higher dilation could be used which would allow the model to learn even sparser representations of the inputs and potentially increase the signal-to-noise ratio even more.

Bibliography

Abdelhameed, A. M., Daoud, H. G., & Bayoumi, M. (2018). Epileptic Seizure Detection using Deep Convolutional Autoencoder, In *2018 ieee international workshop on signal processing systems (sips)*, IEEE. https://doi.org/10.1109/SiPS.2018.8598447

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Ahmadi, M., & Quian Quiroga, R. (2013). Automatic denoising of single-trial evoked potentials. *NeuroImage*, *66*, 672–680. https://doi.org/10.1016/j.neuroimage.2012.10.062

Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). Layer Normalization, arXiv 1607.06450. http://arxiv.org/abs/1607.06450

Ball, T., Kern, M., Mutschler, I., Aertsen, A., & Schulze-Bonhage, A. (2009). Signal quality of simultaneously recorded invasive and non-invasive EEG. *NeuroImage*, *46*(3), 708–716. https://doi.org/10.1016/j.neuroimage.2009.02.028

Bell, A. J., & Sejnowski, T. J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bengio, Y. (2012). Practical Recommendations for Gradient-Based Training of Deep Architectures, In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. https://doi.org/10.1007/978-3-642-35289-8_26

Bengio, Y., Boulanger-Lewandowski, N., & Pascanu, R. (2013). Advances in optimizing recurrent networks, In *2013 ieee international conference on acoustics, speech and signal processing*, IEEE. https://doi.org/10.1109/ICASSP.2013.6639349

Bengio, Y., & Delalleau, O. (2011). On the Expressive Power of Deep Architectures, In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. https://doi.org/10.1007/978-3-642-24412-4_3

Bengio, Y., & Lecun, Y. (2007). Scaling Learning Algorithms toward AI, In *Large-scale kernel machines*. The MIT Press. https://doi.org/10.7551/mitpress/7496.003.0016

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157–166. https://doi.org/10.1109/72.279181

Berkovsky, S., & Freyne, J. (2010). Group-based recipe recommendations, In *Proceedings of the fourth acm conference on recommender systems - recsys '10*, New York, New York, USA, ACM Press. https://doi.org/10.1145/1864708.1864732

Bijma, F., de Munck, J. C., Huizenga, H. M., & Heethaar, R. M. (2003). A mathematical approach to the temporal stationarity of background noise in MEG/EEG measurements. *NeuroImage*, *20*(1), 233–243. https://doi.org/10.1016/S1053-8119(03)00215-5

Binnie, C. D., & Prior, P. F. (1994). Electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, *57*(11), 1308–1319. https://doi.org/10.1136/jnnp.57.11.1308

Cakir, E., Parascandolo, G., Heittola, T., Huttunen, H., & Virtanen, T. (2017). Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *25*(6), arXiv 1702.06286, 1291–1303. https://doi.org/10.1109/TASLP.2017.2690575

Campos Viola, F., Thorne, J., Edmonds, B., Schneider, T., Eichele, T., & Debener, S. (2009). Semi-automatic identification of independent components representing

EEG artifact. *Clinical Neurophysiology*, *120*(5), 868–877. https://doi.org/10.1016/j.clinph.2009.01.015

Casarotto, S., Bianchi, A. M., Cerutti, S., & Chiarenza, G. A. (2004). Principal component analysis for reduction of ocular artefacts in event-related potentials of normal and dyslexic children. *Clinical Neurophysiology*, *115*(3), 609–619. https://doi.org/10.1016/j.clinph.2003.10.018

Cigizoglu, H. K., & Alp, M. (2006). Generalized regression neural network in modelling river sediment yield. *Advances in Engineering Software*, *37*(2), 63–68. https://doi.org/10.1016/j.advengsoft.2005.05.002

Crespo-Garcia, M., Atienza, M., & Cantero, J. L. (2008). Muscle Artifact Removal from Human Sleep EEG by Using Independent Component Analysis. *Annals of Biomedical Engineering*, *36*(3), 467–475. https://doi.org/10.1007/s10439-008-9442-y

Croft, R. J., Chandler, J. S., Barry, R. J., Cooper, N. R., & Clarke, A. R. (2005). EOG correction: A comparison of four methods. *Psychophysiology*, *42*(1), 16–24. https://doi.org/10.1111/j.1468-8986.2005.00264.x

Debener, S., Kranczioch, C., & Gutberlet, I. (2009). EEG Quality: Origin and Reduction of the EEG Cardiac-Related Artefact, In *Eeg - fmri*. Berlin, Heidelberg, Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87919-0_8

de Cheveigné, A., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience Methods*, *165*(2), 297–305. https://doi.org/10.1016/j.jneumeth.2007.06.003

Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. *ICLR Workshop*, (1), 2013–2016.

Engel, A. K., Moll, C. K. E., Fried, I., & Ojemann, G. A. (2005). Invasive recordings from the human brain: clinical insights and beyond. *Nature Reviews Neuroscience*, *6*(1), 35–47. https://doi.org/10.1038/nrn1585

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2014). Did I Do That? Abnormal Predictive Processes in Schizophrenia When Button Pressing to De-

liver a Tone. *Schizophrenia Bulletin*, *40*(4), 804–812. https://doi.org/10.1093/schbul/sbt072

Ghandeharion, H., & Erfanian, A. (2010). A fully automatic ocular artifact suppression from EEG data using higher order statistics: Improved performance by wavelet analysis. *Medical Engineering & Physics*, *32*(7), 720–729. https://doi.org/10.1016/j.medengphy.2010.04.010

Ghosh, R., Sinha, N., & Biswas, S. K. (2019). Automated eye blink artefact removal from EEG using support vector machine and autoencoder. *IET Signal Processing*, *13*(2), 141–148. https://doi.org/10.1049/iet-spr.2018.5111

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research*, *9*, 249–256.

Gondara, L. (2016). Medical Image Denoising Using Convolutional Denoising Autoencoders, In *2016 ieee 16th international conference on data mining workshops (icdmw)*, IEEE. https://doi.org/10.1109/ICDMW.2016.0041

Harender, & Sharma, R. K. (2017). EEG signal denoising based on wavelet transform, In *2017 international conference of electronics, communication and aerospace technology (iceca)*, IEEE. https://doi.org/10.1109/ICECA.2017.8203645

Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, *69*(3), 635–640. https://doi.org/10.1093/biomet/69.3.635

Helal, M. A., Eldawlatly, S., & Taher, M. (2017). Using Autoencoders for Feature Enhancement in Motor Imagery Brain-Computer Interfaces, In *Biomedical engineering*, Calgary,AB,Canada, ACTAPRESS. https://doi.org/10.2316/P.2017.852-052

Heydari, E., & Shahbakhti, M. (2015). Adaptive wavelet technique for EEG de-noising, In *2015 8th biomedical engineering international conference (bmeicon)*, IEEE. https://doi.org/10.1109/BMEiCON.2015.7399503

Himani, A., Tandon, O. P., & Bhatia, M. S. (1999). A study of P300-event related evoked potential in the patients of major depression. *Indian Journal of Physiology and Pharmacology*, *43*(3), 367–372.

Hu, L., Xiao, P., Zhang, Z., Mouraux, A., & Iannetti, G. (2014). Single-trial time–frequency analysis of electrocortical signals: Baseline correction and beyond. *NeuroImage*, *84*, 876–887. https://doi.org/10.1016/j.neuroimage.2013.09.055

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626–634. https://doi.org/10.1109/72.761722

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv 1502.03167. http://arxiv.org/abs/1502.03167

Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, *46*(4-5), 287–305. https://doi.org/10.1016/j.neucli.2016.07.002

Jia, Y., Zhou, C., & Motani, M. (2017). Spatio-temporal autoencoder for feature learning in patient data with missing observations, In *2017 ieee international conference on bioinformatics and biomedicine (bibm)*, IEEE. https://doi.org/10.1109/BIBM.2017.8217773

Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M., Iragui, V., & Sejnowski, T. (1998). Removing electroencephalographic artifacts: comparison between ICA and PCA, In *Neural networks for signal processing viii. proceedings of the 1998 ieee signal processing society workshop (cat. no.98th8378)*, IEEE. https://doi.org/10.1109/NNSP.1998.710633

Jung, T.-P., Makeig, S., Humpries, C., Lee, T.-W., McKeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), S0048577200980259. https://doi.org/10.1017/S0048577200980259

Kang, D., & Zhizeng, L. (2012). A Method of Denoising Multi-channel EEG Signals Fast Based on PCA and DEBSS Algorithm, In *2012 international conference on computer science and electronics engineering*, IEEE. https://doi.org/10.1109/ICCSEE.2012.105

Kaushal, G., Singh, A., & Jain, V. (2016). Better approach for denoising EEG signals, In *2016 5th international conference on wireless networks and embedded systems (wecon)*, IEEE. https://doi.org/10.1109/WECON.2016.7993455

Kiamini, M., Alirezaee, S., Perseh, B., & Ahmadi, M. (2009). Elimination of Ocular Artifacts from EEG signals using the wavelet transform and empirical mode decomposition, In *2009 6th international conference on electrical engineering/electronics, computer, telecommunications and information technology*, IEEE. https://doi.org/10.1109/ECTICON.2009.5137235

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, arXiv 1412.6980, 1–15. http://arxiv.org/abs/1412.6980

Klem, G. H., Lüders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalography and clinical neurophysiology. Supplement*, *52*(2), 3–6. http://www.ncbi.nlm.nih.gov/pubmed/10590970

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal*, *37*(2), 233–243. https://doi.org/10.1002/aic.690370209

Leite, N. M. N., Pereira, E. T., Gurjao, E. C., & Veloso, L. R. (2018). Deep Convolutional Autoencoder for EEG Noise Filtering, In *2018 ieee international conference on bioinformatics and biomedicine (bibm)*, IEEE. https://doi.org/10.1109/BIBM.2018.8621080

Li, J., Struzik, Z., Zhang, L., & Cichocki, A. (2015). Feature learning from incomplete EEG with denoising autoencoder. *Neurocomputing*, *165*arXiv 1410.0818, 23–31. https://doi.org/10.1016/j.neucom.2014.08.092

Liu, J. Y., & Yang, Y. H. (2019). Denoising Auto-Encoder with Recurrent Skip Connections and Residual Regression for Music Source Separation. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, arXiv 1807.01898, 773–778. https://doi.org/10.1109/ICMLA.2018.00123

Longo, L. (2011). Human-Computer Interaction and Human Mental Workload: Assessing Cognitive Engagement in the World Wide Web, In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. https://doi.org/10.1007/978-3-642-23768-3_43

Longo, L. (2014). Formalising Human Mental Workload as a Defeasible Computational Concept.

Longo, L. (2015). Designing Medical Interactive Systems Via Assessment of Human Mental Workload, In *2015 ieee 28th international symposium on computer-based medical systems*, IEEE. https://doi.org/10.1109/CBMS.2015.67

Longo, L. (2018a). Experienced mental workload, perception of usability, their interaction and impact on task performance (S. Federici, Ed.). *PLOS ONE*, *13*(8), e0199661. https://doi.org/10.1371/journal.pone.0199661

Longo, L. (2018b). On the Reliability, Validity and Sensitivity of Three Mental Workload Assessment Techniques for the Evaluation of Instructional Designs: A Case Study in a Third-level Course, In *Proceedings of the 10th international conference on computer supported education*, SCITEPRESS - Science; Technology Publications. https://doi.org/10.5220/0006801801660178

Longo, L., & Orru, G. (2019). An Evaluation of the Reliability, Validity and Sensitivity of Three Human Mental Workload Measures Under Different Instructional Conditions in Third-Level Education, In *Communications in computer and information science*. https://doi.org/10.1007/978-3-030-21151-6_19

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, *94*(20), 10979–10984. https://doi.org/10.1073/pnas.94.20.10979

Marchi, E., Vesperini, F., Eyben, F., Squartini, S., & Schuller, B. (2015). A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks, In *2015 ieee international conference on acoustics, speech and signal processing (icassp)*, IEEE. https://doi.org/10.1109/ICASSP.2015.7178320

Marinković, K. (2004). Spatiotemporal Dynamics of Word Processing in the Human Cortex. *The Neuroscientist*, *10*(2), 142–152. https://doi.org/10.1177/1073858403261018

Miller, D. M., Kaminsky, E. J., & Rana, S. (1995). Neural network classification of remote-sensing data. *Computers & Geosciences*, *21*(3), 377–386. https://doi.org/10.1016/0098-3004(94)00082-6

Min, B.-K., & Herrmann, C. S. (2007). Prestimulus EEG alpha activity reflects prestimulus top-down processing. *Neuroscience Letters*, *422*(2), 131–135. https://doi.org/10.1016/j.neulet.2007.06.013

Moustafa, K., & Longo, L. (2019). Analysing the Impact of Machine Learning to Model Subjective Mental Workload: A Case Study in Third-Level Education (L. Longo & M. C. Leva, Eds.). In L. Longo & M. C. Leva (Eds.), *International symposium on human mental workload: Models and applications*. Cham, Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_6

Nair, D. R., Burgess, R., McIntyre, C. C., & Lüders, H. (2008). Chronic subdural electrodes in the management of epilepsy. *Clinical Neurophysiology*, *119*(1), 11–28. https://doi.org/10.1016/j.clinph.2007.09.117

Nakamura, W., Anami, K., Mori, T., Saitoh, O., Cichocki, A., & Amari, S. (2006). Removal of Ballistocardiogram Artifacts From Simultaneously Recorded EEG and fMRI Data Using Independent Component Analysis. *IEEE Transactions on Biomedical Engineering*, *53*(7), 1294–1308. https://doi.org/10.1109/TBME.2006.875718

Nguyen, H.-A. T., Do, A. T., Le, T. H., & Bui, T. D. (2019). A deep sparse autoencoder method for automatic EOG artifact removal, In *2019 19th international conference on control, automation and systems (iccas)*, IEEE. https://doi.org/10.23919/ICCAS47443.2019.8971645

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. https://doi.org/10.1016/S1388-2457(00)00527-7

Orru, G., & Longo, L. (2019). The Evolution of Cognitive Load Theory and the Measurement of Its Intrinsic, Extraneous and Germane Loads: A Review (L. Longo & M. C. Leva, Eds.). In L. Longo & M. C. Leva (Eds.), *International symposium on human mental workload: Models and applications*. Cham, Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_3

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Peng, H., Hu, B., Shi, Q., Ratcliffe, M., Zhao, Q., Qi, Y., & Gao, G. (2013). Removal of Ocular Artifacts in EEG—An Improved Approach Combining DWT and ANC for Portable Applications. *IEEE Journal of Biomedical and Health Informatics*, *17*(3), 600–607. https://doi.org/10.1109/JBHI.2013.2253614

Qiu, Y., Zhou, W., Yu, N., & Du, P. (2018). Denoising Sparse Autoencoder Based Ictal EEG Classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(9), 1–1. https://doi.org/10.1109/TNSRE.2018.2864306

Rifai, S., Vincent, P., Muller, X., Glorot, X., & Bengio, Y. (2011). Contractive auto-encoders: Explicit invariance during feature extraction. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, (1), 833–840.

Rizzo, L., & Longo, L. (2017). Representing and inferring mental workload via defeasible reasoning: A comparison with the NASA task load index and the workload profile, In *1st workshop on advances in argumentation in artificial intelligence*.

Rizzo, L., & Longo, L. (2019). Inferential models of mental workload with defeasible argumentation and non-monotonic fuzzy reasoning: A comparative study, In *Proceedings of the 2nd workshop on advances in argumentation in artificial intelligence*.

Rousselet. (2010). Healthy aging delays scalp EEG sensitivity to noise in a face discrimination task. *Frontiers in Psychology*, *1*(July), 1–14. https://doi.org/10.3389/fpsyg.2010.00019

Rousselet, G. A., Husk, J. S., Pernet, C. R., Gaspar, C. M., Bennett, P. J., & Sekuler, A. B. (2009). Age-related delay in information accrual for faces: Evidence from

a parametric, single-trial EEG approach. *BMC Neuroscience*, *10*(1), 114. https://doi.org/10.1186/1471-2202-10-114

Roy, V., & Shukla, S. (2015). Mth Order FIR Filtering for EEG Denoising Using Adaptive Recursive Least Squares Algorithm, In *2015 international conference on computational intelligence and communication networks (cicn)*, IEEE. https://doi.org/10.1109/CICN.2015.85

Scott Makeig, Bell, A. J., Jung, T.-P., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic signals, In *Proceedings of the 8th international conference on neural information processing systems*, MIT Press. https://papers.nips.cc/paper/1091-independent-component-analysis-of-electroencephalographic-data.pdf

Sheoran, M., Kumar, S., & Chawla, S. (2015). Methods of denoising of electroencephalogram signal: a review. *International Journal of Biomedical Engineering and Technology*, *18*(4), 385. https://doi.org/10.1504/IJBET.2015.071012

Soylu, F., Rivera, B., Anchan, M., & Shannon, N. (2019). ERP differences in processing canonical and noncanonical finger-numeral configurations. *Neuroscience Letters*, *705*(April), 74–79. https://doi.org/10.1016/j.neulet.2019.04.032

Srivastava, G., Crottaz-Herbette, S., Lau, K., Glover, G., & Menon, V. (2005). ICA-based procedures for removing ballistocardiogram artifacts from EEG data acquired in the MRI scanner. *NeuroImage*, *24*(1), 50–60. https://doi.org/10.1016/j.neuroimage.2004.09.041

Supratak, A., Ling Li, & Yike Guo. (2014). Feature extraction with stacked autoencoders for epileptic seizure detection, In *2014 36th annual international conference of the ieee engineering in medicine and biology society*, IEEE. https://doi.org/10.1109/EMBC.2014.6944546

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions, In *2015 ieee conference on computer vision and pattern recognition (cvpr)*, IEEE. https://doi.org/10.1109/CVPR.2015.7298594

Trigeorgis, G., Ringeval, F., Brueckner, R., Marchi, E., Nicolaou, M. A., Schuller, B., & Zafeiriou, S. (2016). Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network, In *2016 ieee international conference on acoustics, speech and signal processing (icassp)*, IEEE. https://doi.org/10.1109/ICASSP.2016.7472669

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio, arXiv 1609.03499, 1–15. http://arxiv.org/abs/1609.03499

van de Velde, M., van Erp, G., & Cluitmans, P. J. (1998). Detection of muscle artefact in the normal human awake EEG. *Electroencephalography and Clinical Neurophysiology*, *107*(2), 149–158. https://doi.org/10.1016/S0013-4694(98)00052-2

Vaseghi, & V., S. (2001). Noise and Distortion, In *Advanced digital signal processing and noise reduction*. Chichester, UK, John Wiley & Sons, Ltd. https://doi.org/10.1002/0470841621.ch2

Verleger, R. (1991). The instruction to refrain from blinking affects auditory P3 and N1 amplitudes. *Electroencephalography and Clinical Neurophysiology*, *78*(3), 240–251. https://doi.org/10.1016/0013-4694(91)90039-7

Vigário, R. N. (1997). Extraction of ocular artefacts from EEG using independent component analysis. *Electroencephalography and Clinical Neurophysiology*, *103*(3), 395–404. https://doi.org/10.1016/S0013-4694(97)00042-8

Vincent, J. L., Larson-Prior, L. J., Zempel, J. M., & Snyder, A. Z. (2007). Moving GLM ballistocardiogram artifact reduction for EEG acquired simultaneously with fMRI. *Clinical Neurophysiology*, *118*(5), 981–998. https://doi.org/10.1016/j.clinph.2006.12.017

Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders, In *Proceedings of the 25th international conference on machine learning - icml '08*, New York, New York, USA, ACM Press. https://doi.org/10.1145/1390156.1390294

Wan, R., Mei, S., Wang, J., Liu, M., & Yang, F. (2019). Multivariate Temporal Convolutional Network: A Deep Neural Networks Approach for Multivariate Time Series Forecasting. *Electronics*, *8*(8), 876. https://doi.org/10.3390/electronics8080876

Wen, T., & Zhang, Z. (2018). Deep Convolution Neural Network and Autoencoders-Based Unsupervised Feature Learning of EEG Signals. *IEEE Access*, *6*, 25399–25410. https://doi.org/10.1109/ACCESS.2018.2833746

Xiong, P., Wang, H., Liu, M., Zhou, S., Hou, Z., & Liu, X. (2016). ECG signal enhancement based on improved denoising auto-encoder. *Engineering Applications of Artificial Intelligence*, *52*, 194–202. https://doi.org/10.1016/j.engappai.2016.02.015

Yang, B., Duan, K., Fan, C., Hu, C., & Wang, J. (2018). Automatic ocular artifacts removal in EEG using deep learning. *Biomedical Signal Processing and Control*, *43*, 148–158. https://doi.org/10.1016/j.bspc.2018.02.021

Yang, B., Duan, K., & Zhang, T. (2016). Removal of EOG artifacts from EEG using a cascade of sparse autoencoder and recursive least squares adaptive filter. *Neurocomputing*, *214*, 1053–1060. https://doi.org/10.1016/j.neucom.2016.06.067

Yin, Z., & Zhang, J. (2016). Recognition of Cognitive Task Load levels using single channel EEG and Stacked Denoising Autoencoder, In *2016 35th chinese control conference (ccc)*, IEEE. https://doi.org/10.1109/ChiCC.2016.7553961

Zheng, Y., Liu, Q., Chen, E., Ge, Y., & Zhao, J. L. (2014). Time series classification using multi-channels deep convolutional neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8485 LNCS*, 298–310. https://doi.org/10.1007/978-3-319-08010-9_33

Zinovyev, A., Kairov, U., Karpenyuk, T., & Ramanculov, E. (2013). Blind source separation methods for deconvolution of complex signals in cancer biology. *Biochemical and Biophysical Research Communications*, *430*(3), 1182–1187. https://doi.org/10.1016/j.bbrc.2012.12.043

Appendix A

Additional content

Electrode Placement Sites									
Pre-frontal	Frontal	Frontal Central	Central	Temporal	Central Parietal	Temporal Parietal	Parietal	Occipital	Ground
Fp1	F3	FC1	C3	T7	CP1	TP10	P3	O1	Fz
Fp2	F4	FC2	C4	T8	CP2		P4	O2	Oz
	F7	FC5			CP5		P7		Pz
	F8	FC6			CP6		P8		

Table A.1: Common electrode placement sites for primary and secondary datasets

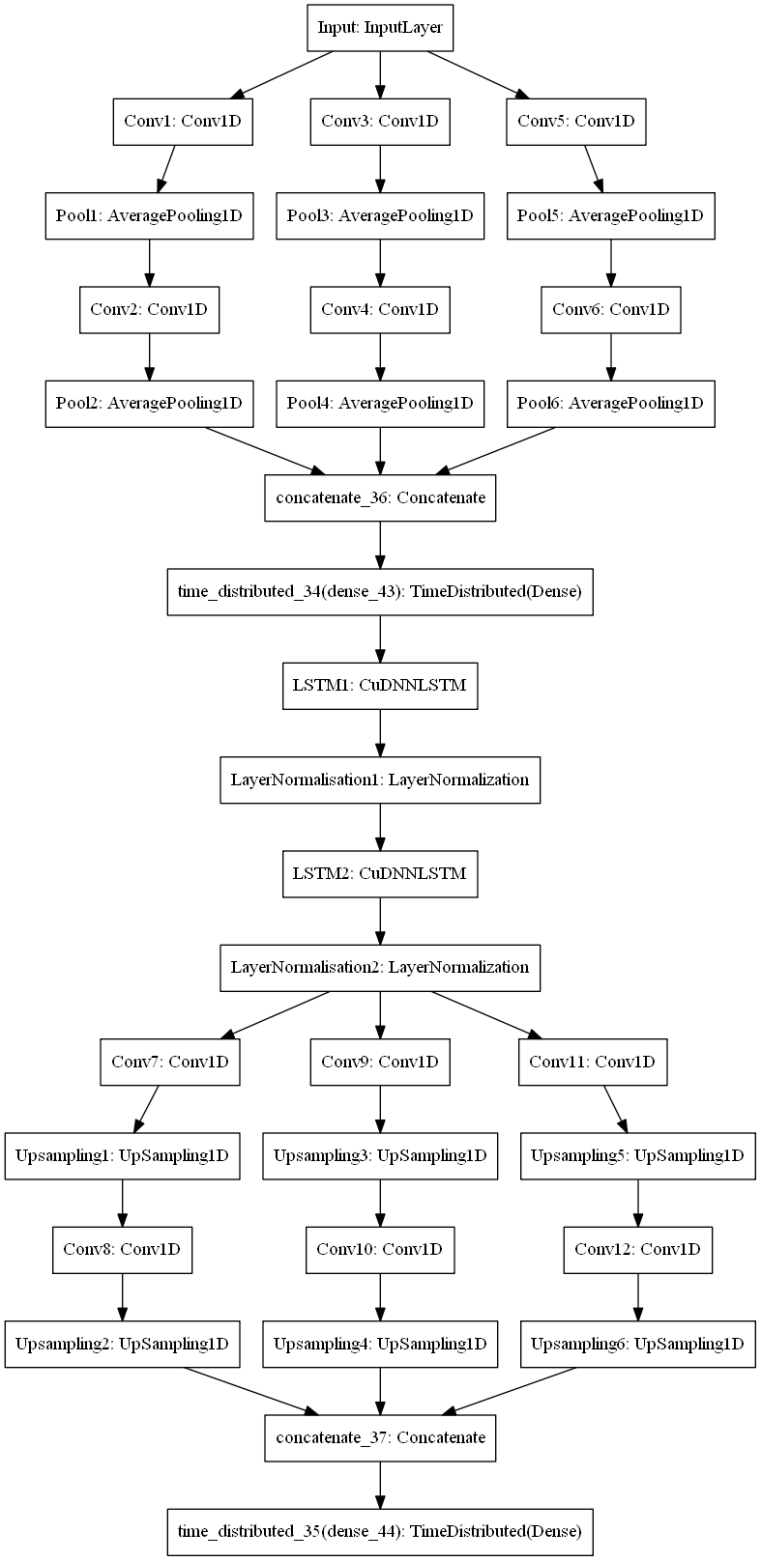


Figure A.1: Technical model of proposed CR-AE

		Hyperparameters		
Stage	Layer	Block 1	Block 2	Block 3
Encoder	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	AveragePooling1D	Pool size = 3	Pool size = 3	Pool size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		
	CuDNNLSTM	Units = 150		
	LayerNormalisation			
Decoder	CuDNNLSTM	Units = 150		
	LayerNormalisation			
	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	UpSampling1D	Size = 3	Size = 3	Size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		

Table A.2: Chosen hyper-parameters for architecture 1

		Hyperparameters		
Stage	Layer	Block 1	Block 2	Block 3
Encoder	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	AveragePooling1D	Pool size = 3	Pool size = 3	Pool size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		
	CuDNNLSTM	Units = 150		
	LayerNormalisation			
	CuDNNLSTM	Units = 150		
	LayerNormalisation			
Decoder	CuDNNLSTM	Units = 150		
	LayerNormalisation			
	CuDNNLSTM	Units = 150		
	LayerNormalisation			
	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	UpSampling1D	Size = 3	Size = 3	Size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		

Table A.3: Chosen hyper-parameters for architecture 2

		Hyperparameters		
Stage	Layer	Block 1	Block 2	Block 3
Encoder	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	AveragePooling1D	Pool size = 3	Pool size = 3	Pool size = 3
	Conv1D	Dilation rate = 2 All other parameters as above	Dilation rate = 2 All other parameters as above	Dilation rate = 2 All other parameters as above
	AveragePooling1D	Pool size = 3	Pool size = 3	Pool size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		
	CuDNNLSTM	Units = 150		
	LayerNormalisation			
Decoder	CuDNNLSTM	Units = 150		
	LayerNormalisation			
	Conv1D	Dilation rate = 2 All other parameters as below	Dilation rate = 2 All other parameters as below	Dilation rate = 2 All other parameters as below
	UpSampling1D	Size = 3	Size = 3	Size = 3
	Conv1D	Filters = 28 Kernel size = 3 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 5 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'	Filters = 28 Kernel size = 7 Strides = 1 Padding = 'causal' Dilation rate = 1 Activation = 'linear'
	UpSampling1D	Size = 3	Size = 3	Size = 3
	Concatenate			
	TimeDistributed(Dense)	Units = 28		

Table A.4: Chosen hyper-parameters for architecture 3

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.041	0.099	0.140	0.028	-1.536	-1.564
	2	-0.020	0.264	0.285	-0.003	-1.037	-1.034
	3	-0.014	0.380	0.395	0.026	-0.809	-0.835
	4	0.012	0.534	0.521	0.035	-0.631	-0.666
Trial	5	0.039	0.656	0.617	0.025	-0.437	-0.463
	6	0.016	0.757	0.741	0.034	-0.261	-0.295
	7	-0.009	0.882	0.890	0.037	-0.108	-0.145
	8	0.015	1.034	1.019	-0.004	0.115	0.119
	9	0.020	1.257	1.237	-0.004	0.446	0.451
	Mean	0.002	0.651	0.649	0.019	-0.473	-0.492

Table A.5: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 2 trial sampling

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.032	0.051	0.083	-0.045	-1.626	-1.582
	2	-0.005	0.212	0.217	-0.052	-1.128	-1.076
	3	-0.002	0.328	0.330	0.004	-0.918	-0.922
	4	0.018	0.463	0.445	0.037	-0.728	-0.765
Subject	5	0.015	0.581	0.567	0.031	-0.562	-0.593
	6	0.039	0.713	0.674	0.042	-0.387	-0.429
	7	0.028	0.849	0.821	0.041	-0.223	-0.264
	8	0.002	1.002	1.001	0.069	0.028	-0.041
	9	0.031	1.218	1.187	0.105	0.300	0.195
	Mean	0.010	0.602	0.592	0.026	-0.583	-0.608

Table A.6: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 2 subject sampling

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.016	0.058	0.074	-0.021	-1.716	-1.694
	2	-0.018	0.212	0.230	-0.019	-1.174	-1.155
	3	-0.031	0.356	0.386	-0.037	-0.897	-0.860
	4	-0.040	0.490	0.530	-0.041	-0.678	-0.637
Trial	5	-0.029	0.615	0.644	-0.032	-0.498	-0.466
	6	-0.031	0.720	0.751	-0.061	-0.339	-0.278
	7	-0.047	0.867	0.914	-0.062	-0.161	-0.099
	8	-0.035	1.041	1.075	-0.018	0.078	0.096
	9	-0.003	1.271	1.275	0.009	0.446	0.437
	Mean	-0.028	0.626	0.653	-0.031	-0.549	-0.517

Table A.7: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 3 trial sampling

		SNR			PSNR		
Split	Decile	PCA vs Raw	CR-AE vs Raw	Difference	PCA vs Raw	CR-AE vs Raw	Difference
	1	-0.035	0.040	0.075	-0.032	-1.695	-1.663
	2	-0.040	0.195	0.235	-0.007	-1.180	-1.174
	3	-0.047	0.341	0.388	-0.026	-0.912	-0.886
	4	-0.027	0.467	0.493	-0.004	-0.710	-0.707
Subject	5	0.001	0.586	0.586	-0.036	-0.543	-0.508
	6	0.009	0.730	0.721	-0.049	-0.400	-0.351
	7	-0.023	0.844	0.867	-0.005	-0.216	-0.212
	8	-0.013	1.018	1.031	-0.008	-0.030	-0.022
	9	-0.011	1.286	1.297	0.024	0.254	0.230
	Mean	-0.021	0.612	0.633	-0.016	-0.604	-0.588

Table A.8: SNR and PSNR HD quantile differences for PCA and the proposed CR-AE — condition 3 subject sampling

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.096	0.099	0.195	-0.039	-1.536	-1.497
	2	-0.085	0.264	0.350	-0.075	-1.037	-0.962
	3	-0.098	0.380	0.479	-0.095	-0.809	-0.714
	4	-0.106	0.534	0.639	-0.106	-0.631	-0.525
Trial	5	-0.114	0.656	0.770	-0.122	-0.437	-0.316
	6	-0.117	0.757	0.874	-0.141	-0.261	-0.121
	7	-0.135	0.882	1.016	-0.165	-0.108	0.057
	8	-0.142	1.034	1.175	-0.193	0.115	0.309
	9	-0.173	1.257	1.429	-0.245	0.446	0.691
	Mean	-0.118	0.651	0.770	-0.131	-0.473	-0.342

Table A.9: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 2 trial sampling

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.109	0.051	0.160	-0.099	-1.626	-1.527
	2	-0.099	0.212	0.311	-0.123	-1.128	-1.005
	3	-0.106	0.328	0.434	-0.141	-0.918	-0.777
	4	-0.107	0.463	0.570	-0.152	-0.728	-0.576
Subject	5	-0.116	0.581	0.698	-0.162	-0.562	-0.399
	6	-0.104	0.713	0.816	-0.179	-0.387	-0.207
	7	-0.105	0.849	0.954	-0.188	-0.223	-0.034
	8	-0.115	1.002	1.118	-0.215	0.028	0.243
	9	-0.124	1.218	1.342	-0.246	0.300	0.546
	Mean	-0.110	0.602	0.712	-0.167	-0.583	-0.415

Table A.10: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 2 subject sampling

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.044	0.058	0.101	-0.024	-1.716	-1.691
	2	-0.049	0.212	0.261	-0.034	-1.174	-1.140
	3	-0.055	0.356	0.410	-0.067	-0.897	-0.830
	4	-0.054	0.490	0.544	-0.073	-0.678	-0.605
Trial	5	-0.062	0.615	0.677	-0.089	-0.498	-0.409
	6	-0.055	0.720	0.775	-0.112	-0.339	-0.227
	7	-0.065	0.867	0.932	-0.120	-0.161	-0.041
	8	-0.089	1.041	1.130	-0.162	0.078	0.240
	9	-0.139	1.271	1.410	-0.225	0.446	0.671
	Mean	-0.068	0.626	0.694	-0.101	-0.549	-0.448

Table A.11: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 3 trial sampling

		SNR			PSNR		
Split	Decile	ICA vs Raw	CR-AE vs Raw	Difference	ICA vs Raw	CR-AE vs Raw	Difference
	1	-0.045	0.040	0.085	0.004	-1.695	-1.698
	2	-0.057	0.195	0.252	-0.031	-1.180	-1.150
	3	-0.056	0.341	0.397	-0.046	-0.912	-0.866
	4	-0.062	0.467	0.529	-0.060	-0.710	-0.650
Subject	5	-0.057	0.586	0.643	-0.072	-0.543	-0.471
	6	-0.065	0.730	0.795	-0.092	-0.400	-0.308
	7	-0.067	0.844	0.911	-0.113	-0.216	-0.103
	8	-0.088	1.018	1.107	-0.139	-0.030	0.108
	9	-0.130	1.286	1.416	-0.195	0.254	0.449
	Mean	-0.070	0.612	0.681	-0.083	-0.604	-0.521

Table A.12: SNR and PSNR HD quantile differences for ICA and the proposed CR-AE — condition 3 subject sampling

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
	1	-0.059	0.099	0.158	0.006	-1.536	-1.543
	2	-0.033	0.264	0.298	0.021	-1.037	-1.058
	3	-0.040	0.380	0.421	-0.005	-0.809	-0.804
	4	-0.036	0.534	0.570	0.003	-0.631	-0.634
Trial	5	-0.019	0.656	0.674	-0.024	-0.437	-0.413
	6	-0.026	0.757	0.783	-0.011	-0.261	-0.250
	7	-0.039	0.882	0.920	-0.021	-0.108	-0.087
	8	-0.027	1.034	1.061	-0.042	0.115	0.157
	9	-0.081	1.257	1.337	-0.037	0.446	0.484
	Mean	-0.040	0.651	0.691	-0.012	-0.473	-0.461

Table A.13: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 2 trial sampling

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
	1	-0.029	0.051	0.080	-0.097	-1.626	-1.530
	2	0.001	0.212	0.210	-0.049	-1.128	-1.079
	3	-0.022	0.328	0.350	-0.020	-0.918	-0.898
	4	-0.003	0.463	0.466	-0.001	-0.728	-0.727
Subject	5	-0.018	0.581	0.599	0.007	-0.562	-0.568
	6	0.006	0.713	0.706	0.032	-0.387	-0.419
	7	-0.004	0.849	0.854	0.016	-0.223	-0.239
	8	-0.022	1.002	1.024	0.020	0.028	0.008
	9	-0.055	1.218	1.273	0.071	0.300	0.229
	Mean	-0.016	0.602	0.618	-0.002	-0.583	-0.580

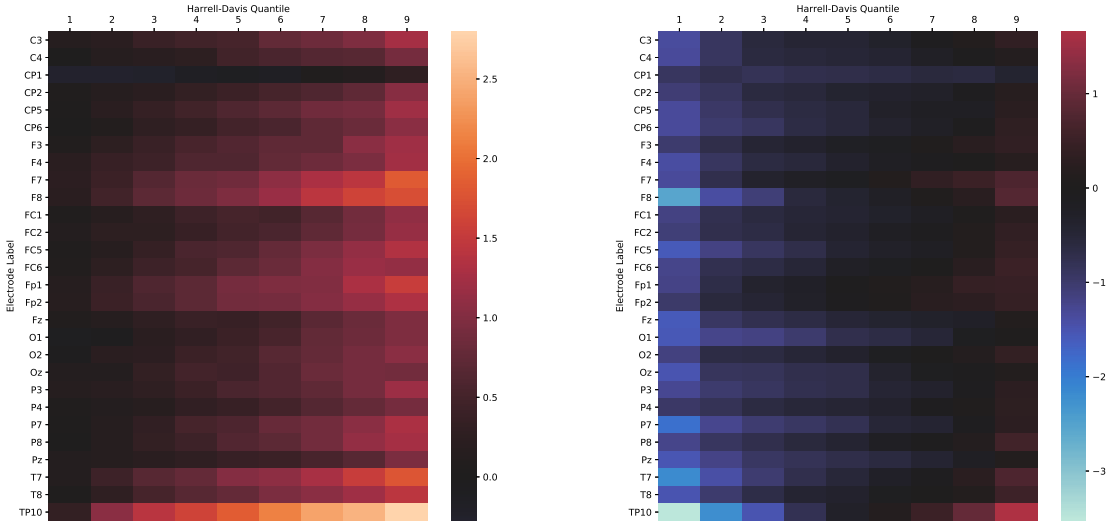
Table A.14: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 2 subject sampling

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
	1	-0.062	0.058	0.119	0.072	-1.716	-1.788
	2	-0.056	0.212	0.268	-0.006	-1.174	-1.168
	3	-0.064	0.356	0.420	0.000	-0.897	-0.897
	4	-0.077	0.490	0.567	-0.032	-0.678	-0.647
Trial	5	-0.080	0.615	0.695	-0.045	-0.498	-0.453
	6	-0.091	0.720	0.811	-0.077	-0.339	-0.262
	7	-0.086	0.867	0.953	-0.103	-0.161	-0.058
	8	-0.081	1.041	1.121	-0.052	0.078	0.131
	9	-0.047	1.271	1.318	-0.014	0.446	0.460
	Mean	-0.071	0.626	0.697	-0.029	-0.549	-0.520

Table A.15: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 3 trial sampling

		SNR			PSNR		
Split	Decile	AE vs Raw	CR-AE vs Raw	Difference	AE vs Raw	CR-AE vs Raw	Difference
	1	-0.061	0.040	0.101	-0.023	-1.695	-1.672
	2	-0.067	0.195	0.262	-0.031	-1.180	-1.149
	3	-0.063	0.341	0.404	-0.030	-0.912	-0.882
	4	-0.066	0.467	0.533	-0.004	-0.710	-0.706
Subject	5	-0.060	0.586	0.646	-0.033	-0.543	-0.511
	6	-0.050	0.730	0.779	-0.072	-0.400	-0.328
	7	-0.066	0.844	0.909	-0.045	-0.216	-0.172
	8	-0.053	1.018	1.071	-0.041	-0.030	0.011
	9	-0.015	1.286	1.301	-0.014	0.254	0.267
	Mean	-0.056	0.612	0.667	-0.032	-0.604	-0.571

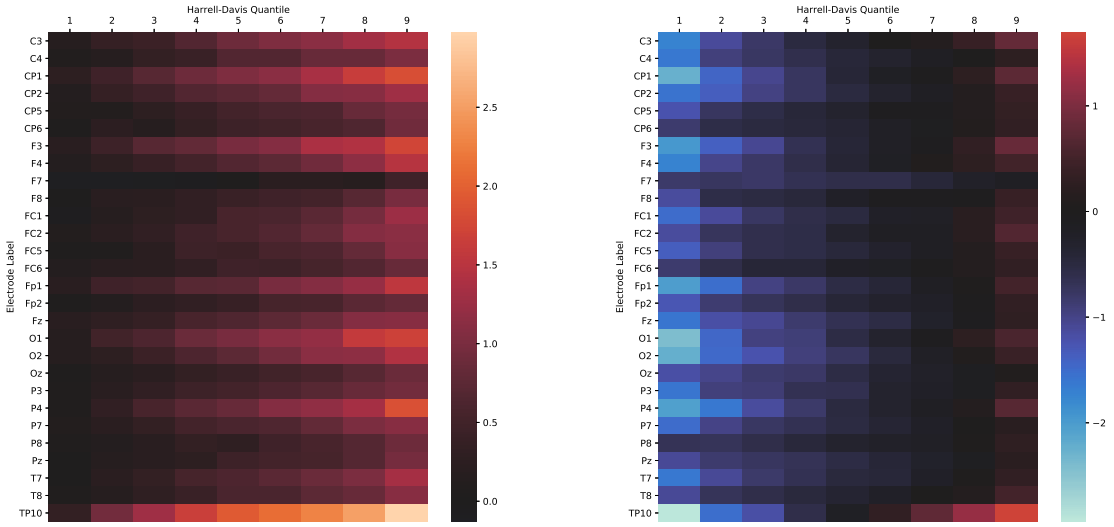
Table A.16: SNR and PSNR HD quantile differences for AE and the proposed CR-AE — condition 3 subject sampling



(a) CR-AE vs Raw - SNR condition 1

(b) CR-AE vs Raw - PSNR condition 1

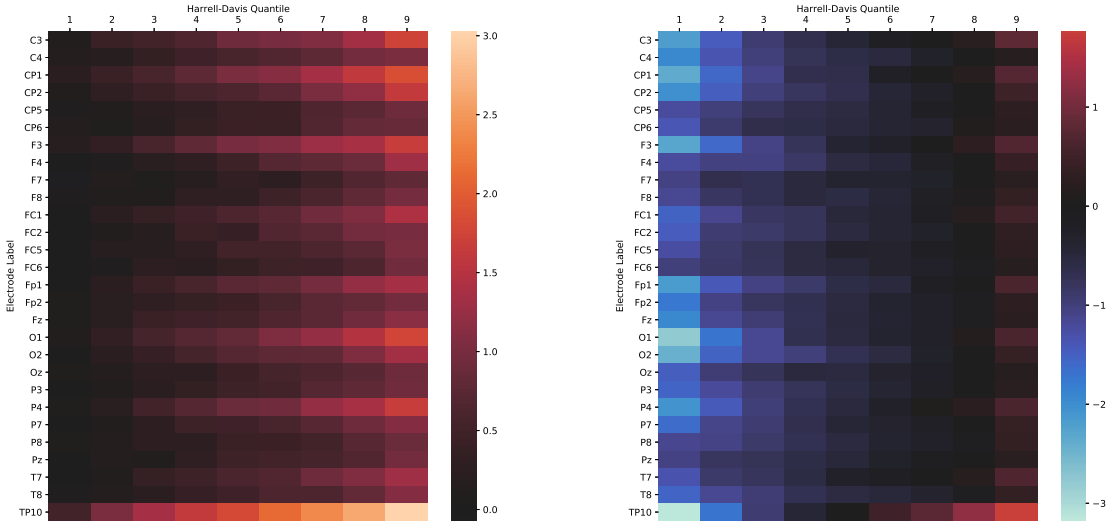
Figure A.2: Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 1



(a) CR-AE vs Raw - SNR condition 2

(b) CR-AE vs Raw - PSNR condition 2

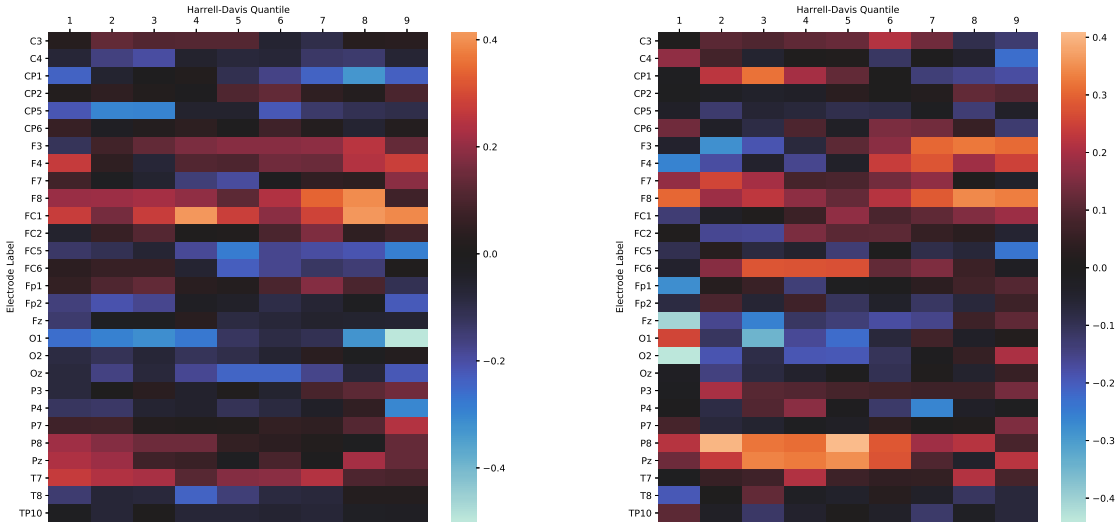
Figure A.3: Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 2



(a) CR-AE vs Raw - SNR condition 3

(b) CR-AE vs Raw - PSNR condition 3

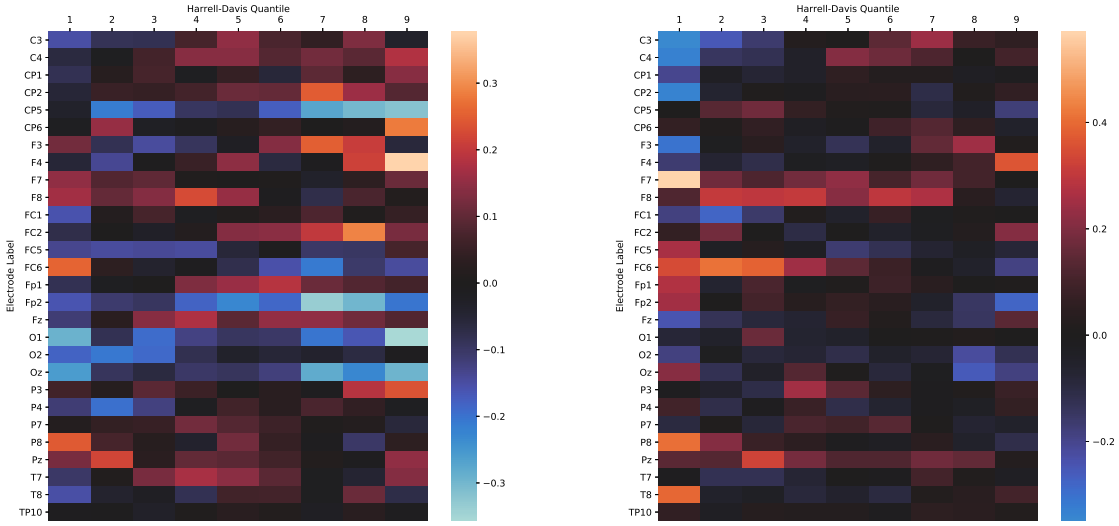
Figure A.4: Heat-maps of SNR and PSNR HD quantile differences at channel level for CR-AE reconstructed signals compared to original signals — condition 3



(a) PCA vs Raw - SNR condition 1

(b) PCA vs Raw - PSNR condition 1

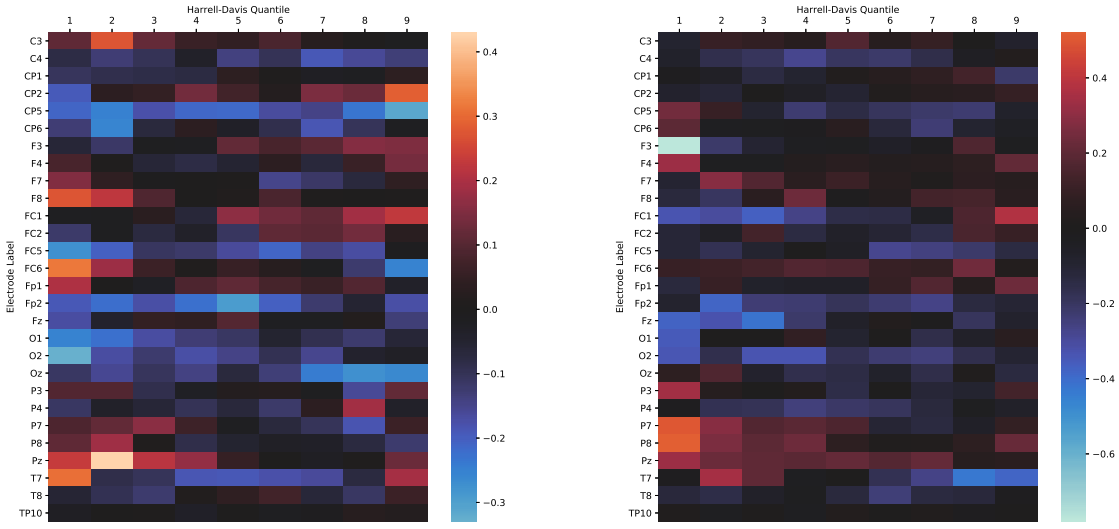
Figure A.5: Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 1



(a) PCA vs Raw - SNR condition 2

(b) PCA vs Raw - PSNR condition 2

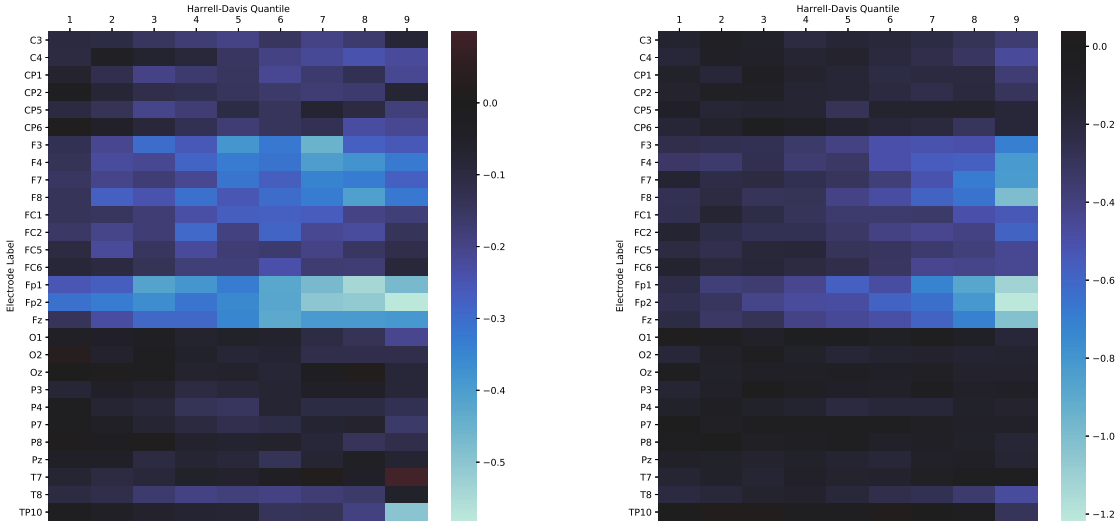
Figure A.6: Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 2



(a) PCA vs Raw - SNR condition 3

(b) PCA vs Raw - PSNR condition 3

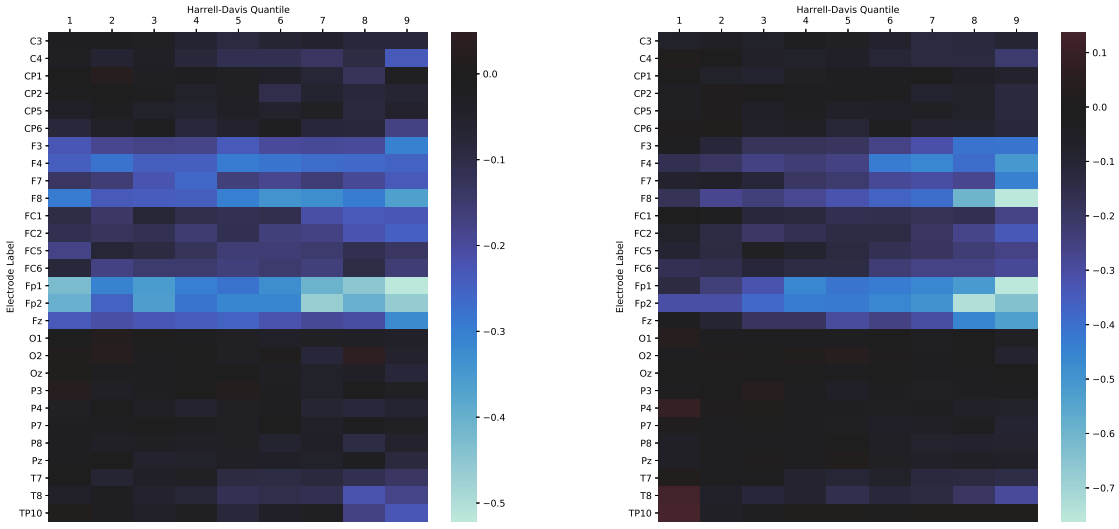
Figure A.7: Heat-maps of SNR and PSNR HD quantile differences at channel level for PCA reconstructed signals compared to original signals — condition 3



(a) ICA vs Raw - SNR condition 1

(b) ICA vs Raw - PSNR condition 1

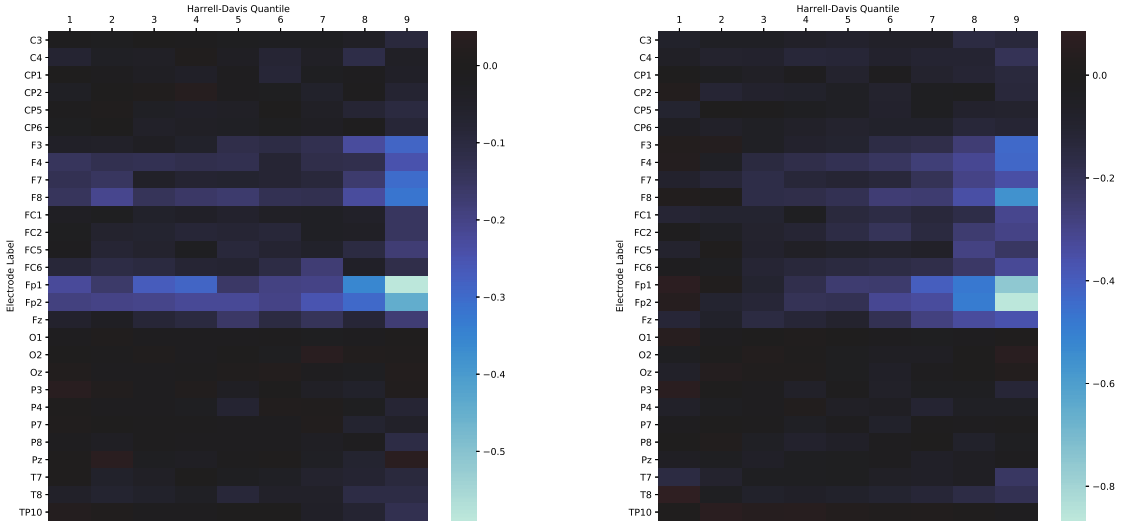
Figure A.8: Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 1



(a) ICA vs Raw - SNR condition 2

(b) ICA vs Raw - PSNR condition 2

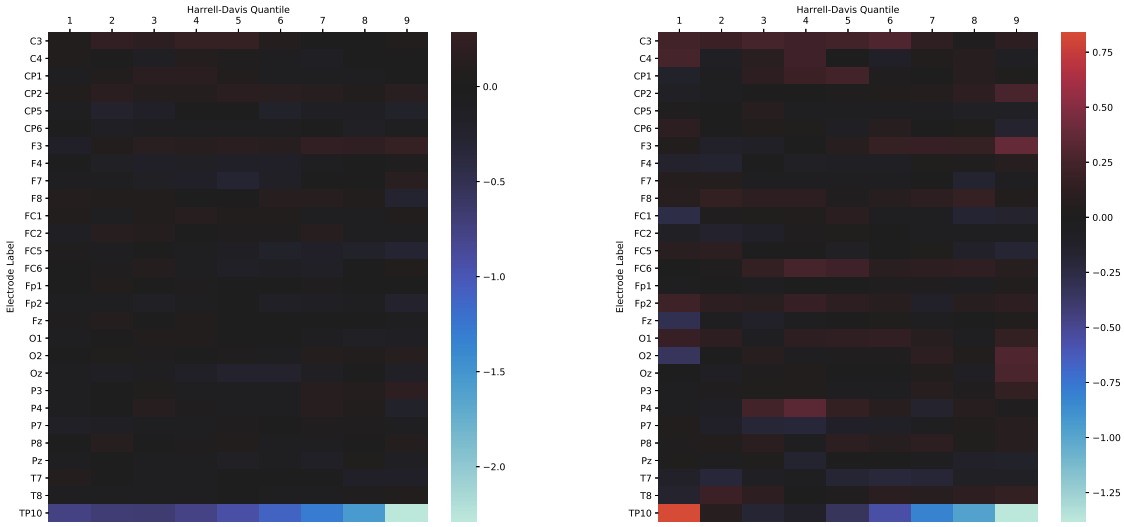
Figure A.9: Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 2



(a) ICA vs Raw - SNR condition 3

(b) ICA vs Raw - PSNR condition 3

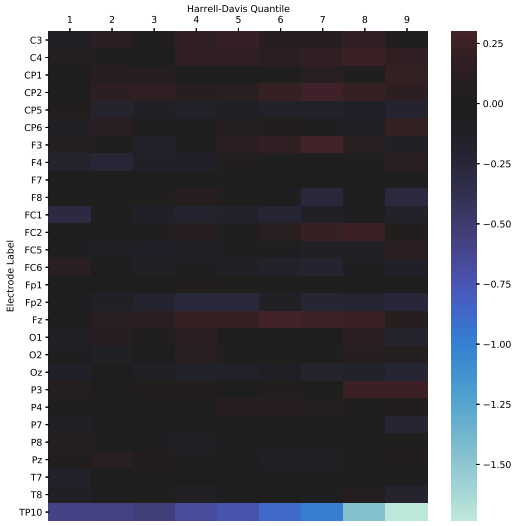
Figure A.10: Heat-maps of SNR and PSNR HD quantile differences at channel level for ICA reconstructed signals compared to original signals — condition 3



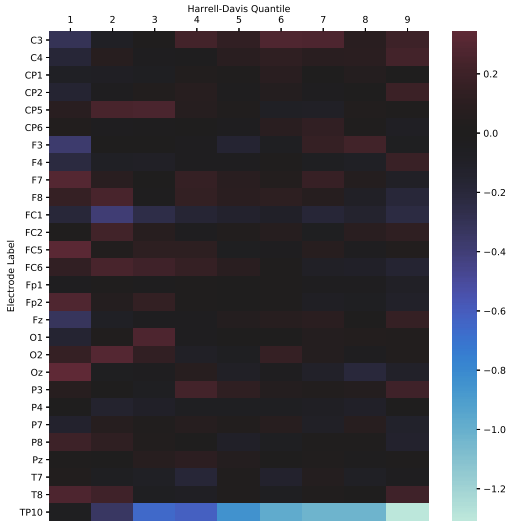
(a) AE vs Raw - SNR condition 1

(b) AE vs Raw - PSNR condition 1

Figure A.11: Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 1

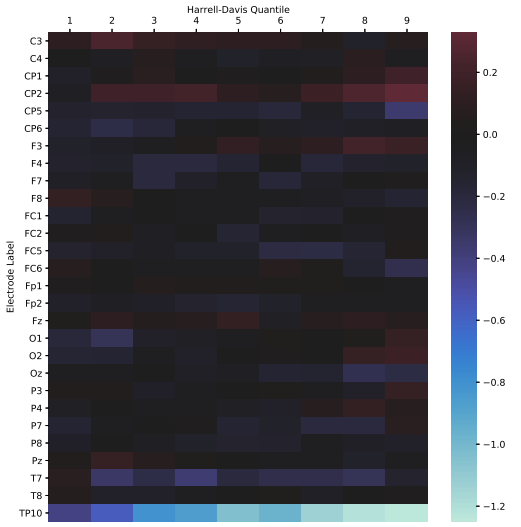


(a) AE vs Raw - SNR condition 2

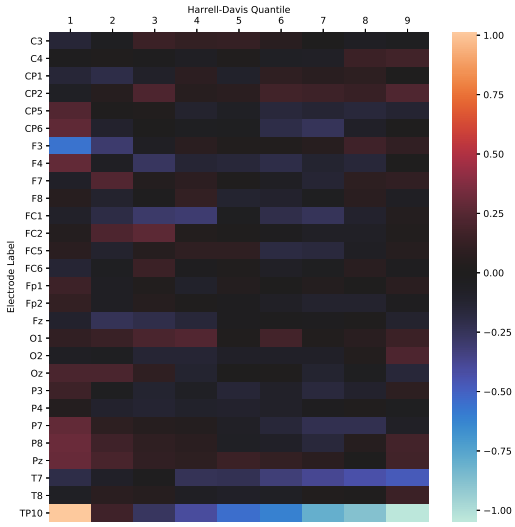


(b) AE vs Raw - PSNR condition 2

Figure A.12: Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 2



(a) AE vs Raw - SNR condition 3



(b) AE vs Raw - PSNR condition 3

Figure A.13: Heat-maps of SNR and PSNR HD quantile differences at channel level for AE reconstructed signals compared to original signals — condition 3