Investigating the Impact of an EEG Device's Quality on the Accuracy of Imagined Speech Classification

William Prior

MSc in Software Systems
The University of Bath
2021-2022

Investigating the Impact of an EEG Device's Quality on the Accuracy of Imagined Speech Classification

Submitted by: William Prior

Copyright

Declaration

This dissertation is submitted to the University of Bath in accordance with the requirements of the degree of Master of Science in the Department of Computer Science. No portion of the work in this dissertation has been submitted in support of an application for any other degree or qualification of this or any other university or institution of learning. Except where specifically acknowledged, it is the work of the author.

Abstract

This Study investigates the extent at which it is possible to achieve similar Classification accuracy's from data produced from a lower quality EEG with 14-channels and a 256Hz sampling rate in the FEIS dataset (Clayton et al., 2020) vs that of the a higher quality EEG with 62-channels and a 1000Hz sampling rate in the Kara One Dataset (Zhao and Rudzicz, 2015). These were evaluated across 4 binary tasks using the SVM, Random Forest and KNN machine learning classifiers and a range of deep and non-deep learning feature extraction methods to understand which is optimal. The results show on the whole the disparity between the device accuracy's to be rather small in all of the binary tasks bar one. The Random Forest algorithm appears to be the classifier which obtains the highest accuracy's across the two results with the non-deep learning feature extractor methods outperforming that of the Siamese Neural Network deep learning method.

Contents

List of Figures

List of Tables

Chapter 1

Code Repository

Please find all the code that was produced during this project in the following GitHub Repository: https://github.com/wp342/Investigating-the-Impact-of-an-EEG-Device-s-Quality-on-the-Accuracy-of-Imagined-Speech-Classificatio.git

Chapter 2

Introduction

Imagined speech is a brain process that sends signals to represent speech prompts (phonemes, vowels, words, etc.) without the use of any auditory stimulation or tongue, lips or hand expressions (Bakhshali et al., 2020). This has many applications including in medicine to help people with disorders such as amyotrophic lateral sclerosis (ALS), muscular dystrophies, brainstem stroke communicate. In these cases the patients have normal brain function but are unable to move their muscles effectively that allow them to speak to those around them. It also has applications in military situations, with particular uses in reconnaissance and special operations. Where communicating silently through the use of electroencephalography (EEG) signals of intended speech instead of vocalisation or body gestures would be incredibly beneficial when having to move silently. So much so that the US Defence Advanced Research Projects Agency (DARPA) started a projected called "silent talk" looking into systems such as this (Bogue, 2010).

Most current literature in the imagined speech classification field discuss' the use of EEG devices which are high-grade scientific pieces of equipment such as a 64-channel Neuroscan Quick-cap, sampling at 1000 Hz, which cost, size and set up make it quite impractical for everyday use. With the increasing availability of the cheaper commercial grade EEG's, such as that used in the FEIS dataset collection (Clayton et al., 2020), further research needs to be done in the field of "lower quality" EEG equipment. Many of which have the bonus' of being more user friendly with the emotive epoc+ (Emotiv, n.d.) boasting set up times of 3-5 minutes and wireless connectivity to the data recording device making it perfect to use day to day. With cheaper and lower quality however also comes less channels, lower sampling rate and poorer performing electrodes, with 14-channels and 256 Hz being used in the most recently mentioned device. However, in the FEIS study there was a comparison of an Support Vector Machine (SVM) model, trained to predict three binary phone classes between the dataset collected in the paper and Kara One(Zhao and Rudzicz, 2015). The outcome showed the model that was trained with the kara one data to be, overall, more accurate in its prediction of the classes but not by a large magnitude. This shows the gap between the two grades of EEG not to be as big as would have been thought – leaving the potential for the right algorithm, that could deal better with noisier signals, to make this difference in performance negligible. Bringing the idea of using such technology live, and on the go, that little bit closer to reality. Therefore, the motivation behind this study is to investigate into potential algorithms or deep-learning system that can bridge the imagined speech classification accuracy gap between the high and low fidelity equipment, allowing this cheaper technology to be used in everyday practical applications.

Chapter 3

Literature and Technology Survey

Brain signals can be captured in a range of ways including invasive, where electrodes penetrate the skin, or non invasive measures. With invasive sampling boasting advantages such as higher information transfer rates and better signal quality, it also has clear disadvantages surrounding the surgical risk and cost in placing the given electrodes. Non-invasive methods include magnetoencephalography (MEG), electrocorticography (ECoG), positron emission tomography (PET), functional magnetic resonance imaging (fMRI) and optical imaging. However, these all have the hindrance of them either being too expensive or immobile leaving them not practical for real life situations (Wolpaw et al., 2002). Electroencephalography (EEG) systems on the other hand, have brought much excitement into this field in recent years. Providing this non-invasive mechanism for measuring electrical signals near the surface of the brain that is cheap, more portable, and has high time resolution.

3.1 EEG

EEG devices are non-invasive systems which measure the brains electric fields via multiple electrodes (channels) placed in various position on the scalp. The electrodes record voltage potentials which are a result of current flowing in and around the neurons on the brains surface. The EEG is said to be able to detect two types of these electrical signals in Action Potentials and Postsynaptic Potentials (PSPs). Action Potentials are very short duration, tiny magnitude, voltage spikes usually about one millisecond, that travel along a neurons axon to the axon terminal (Biasiucci, Franceschiello and Murray, 2019). These neurons are usually arranged perpendicular to the surface of the brain with dendrites at the top and axon terminal at the bottom or vice versa (see figure 3.1). This makes it hard for EEGs to detect these signals due to the potentials created cancelling each other out as a consequence of their opposing orientation. Therefore, the only time it can detect such a signal is when in a certain area of the brain each individual neuron is in the same orientation and fires at the exactly same time to create a potential large enough to be detectable. Hence, most of the EEG signals come from the PSP, these are triggered by action potentials which cause neurotransmitters (chemicals which diffuse across the gap between neurons) to be released by the synaptic vesicles of the pre-synaptic neuron (Evans, 2021). These build up on the receptors of the membrane of the postsynaptic cell. After a certain build up threshold, an ion channel forms creating a change in potential across the cells membrane. These have a larger duration between tens and hundreds of milliseconds and are generally confined to the dendrites of the cell as show in in figure 3.1.

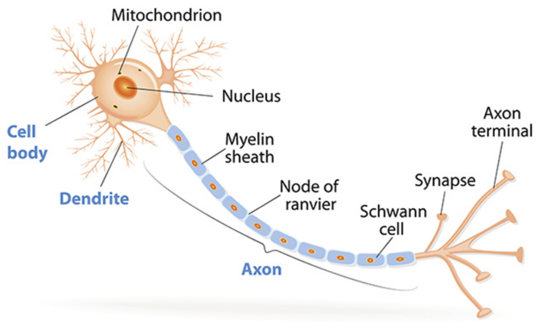

Figure 3.1: Diagram of Brain Neuron (of Queensland: Queensland Brain Institute, n.d.)

These dendrites are usually organised in parallel to the brain surface making the signals easier to record by surface electrodes (Hypotheses, n.d.).

3.1.1 Noise Sources in EEG

The signals produced from an EEG are very noisy with signal distortion. Also known as artifacts, this noise can be located from two main sources, the participant themselves and the equipment and environment. This physiological noise from the participant can be caused by various noise generators including the cardiac signal (electrocardiogram, ECG), muscle contraction (Electromyogram, EMG) and ocular movement caused by moving the eye and blinking (electrooculogram, EOG). As the EEG channels are placed on the surface of the skin there is also the added skin potentials to consider with the insulating factor of the skin, and ionic potential of the sweat glands (Repov, 2010). The environment can cause noise with electrical interference from alternating current (AC) power lines, lighting and large electronic equipment in the vicinity such as computers, displays and TVs, mobile phones etc. Noise can also be created from the EEG itself, this can come from movement in its wires, poor electrode contact for example if there is hair in the way or the pads are too saturated or not saturated enough with the conducting liquid being used (Teplan et al., 2002).

There are many ways to get rid of this noise from signal but the easiest way is to eliminate it from the test environment in the first place. In the case of physiological sources this is not as easily done especially in the cases of ECG and skin potentials. However, there are some cues that can be used to mitigate the other sources such as for EOG, informing the test participant to focus on a particular point and not blink during the short critical times of recording. Setting the participant up in a comfy seating position before testing to reduce EMG and scratching the channel positions on the participants skin with an hypodermic needle or abrasive creme before applying the electrodes to reduce the skin potential and increase the signal to noise ratio (SNR). As for environmental factors these can be more easily removed with the use of Faraday cages and removing any AC equipment only using equipment that used direct current (DC). Although, this may require advanced planning and costly building.

3.2 Data Collection

Being a very new field, the number of open source EEG data-sets available for imagined speech are still very limited. With no universal paradigm that is said to give the best results currently found, each data set has been sampled in different ways. This includes the words or sounds

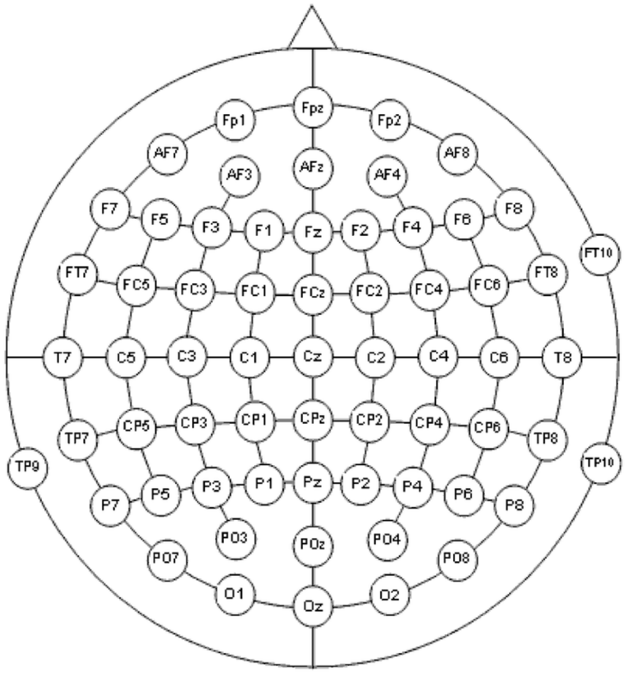

Figure 3.2: Diagram of 10-20 electrode placement of 64 channel EEG electrodes on the head of a participant (Demin, Panischev and Demina, 2016)

being recorded, alongside participant stimulus, the experimental environment and the EEG devices used to carry these out.

3.2.1 ASU

Referred to in the literature as the 'ASU dataset' (Nguyen, Karavas and Artemiadis, 2017), consists of 15 healthy subjects between the age of 22 and 32, with an 11: 4, male:female split. The recorded word/phoneme set entailed three groups, short words - 'in', 'out' and 'up', long words - 'cooperate' and 'independent' and three vowels /a/, /i/ and /u/. This data set focused on imagined speech only, meaning the participants were instructed to pronounce the words internally and to avoid any muscle movements or outward vocalization. The paradigm consisted of an hour-long session where a single group was recorded i.e. the list of short words. Each session consisted of 100 trials per word, these would be prompted randomly on a screen to keep the participant engaged. A series of periodic beeps in intervals T was used to instil a rhythm into the participant, in which they should be imagining the word pronunciations. Each cue lasts 7 x T s, and the trial ended with a rest period of approximately 2 seconds in-between queues. The T value varied depending on the length of the word from 1 – 1.4 seconds. The ASU dataset was recorded using a 64 channel Brainproducts ActiCHamp amplifier system using the 10-20 system as shown in figure 3.2 sampling at 1000Hz.

3.2.2 Kara One

The Kara One dataset (Zhao and Rudzicz, 2015) comprises of 12 healthy participants with a mean age of 27.4, with an 4:8 male:female split. The recorded content consists of 7 phonemic/syllabic prompts (/iy/, /uw/, /piy/, /tiy/, /diy/, /m/, /n/) and 4 words derived

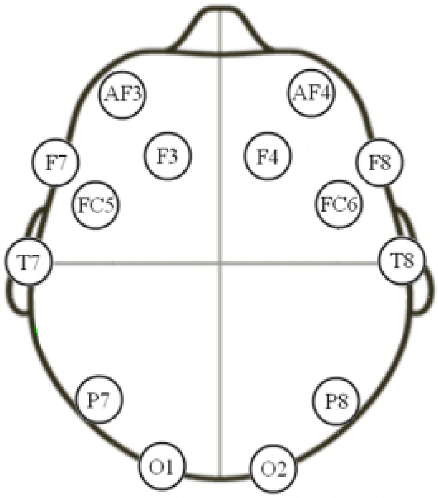

Figure 3.3: Diagram of 10-20 electrode placement of a 14 channel EEG electrodes on the head of a participant (Moctezuma and Molinas, 2019)

from Kents list of phonetically similar pairs such as pat, pot, knew, and gnaw. It was recorded in an office environment with the using a 64-channel Neuroscan Quick-cap EEG device arranged in the 10-20 system, sampling at 1000 Hz and a Microsoft Kinect (v.1.8) camera (to measure facial and audio signals). The paradigm was split into three different sections, thinking, speaking and hearing in which the participant was placed in front of a screen which they had 5 second intervals of rest, on-screen visual and audio stimulus (hearing), imagined pronunciation of the stimulus (thinking), and finally actually speaking the stimulus (speaking). Each prompt was presented 12 times totalling 132 trials with the participants given time to rest every 40 trials.

3.2.3 FEIS

The FEIS dataset (Clayton et al., 2020) comprises of 21 English speaking participants. 16 English phonemes were recorded that represented a balanced categorical spread of binary phonological features, these can be best described in figure 3.3. The FEIS data set was recorded using the mobile 14-channel Emotiv EPOC+ EEG device, using the 10-20 system, at a sampling frequency of 256 Hz, using a screen for visual prompt and 16 audio prompts pre-recorded in the participants own voices to be used a stimulus during the recording. Participants performed mainly the same process as that of Kara one with the same timing intervals. However, their environment was different, instead they were made to sit alone in a comfy chair in a hemi-anechoic chamber. This was to try and mitigate any brainwave contamination from external stimuli in the participants immediate surroundings. Also, instead of each phoneme being heard once, the participant was listening to a steady repetition of the same word at a steady rate 5 times, which they were to repeat in the final two stages at this same tempo. This was repeated 10 times per phoneme totalling 160 trials per participant.

3.2.4 Analysis

The ASU data set has some great qualities, with its large amount of samples per word, its good use of visual cues and its use of the audio rhythm to help the participant create the steady rate in which they should be thinking about the given experimental word. This has the bonus of being able to split each sample up further to create even more data per word recorded. It however does not mention much about the procedures it used to mitigate any external/environmental or physiological noise, this could potentially mean that the data would require large amounts of artifact removal when it came to the signal processing. The Kara One paradigm made efforts to reduce physiological factors by asking participants to move as little as possible during the trials, but as it was not just focusing on the imagined speech meant that fewer trials per word/phoneme were recorded making the data set slightly smaller. It was also recorded in an office environment meaning that it is highly likely that the external factors of EM interference from computers etc. close by could have caused noise in the recordings. Unlike the ASU data set, it did not randomise the order of the words displayed meaning the participants potentially could have been less engaged - this could have been one of the reasons some participants fell asleep during the experiments. Finally, it also did not use a rhythmic stimuli like ASU this makes it harder when looking solely at the imagined speech to assume at what points the participant was thinking of the given stimulus and if it was repeated and the rate of this repetition. The experimental paradigm used in the FEIS data collection is shown to be a combination of both Kara One with the audio prompt and ASU with the sense of rhythm provided by the repeated prompts. With the added feature of the participants own voice helping the participant know exactly what they need to think, in what tone and at the correct rate. Like ASU, the repetition gives the added advantage of more data per trial making it ideal for an investigation into deep learning which generally performs better the more data it consumes during training. FEIS reduces artifact sources, in a couple of ways. The participant was put in a comfortable chair, in a room on their own in a hemi-aneochic sitting to reduce EMG, audio or visual stimuli, and to reduce any electromagnetic interference from external sources.

3.3 Pre-Processing

To remove any remaining unavoidable artefacts and augment the data into a form in which the classification algorithms can classify it more easily, pre-processing is carried out on the data.

3.3.1 Filtering and Artifact Removal

EEG signals can be classified by their frequency into 5 main groups. Delta band (0.1-4Hz) usually associated with the brain while sleeping, Theta band (4 - 8Hz) associated with subconscious activity, most commonly be observed when in deep relaxation and meditation. The Alpha band (8-13Hz), bridges between the conscious and subconscious mind, this can be most commonly found in adults who are awake with closed eyes. The Beta band (13 - 30Hz) concerns behaviour and actions, i.e. what we see, hear, touch, smell, taste, talking, problem solving, judgment, and decision making. The Gamma band (30-100 Hz) is associated with perception and consciousness however these usually range between 30-70Hz (Kumar and Bhuvaneswari, 2012).

With this in mind many EEG signals are filtered with band pass filters within some of these ranges. For example, the investigation carried out in the Kara one paper (Zhao and Rudzicz, 2015) showed them to band-pass their data between 1 and 50 Hz with the extra pre-processing of the mean values being subtracted from each channel and a small Laplacian filter was applied to each trial, using the neighbourhood of adjacent channels. Meanwhile, in another paper looking into the use of Siamese Neural Networks, they decided to filter between 2 and 40 Hz using a finite impulse response (FIR) filter (Lee, Lee and Lee, 2020). And in another which looked into Deep learning and common spatial patterns they investigated using filtered data between 8-70 Hz, using a notch filter at 60 Hz to remove line noise (Panachakel, Ramakrishnan and Ananthapadmanabha, 2020).

To remove any further remaining artefacts there are other methods that can be applied including independent component analysis (ICA), principal component analysis (PCA) and adaptive filtering (Jiang, Bian and Tian, 2019). ICA can be used to reduce the dimensionality of data alongside finding all sources of the data that are statistically independent. Going on the assumption that all biological processes run independently, ICA could be used to split all the components from different biological sources. This in turn could be useful when removing artefacts from the EEG signals (Brigham and Kumar, 2010). Many of these have been implemented into automatic artefact identification packages that can be used. For example, the python package mne has an automatic ECG identifier (MNE, 2019) and Matlab's EEGLAB (for Computational Neuroscience, n.d.) has a easy to use GUI that can perform similar actions.

3.3.2 Feature Extraction

Linear and Non Linear Features

In both the Kara One (Zhao and Rudzicz, 2015) and the FEIS (Clayton et al., 2020) papers they make use of list of linear and non linear features by windowing the data into windows roughly 10% the size the data set with a 50% overlap. For ease of writing from now on the linear and non-linear feature extraction methods will be referred to as either linear features of (non-) linear features. On each window created a series of linear features were calculated. In Kara ones case, these consisted of the mean, median, stand deviation, variance, maximum, minimum, maximum+minimum and vice versa, the sum, spectral entropy, energy, skewness and kurtosis. Looking into the code of the FEIS data set they also used similar metrics plus absolute mean, absolute maximum ,absolute min, curve length, integral, sample entropy, permutation entropy, Singular Value Decomposition entropy, approximate entropy, Petrosian fractal dimension, Katz Fractal Dimension, Higuchi Fractal Dimension, root mean square, Detrended fluctuation analysis (DFA). These metrics were then fed into a range of classification algorithms to analyse the results.

Mel Frequency Cepstral Coefficients (MFCC)

More commonly used in automatic speech recognition (Muda, Begam and Elamvazuthi, 2010), MFCCs have starting to gain success in imagined speech classification. (Cooney, Folli and Coyle, 2018), showed its successes against that of linear and non linear features showing it to be overall substantially more effective when using it with both a decision tree and an support vector machine. The data used was windowed and for each, 13 MFCC coefficients were extracted.

MFCCs are based on the fact human hearing perceptions are not able to perceive frequencies over 1KHz and therefore focusses on the human ears critical bandwidth frequency. The method uses two filters which are spaced linearly below 100Hz and logarithmic from there on. The process involves a series of filtering windowing, Fast Fourier Transform (FFT) and discrete cosine transform (DCT). With the final part of the process involving converting the log Mel spectrum into the time domain using the DCT as shown in the equation:

$$C_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{n} m_j cos(\frac{\pi i}{N}(j - 0.5))$$

Where N is the number of filterbank channels

Deep Learning

(Lee, Lee and Lee, 2020) showed that deep neural networks don't just have to be the end classifier, with their use of a Siamese Neural network (SNN) to feature extract and reduce the dimensionality of the data in a way that improved the accuracy of the KNN algorithm for a multi-class classification system. Siamese Neural networks consist of the same neural network being implemented in parallel and then combined in the final layer. The structure of the neural network used in this paper can be seen in the figure 3.1 with relu activation functions being used at every fully connected and convoluted layer.

Table 3.1: The Neural network architecture of one of the networks (one of the embedding networks) in the SNN, found in the literature(Lee, Lee and Lee, 2020)

Layer	Input	Output	Kernel
Conv1	6 x 512 x 1	2 x 103 x 64	5 x 5
Maxpool1	2 x 103 x 64	1 x 52 x 64	2 x 2
Conv2	1 x 52 x 64	1 x 52 x 128	1 x 3
Maxpool2	1 x 52 x 128	1 x 52 x 128	1 x 3
Conv3	1 x 52 x 128	1 x 52 x 128	1 x 2
Maxpool3	1 x 52 x 128	1 x 52 x 128	1 x 2
Flatten	1 x 52 x 128	6656	–
FC1	6656	1024	–
FC2	1024	512	–
FC3	512	256	–
FC4	256	8	–

The network was trained using the ADAM optimizer and the contrastive loss function. Used instead of the normal cross entropy function used in most deep learning models as in this case the Siamese network takes in two input samples. The contrastive loss then tries to class each by how close they are to each other – if they are close, they are of the same class, if they are different, they are further apart. This contrastive loss can be calculated using the following equation:

$$L(x_1, x_2, y) = \frac{1}{2}yD^2 + \frac{1}{2}(1 - y)max(m - D, 0)^2$$

Where x_1 and x_2 signify the two inputs to the Siamese network, y the label either being a 1 or 0 depending if they are the same class or different respectively. D denoting the euclidean distance between the two inputs and m denoting the margin which must be greater than

0. This falls into the hands of the K-Nearest Neighbour (KNN) classifier which was used on the output data of the embedding networks of the SNN later on. As the SNN takes in two inputs it is meant that the data needs to be paired together increasing the size of the training set, a great advantage for when working with small datasets. In this paper they used a dataset consisting of 6 Spanish words and 5 vowels, recorded using 18 channels using Ag-Ag Cl electrodes directly attached to the scalp sampling at 1,024 Hz. Before it could be fed into the network it had to be down sampled to 128 Hz. The output KNN used in this case used a K value of 5, the classification of which gives an average accuracy of 9.93% higher than any other model that has used the same dataset proving it to be an effective method. The main advantages of this form of network are the way it can reduce high dimensional data and learn effectively on small data sets. Another paper was brought out a year later with further developments in this area as can be seen in appendix A.1

3.3.3 Channel Selection/ Dimensionality Reduction

Due to the state of the data produced from an EEG with multiple channels and high numbers of samples, there can often be large amounts of data. This can lead to long training times or the data to be in the wrong format for the given classification model. Therefore, dimensionality reduction methods can be used to fix this issue and in some cases increase the output accuracy by removing any of the irrelevant data before training starts.

Pearson's Correlation Coefficient

Denoted by r the Pearson's correlation coefficient is a measure of linear association between two variables, x and y. This coefficient's magnitude can be between -1 and 1 with the former indicating a complete negative correlation and the latter indicating a complete positive correlation and 0 indicating no correlation at all (the bmj, n.d.), (Adler and Parmryd, 2010). This can be calculated using the following:

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2(y-\bar{y})^2}}$$

This was used in both the FEIS (Clayton et al., 2020) and Kara one (Zhao and Rudzicz, 2015) papers to reduce the dimensionality of the data by ranking the features they extracted by this correlation features and then picking the first $N \in [5:100]$ features with the highest rank.

Common Spatial Patterns

Common spatial patterns (CSP) was initially used for classification of multichannel EEG for imagined hand movements (Wang, Gao and Gao, 2005). It was then also experimented with the imagination of Geometric objects (Costa, Patow and Azorin, 2019) and now more recently with spoken words and imagined speech (Panachakel, Ramakrishnan and Ananthapadmanabha, 2020). The main idea of it is to create a linear transformation matrix consisting of a weight for each channel of the EEG. This matrix is to be projected on the signal set to select a given number of channels that maximise the variance between two-classes, reducing its dimensionality. The method used to do this consists of simultaneous diagonalization of the covariance matrices of both classes, to maximise the following equation:

$$J(w) = \frac{w^T X_1 X_1^T w}{w^T X_2 X_2^T w} = \frac{w^T C_1 w}{w^T C_2 w}$$

Where T is the matrix transpose, matrix X_i contains the EEG signals of class i, with data samples as columns and channels as rows, w is the spatial filter and C_i is the spatial covariance matrix of class i.The output giving a matrix of (number of channels x 1) the channels associated with the top N values in this matrix are then selected and used for classification. (Panachakel, Ramakrishnan and Ananthapadmanabha, 2020) experimented with the number of channels that should be selected after running the CSP and found that for a 64 channel EEG selecting 9 channels, which were later feature extracted via discrete wavelet transforms and classified through a deep neural network (DNN), actually increased the classification accuracy alongside reducing the datas dimensionality with a large drop off of over 5% either side of this value.

3.4 Signal Classification

Once the data has been put into the appropriate form it is now ready to be classified. The classifiers train on the labelled data and then once trained predicts non-labelled data. There are many types of classifiers including the basic Machine learning methods in the SVM, KNN and Random Forest and also more complex deep learning methods. The latter of which being explained in appendix A.2.

3.4.1 SVMs

SVMs have been used in a range of papers for imagined speech classification due to their significant accuracy but comparably low computation power. The main parameter which defines the accuracy of an SVM is the kernel choice. Research shows there to be a large range of kernels including, quadratic, radial basis function (rbf), linear and Gaussian. The Kara One paper (Zhao and Rudzicz, 2015) shows the quadratic kernel to produce an average classification of 55.4% across all subjects, modalities, and classes. (Cooney, Folli and Coyle, 2018) showed the linear kernel to not be as effective in this area, although it does highlight the importance of applying the appropriate feature extraction method. With an obvious difference between using linear and non- linear vs using the Mel Frequency Cepstral Coefficients (MFCC) method. Another paper by Matsumoto and Hori (Matsumoto and Hori, 2014) experiments with the use of a Gaussian kernel which managed to gain an average classification accuracy across vowels of 77%. Unlike the previous two however, it was carried out using an EEG of 128 channels compared to the 64 of the previous. The dataset is also slightly different with it being a smaller size identifying vowels only rather than a range of phonemes and constants, which would have an effect on the variance of the results and hence the average accuracy. Taking a wider view on literature it is clear that SVM is one of the more popular machine learning algorithms used in the field due to its constant respectable classification results.

3.4.2 K-Nearest Neighbour (KNN) and Random Forest

K-nearest Neighbour (KNN) and Random Forest were initially researched through a paper by Akshi and Rao (Akshi and Rao, 2021). This outlined the classification of 16 similar phonemes plus that of the words “yes” and “no”. A single participant used the same EEG device as the

FEIS data set to collect the data. For both word and phoneme data sets the KNN algorithm used a K value of 3 and the random forest used 100 and 1000 trees respectively. The data was fed into the two algorithms with three different forms of feature extraction – the initial used a mixture of linear and non linear features similar to that mentioned in section 3.3.2, the second used a 154-coefficient level 4 Discrete wavelet transform and set three was the combination of the first two. The Random Forest algorithm seemed to dominate average accuracy in all 3 forms of the data. The phoneme classifications top result happening mostly with the first set of features. With some comparable and even some improvements to the results found in initial Kara one and FEIS papers. This paper however only uses a single participant meaning there would be less variance in the results. An investigation into the value of k in the KNN algorithm was investigated in (Hashim, Ali and Mohd-Isa, 2018). It showed the classification, in a linear increase of k in increments of 1, to be at its maximum when k=3, this is supported by the same value being used in the previous paper and for the cosine function to be used as its distance metric. Nevertheless, this papers data set contains only two words of yes and no – therefore the likelihood of there being a different answer to a data set where there are more data categories is very likely. Random forest was also found to be used in Moctezuma et al. (2018) in which they are identifying 5 words in Spanish, with the added help of 10-folds cross validation. An 85% accuracy was achieved with a python implementation using the library scikit-learn with no specified number of trees (default is either 10 or 100 depending on the version) with a maximum depth of 5 per tree.

Chapter 4

Hypotheses

4.1 Main Hypothesis

Explore the extent to which it is possible to achieve similar classification accuracy's for imagined speech data collected from a lower quality EEG device vs data collected using a higher fidelity EEG using the same methods.

4.1.1 Sub-Hypothesis

Data produced from deep Learning feature extraction methods will out perform data produced from non-deep learning feature extraction methods when run through the same machine learning classifiers on imagined speech Data.

4.1.2 Sub-Hypothesis

Using Common Spatial Patterns and Pearson correlations to reduce data dimensionality can increase machine learning algorithms accuracy for imagined speech classification.

4.1.3 Sub-Hypothesis

Support vector Machines (SVM) will out perform Random Forest and K-Nearest Neighbour in classification accuracy of imagined speech data.

Chapter 5

Methodology

5.1 Overview

The overall methodology can be split into two main sections which were applied in the same way to both the Kara One (Zhao and Rudzicz, 2015) and the FEIS (Clayton et al., 2020) data sets. The first was initial tuning of the machine learning variables with a range of non-deep learning feature extraction inputs to find the optimal values and gain insight on sub-hypothesis 4.1.3.Using parameters mentioned in the literature as a good starting point. These optimal values were then used to help prove sub-hypothesis 4.1.2 in which a range of dimensionality reduction/ channel selection techniques were carried out on the optimal values for each optimal configuration found for each ML classifier.

Meanwhile, a deep learning feature extraction method was being developed and trained in the form of a Siamese Neural network to investigate 4.1.1. This was then tested using the same machine learning methods used in the previous section in some cases using the parameters obtained from the initial parameter sweep.

The results from each data set for all of the mentioned stages were then analysed and the Kara One and FEIS datsets were then compared to look into the extent of the main hypothesis 4.1.

With the large variance in the EEG signals from participant to participant each of the following stages mentioned were performed on per participant basis.

5.2 Data Collection

5.2.1 Initial Proposal

Following the main Hypothesis the initial idea was to perform a preliminary study on a higher fidelity EEG device data set that samples at a high frequency (Kara One) and then collect the same words/ phonemes in a new data set using the lower sampling frequency emotiv EPOC+ (Emotiv, n.d.) device using the following experimental paradigm as seen in appendix B.1. However, due to availability of the device and issues other colleagues were experiencing with it, this did not happen and the contingency plan was used instead.

5.2.2 Data Used

The preliminary study was still carried out using the Kara One data set using this as the 'high fidelity' data set due to its high sample rate and large number of channels. With FEIS already shown to make comparisons with Kara One, it seemed logical to use this data set as the 'lower fidelity' data set to perform the machine learning on.

As this study is only into the imagined speech of the data only the 'thinking' data was extracted from the data sets and split into four binary groups similar to those found in the FEIS and Kara One papers. These consisted of vowel-only vs consonant (C/V), presences of nasal (± Nasal), presence of bilabial (± Bilab.), and presence of a back vowel (± Back). These can be split up as seen in the table below:

Table 5.1: Kara One Binary Data Classes used for this investigation

Classes	+	−
C/V	/m/, /n/	/uw/, /iy/
± Nasal	/m/, /n/, gnaw	/tiy/, /piy/, pat
± Bilab.	/piy/	/tiy/
± Back	/uw/	/iy/

Table 5.2: FEIS Binary Data Classes used for this investigation

Classes	+	−
C/V	/m/, /n/	/u/, /i/
± Nasal	/m/, /n/, /ŋ/	/t/, /ʃ/, /p/
± Bilab.	/p/	/k/
± Back	/u/	/i/

Analysing tables 5.1 and 5.2 you will notice how they have both been given the same number and as similar as possible words/phonemes. This is to allow the two data sets to be compared more fairly and with similar sized training sets.

5.3 Pre-processing

5.3.1 Loading the Data Sets

Kara One

The trials are noted in the literature to be 5 seconds in length, at 1000Hz, proceeding to give 5000 samples. However, some trials were less then this with some in the low 4800s samples. To make the entire data-set uniform size, some experimentation with running the raw data through the three mentioned machine learning algorithms was undertaken with zero padded data to 5000 samples and truncating the data to 4750 to allow ease in windowing. It was found that the best results appeared from the truncated version of the data so was used from here onward.

FEIS

As all data in FEIS was uniform nothing was changed and the full 5000 samples per trial was used. However, due to the nature of FEIS' paradigm with its assumed rhythmic repetition of the imagined speech, the data can be split up further to create a larger data set. Therefore, each 5 second sample of the words/phonemes being imagined was split into 5 2000 ms windows and used as separate inputs to the feature extraction methods. In this case, as there was overlap, they still had to be kept as a group before they were split into test and train to make sure there was no data bleed between results.

5.3.2 Filtering

Before all feature extraction methods, both data sets were filtered between 7 and 70 Hz to cover the range of frequencies that the brain is awake and thinking and therefore not unnecessarily remove any useful information, as mentioned in section 3.3.1, with a notch filter at 50 and 60 Hz to remove line noise. Following the approach of (Panachakel, Ramakrishnan and Ananthapadmanabha, 2020), this range was also chosen with the forward thought of experimenting with CSP, this seeming logical with their successes using this range. The the notch filtering was performed more for the Kara One data's benefit with the Emotiv EPOC+ device doing this automatically, therefore giving little effect on the FEIS data.

In the case of the deep learning feature extraction the Kara One data was also down sampled to the same sample frequency as FEIS at 256 Hz before it was filtered to decrease training time.

5.3.3 Feature Extraction

To augment the data into a form that would give better results for the classifiers, feature extraction was carried out. Three types of feature extraction was used, these can be grouped into two groups of non-deep learning consisting of linear and non linear features and MFCCs and deep learning feature extraction consisting of the Siamese neural network. The latter also performing dimensionality reduction at the same time.

Non-Deep Learning Feature Extraction

With the thinking that splitting the EEG signals into statistically independent components before applying the feature extraction methods in an attempt to make each signals features more prominent to the proceeding methods.The linear/non-linear and MFCC feature extracted data was output in two forms, one which performed ICA on each trial before the feature extraction and one which only carried out the feature extraction method. The ICA was configured using sklearns FastICA (scikit learn, n.d.a), with the number of components set to the number of channels of the dataset being used. It therefore was not used to reduce the data dimensions but as a tool to try and fix the cocktail party problem of signals, separating them into their respective sources. A complete list of feature extraction methods carried out for both datasets can be seen in table 5.3

Linear and non-linear Features: Using the process in both (Clayton et al., 2020) and (Zhao and Rudzicz, 2015), each data set was windowed with a window size of approximately 10% of the data length with a 50% overlap. This being 500 samples in the case of Kara

Table 5.3: List of feature extraction methods carried out using non-deep learning methods

FEIS	Kara One
Linear Only	Linear Only
Linear with pre ICA	Linear with pre ICA
Split Linear Only	MFCC only
Split Linear with pre ICA	MFCC with pre ICA
MFCC only	-
MFCC with pre ICA	-
Split MFCC only	-
Split MFCC with pre ICA	-

One, ignoring the truncation, and 128 samples in FEIS. Using the linear features described in section 3.3.2 and much of the code produced for the FEIS dataset, 27 features were extracted alongside their deltas and double deltas per window leaving there with 81 features in total this gave a trial size of (81 features x 16 windows x 62 channels) which was reshaped to (81 x 992) for the Kara One data set and (81 features x 17 windows x 14 channels) which was reshaped to (81 x 238)for the FEIS dataset. Notice how there was one less window in the Kara one data due to the truncation that was mentioned in section 5.3.1.

MFCCs: As mentioned in the Linear Features each data set was windowed with previously stated window sizes and 50% window overlap. MFCCs were then calculated on each channel in each window with a MFCC calculated produced by James Lyon (Lyons, 2020). Like (Cooney, Folli and Coyle, 2018), this managed to produce 13 coefficients per channel per window for the Kara One and FEIS data leaving it with a data shape of (62 channels x 13 coefficients x 18 windows) reshaped to (62 x 234) and (14 Channels x 13 coefficients x 19 windows) reshaped to (14 x 247) respectively.

Siamese Neural Network (SNN)

The Siamese Neural Network used was made up of two embedding deep neural networks with the same kernel sizes, number of filters used in each layer as the architecture found (Lee, Lee and Lee, 2020) and stated in table 3.1. However, due to the input size of the data varying due to the truncation of the Kara One data and the use of CSP in some cases the architecture of the network changes depending on the data used, therefore the generalised architecture of the embedding networks can be as seen in 5.4. These two identical networks are brought together in a final layer in which the euclidean distance between the two outputs are calculated and the contrastive loss function is used on this output.

The input data was trialed in two forms. The first consisted of filtered raw data (and also downsampled in Kara One's case). This led to trial sizes of (14 x 1280) for FEIS and input size of (62 x 1216) for Kara One. This following the success of the classification using the feature extraction architecture found in (Lee, Lee and Lee, 2020), it was decided it would be best to see how the new data would perform using the same layer structure and parameters.

The second data augmentation took a further step and applied CSP after the appropriate filtering. The first 9 channels that gave the largest variance between the two classes was then chosen in accordance to the successes found using this value in the (Panachakel,

Table 5.4: The general embedding neural network architecture used for the feature extraction

Layer	Input	Output	Kernel
Conv1	$N_l x S_l x1$	$N_1 x S_1 x64$	5 x 5
Maxpool1	$N_1 x S_1 x64$	$N_2 x S_2 x64$	2 x 2
Conv2	$N_2 x S_2 x64$	$N_2 x S_2 x128$	1 x 3
Maxpool2	$N_2 x S_2 x128$	$N_2 x S_2 x128$	1 x 3
Conv3	$N_2 x S_2 x128$	$N_2 x S_2 x128$	1 x 2
Maxpool3	$N_2 x S_2 x128$	$N_2 x S_2 x128$	1 x 2
Flatten	$N_2 x S_2 x128$	$(N_2 * S_2 * 128)$	–
FC1	$(N_2 * S_2 * 128)$	1024	–
FC2	1024	512	–
FC3	512	256	–
FC4	256	8	–

Ramakrishnan and Ananthapadmanabha, 2020) paper. This paper although did not use a SNN, did use a DNN for its investigation and was intriguing to see if the same results could carry over to its Siamese brother. Therefore the trial sizes for this investigation was (9 x 1280) for FEIS and input size of (9 x 1216).

One of the main reasons for investigating the use of the SNN was its ability to learn from small data sets by pairing up the data with each other to increase the dataset size from the same amount of original data. Therefore, before the trial lists could be input into the SNN, each trial was then paired with every other trial in the class set apart from itself, increasing the size of the class set. If the pairs were of the same class then the pair are labeled with a 1 and if they are of different classes they are labelled 0. Each pair and label are then fed into the Siamese Neural network for training. One advantage of neural networks in this case is they take in multidimensional data, therefore no flattening formatting had to be done to the data before it was fed into the model. The model was trained for 1000 epochs using a batch size of 100, saving the model every 100 epochs. The 1000 epochs was again deduced from (Lee, Lee and Lee, 2020) with a slightly reduced batch size due to the small amount of data produced even after pairing. This was to allow for at least 2 full batches per epoch to make sure the network weights were getting update more than once per epoch. Due to the batchsize and the number of epochs needed to be run, the models were trained and analysed on Google Collab. Once the model was fully trained the output of the embedding networks were then used to train the machine learning classifiers leaving each trial an 8 dimensional array.

5.4 Classification

For classification the data was split 80, 20 in train and test using 5-fold cross validation. Apart from the deep learning method all of the data was split after feature extraction. Therefore, both the linear and MFCCs were calculated on each individual trial to not allow for any data bleed between test and train. Support Vector Machine (SVM), Random Forest and k-Nearest Neighbour (kNN) were implemented using sciKit Learns SVC (scikit learn, n.d.d), RandomForestClassifier (scikit learn, n.d.c), KNeighborsClassifier (scikit learn, n.d.b) packages.

5.4.1 Linear and MFCC data Classification

As these packages only took in 2 dimensional data, each individual linear and MFCC trial were concatenated into one list before running them through the algorithms. There was some variation in the kernels, number of trees and distance metric used to work out what gave the optimal result. The research shown in section 3.4.1, showed a range of kernels to be effective on the results, therefore a range of kernels were tested. The quadratic kernel was used with its success shown in the Kara One paper, this was configured using SVC's 'poly' kernel with an extra parameter required to enter the degree in the setup. The Gaussian/rbf kernel also had some merit in (Matsumoto and Hori, 2014), there this was also implemented alongside the linear kernel which was said to not perform quite as well so was to be used as a baseline set of values. The random forest needed the number of trees/n_estimators to be decided. The literature showed values being used from 10 to 1000, although after some inital investigation there was not much difference in output between 500 and 1000 estimators. Therefore, to save computation time allowing more runs to be performed the number of estimated was capped at 500 and the increments from 10 can be seen in table. (Moctezuma et al., 2018) points the best distance metric to be the cosine distance - this was the first to be tested using the KNN and was later added to by the metrics shown in table 5.5.

Table 5.5: Machine Learning Parameter Testing

ML method	Parameter Type	Parameters Tested
SVM	Kernel	poly, rbf, linear
Random Forest	Number of Trees	10, 50, 100, 500
KNN	Distance Metric	cosine, euclidean, manhattan, minkowski

Table 5.5 shows the Parameter types which were implemented on the Linear and MFCC features using 5 fold leave one out cross validation, with the k value for the KNN algorithms set at 3 as advised from the literature as seen in section 3.4.2. Once this was complete, the parameter from each machine learning method with the highest total classification accuracy was chosen to investigate the effect of dimensionality reduction using Pearson Correlation and Common spatial Patterns and the classification accuracy with both feature extraction datasets. Each dimensionality reduction method was tested with values starting from 3 features/channels, depending on the feature extraction method, up to the maximum number of features/channels that were available, this being the number of channels of the data set in MFCCs case and 81 in the (non-)linear data.

5.4.2 SNN Data

To test the output data of the SNN, the original SNN training data was paired up with itself and run through the SNN, while it was not in training mode. As both the embedding networks are identical they both output the same value for the same input, therefore the output of one of the embedding networks was taken as the feature extracted data. After running the class set through the SNN a list of trials with shape 8 is gained. This training data is then fed into the three machine learning algorithms separately with its original classification labels to train them. The test data is then run through the SNN in the same process, the output of which is input into the trained machine learning algorithms without a label, these then try and classify this test data.

Due to the length of time and computation it takes to train these models, there was less room for testing a range of machine learning parameters. Therefore, after the research displayed in section 3.4, it was decide that the would SVM use the (poly) quadratic kernel, Random Forest should use 100 trees, and KNN should use the cosine distance metric with the k value of 5, the k value coming from that being used in (Lee, Lee and Lee, 2020). As the SNN model was saved every 100 epochs this process was completed 10 times, between 100 and 1000 epochs. This outlines the models effectiveness at extracting features that the machine learning models can identify as it is exposed more and more to the training data. 5 fold cross validation was used, meaning 5 different SNN models were trained and tested for each 100^{th} epoch per data class.

Chapter 6

Results

In this results section we present the findings of the initial machine learning parameter investigation across the feature extraction method as seen in Table 6.1, and figures 6.1, 6.2 with the remaining graphed results found in appendix C. The Channel selection/ dimensionality reduction results are then presented in table 6.2 and figures 6.3, 6.4, 6.5, and 6.6 with the remaining results found in appendix D.The next presented is that of the feature extraction investigation, with results located in table 6.3, and figures 6.7, 6.8, 6.9, 6.10 and 6.11. Finally the Comparative results between the Kara One and FEIS data set can be seen in figures 6.12, 6.13 and 6.14.

As all tasks carried out in this investigation were binary, and each test carried out had an even distribution of both classes the chance accuracy in this case is 50% therefore anything above this is considered above chance.

Table 6.1: Top performing Average Classification Accuracy (std.) with associated feature extraction method and ML parameter of the three ML methods across the two datasets

ML method	FEIS	Kara One
KNN	ICA MFCC Manhattan: 52.62% (1.56)	MFCC Manhattan: 53.79% (3.04)
Random Forest	ICA MFCC 50: 50.73% (0.71)	MFCC 100: 55.09% (3.42)
SVM	ICA MFCC $Poly^2$: 52.08%(0.94)	MFCC rbf: 54.57% (1.92)

Table 6.2: Best performing Total Average Classification Accuracy (std.) of associated non-deep learning feature extraction method of the three ML methods across the two datasets, after channel selection (See table 6.1 for ML parameters)

ML method	FEIS	Kara One
KNN	ICA MFCC: 52.62% (1.56)	MFCC Pearson 6: 55.19% (3.77)
Random Forest	ICA MFCC CSP 6: 52.88% (1.65)	MFCC CSP 33: 55.64% (3.43)
SVM	ICA MFCC Pearson 13: 52.57% (1.05)	MFCC CSP 6: 54.69% (2.67)

Table 6.3: Total Average across Four Binary Tasks of Top Performing Feature Extraction methods classification Accuracy

Feat. Extract.	FEIS	Kara One
(Non-)Linear	R.Forest(10) CSP(67): 52.61% (2.51)	R.Forest(100) CSP(25): 55.32% (4.89)
MFCC	ICA R.Forest(50) CSP(6): 52.88% (1.66)	R.Forest(100) CSP(33): 55.65% (3.42)
SNN	epochs(600) KNN: 52.06% (1.48)	epochs(100) KNN : 52.57% (2.17)

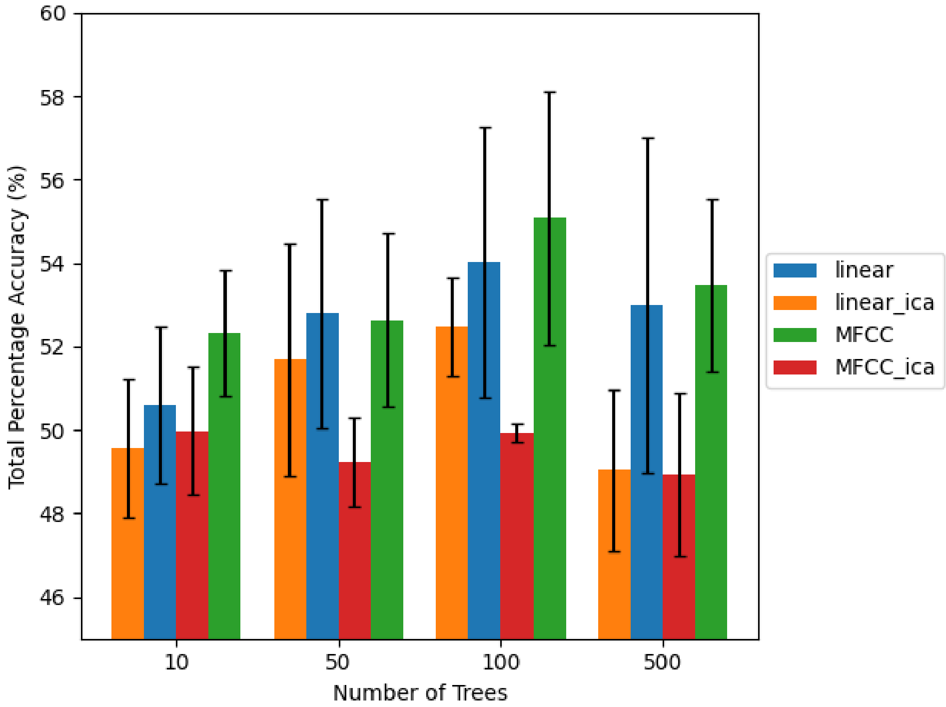

Figure 6.1: Comparison of the varying number of trees/estimators, and feature extracted data used by and input to the random forest classifier against the total average classification accuracy across all participants of the four imagined speech test classes of the Kara One Dataset

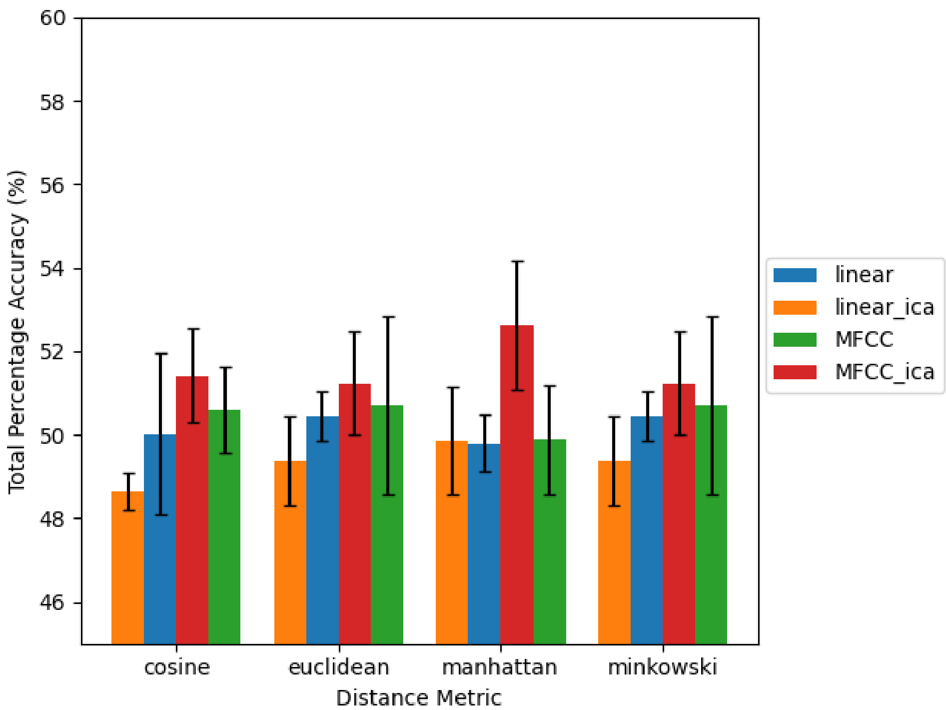

Figure 6.2: Comparison of the different distance metrics, and feature extracted data used by and input to the KNN classifier using a K value of 3 against the total average classification accuracy across all participants of the four imagined speech test classes of the FEIS Dataset

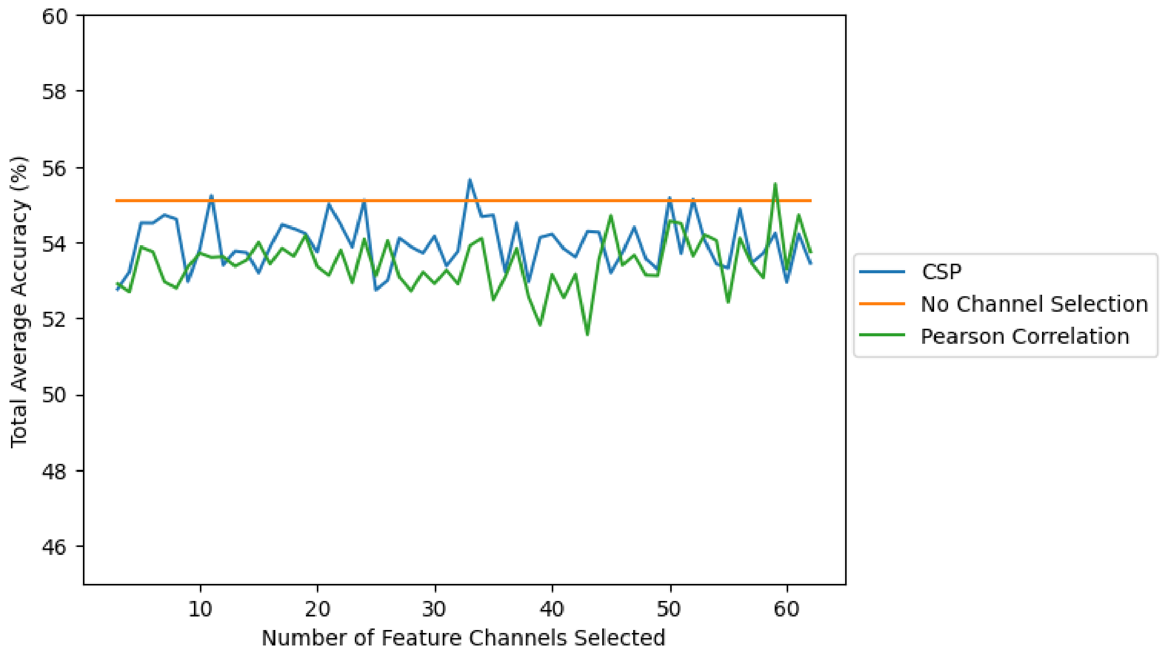

Figure 6.3: Graph of how the total average classification accuracy across the four binary classes of the top performing feature extractor (MFCC) and machine learning algorithm (random forest 100 trees/estimators) for the Kara One dataset changes with the increase in feature channel selection using CSP and pearson correlation methods

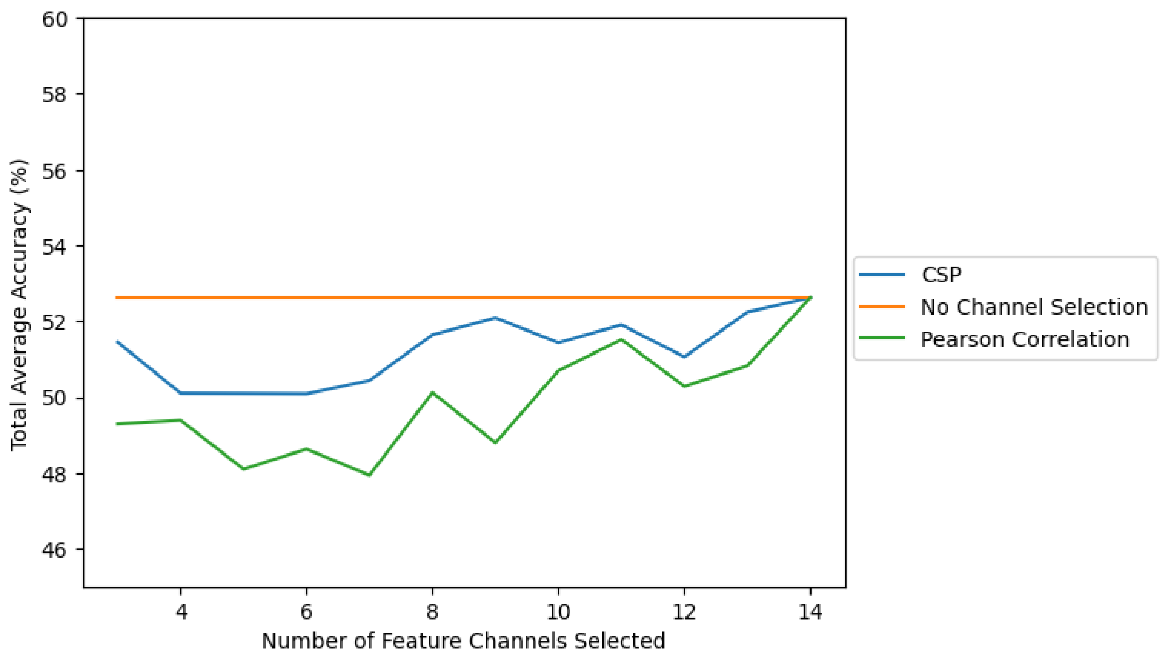

Figure 6.4: Graph of how the total average classification accuracy across the four binary classes of the top performing feature extractor (ICA MFCC) and machine learning algorithm (KNN manhattan distance metric) for the FEIS dataset changes with the increase in feature channel selection using CSP and pearson correlation methods

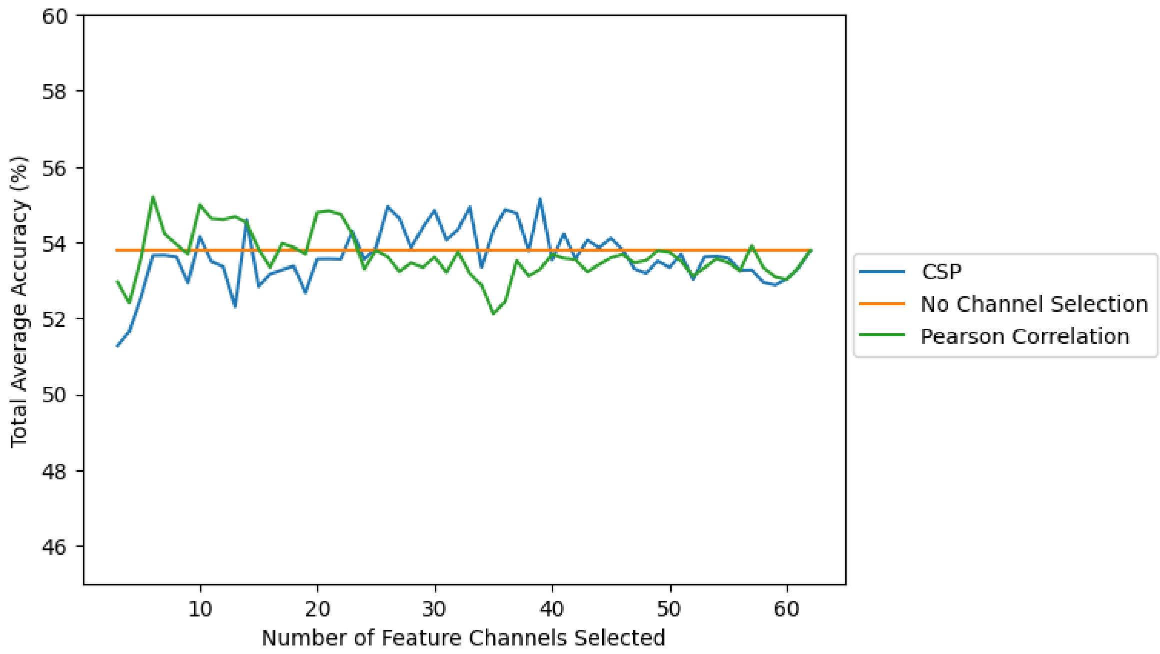

Figure 6.5: Graph of how the total average classification accuracy across the four binary classes of the feature extractor (ICA MFCC) and machine learning algorithm (KNN using a k value of 3 and the Manhattan distance metric) for the Kara One dataset changes with the increase in feature channel selection using CSP and pearson correlation methods

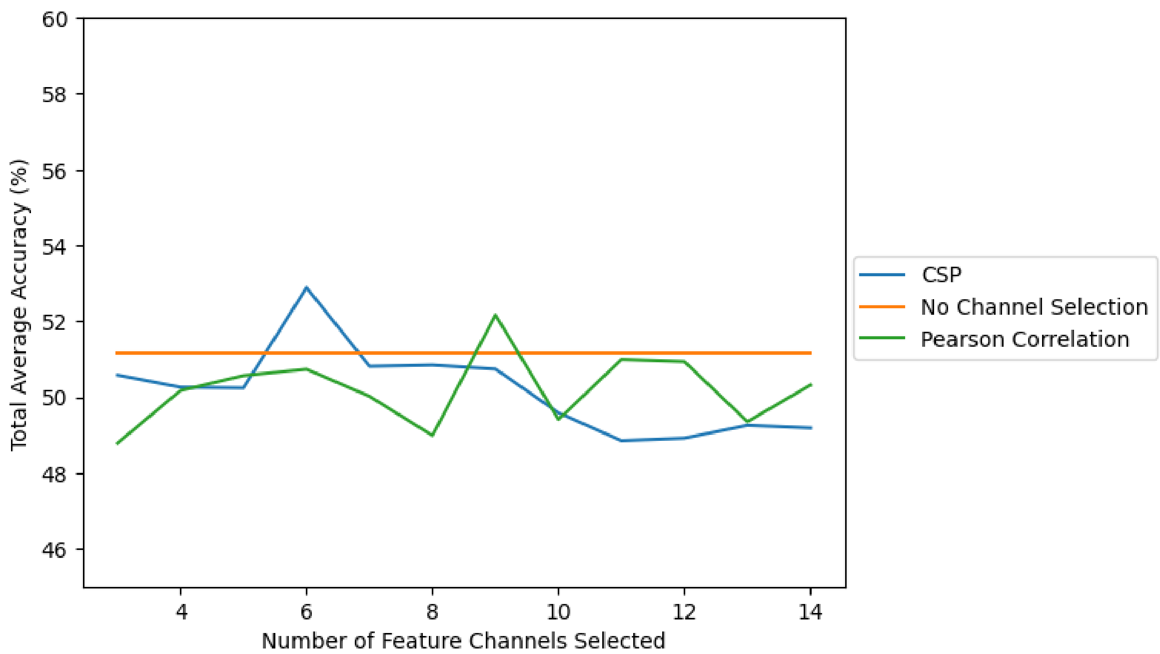

Figure 6.6: Graph of how the total average classification accuracy across the four binary classes of the feature extractor (ICA MFCC) and machine learning algorithm (Random Forest using 50 estimators) for the FEIS dataset changes with the increase in feature channel selection using CSP and pearson correlation methods

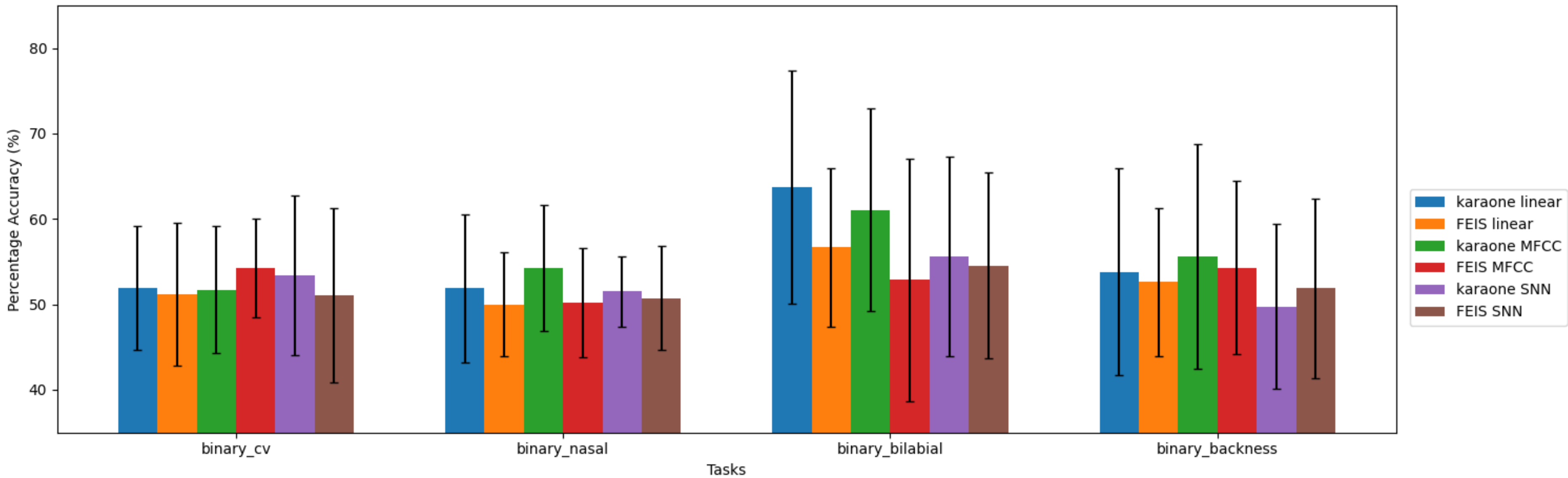

Figure 6.7: Comparison across FEIS and Kara One datasets on Top Performing Feature Extraction methods classification per Task (see Table 6.3 for ML algorithm and further parameters)

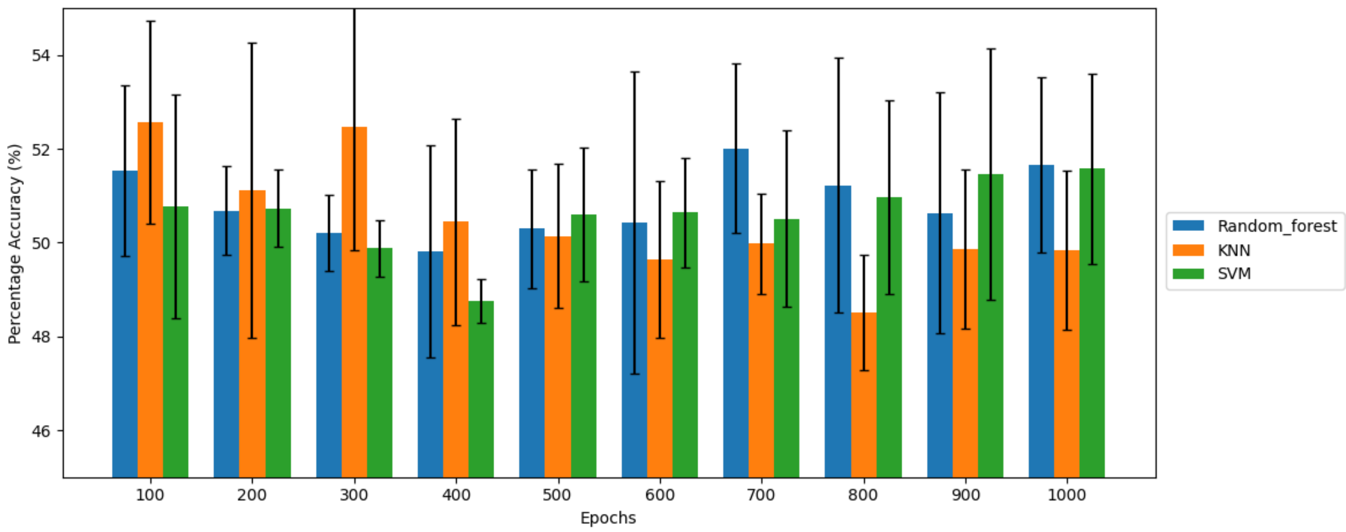

Figure 6.8: SNN trained on Kara One total average ML accuracy's as the number of epochs increased

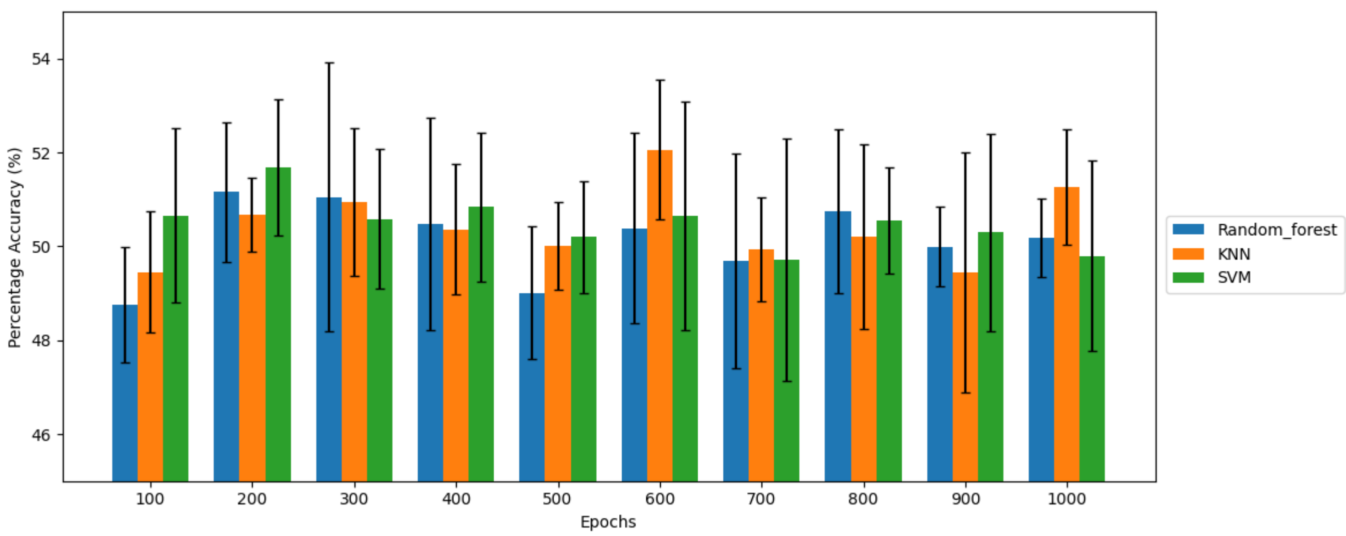

Figure 6.9: SNN trained on FEIS total average ML accuracy's as the number of epochs increased

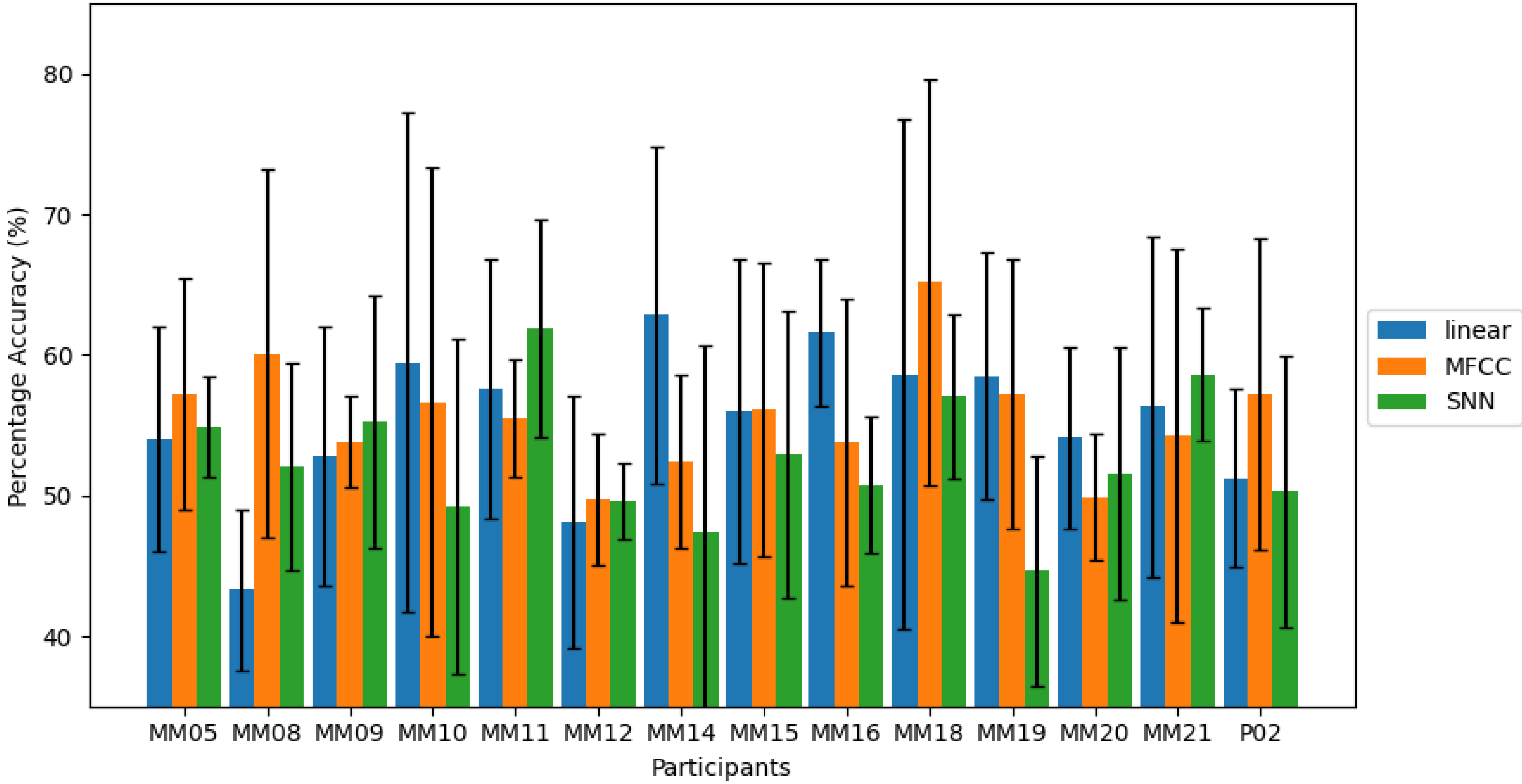

Figure 6.10: Comparison of the average accuracy across the four binary groups of the top feature extraction methods classification per Participant of the Kara One dataset (see table 6.3 further classification information)

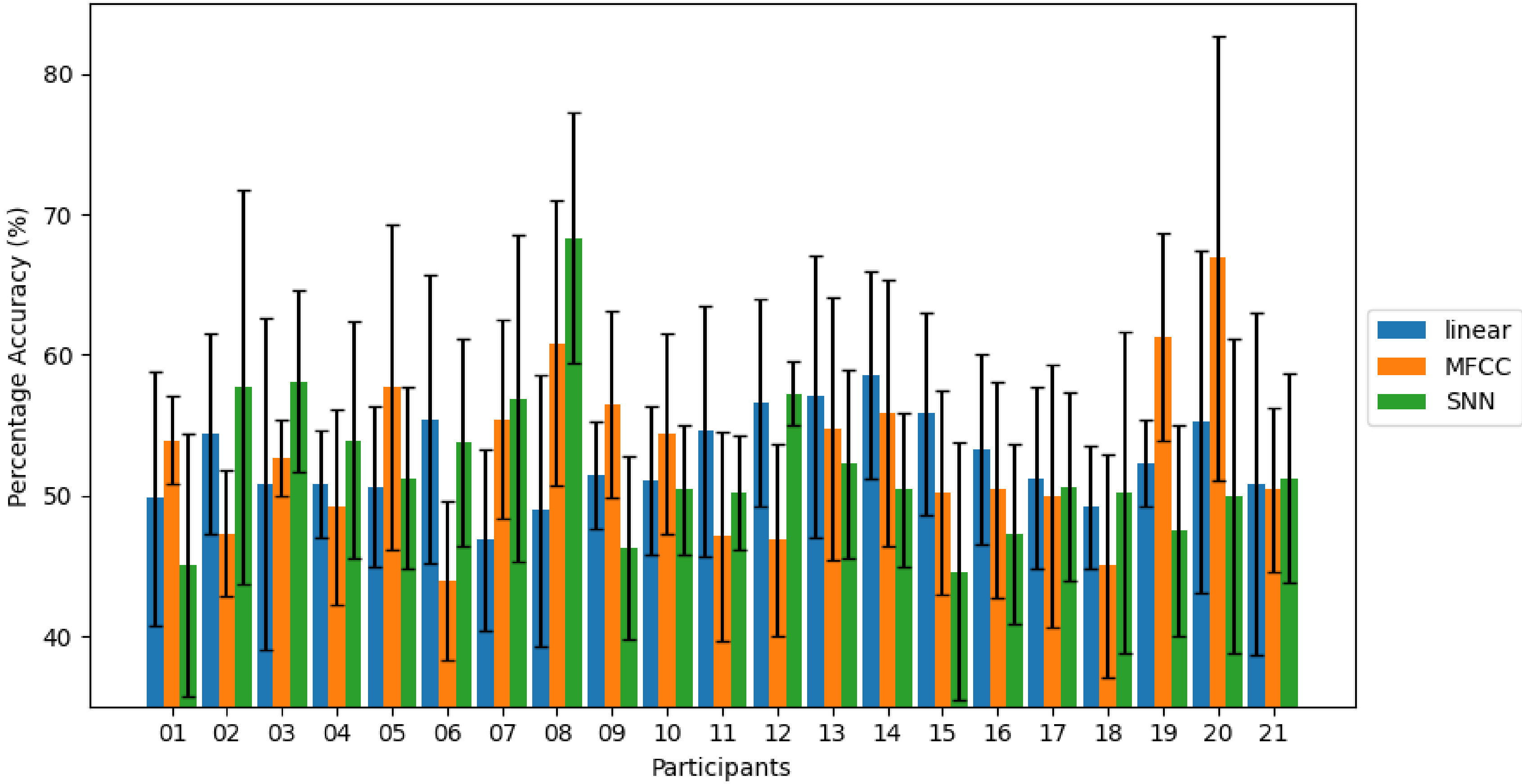

Figure 6.11: Comparison of the average accuracy across the four binary groups of the top feature extraction methods classification per Participant of the FEIS dataset (see table 6.3 further classification information)

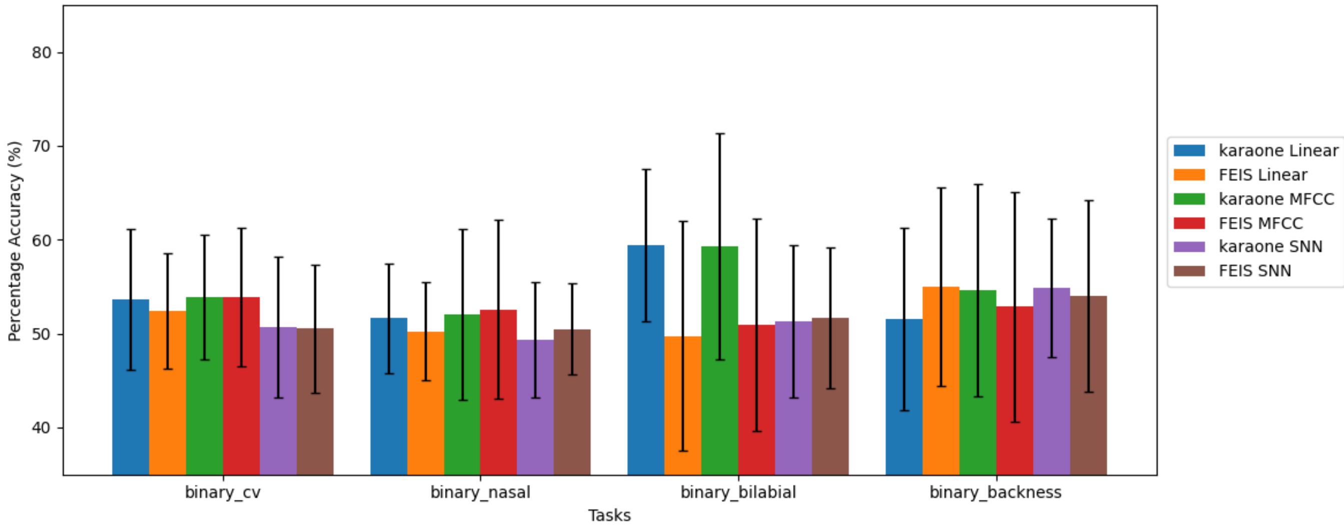

Figure 6.12: Comparison of Kara One Vs FEIS of the best individual performing feature extraction method for SVM

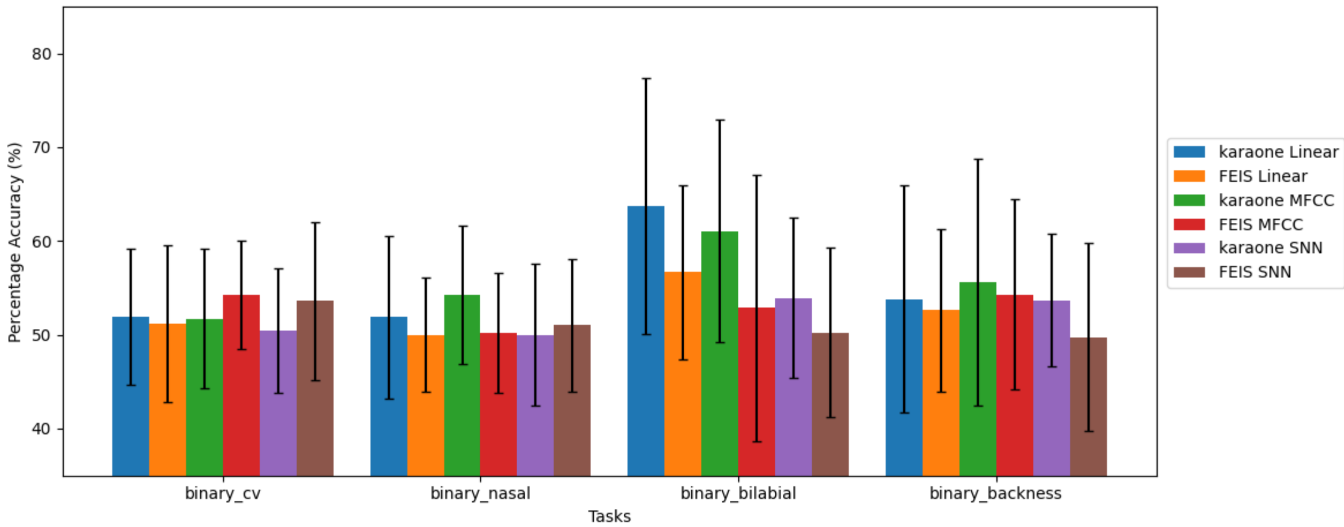

Figure 6.13: Comparison of Kara One Vs FEIS of the best individual performing feature extraction method for Random Forest

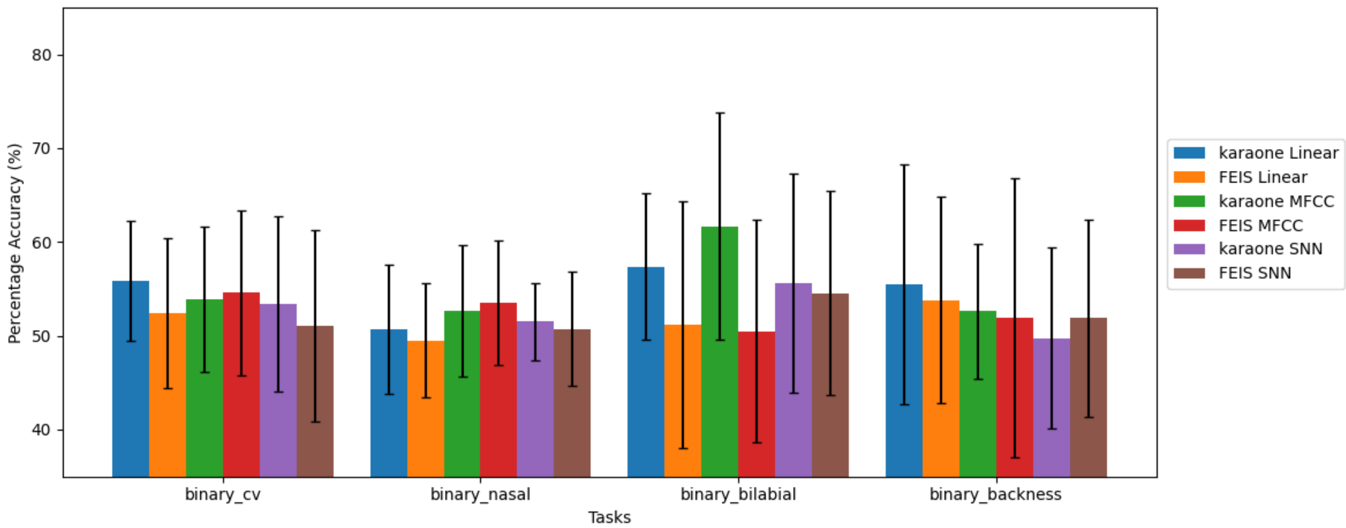

Figure 6.14: Comparison of Kara One Vs FEIS of the best individual performing feature extraction method for KNN

Chapter 7

Discussion

7.1 Machine Learning Classifier Tuning

After completing the methodology outlined in section 5.4.1, the classification accuracy's across the four binary classes/tasks were averaged and the greatest values for each Machine Learning algorithm were input into table 6.1. These showed that the Kara One dataset performed best using the Random Forest algorithm, obtaining an accuracy of 55.09% compared to that of SVMs 54.57%. Considering this is binary classification and there is even numbers of test classes in each test carries out, the chance value for these results would be 50%. Therefore this difference of 0.52% is actually quite significant, in proving that SVM may not be the best classifier for imagined speech. Analysing this result further using Figure 6.1, it shows that Random Forest achieved this accuracy using 100 trees/ estimators, on the MFCC extracted data. Both the linear and MFCC data peak at this value enforcing the fact this is the best value for this data set. Comparing all the feature extraction methods in figure 6.1, its clear that MFCC is the method that achieves the greatest accuracy apart from when the tree/estimator value is 50 where the linear/non-linear data seems to do better. It also shows that the ICA applied seems to have a negative effect on the data, most significantly visible in the MFCC data as all the feature extraction carried out without it seemed to do significantly better with an approximately 5% difference between the MFCC values when 100 estimators are used.

Table 6.1 also shows the FEIS dataset to perform best using the KNN algorithm achieving a classification accuracy of 52.62% with SVM again over .5% thus disproving sub-hypothesis 4.1.3. The KNN algorithm used the Manhattan distance metric and the MFCC feature extracted data as can be seen in figure 6.2. The manhattan distance was a new result with a struggle to find its use in the literature when associated with imagined speech, with the second best result in the cosine distance the more widely used distance metric. FEIS also achieved another interesting result with the outcome being significantly better from the data that performed ICA before the coefficients were extracted, the complete opposite to that of the Kara One data. This seemed to be the case for all the top performing FEIS data across the ML methods as seen in table 6.1. This could be because the number of channels used in the FEIS data and hence the number of components asked of ICA is significantly less, making it easier to split the signal into a smaller number of statistically independent features rather then potentially looking for components which may not exist when a larger component value is used in Kara Ones case.

The Split FEIS data as mentioned in section 5.3.1, seemed to achieve some peculiar results across classification algorithms all showing to have the same low values as shown in appendix C.5 and C.6. The source of which could potentially have been some sort of error in the set up, however with sufficient results already obtained, the non split results were used in future analysis and testing instead.

Both of the non-split results showed to disprove the sub-hypothesis 4.1.3, as both data sets proved that alternative machine learning classifiers obtained greater classification accuracy's then that of the SVM implementation. However, it cant be ignored that the SVM had the lowest standard deviation by a large amount for the Kara One data and not far off this in the FEIS data in table 6.1. This greater consistency in output accuracy's across tasks shows it to be better with a range of input data which with the high complexity and variation of EEG data between participants may be one of the reasons why it is so widely used as a baseline of investigation in much of the imagined speech literature.

7.2 Optimisation Using Pearson Correlation and CSP

Figure 6.3, shows the outcome of the channel selection investigation that was carried out using the optimal feature extraction method on the kara one dataset and machine learning algorithm as shown in Table 6.1. You can see from the peaks above the baseline classification in the graph that both the Pearson correlation and the CSP method manage to increase the total accuracy of the classifier. The greatest accuracy comes from the CSP method when 33 feature channels are selected increasing the total accuracy to 55.6% from 55.1% showing these dimensionality reduction methods to be effective at increasing accuracy levels while also having other benefits such as reducing training time and the data storage required.

Testing the feature channel selection for FEIS dataset, was carried out using the ICA MFCC feature extraction and the KNN algorithm with the manhattan distance metric as this was the top performer for the dataset from the previous section. This had disappointing results as seen in figure 6.4, showing both the feature channel selection methods to only match that of the accuracy of the original run when all the channels are selected.

However, applying these methods to the rest of the configurations in table 6.1, gives some more positive results. Figure 6.5, shows Kara ones KNN MFCC data to be particularly sensitive to the channel selection showing improvements to its baseline accuracy for almost all of the range of 5-40 channels selected. With the pearson correlation showing more of a positive influence on the first 25 channels selected and CSP showing classification improvements between 25 and 40 channels selected. These values produced however still do not beat that of the improvements made to the random forest algorithm. The FEIS random forest results shown in figure 6.6, help to prove that these can be used to increase classification accuracy with FEIS' largest result coming from this, using 6 channels, leap frogging all other results for the dataset.

Comparing tables 6.1 and 6.2, besides the FEIS KNN algorithm it is clear to see these channel selection methods have increased the accuracy of the classifiers, making a strong case that this sub-hypothesis 4.1.2 is proven.

7.3 Feature Extraction Analysis

To analyse the feature extraction methods and prove or disprove sub-hypothesis 4.1.1, it seemed logical to use and compare the top performing classification configuration across all three feature extraction methods and the two datasets, the outcomes of which can be seen in table 6.3. With the majority of the contents origin already explained it leaves only the SNN data - which can be seen from figures 6.8 and 6.9. You can notice that the best performing with the largest total average accuracy score is the KNN algorithm after the model had run 100 epochs for Kara One and the same classification algorithm although this time after 600 epochs for the FEIS data set. The highest performing algorithm is no surprise in being the KNN, proving its choice of use in (Lee, Lee and Lee, 2020), with the algorithm having the ideal features to tailor to data that is being spread out by the SNN's contrastive loss function. When introducing CSP to the data before the data is trained the results seem to be slightly different with the random forest algorithm obtaining the highest classification results for both data sets at 800 epochs as can be seen in appendix E. These data peaks performed worse then that of the data which had no CSP applied to it, with the Kara one data obtaining an accuracy of 51.24% with a std. of 1.75 and the FEIS data surprisingly performed better with an average total accuracy of 51.95% with an std. of 1.3. Due to the lower nature of these peak accuracy scores compared to that without the CSP, this data was not used for any further analysis.

Taking table 6.3 at face value its clear that the sub-hypothesis 4.1.1 of the deep learning method outperforming all non-deep learning feature extraction methods is not the case when using the SNN. Instead it is the MFCC feature extraction method that appears to achieve the highest accuracy and also does not have the largest standard deviation, indicating the spread of data is closer around this average and therefore has a more consistent performance across the binary classes.

Diving deeper into this data and looking at it from a per task perspective as is shown in figure 6.7 you will notice that the MFCC method is not the top achiever all the time. This figure showing the average classification across all the participants per task for each of the set ups seen in table 6.3. The most obvious can be seen in the binary bilabial ($\pm$ Bilab.) result, where there linear data achieves the highest classification accuracy in both datasets. However, the standard deviation seems to be larger in the case of the Kara One linear data compared to that of the MFCC, again hinting at MFCCs better performance. Another surprising result is the FEIS MFCC data seemed to perform the best out of all the data displayed for binary CV (C/V) with a very small error bar compared to the rest of the data. This shows that even lower fidelity data can achieve good results - a positive outcome in proving this papers main hypothesis 4.1.

The SNN feature extractors results don't seem to be as encouraging, with the only interesting outcomes being it out performing other Kara One data on the binary cv classes however this is overshadowed by the previous point. The SNN Kara One data seemed to really struggle with the binary backness ($\pm$ Back.) with the FEIS dataset achieving better results then this.

Figures 6.10 and 6.11 highlight how much EEG data can vary between participants leading to some signals being more susceptible to some methods rather then others. Figure 6.10, has some quite large peaks from the MFCC data on the MM18 participant in particular achieving an accuracy of 65.17% followed closely by some other peaks in the Linear data at 62.82% and 61.60% and surprisingly a large peak in the MM11 SNN data achieving a peak of

61.92%. However, apart from these the trends of the graphs seem to be as expected with the MFCC data achieving the greatest number of highest accuracy's per participant and the least number of lowest accuracy's. The Linear data has a medial spread and the SNN data has the least peaks per participant and the most lowest accuracy classifications. The FEIS data (shown in figure 6.11) on the other hand has a different distribution with the largest peak coming from the SNN data at 68.33% from participant 08, which is followed by peaks from the MFCC data at 19 and 20 at 61.25% and 66.88% respectively. Nevertheless the most interesting part about this is that per participant the SNN data achieves the best classification accuracy the most times at 8 times with linear and MFCC only achieving this feat 7 and 6 times. The SNN data achieves such a low overall accuracy because it also has the most worst accuracy performances at 9 compared to the other two 8 for MFCC and 4 for linear. With so many high and medial accuracy's achieved from the linear dataset its surprising that in this case it is not the top performer.

7.4 Kara One Vs FEIS dataset Accuracy's

To investigate the papers main hypothesis 4.1, the highest achieving feature extraction methods for each dataset were plotted in 3 graphs per classification method as seen by figures 6.12, 6.13 and 6.14. You will notice that apart from the large values achieved during the binary bialbial ($\pm$ Bilab.) that the classification accuracy between FEIS and Kara One is not too dissimilar. It seems on the whole the higher fidelity Kara One to be the slightly more accurate, but this was to be expected with the data being higher in quality with more channels and higher sampling rate used. Nevertheless, there were still instances in which the lower fidelity FEIS data set obtained higher task classification accuracy using the same methods, the most prominent case being in the binary CV classification in the Random forest testing in figure 6.13. This shows FEIS to outperform all three versions of the Kara One data in both the SNN and MFCC data by a minimum of over 2.2%. This is also the case as seen in figure 6.14, in the binary nasal task, in which the FEIS MFCC data outperforms all other data tested in this task, with the same also occuring in figure 6.12 but this time to a lesser extent.

Therefore, the extent the classification accuracy's between the Kara One and FEIS data sets are similar is very high in the binary CV, nasal and Backness tasks however they are very far apart in the binary bilabial category. These binary bilabial results seem to be one of the main influences of the large difference between total average accuracy scores which can be demonstrated in tables 6.1 and 6.2

Chapter 8

Conclusions

8.1 Hypothesis Conclusions Summary

Main Hypothesis Conclusion:

- The extent the classification accuracy's between the higher fidelity Kara One and lower fidelity FEIS data sets are similar is very high in the binary CV, nasal and Backness tasks however they are very far apart in the binary bilabial category.

Sub-Hypothesis Conclusions:

- The deep learning feature extraction method used in this study in the form of a Siamese neural network did not out perform data produced from non-deep learning feature extraction methods when run through the same machine learning classifiers on imagined speech Data. The MFCC data and (non-)linear feature extraction methods both outperforming the classification of the SNN.
- Common Spatial Patterns and Pearson correlations can be used effectively to increase machine learning algorithms accuracy for imagined speech classification at the same time as reducing the date's dimensionality. However the magnitude of this reduction could vary data set to data set.
- Support vector Machines (SVM) will not out perform Random Forest and K-Nearest Neighbour in classification accuracy of imagined speech data in this study. However, it did have one of the lower standard deviations within the results showing it to be consistent with its accuracy across the tasks given.

8.2 Data Set Comparisons

The investigation shows that the Kara One and FEIS data set can produce some results that are very closely comparable in the case of the C/V $\pm$ Nasal. and $\pm$ Back. tasks with Kara one data showing a particularly high classification result surrounding the $\pm$ Bilab. result. This on the whole is a good result considering, the difference in the individual words and phonemes used, the difference in size with this being the number of participants trials and also length of the recordings and used experimental paradigm. In future, it would be interesting to be able to reduce this uncertainty between the two data sets and be able to observe a direct comparison between two data sets using different fidelity of equipment. This could be carried out by performing a custom investigation using the same experimental paradigm on the same set of people imagining the same words and phonemes, to really fully see the extent of the gap between the two qualities of EEG.

8.3 Deep Learning Vs Non-Deep Learning Feature Extraction

The results showed that it was the non-deep learning feature extraction method that on the whole seemed to provide better accuracy scores more specifically the MFCC method for the Kara One data and the ICA MFCC method for the FEIS data. This was a disappointing result due to positive signs from the literature showing the SNN to be a great method of extracting information that was easily identified by classification algorithms at the same time as reducing its dimensionality. However, there could be many reasons for why this occurred, such as the network being used not working well with the size of the input data with the literature implementation (Lee, Lee and Lee, 2020) using data with a far smaller dimensionality especially in the case of Kara One. This implementation was also used for multi-class identification meaning it may not be as good with binary classes. Furthermore, its likely as they were training it for multi-class use that it had more data to train on and making it more effective at extracting the data features. Therefore, in the future it would be good to try testing more with the architecture parameters to allow it to deal better with larger input sizes.

8.4 Does Channel Selection/ Dimensionality Reduction Optimise?

With only the FEIS ICA MFCC data and the SNN data not seeing any benefits from the channel selection it is believed that this is an effective method at gaining a few extra percent when classifying imagined speech. From the results it looks that CSP is the slightly better of the two methods as it achieved the highest accuracy in more configurations that it ran on. As latter of the two negative results was only carried out using once using 9 channels. It is therefore believed if this was repeated on a range of channels that there would be some increase in overall classification accuracy, this would therefore be a good source of future development in this area.

8.5 Machine learning Classifier Comparisons

With the initial classification results on the non-deep learning feature extracted data showing the KNN and random forest to be the best classifiers out of the three ML algorithms. However, after the CSP took place it seemed the random forest became the dominant model that achieved the highest accuracy's for both models. The same trend occurred with the SNN data in which the random forest became the more dominant classifier after CSP with it previously being KNN, potentially showing a link between the channel selection method and random forest. With no mention of the SVM classifier at the top end of the accuracy it did boast the smallest standard deviation when all the top performing channel selection results were analysed - showing it to be a more consistent classifier then the rest over a range of tasks.

Bibliography

Adler, J. and Parmryd, I., 2010. Quantifying colocalization by correlation: The pearson correlation coefficient is superior to the mander's overlap coefficient. *Cytometry. part a : the journal of the international society for analytical cytology* [Online], 77, pp.733–42. Available from: https://doi.org/10.1002/cyto.a.20896.

Akshi and Rao, M., 2021. Decoding imagined speech using wearable eeg headset for a single subject [Online]. *2021 ieee international conference on bioinformatics and biomedicine (bibm)*. pp.2622–2627. Available from: https://doi.org/10.1109/BIBM52615.2021.9669361.

Bakhshali, M.A., Khademi, M., Ebrahimi-Moghadam, A. and Moghimi, S., 2020. Eeg signal classification of imagined speech based on riemannian distance of correntropy spectral density. *Biomedical signal processing and control* [Online], 59, p.101899. Available from: https://doi.org/https://doi.org/10.1016/j.bspc.2020.101899.

Biasiucci, A., Franceschiello, B. and Murray, M.M., 2019. Electroencephalography. *Current biology* [Online], 29(3), pp.R80–R85. Available from: https://doi.org/https://doi.org/10.1016/j.cub.2018.11.052.

bmj the, n.d. *11. Correlation and Regeression* [Online]. Available from: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression [Accessed 2022-09-04].

Bogue, R., 2010. Brain-computer interfaces: control by thought. *Industrial robot: An international journal*.

Brigham, K. and Kumar, B.V.K.V., 2010. Imagined speech classification with eeg signals for silent communication: A preliminary investigation into synthetic telepathy [Online]. *2010 4th international conference on bioinformatics and biomedical engineering*. pp.1–4. Available from: https://doi.org/10.1109/ICBBE.2010.5515807.

Clayton, J., Wellington, S., Valentini-Botinhao, C. and Watts, O., 2020. EnglishDecoding imagined, heard, and spoken speech: Classification and regression of eeg using a 14-channel dry-contact mobile headset [Online]. *Proceedings interspeech 2020*. International Speech Communication Association, vol. 2020-October, pp.4886–4890. Publisher Copyright: © 2020 ISCA; Interspeech 2020, INTERSPEECH 2020 ; Conference date: 25-10-2020 Through 29-10-2020. Available from: https://doi.org/10.21437/Interspeech.2020-2745.

Computational Neuroscience, S.C. for, n.d. *What is EEGLAB?* [Online]. Available from: https://sccn.ucsd.edu/eeglab/index.php [Accessed 2022-09-04].

Cooney, C., Folli, R. and Coyle, D., 2018. Mel frequency cepstral coefficients enhance imagined speech decoding accuracy from eeg [Online]. *2018 29th irish signals and systems conference (issc)*. pp.1–7. Available from: https://doi.org/10.1109/ISSC.2018.8585291.

Costa, F., Patow, G. and Azorin, J., 2019. Common spatial pattern for the classification of imagined geometric objects [Online]. pp.152–155. Available from: https://doi.org/10.5220/0007934601520155.

Demin, S., Panischev, O. and Demina, N., 2016. Cross mff—analysis in studying the obsessive-compulsive disorder. *Journal of physics: Conference series* [Online], 741, p.012073. Available from: https://doi.org/10.1088/1742-6596/741/1/012073.

Emotiv, n.d. *Emotiv EPOC+* [Online]. Available from: https://www.emotiv.com/epoc/ [Accessed 2022-09-04].

Evans, G., 2021. What is a neuron? function, parts, structure, and types. *Simple psychology* [Online]. Available from: https://www.simplypsychology.org/neuron.html.

Hashim, N., Ali, A. and Mohd-Isa, W.N., 2018. *Word-based classification of imagined speech using eeg* [Online], pp.195–204. Available from: https://doi.org/10.1007/978-981-10-8276-4_19.

Hypotheses, n.d. *EEG* [Online]. Available from: https://blricrex.hypotheses.org/ressources/eeg [Accessed 2022-09-03].

Jiang, X., Bian, G.B. and Tian, Z., 2019. Removal of artifacts from eeg signals: a review. *Sensors*, 19(5), p.987.

Kumar, J.S. and Bhuvaneswari, P., 2012. Analysis of electroencephalography (eeg) signals and its categorization–a study. *Procedia engineering* [Online], 38, pp.2525–2536. INTERNATIONAL CONFERENCE ON MODELLING OPTIMIZATION AND COMPUTING. Available from: https://doi.org/https://doi.org/10.1016/j.proeng.2012.06.298.

Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P. and Lance, B.J., 2016. Eegnet: A compact convolutional network for eeg-based brain-computer interfaces. *Corr* [Online], abs/1611.08024. 1611.08024, Available from: http://arxiv.org/abs/1611.08024.

learn scikit, n.d.a. *sklearn.decomposition.FastICA* [Online]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.FastICA.html [Accessed 2022-09-05].

learn scikit, n.d.b. *sklearn.ensemble.KNeighborsClassifier* [Online]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html [Accessed 2022-09-05].

learn scikit, n.d.c. *sklearn.ensemble.RandomForestClassifier* [Online]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html [Accessed 2022-09-05].

learn scikit, n.d.d. *sklearn.svm.SVC* [Online]. Available from: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html [Accessed 2022-09-05].

Lee, D.Y., Lee, M. and Lee, S.W., 2020. Classification of imagined speech using siamese neural network [Online]. *2020 ieee international conference on systems, man, and cybernetics (smc)*. pp.2979–2984. Available from: https://doi.org/10.1109/SMC42975.2020.9282982.

Lee, D.Y., Lee, M. and Lee, S.W., 2021. Decoding imagined speech based on deep metric learning for intuitive bci communication. *Ieee transactions on neural systems and rehabilitation engineering* [Online], 29, pp.1363–1374. Available from: `https://doi.org/10.1109/TNSRE.2021.3096874`.

Lyons, J., 2020. *python$_s$peech$_f$eatures*[*Online*].

Matsumoto, M. and Hori, J., 2014. Classification of silent speech using support vector machine and relevance vector machine. Applied soft computing *[Online], 20, pp.95–102. Hybrid intelligent methods for health technologies. Available from:* `https://doi.org/https://doi.org/10.1016/j.asoc.2013.10.023`.

MNE, 2019. Introduction to artifacts and artifact detection *[Online]. Available from:* `https://mne.tools/0.17/auto_tutorials/plot_artifacts_detection.html` *[Accessed 2022-09-04].*

Moctezuma, L. and Molinas, M., 2019. Subject identification from low-density eeg-recordings of resting-states: A study of feature extraction and classification.

Moctezuma, L., Molinas, M., Torres-García, A., Villaseñor-Pineda, L. and Carrillo, M., 2018. Towards an api for eeg-based imagined speech classification.

Muda, L., Begam, M. and Elamvazuthi, I., 2010. Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques [Online]. [Online]. Available from: `https://doi.org/10.48550/ARXIV.1003.4083`.

Nguyen, C.H., Karavas, G.K. and Artemiadis, P., 2017. Inferring imagined speech using EEG signals: a new approach using riemannian manifold features. Journal of neural engineering *[Online], 15(1), p.016002. Available from:* `https://doi.org/10.1088/1741-2552/aa8235`.

Panachakel, J.T., Ramakrishnan, A. and Ananthapadmanabha, T., 2020. A novel deep learning architecture for decoding imagined speech from eeg. arxiv preprint arxiv:2003.09374*.*

Queensland: Queensland Brain Institute, T.U. of, n.d. Axons: the cable transmission of neurons *[Online]. Available from:* `https://qbi.uq.edu.au/brain/brain-anatomy/axons-cable-transmission-neurons` *[Accessed 2022-09-03].*

Repov, G., 2010. Dealing with noise in eeg recording and data analysis.

Saha, P., Abdul-Mageed, M. and Fels, S.S., 2019. SPEAK YOUR mind! towards imagined speech recognition with hierarchical deep learning. Corr *[Online], abs/1904.05746.* `1904.05746`*, Available from:* `http://arxiv.org/abs/1904.05746`.

Teplan, M. et al., 2002. Fundamentals of eeg measurement. Measurement science review*, 2(2), pp.1–11.*

Wang, Y., Gao, S. and Gao, X., 2005. Common spatial pattern method for channel selelction in motor imagery based brain-computer interface [Online]. 2005 ieee engineering in medicine and biology 27th annual conference*. pp.5392–5395. Available from:* `https://doi.org/10.1109/IEMBS.2005.1615701`.

Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G. and Vaughan, T.M., 2002. Brain–computer interfaces for communication and control. Clinical neurophysiology *[On-*

line], 113(6), pp.767–791. Available from: *https://doi.org/https://doi.org/10.1016/S1388-2457(02)00057-3.*

Zhao, S. and Rudzicz, F., 2015. Classifying phonological categories in imagined and articulated speech. 2015 ieee international conference on acoustics, speech and signal processing (icassp). *IEEE, pp.992–996.*

Appendix A

Extra Literature Review and Technology Survey

A.1 Further Deep learning Feature extraction Developments

Building on the above ideas, (Lee, Lee and Lee, 2021), implemented a new SNN network this time pre-processing the data and extracting the instantaneous frequency and spectral entropy of each trial. The respective extracted data from each channel was then concatenated into two streams in the form of (2 x 520 x 1) and fed into the SNN in pairs. The training system for this network was more complex with there being a 'pretraining' in which they set the batch size to 500 for 1000 iterations then for the actual 'training' they reduced it to 100 for another 5000 iterations. The embedding data was classified using the KNN algorithm again using the K value of 5.

A.2 Deep Learning Classifiers

Deep learning can be implemented into imagined speech classification, in a way that makes it more efficient and higher performing then standard machine learning classifiers. One of the first created specifically for use in EEG is EEGNet (Lawhern et al., 2016). Its compact nature brings many advantages in it can be trained on limited data and is able to produce neurophysiologically interpretable features. This was compared against the most recent networks at the time such as ShallowConvNet, DeepConvNet and xDAWN+RG with its more complex version EEGNet-8, 2 outperforming them all. More recently other more complex architectures have been created including the already mentioned in Panachakel, Ramakrishnan and Ananthapadmanabha (2020) who made use of CSP and a large filtered frequency band. They implemented a network with four hidden dense layers. Three of which use the relu activation function and the final using a hyperbolic tangent function. Outputting 9 values using a sigmoid function with the final classified version to be decided via hard voting. Testing it on the ASU data set, the outcome showed it to have comparable results to that of the current machine learning algorithms, with the thought that using more dense data for example 128 channel compared to the current 64 would give better accuracy. Hierarchical Deep Learning (Saha, Abdul-Mageed and Fels, 2019) has also proved to be a very effective method for multi-class identification, with the to be discussed obtaining an average accuracy of 84.27% on the Kara one data set grossly outperforming the original investigations initial benchmarks.

Appendix B

Extra Methodology

B.1 The original experimental paradigm

The participants will be encouraged to blink as little as possible.
Before the Experiment:
Each participant will record themselves saying all of the words/phonemes in a steady period 5 times. This will be replayed to them as the stimulus before they begin their current trial. The participant will be put into a room on their own in a hemianechoic chamber and it will be made sure that they have a comfortable seating position.
During the experiment:

1. The participant would have a rest state for 3 seconds, during this time they are encouraged to clear their mind.
2. The participant would then be shown the speech being tested in a random order and would also hear the pre-recorded speech being repeated in a periodic fashion for 5 seconds
3. There would then be a 5 second imagined speech state in which the participant tries to think of the prompt in the exact same way it was pronounced to them through the audio stimulus. The visual cue would also be pulsing at the required tempo to further aid the participants rhythm.
4. The same would then be recorded for speaking with this being used in the case of contingency.

Each 5 second "epoch" will be repeated 20 times per word giving a total of 20 x 11 = 220 trials per participant.

Appendix C

Remaining ML Variable Sweeping Graphs

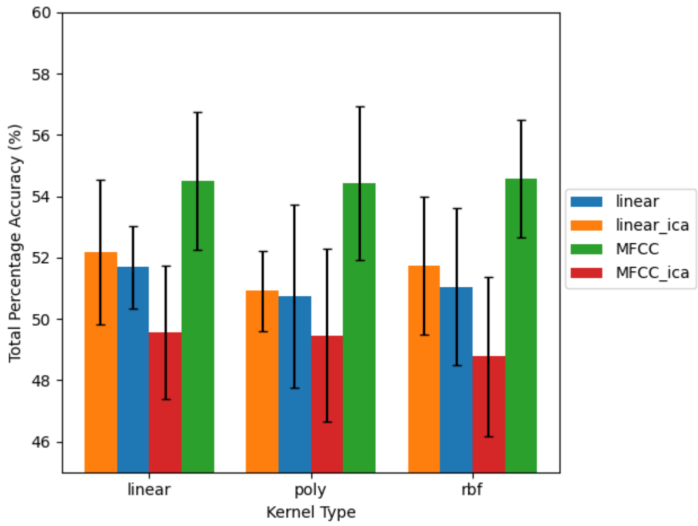

Figure C.1: Comparison of the varying kernels, and feature extracted data used by and input to the SVM against the total average classification accuracy across all participants of the four imagined speech test classes of the Kara One Data set

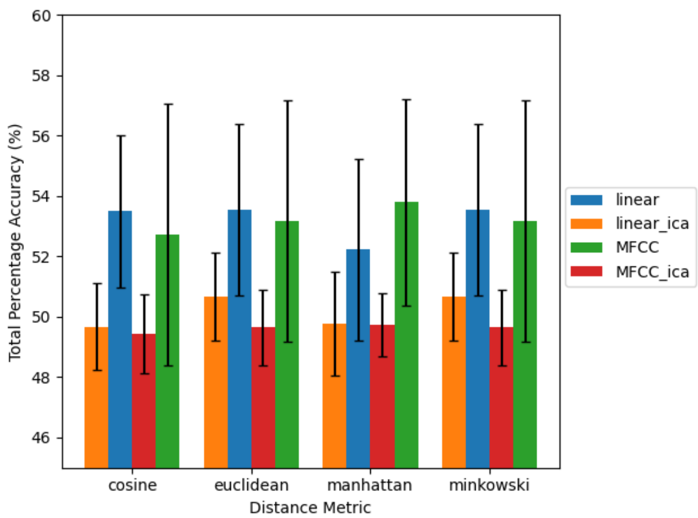

Figure C.2: Comparison of the varying distance metrics, and feature extracted data used by and input to the KNN classifier against the total average classification accuracy across all participants of the four imagined speech test classes of the Kara One Data set

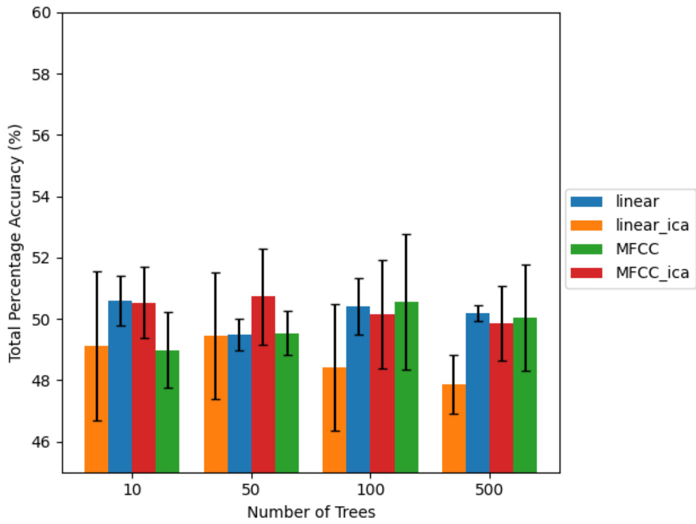

Figure C.3: Comparison of the varying number of trees/estimators, and feature extracted data used by and input to the random forest classifier against the total average classification accuracy across all participants of the four imagined speech test classes of the FEIS Data set

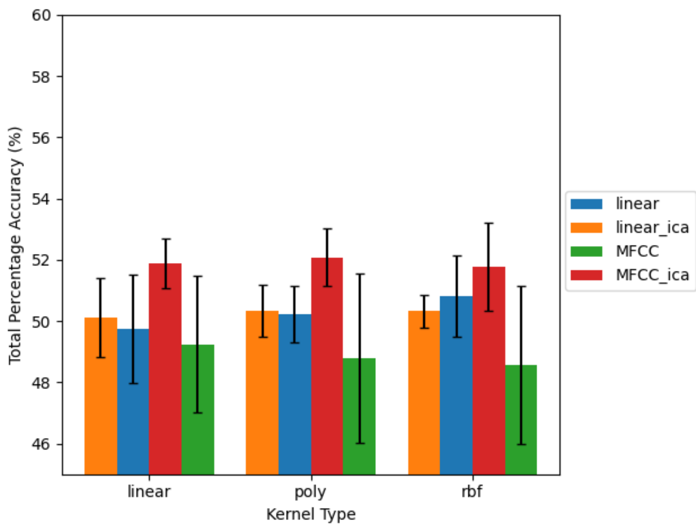

Figure C.4: Comparison of the varying the kernels, and feature extracted data used by and input to the SVM classifier against the total average classification accuracy across all participants of the four imagined speech test classes of the FEIS Data set

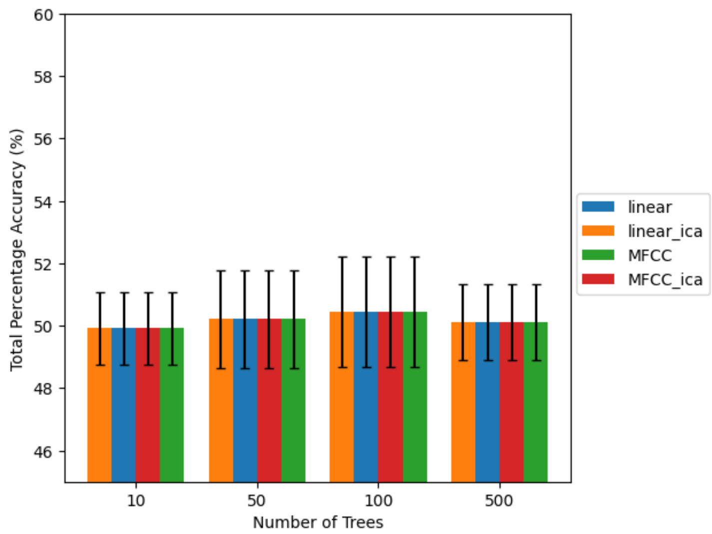

Figure C.5: Comparison of the varying number of trees/estimators, and feature extracted data used by and input to the Random Forest against the total average classification accuracy across all participants of the four imagined speech test classes of the split FEIS Data set

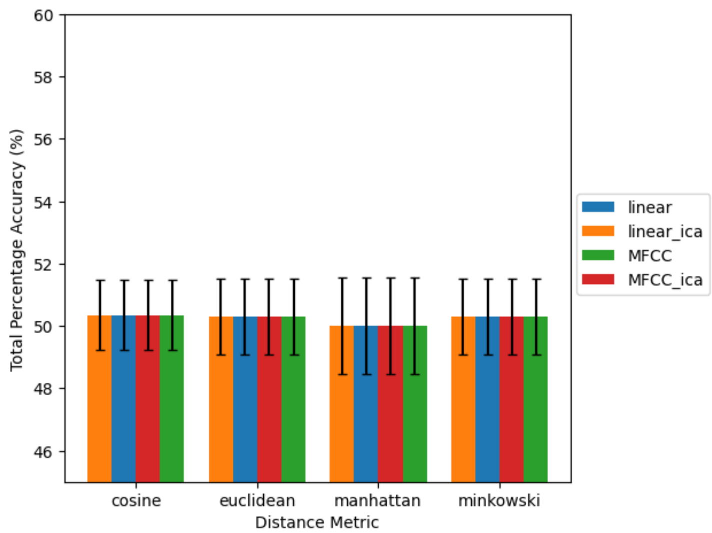

Figure C.6: Comparison of the varying the distance metric, and feature extracted data used by and input to the KNN against the total average classification accuracy across all participants of the four imagined speech test classes of the split FEIS Data set

Appendix D

Remaining Channel Selection/Dimensionality Reduction Investigation Graphs

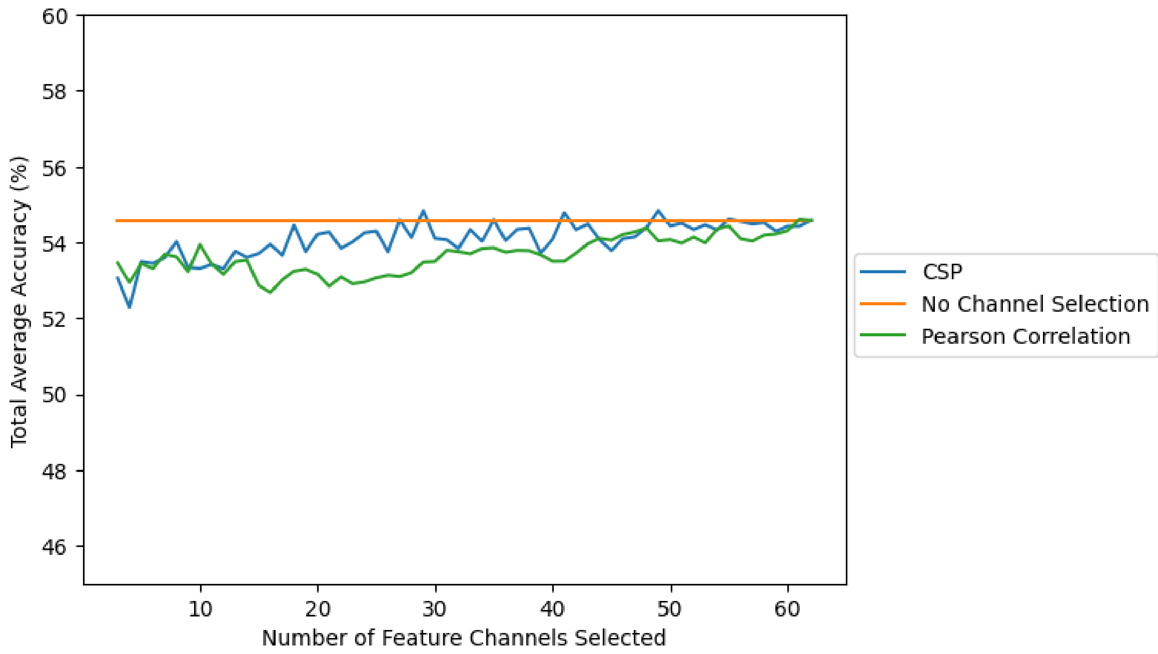

Figure D.1: Graph of how the total average classification accuracy across the four binary classes of the top performing feature extractor (MFCC) and machine learning algorithm (SVM rbf kernel) for the Kara One dataset changes with the increase in feature channel selection using CSP and Pearson correlation methods

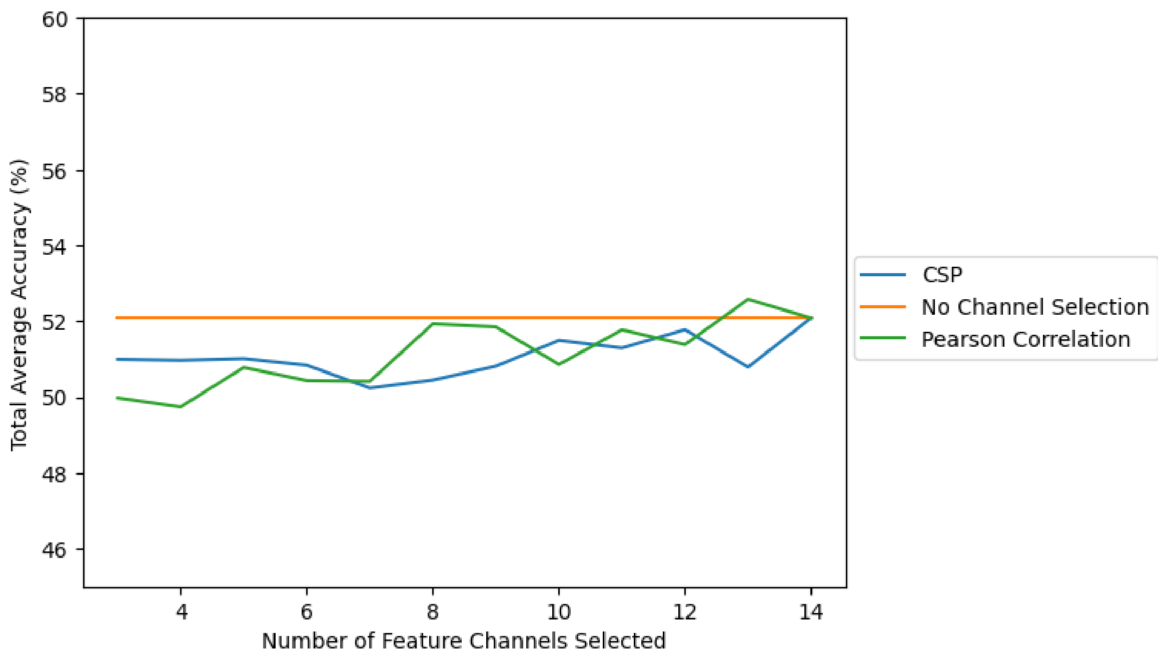

Figure D.2: Graph of how the total average classification accuracy across the four binary classes of the top performing feature extractor (ICA MFCC) and machine learning algorithm (SVM poly (quadratic) kernel) for the Kara One dataset changes with the increase in feature channel selection using CSP and pParson correlation methods

Appendix E

CSP SNN classification Results

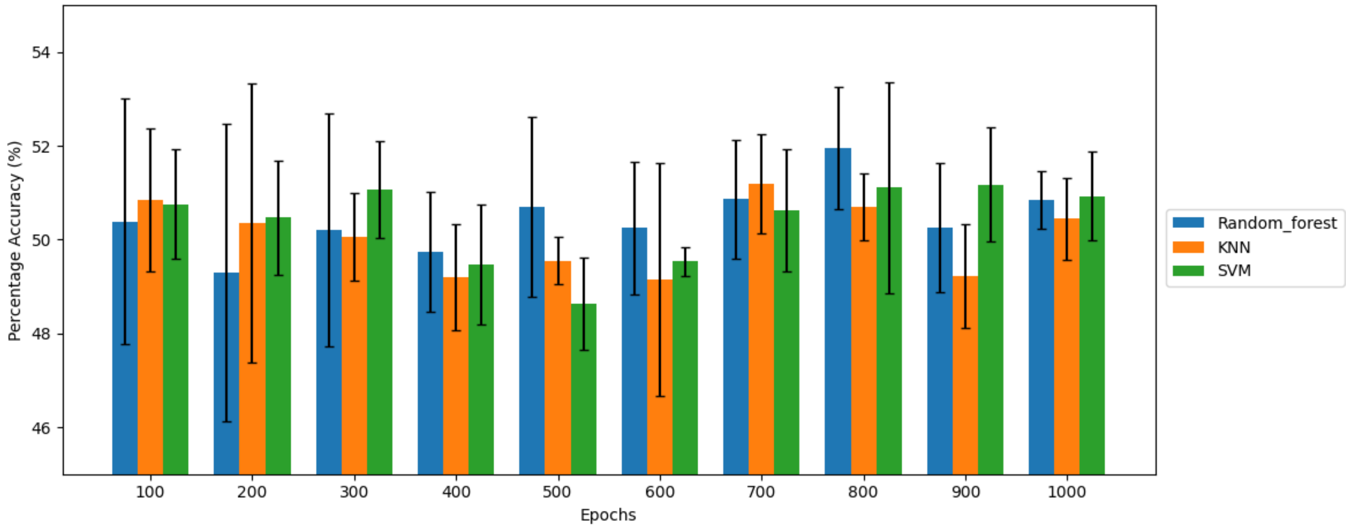

Figure E.1: SNN trained on 9 selected channels using CSP of the FEIS, total average ML accuracy's as the number of epochs increased

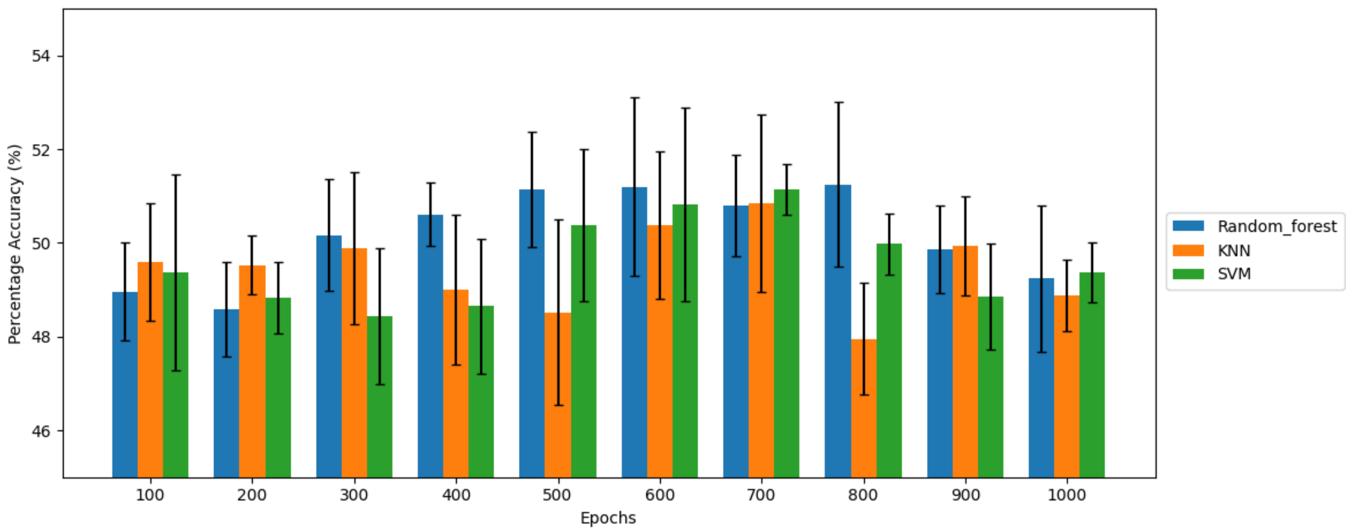

Figure E.2: SNN trained on 9 selected channels using CSP of the Kara One, total average ML accuracy's as the number of epochs increased