Abstract

Working Memory (WM) is associated with short-term memory and the control over information intended for long-term memory processing. This study explored WM memory success and failure by investigation of electroencephalography (EEG) data of 130 healthy subjects carrying out a digit span task. Frequency power throughout trials for each traditional frequency band was obtained (delta, theta, alpha, beta, gamma and high gamma), and these were used as individual datasets. Machine learning approaches were applied to each dataset to examine the potential role of each frequency band in WM processes, and to build a successful classifier that can classify items that were remembered or not remembered. Three variations of k-nearest neighbour algorithm have been built to enhance classification accuracy and the understanding of feature importance.

Each electrode that recorded the patients' neural oscillations during the task was used as individual features in this experiment. Frequency power, therefore, was computed for each feature. In the first classification method (KNN), all electrodes or features were used for the classification. The second classification method used those features only, that were selected by a feature selection algorithm designed for KNN, neighbourhood component analysis (NCA). In the third classification method, a random subspace ensemble method was used with KNN base learners.

Random subspace ensemble method was found to be the most successful classifier, with an average overall accuracy of 69.1 % throughout the six datasets, while KNN using all features resulted an average 59.1 % accuracy. The classification with selected features by NCA, has improved KNN classification but has not reached the accuracy of the random subspace ensemble, with an average classification accuracy of 62. 8%.

The difference in classification accuracies between different frequency bands did not show major differences, and therefore, it was not possible to associate any specific frequency band with the success of WM status classification. However, the different methods applied have highlighted that some features might be the key for the successful classification between remembered and not remembered items. The success of random subspace ensemble method seems to combine the 'knowledge' of those features important to both remembered and not remembered memory items.

Word Count:

9179

Acknowledgement

I would like to thank my supervisor, Dr Di Bernardi Luft, to get me this far and teach me everything I learned through this project.

I would also like to thank Andrea for every night spent awake with me studying, to Ferenc to help with all my computational questions,

And I wish to thank most my Dad who made all this possible.

Thank you

Table of Contents

List of Figures and Tables

Introduction

Working memory (WM) is a multicomponent cognitive system that is believed to influence comprehension, learning and reasoning (Baddeley, 2010). WM is associated with short-term memory and the control over information intended for long-term memory processing. WM capacity has been widely studied in the field of psychology and neuroscience and was associated with a wide range of cognitive behaviours (Conway *et al.*, 2005). Understanding the underlying neural oscillation during WM processes has a great potential to have a better understanding of complex behaviours. To investigate this question, multiple studies were conducted related to WM and neural oscillatory activity, many of these using electroencephalography (EEG) (Gevins *et al.*, 1997, Onton, Delorme and Makeig, 2005, Sauseng *et al.*, 2005, Sauseng *et al.,* 2010). The most commonly discussed frequency ranges across neurological studies are; delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (13-30 Hz), gamma (30-45 Hz), high Gamma (55-80 Hz). However, the connection between neural oscillations and cognitive processes is often difficult to interpret, but these traditional frequency ranges were shown to be associated with neural activities and blood-oxygen level changes (Meltzer *et al*., 2007).

Several WM studies suggest that theta, alpha and gamma frequencies could have specific roles in WM processes, such as maintain or inhibit information retention (Roux and Uhlhaas, 2014). Palva *et al*., (2010) proposed increasing synchronicity between alpha, beta and gamma frequency power ranges, during increased WM load. Other studies also reported theta, alpha and gamma power changes related to WM, for example, increased gamma and theta power coupled with increased WM load (Howard, 2003, Meltzer *et al.*, 2007). Theta was consistently associated with WM processes (Gevins *et al.*, 1997, Sarnthein *et al*., 1998, Sammer *et al.*, 2007). Studies found a frontal-midline increase in theta rhythm during WM load, reporting that these brainwaves increase with task difficulty and memory load in the frontal areas of the brain, therefore theta waves may play an important role in encoding and retention (Sauseng *et al*., 2010). While alpha waves initially increased during WM load, it was also reported to decrease with increasing task difficulty (Gevins *et al.*, 1997). In a general term, alpha waves were repeatedly reported to reflect inhibitory activity (Tuladhar *et al*., 2007). Although gamma and beta waves are less commonly reported among WM studies, there is evidence that supports a gamma-beta bursts underlie WM item retention (Onton, Delorme and Makeig, 2005). It was also reported that these

gamma bursts, nested in a theta phase, represent and mark those individual items that should be kept in WM (Canolty *et al*., 2006, Chaieb *et al.*, 2015).

Neural oscillations and cognitive processes are, however, multifaceted, with multiple influencing factors and variables. A direct statistical comparison between the differences in frequency powers fails to solve more complex cognitive tasks, such as the process of working memory success or failure. However, machine learning approaches might be able to highlight complex correlations and clustering and therefore, a computational approach could make it possible to differentiate between the relationship between frequency power, neurological scene and memory success.

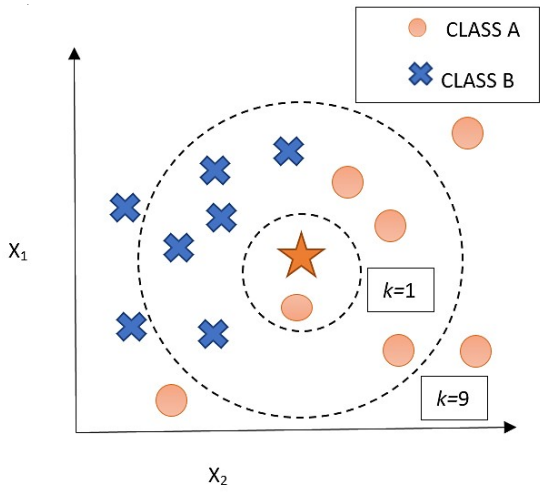
The aim of this experiment is to tie together the known information about WM, uses oscillatory power datasets with advanced machine learning approaches to predict WM success and failure. By understanding WM failure and success, the underlying neurological processes and their importance might be revealed.

The main machine learning algorithm in this experiment is a k-nearest neighbour algorithm. The variations of this algorithm were explored throughout the experiment in order to find the most successful classifier.

KNN is one of the simplest, yet stable classification method that competes with algorithms with more complexity (Abu Alfeilat *et al*., 2019). Despite its simplicity, KNN is reported to be successful on high-level classification tasks (Yao and Ruzzo, 2006).

KNN is a non-parametric supervised algorithm. KNN is often referred as a 'lazy' algorithm as it simply uses the similarity between *k* number of labelled datapoint surrounding the unlabelled datapoints for prediction. This means that the algorithm simply 'remembers' the training instances and predictions are made by comparing the prediction datapoint to the known training instances. The *k* in *k*NN is the parameter that determines how many nearest training points are involved in prediction. If *k* is 10, the prediction of any datapoint will happen by finding the 10 nearest 'neighbours' to this point from the training datapoints (various distance measures can also be used to determine which points are the nearest 'neighbours'). Then the prediction datapoint's class is determined by majority voting. If most of the surrounding training data points belong to one class, the algorithm will predict that class. In binary classification therefore, it's possible to have a tie between voting if the *k* parameter is an even number. The basic ideology of KNN classification method is represented in the picture 1.

Picture 1.: K-Nearest Neighbour Classification



Picture 1: The basic concept of K-nearest algorithm (KNN). Classification can vary as a result of changing parameter *k*. This example shows how the class of a previously unseen datapoint (star on the picture) is determined by KNN. On this representation, if k=1, class A would be predicted, while if k=9 Class B would be predicted for the unseen datapoint, by analysing the surrounding nearest neighbours.

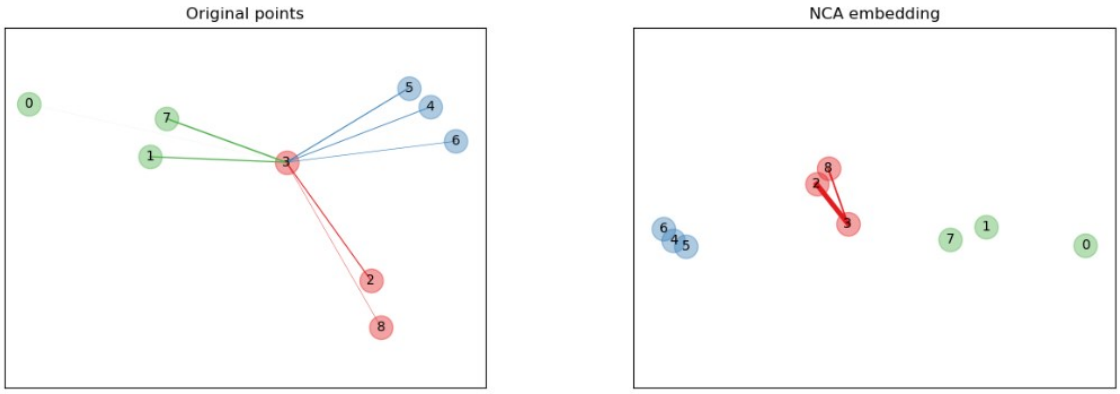
As parameter *k* can be set to various integers, the decision boundary of this algorithm is very flexible. Both smooth, almost linear edges can be achieved, as well as fine, rough decision boundaries. Generally, the lower parameter *k* is, the decision boundary is finer, while a higher *k* would indicate a decision boundary with a smoother edge.

Although, Abu Alfeilat *et al.* (2019) reported that KNN (with the most commonly used distance measures) are relatively tolerant to noise in the data, the curse of dimensionality is a well-known possibility with KNN (Pestov, 2013). This means a lower feature space might benefit this algorithm. Not only the curse of dimensionality, but the above-mentioned biological observations also support the potential benefits of feature reduction. Frequency power changes related to WM were shown near specific neural sites (e.g: frontal-midline, frontal-parietal), or in this case, surrounding different features, but potentially not through all feature simultaneously. Features in this experiment are EEG electrodes at different locations on the skull.

Therefore, from a wide variety of available feature selection methods, a novel approach was applied, neighbourhood component analysis (NCA). NCA was created for dimensionality reduction in KNN (Goldberger *et al.*, 2005). NCA works by computing a quadratic distance measure between datapoints, using stochastic neighbour assignment. The goal is to learn a distance metric that maximizes the classification accuracy of the nearest neighbours. By restricting this metrics to be low, dimensionality can be reduced. NCA carries out multiple classifications, using random neighbour assignments, and

learns from the accuracies derived from a leave-one-out KNN score. NCA maximises the expected number of correctly classified datapoints by maximizing the stochastic variant of the leave-one-out KNN score on the training data (Goldberger *et al.*, 2005). This idea is represented on Picture 2.

Picture 2. Neighbourhood Component analysis



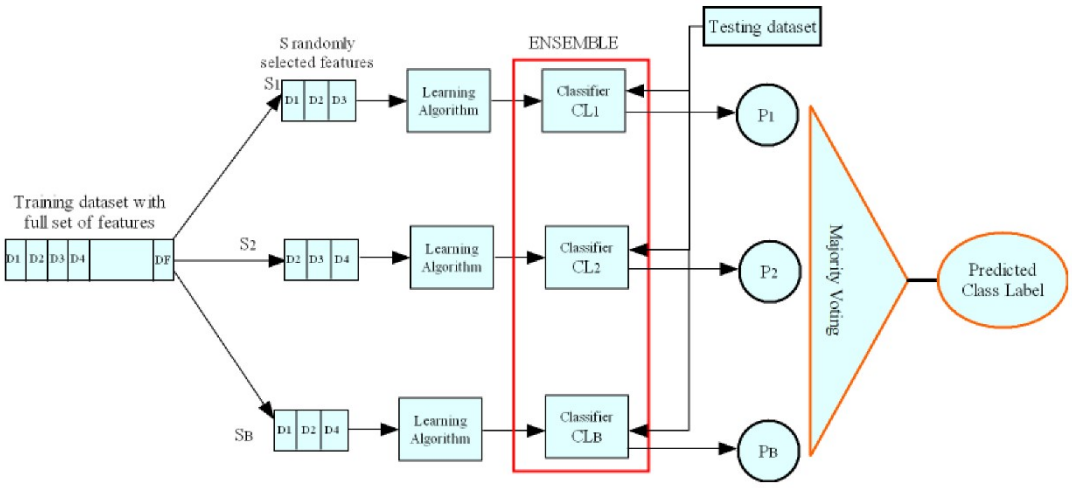
Picture 2: The picture shows the basic representation of how neighbourhood component analysis (NCA) supports the most optimal distance measures for classification. To classify point 3 correctly as class 'red', the most 'important' features are points 2 and 8, therefore the NCA embedded distances with the selected features place these points near the point 3, while other features do not support the classification as well, and distance hence is increased between.

NCA was one of the chosen approaches to improve classification accuracies by KNN. The other method that was a random subspace ensemble (RSE) with KNN 'base' or 'weak' learners.

RSE machine learning approach was designed for high dimensional data; therefore, it was expected to be less influenced by the 'noisy' nature of EEG data (Ho, 1998). An RSE is based on three key steps; random feature selection, random subspace classification, and final classification. The algorithm is trained by several learning cycles, or subspaces, and during each cycle, a prior classification is completed with the randomly selected features only. These subspaces apply a base or 'weak' classification on the entire training dataset, but at each learning cycle only randomly selected features. This experiment used KNN as a base learner. This means that at each learning cycle, random features are selected as the 'knowledge' of the algorithm. Prediction datapoints are classified with the prior described KNN method, but only with these selected features. Each subspaces' classification result counts towards the learning curve for the final trained model which is used for prediction (Ho, 1998). Therefore, the same prediction trial will be classified multiple times (as many times as learning cycles run)

and decision is made again by majority voting. The ideas described are summarized on Picture 3.

Picture 3.: Random subspace ensemble classification



Picture 3.: The picture summarizes the training and testing process of random subspace ensemble method. *Source:* Derhab, A., Guerroumi, M., Gumaei, A., Maglaras, L., Ferrag, M., Mukherjee, M. and Khan, F. (2019). Blockchain and Random Subspace Learning-Based IDS for SDN-Enabled Industrial IoT Security. *Sensors*, 19(14), p.3119.

Interpretation of the key driving features by the RSE is challenging, as features are picked at random with replacement, and the algorithm is not influenced by the weak classification outcomes, therefore, each feature is chosen equally throughout the learning. This can lead to a better classification accuracy but a more difficult interpretability.

These computational approaches are novel in WM research, with the aim to expand our understanding of WM processes. Classification between WM failure or success will shed light upon the importance of different neural oscillatory activity, computational method, feature selection and WM neural spatial orientation.

Methods and Materials

Data Gathering

67 females and 63 males participated in the experiment (total of 130), with an age spanning between 18 and 32. The participants volunteered to an advertisement and were paid 7.5£ an hour. Confidentiality and anonymity of the participants were ensured by using participant codes. Partakers were informed and consented to follow and complete laboratory protocols. Queen Mary Ethics Committee approved the experiment protocol.

The EEG signals were recorded using a battery-drive system (StarStim, Neuroelectrics, Spain) and 18 PiStim electrodes which were placed in accordance to the extended 10-20 electrode system (Jasper, 1958). These electrodes were P7, T7, F7, C3, F3, P3, Oz, Pz, Cz, Fp1, Fp2, P4, F4, C4, F8, T8 and P8. Two electrodes at the earlobes were used as reference points, and another two electrodes were at the right cheekbone in order to reduce sinusoidal noise at 50 Hz.

Experimental Design

Participants had two 10 minutes sessions, where their working memory was tested whilst their EEG was recorded. At the beginning of each trial, a fixation cross was presented. Working memory was tested by a digit-span task. An increasing series of digits (spanning from 1 to 9) were presented visually in each trial to the participants and then asked to recall all items correctly. The participants had to press on a keyboard the digits they remembered. There was 1 second between each digit presented. On account of an error, the next trial had one less digit to remember. If all the digits were remembered correctly, the next trial exhibited one more digit. Each trial contained different digits, regardless of memory status of the previous trial.

Data Analysis:

In order to analyse the EEG data, MATLAB scripts and toolboxes were used. EEGLAB was used throughout these steps (Delorme and Makeig, 2004); First, the data was re-referenced to the earlobes then high pass filtered at 0.5 Hz. Sinusoidal noise was removed using EEGLAB's CleanLine plugin. EEGLAB's independent component analysis was applied to remove major artefacts, such as blinks (Delorme and Makeig, 2004).

Each participant's recording was sectioned into one second epochs. Epochs of the event of interests, the EEG recordings where the numbers were presented to the participants,

were used for the following steps. After visual and statistical exploration of each recording and each channel, abnormal channels (e.g. channels without data) were interpolated. To be able to analyse the brain oscillatory differences between memory status (whether an item will be remembered or forgotten), Welch's power density estimation was applied on every participants' pre-processed data. The following frequency band power results were acquired; Delta (1-4 Hz), Theta (4-8 Hz), Alpha (8-12 Hz), Beta (13-30 Hz), Gamma (30-45 Hz), High Gamma (55-80 Hz).

These spectral power values were used to classify brain oscillatory differences between working memory success or failure. Therefore, six datasets (one for each frequency power band) were used.

The datasets were then filtered by the experimental digit position, with the aim of keeping those trials only, where participants had to rely on their working memory in the experiment. Therefore, the first 3 digits presented to the participants were excluded from further analysis, where attention and alertness errors were more likely to occur opposed to working memory failure.

Classification Algorithms

All classification was carried out using MATLAB scripts, tailored to each classification method.

K- Nearest Neighbour classification (KNN)

This classification method was used to classify remembered and not remembered trials, using 6 different frequency power datasets. Firstly, all features were fed into the algorithm, then after neighbourhood component analysis, only 6 features were used for classification.

The most optimal parameter *k* was estimated using mean10-fold classification error over three hundred random optimization sample.

In order to avoid inconclusive voting, K was always set to be the best resulting odd number. Both as a weak learner and as an individual classification method, Euclidean distance measures were used.

Random Subspace KNN Ensemble (RSE)

Random subspace ensembles (RSE) was chosen for this experiment as this method was shown to be able to successfully deal with high dimensionality (Ho, 1998). The combination of this with the flexibility of a KNN as a weak/base learner, a nonparametric classification method, can provide a robust, flexible, sensitive and stable classification method, which is essential when classifying EEG data. Furthermore, dimensionality reduction by the random subspace method can potentially aid KNN classification by reducing the noise to be dealt with when calculating the distances between neighbours (Sun, Zhang and Zhang, 2007).

The method of choice for optimum hyperparameters for this type of classifier is still under debate. These parameters (the number of learning cycles, the number of features to be selected at each cycle, and the number of neighbours used by the weak KNN learners) can greatly affect classification performance and stability.

In this experiment, each dataset was randomly sampled three hundred times with replacement, resulting in three hundred different optimizing sets, all with balanced class distributions between remembered and not remembered trials. Three hundred iterations were used throughout the experiment to ensure full data coverage. The sample size of the minority class (which was not remembered trials) often resulted in a small sampled dataset for optimizing, training and predicting; therefore, a relatively higher bootstrap number was chosen. These optimizing sets helped to obtain the most optimal settings for the RSE. 10-fold classification error was calculated for each optimization set and for each hyperparameter. The parameter testing occurred within reasonable scalar values for each hyperparameter (e.g: number of learning cycles; 0-100). The result of each optimization for each parameter was averaged over the three hundred samples. This parameter optimization method, using three hundred random training samples with replacement, was also used to optimize KNN (parameter *k*) and NCA (parameter *lambda*).

Neighbourhood Component Analysis

Neighbourhood component analysis (NCA) was carried out using MATLAB fscnca() function (Yang, Wang and Zuo, 2012). Lambda parameter values were optimized using the same method described in classification optimization, using exact fitting method on all training sets. The datasets were not further standardized by this function. NCA results below 0.5 were not used in further classification.

Overall Classification Workflow

After successful data processing, the classification algorithms' essential parameters were investigated. This process was carried out separately from the classification datasets. As a result of the large imbalance between class distribution in the datasets (90 % of the trials were 'remembered', only 10 % 'not remembered') bootstrap random under-sampling was used to create balanced datasets and to have a better overview of the distribution of the data. 300 iterations were used to randomly sample from the original dataset. This procedure was also followed through the classification steps. During the optimization steps, random bootstrap samples were created with replacement.

After setting the models' parameters, 300 bootstrap iteration was used for the classification of each dataset. This meant that at each iteration, first a prediction dataset was sampled out without replacement (therefore at each iteration, the prediction set is unseen to the algorithm before prediction), then the permutation set was sampled out without replacement, then the training data was sampled out. After each iteration, all datapoints have been replaced and this process was repeated 300 times.

The prediction sample size was 15 % of the not remembered sample group (which was the minority class). The same sample size from the remembered groups was also sampled to create the prediction dataset that had a balanced class distribution.

A permutation dataset was 10 % of the not remembered group, this was also sampled out without replacement. The same number of items from the remembered group. The resulting classes of this dataset was shuffled randomly.

The training dataset was then sampled out from the remaining data. This was 75 % of the not remembered group, and its matching size from the remembered class. The model was then trained on the training data, tested with the prediction dataset and evaluated with the permutation dataset. Datapoints after each iteration have been replaced to the entire dataset.

The classification accuracies and error rates for each bootstrap iteration was recorded. This process was repeated with each frequency power dataset. This method was applied with every classification method.

Data Evaluation

After every iteration of the bootstrap sampling, the classification accuracy, the error rate and the classification's confusion matrix were recorded.

Classification accuracy was measured as the total number of correct predictions divided by the total number of predictions, then the result was multiplied with a 100. Error rate was measured as the total number of incorrect predictions divided by the total number of predictions, then the result was multiplied with a 100.

Confusion matrices were created by recording the true negative, true positive, false positive and false negative predictions. These were added up at each iteration to gain an overall confusion matrix after 300 iterations. The resulting percentages were calculated for each group.

The difference between accuracy means between methods were tested by multiple comparisons. To avoid the increasing type I error by multiple hypothesis testing by t-test, adjusted p-values were calculated using Tukey's honestly significant difference procedure. Each prediction method was tested against each other as well as against all permutation testing. The statistical significance between the permutation tests and the prediction methods allowed to rule out classification by chance.

Results

The result made up of two main sections, the results related to the classification algorithms and the final classification accuracies and comparisons for each frequency and for each classification method.

The first section about the classification algorithms consist of the results of the fine-tuning of the three main model that was used. The three main algorithms were; a random subspace ensemble, and two K-nearest neighbour algorithms (KNN); one to classify each dataset and one to classify datasets with selected features only. The first section also describes the algorithm that was used to select features for the second KNN classification.

In the second section, the results obtained by the three classification methods are presented. This section is further divided to present the results for each frequency dataset that was used.

Classification Algorithms

A) Random Subspace -KNN Ensemble

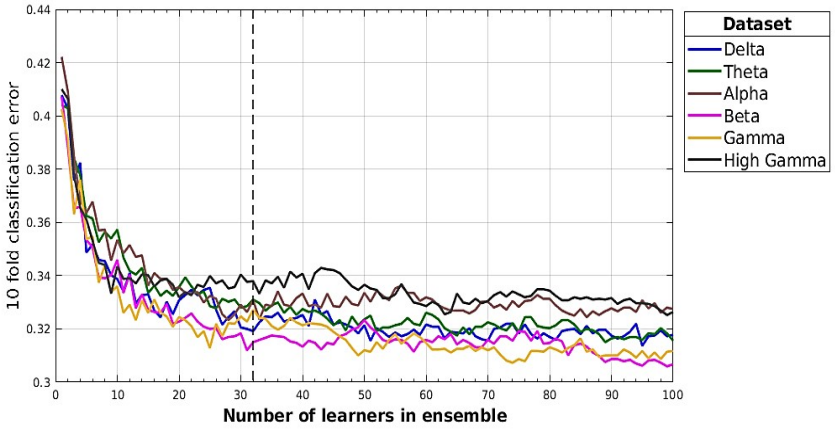
Two main classification algorithms were used; a random subspace ensemble (RSE) based on KNN base learners, and a KNN (without building ensemble), to classify all datasets. Both algorithms' hyperparameters were adjusted to the data.

The six frequency-power datasets resulted to have the same optimal hyperparameters during the optimization steps, therefore it was possible to use a single model with these datasets. The three main parameter for RSE, the number of cycles, the number of predictors to be selected and the number of neighbours were matched across the six datasets. In order to classify the six datasets (consisting of 18 features each, representing the electrodes on the skull) with RSE, the following hyperparameters were found; The most optimal is to use 32 learning cycles, each cycle using 3 randomly picked features for the 'weak' classifications. 32 cycles were used to keep optimal computational power with significant classification error changes. This means after 32 cycles, computational power would have had to be doubled whilst ten-fold classification error would decrease only by 0.01. The selected three features were classified using KNN, with parameter K being 3 (distance; Euclidean). The trained model used the information obtained by the 32 learning cycles.

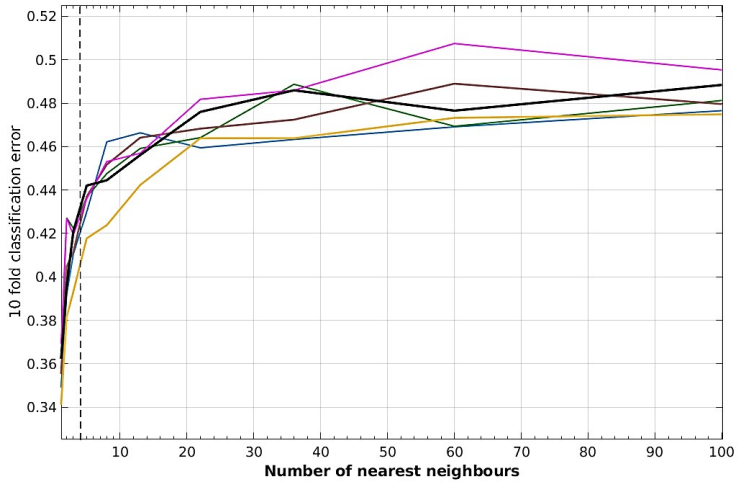
The results of the optimization steps are shown on Figure 1.

Figure 1. Random Subspace Ensemble Optimization Results

A



B



C

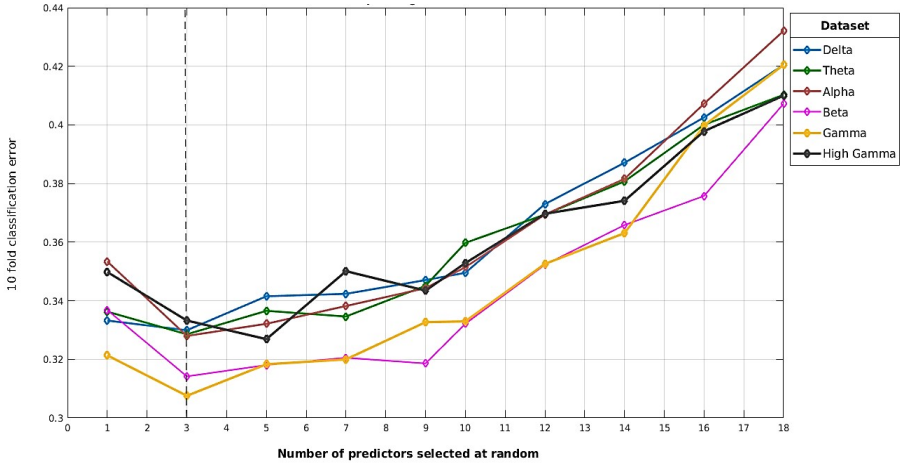


Figure 1.: A:

A shows the relationship between the number of learning cycles in the ensemble and the

resulting classification error. Each dataset (defined in the figures' legend) was bootstrap sampled to create random training sets, then optimization steps were carried out on each dataset using 10-fold classification error. By increasing the number of learning cycles, the classification error decreases for all datasets. The classification error did not meaningfully decrease after 32 cycles; therefore 32 cycles were used in the classification of the above mentioned six datasets, which is highlighted with a dashed line across the graph. **B:** Figure B shows the relationship between the number of neighbours selected (applied to the KNN weak learners) and the ten-fold classification error. Each dataset showed a sharp increase in classification by increasing the number of neighbours used for the classification. A very small value, 3 was therefore used as K parameter (representing the dashed line on the graph). **C:** Figure C shows the relationship between the number of predictors selected at random by the algorithm and the ten-fold classification error of each dataset. This optimization step showed that selecting 3 features randomly at each cycle, result in the lowest classification error (dashed line on Figure C) for each dataset.

B) K- nearest neighbour algorithm (KNN)

As in KNN parameter *k* has to be selected, this parameter was tested with ten-fold classification error, as seen on Figure 1B. This was done 3x300 times, but the same parameters were chosen (one for RSE, twice for KNN with different number of features). As the ten-fold classification error would increase with increasing the number of neighbours, a small number of *k* had to be chosen, without the model being too sensitive to outliers, to noise or to the training data itself. As this classification is a binary classification, to avoid ties during prediction, this parameter was set to be 3. This allowed a slightly smoother decision boundary within the data as well as avoided prediction datapoints where classification decision could not be made.

Neighbourhood Component Analysis

Neighbourhood Component Analysis (NCA) was carried out on each different frequency dataset in order to find the potentially significant features. Each dataset's lambda value was optimized individually. In NCA the values below 0.5 were not used in further classification. The result showed a high importance of channel number 17 and 18 across all datasets. The feature numbers and the underlying channel locations are shown on Figure 2.

Table 1. Summary Table of Neighbourhood Component Analysis

	1st		2nd		3rd		4th		5th		6th	
	Channel/Feature number and Feature Weights											
Delta	T7	1.58	Oz	1.54	F7	1.52	T8	1.34	F4	0.87	P7	0.83
Theta	T7	5.52	T8	5.38	Oz	4.15	F7	4.11	Fp1	3.18	F8	2.76
Alpha	P7	2.40	Fp1	2.20	T7	1.86	F8	1.68	T8	1.65	F7	1.43
Beta	P7	1.23	Fp1	1.12	Oz	1.06	C4	0.97	P8	0.97	T7	0.94
Gamma	FZ	2.42	C3	2.36	F7	2.20	P8	1.18	Cz	1.78	Oz	1.63
High Gamma	P7	2.25	T7	2.15	P8	1.92	14	1.91	P4	1.89	Oz	1.83

Table 1: shows the six highest feature weights found for each frequency dataset. These six features per dataset were used for KNN classification without building an ensemble. The table starts with the highest values on the left temporal side and feature weights decrease towards the right side. Each column shows the resulted channel number on the left and its feature weight next to it on the right side of the column.

The channel numbers and the highest feature weights for each dataset are in the first column on the left. The analysis resulted in high appearance of the posterior temporal (P7) and mid temporal (T7) areas across all datasets. The occipital region (Oz) also appeared on multiple occasion across the datasets within NCA results.

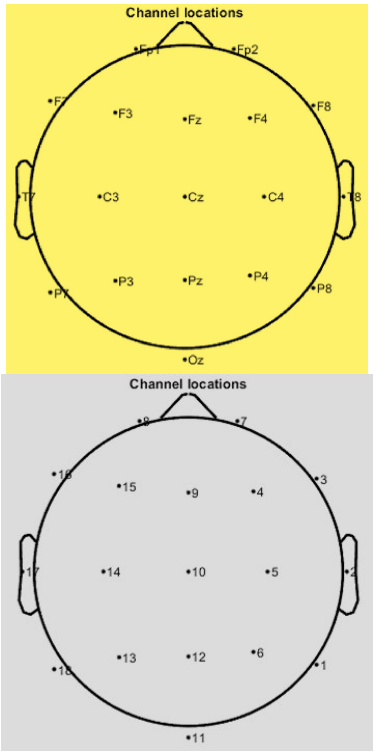


Figure 2. Feature numbers and their underlying spatial locations

Evaluation of classification methods for each frequency and classification method

Delta

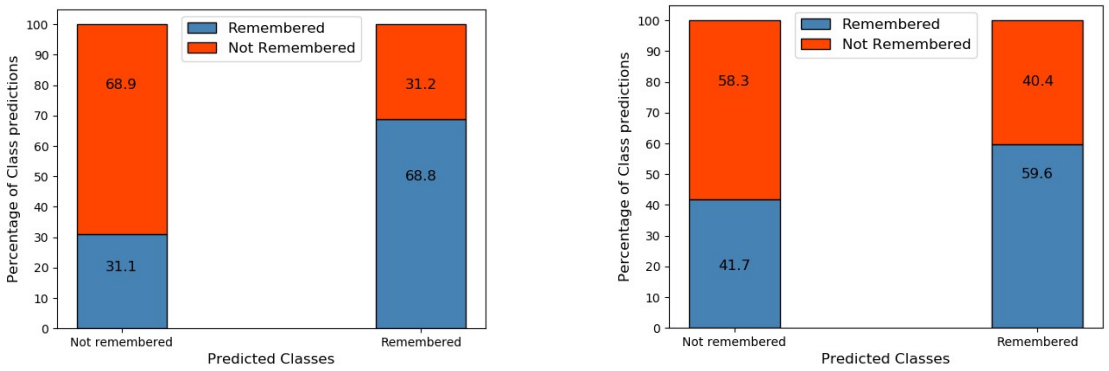
After three hundred iteration of bootstrap and permutation sampling through the delta frequency power dataset, the final mean classification accuracy with RSE was 68.9 per cent with 31.1 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability ($p< 0.01$). Figure 3A shows the percentage distribution of the classification across the classes. 68.9 and 68.8 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 58.9 per cent with 41.1 per cent error rate. The permutation testing resulted in 50.1 per cent ($p<0.01$). Figure 3B shows the percentage distribution of classification across the classes. 59.6 per cent and 58.3 per cent of the trials were correctly classified as remembered and not remembered, respectively.

Delta frequency power dataset classified by KNN, with the identified features only (Table 1, 1st row), resulted in a mean classification accuracy of 60.5 per cent with 39.15 per cent mean error rate. The final mean classification accuracy for the permutation testing resulted in 49.9 per cent, validating the classification reliability ($p< 0.01$). Figure 3C shows the percentage distribution of the classification across the classes. 57.2 and 69.5 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 3. Percentage of Class Predictions With Random Subspace Ensemble and KNN

A B



C D

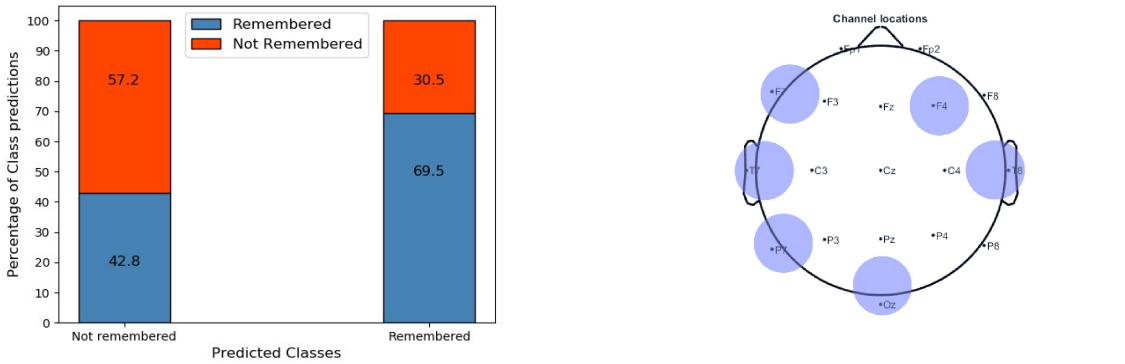


Figure 3**: A, B and C, where** A figure shows the distribution of random subspace ensemble classification accuracy on the delta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the delta power band dataset, RSE was found to be the most successful classifier. Both KNN classifications, as well as RSE were well above 99 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). RSE performs best at achieving higher classification accuracies. The comparison of the algorithm results is shown on Figure 4.

Figure 4: Classification Method Comparison

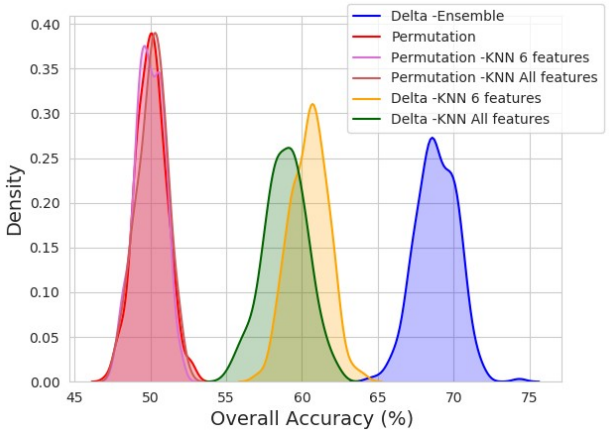


Figure 4.: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the two K-nearest neighbour classifications and Random Subspace Ensemble (RSE) and their permutation testing.

Theta

After three hundred iteration of bootstrap and permutation sampling through the theta frequency power dataset with random subspace ensemble (RSE), the final mean classification accuracy was 69.2 per cent with 30.8 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability (p<0.01). Figure 5A shows the percentage distribution of the classification across the classes. 69.3 and 69.1 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 59.3 per cent with 40.7 per cent error rate. The permutation testing resulted in 50.0 per cent (p<0.01). Figure 5B shows the percentage distribution of classification across the classes. 60.1 per cent and 58.7 per cent of the trials were correctly classified as remembered and not remembered, respectively.

Theta frequency power dataset classified by KNN and with the identified features only (Table 1, 2nd row), resulted in a mean classification accuracy of 60.4 per cent with 39.6 per cent mean error rate. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability (p<0.01). Figure 5C shows the percentage distribution of the classification across the classes. 57.1 and 69.6 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 5.: Percentage of Class Predictions With Random Subspace Ensemble and KNN

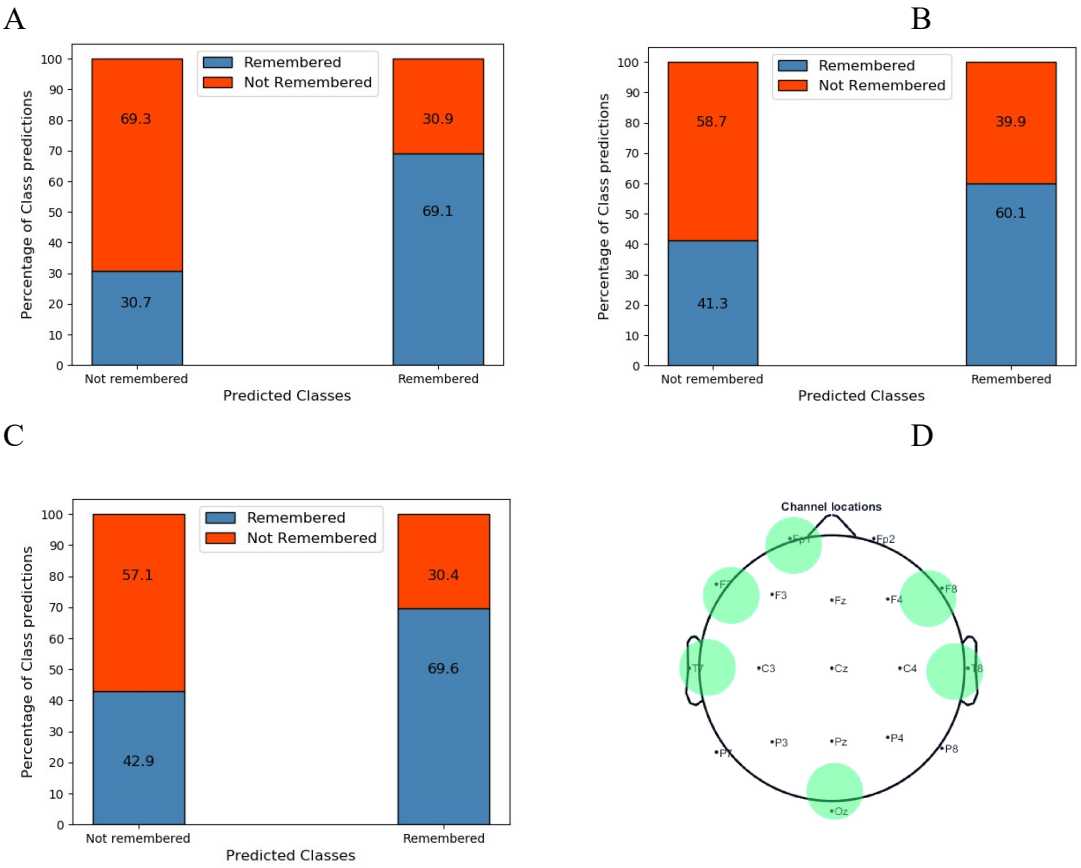


Figure 5**: A B, and C where** A figure shows the distribution of random subspace ensemble classification accuracy on the theta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the theta power band dataset, RSE was found to be the most successful classifier. Both KNN classifiers, and the RSE were well above 95 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). RSE performs best at achieving higher overall accuracies. The comparison of the algorithm results is shown on Figure 6.

Figure 6. Classification Method Comparison

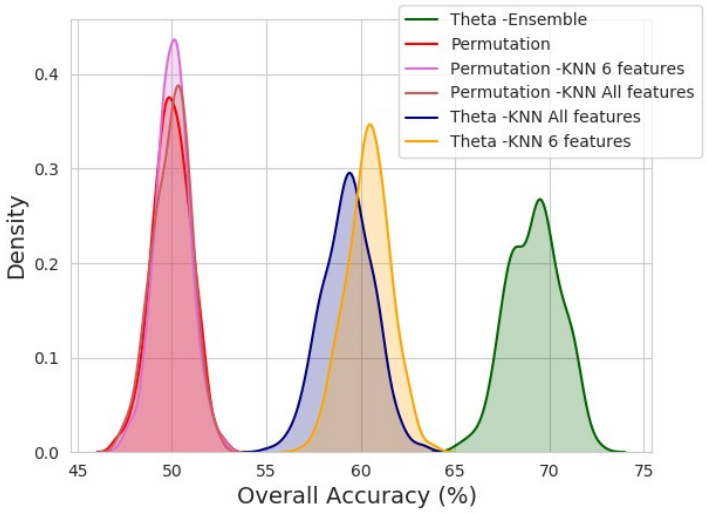


Figure 7.: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the KNN and Random Subspace Ensemble (RSE) and their permutation testing.

Alpha

After three hundred iteration of bootstrap and permutation sampling through the alpha frequency power dataset with random subspace ensemble, the final mean classification accuracy was 69.3 per cent with 30.7 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the permutation testing resulted in 49.8 per cent, validating the classification reliability (p= p<0.01). Figure 7A shows the percentage distribution of the classification across the classes. 69.4 and 69.2 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 59.4 per cent with 40.6 per cent error rate. The permutation testing resulted in 50.0 per cent (p<0.01). Figure 7B shows the percentage distribution of classification across the classes. 60.0 per cent and 58.8 per cent of the trials were correctly classified as remembered and not remembered, respectively.

Alpha frequency power dataset classified by KNN and with the identified features only (Table 1, 3rd row), resulted in a mean classification accuracy of 67.5 per cent with 32.5 per cent mean error rate. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability (p<0.01). Figure 7B shows the percentage distribution of the classification across the classes. 65.8 and 69.6 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 7. Percentage of Class Predictions With Random Subspace Ensemble and KNN

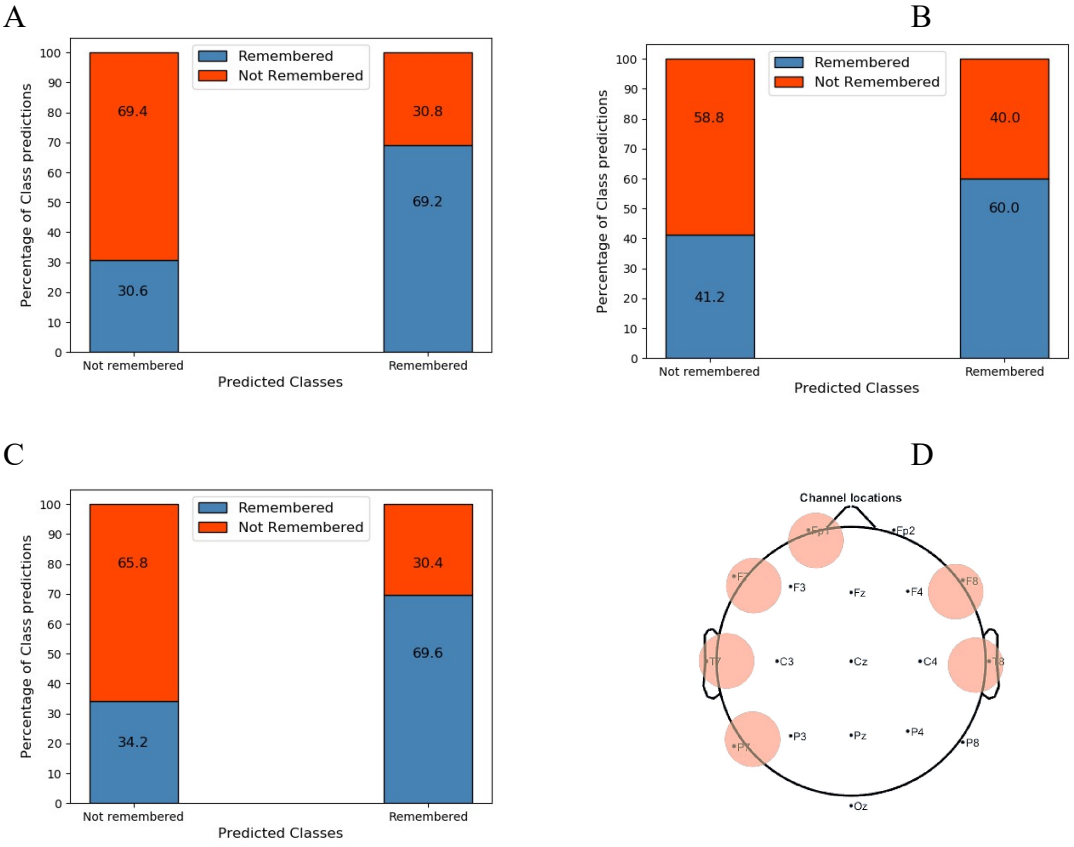


Figure 7**: A B, and C where** A figure shows the distribution of random subspace ensemble classification accuracy on the theta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the

spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the alpha power band dataset, RSE was found to be the most successful classifier, however with a smaller margin than with the other datasets. Both KNN and RSE were well above 95 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). RSE methods performs best at achieving higher overall accuracies. The comparison of the algorithm results is shown on Figure 8.

Figure 8. Classification Method Comparison

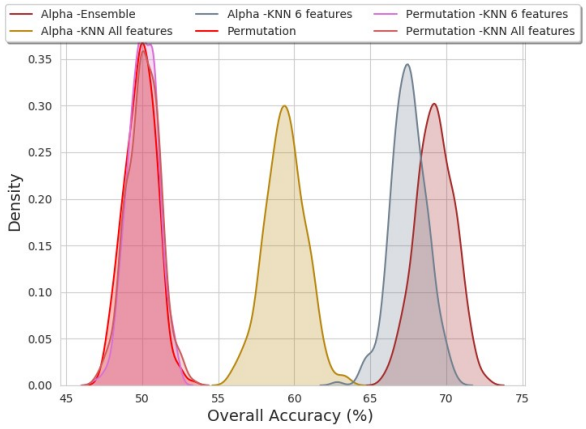


Figure 8.: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the KNN and Random Subspace Ensemble (RSE) and their permutation testing.

Beta

After three hundred iteration of bootstrap and permutation sampling through the Beta frequency power dataset with random subspace ensemble method, the final mean classification accuracy was 69.4 per cent with 30.6 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the

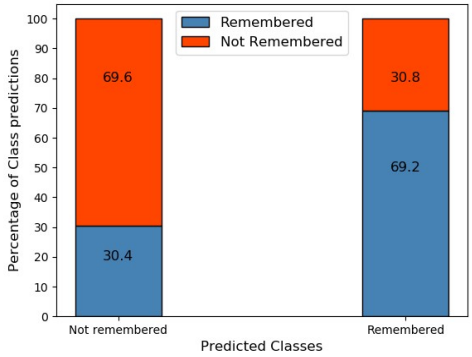
permutation testing resulted in 50.1 per cent, validating the classification reliability (p<0.01). Figure 9A shows the percentage distribution of the classification across the classes. 69.6 and 69.2 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 59.2 per cent with 40.8 per cent error rate. The permutation testing resulted in 50.0 per cent (p<0.01). Figure 9B shows the percentage distribution of classification across the classes. 59.9 per cent and 58.7 per cent of the trials were correctly classified as remembered and not remembered, respectively.

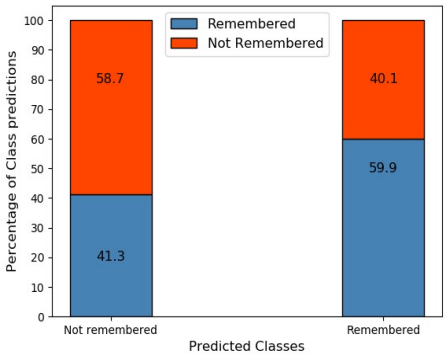
Beta frequency power dataset classified by KNN and with the identified features only (Table 1, 4th row), resulted in a mean classification accuracy of 67.1 per cent with 32.9 per cent mean error rate. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability (p<0.01). Figure 9C shows the percentage distribution of the classification across the classes. 65.5 and 69.1 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 9: Percentage of Class Predictions With Random Subspace Ensemble and KNN

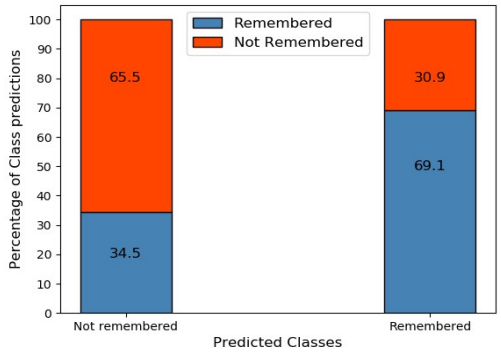
A



B



C



D

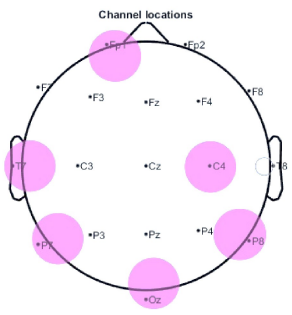


Figure 9**: A B, and C where** A figure shows the distribution of random subspace ensemble classification accuracy on the theta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the beta power band dataset, RSE was found to be the most successful classifier. Both KNN and RSE were well above 95 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). RSE methods performs best at achieving higher overall accuracies. The comparison of the algorithm results is shown on Figure 10.

Figure 10. Classification Method Comparison

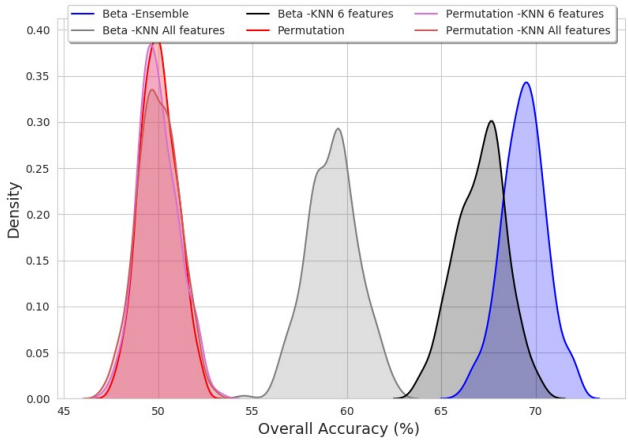


Figure 10.: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the KNN and Random Subspace Ensemble (RSE) and their permutation testing.

Gamma

After three hundred iteration of bootstrap and permutation sampling through the gamma frequency power dataset with random subspace ensemble method, the final mean classification accuracy was 69.0 per cent with 31.0 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability ($p<0.01$). Figure 11A shows the percentage distribution of the classification across the classes. 69.1 and 68.8 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 58.6 per cent with 41.4 per cent error rate. The permutation testing resulted in 50.0 per cent ($p<0.01$). Figure 11B shows the percentage distribution of classification across the classes. 59.3 per cent and 58.0 per cent of the trials were correctly classified as remembered and not remembered, respectively.

Gamma frequency power dataset classified by KNN and with the identified features only (Table 1, 5th row), resulted in a mean classification accuracy of 60.4 per cent with 39.6 per cent mean error rate. The final mean classification accuracy for the permutation testing resulted in 50.0 per cent, validating the classification reliability ($p<0.01$). Figure 11C shows the percentage distribution of the classification across the classes. 57.1 and 69.9 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 11: Percentage of Class Predictions With Random Subspace Ensemble and KNN

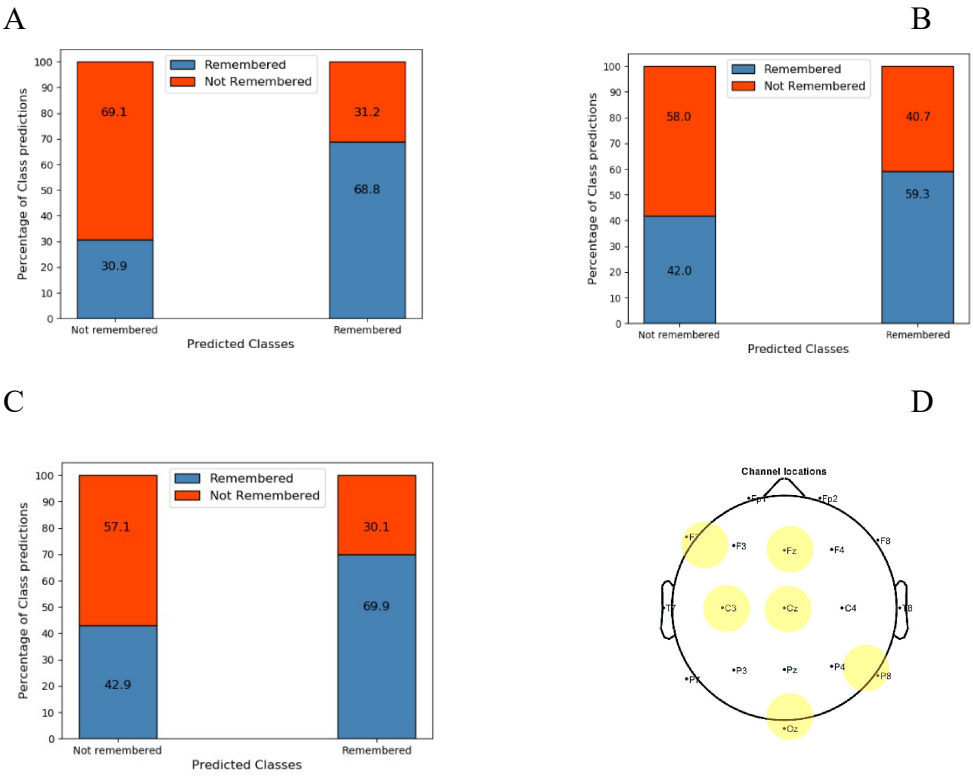


Figure 11**: A B, and C where** A figure shows the distribution of random subspace ensemble classification accuracy on the theta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the gamma power band dataset, RSE was found to be the most successful classifier. Both KNN and RSE were well above 95 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). Both algorithms perform well against permutation testing, where RSE methods performs best at achieving higher overall accuracies. The comparison of the algorithm results is shown on Figure 12.

Figure 12. Classification Method Comparison

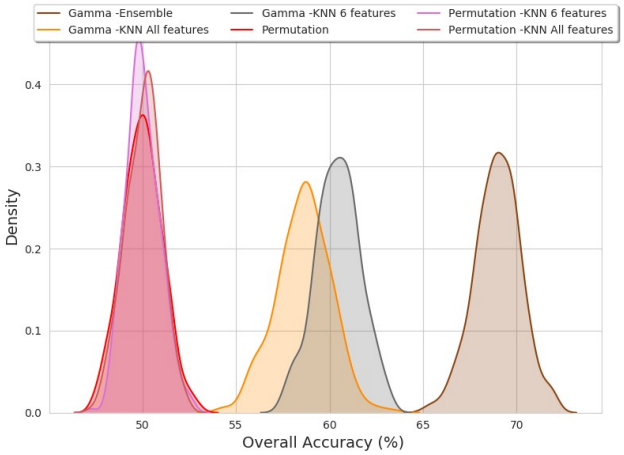


Figure 12.: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the KNN and Random Subspace Ensemble (RSE) and their permutation testing.

High Gamma

After three hundred iteration of bootstrap and permutation sampling through the high ranges of gamma frequency power dataset with random subspace ensemble method, the final mean classification accuracy was 68.9 per cent with 31.10 per cent mean error rate. The RSE model, as described above, used 32 learning cycles, each cycle using 3 randomly picked features with the K parameter being 3. The final mean classification accuracy for the permutation testing resulted in 49.9 per cent, validating the classification reliability (p<0.01). Figure 13A shows the percentage distribution of the classification across the classes. 69.2 and 69.1 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Classification using KNN, but without building an ensemble resulted in a mean classification accuracy of 59.4 per cent with 41.6 per cent error rate. The permutation testing resulted in 50.0 per cent (p<0.01). Figure 13B shows the percentage distribution of classification across the classes. 60.2 per cent and 58.6 per cent of the trials were correctly classified as remembered and not remembered, respectively.

High gamma frequency power dataset classified by KNN and with the identified features only (Table 1, 6th row), resulted in a mean classification accuracy of 60.7 per cent with 39.3 per cent mean error rate. The final mean classification accuracy for the

permutation testing resulted in 49.9 per cent, validating the classification reliability (p<0.01). Figure 13C shows the percentage distribution of the classification across the classes. 57.3 and 70.3 per cent of the trials were correctly classified as not remembered and remembered, respectively.

Figure 13: Percentage of Class Predictions With Random Subspace Ensemble and KNN

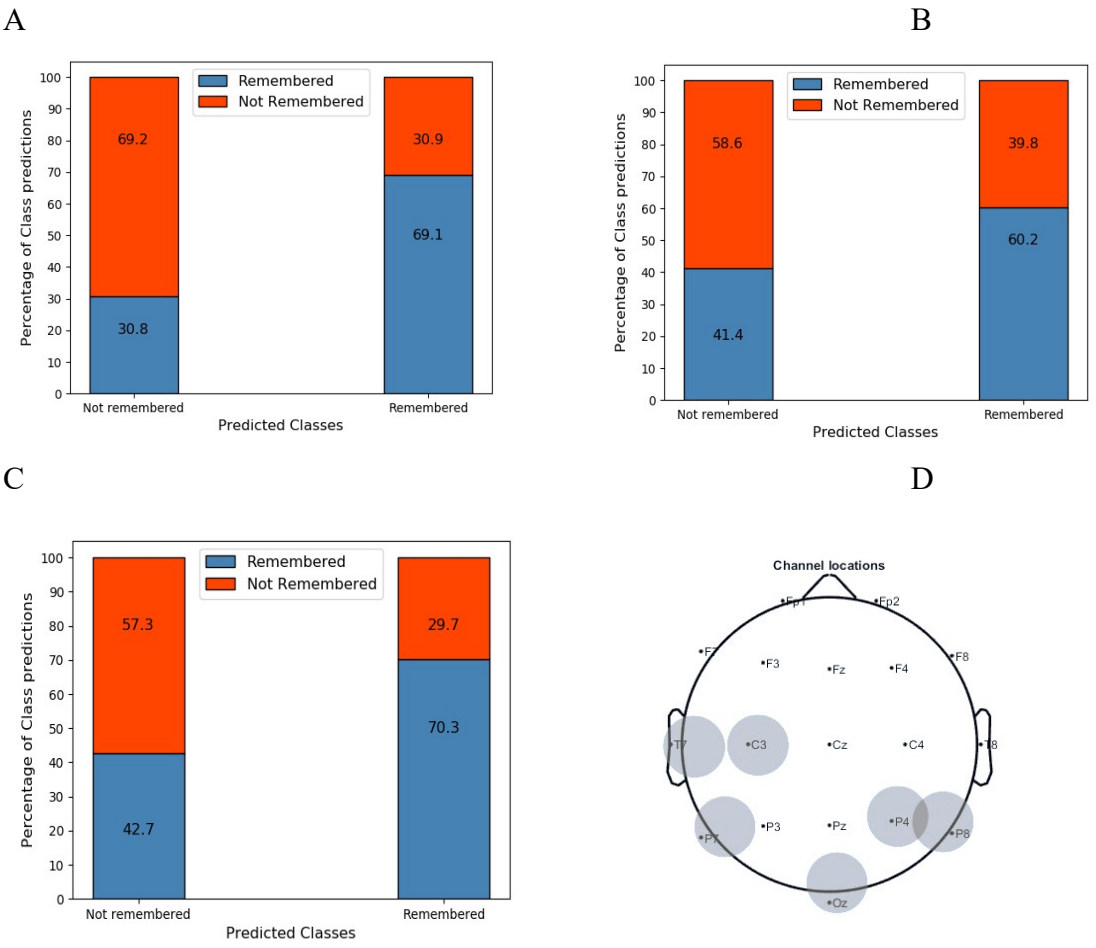


Figure 13**: A B, and C where** A figure shows the distribution of random subspace ensemble classification accuracy on the theta power band dataset, and B shows the classification accuracies of the KNN classification. C shows the classification accuracies achieved with KNN using only selected features by neighbourhood component analysis. The predicted classes are on the x axis, while the percentage of the class predictions are on the y axis. **D** represents the spatial positions of the electrodes of the selected features for the classification. These features were found using neighbourhood component analysis.

Evaluation of Classification Methods

On the high gamma power band dataset, RSE was found to be the most successful classifier. Both KNN and RSE were well above 95 per cent confidence interval of the permutation testing (p<0.01). The difference between means across all classification methods are statistically significant (p<0.01), while the comparison between mean accuracies across permutation testing are not significant, further validating the classification methods used (p=0.99). RSE methods performs best at achieving higher overall accuracies. The comparison of the algorithm results is shown on Figure 14.

Figure 14. Classification Method Comparison

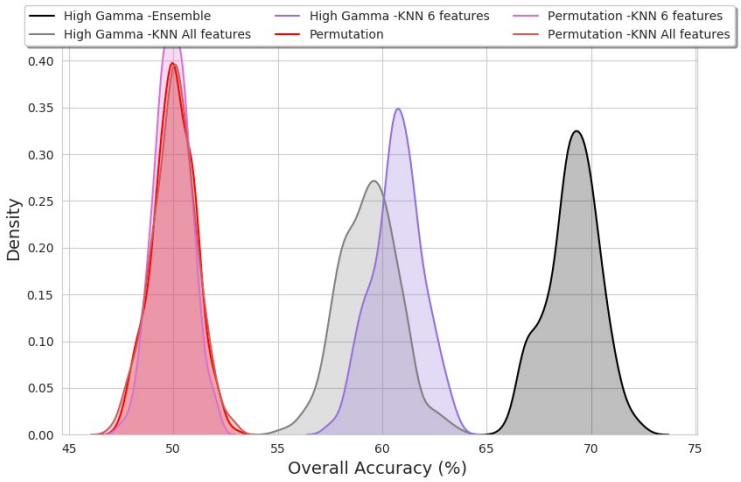


Figure 14: Gaussian kernel density estimate of the classification accuracies after bootstrap sampling across the KNN and Random Subspace Ensemble (RSE) and their permutation testing.

Summary of Results

When looking at the overall classification accuracy, a higher overall classification accuracy was achieved with RSE on every frequency power dataset. The best KNN classification was achieved with the selected features KNN (KNN6) with alpha and beta dataset, closely reaching the accuracy of the ensemble method, with 67.5 and 67.1 per cent overall classification accuracy.

Although RSE achieves better overall classification results, KNN6 classification accuracies outperforms RSE five out of six datasets when predictions are made on remembered items (Table 3.). KNN6 overall classification accuracies are consistently worse as a result of the worse classifications achieved on the not remembered items, while KNN (18 features) performs the worst across all datasets.

The following tables summarizes of the classification results. Table 2. Summarizes the overall average classification results for each frequency band dataset. Table 3 summarizes the correctly classified groups across classification methods.

Table 2.: Summary of the overall average classification results

	Random subspace Ensemble	**K-nearest Neighbour algorithm- 6 features**	**K-nearest Neighbour-algorithm-18 features**
Delta	**68.9**	60.5	58.9
Theta	**69.2**	60.4	59.3
Alpha	**69.3**	67.5	59.4
Beta	**69.4**	67.1	59.2
Gamma	**69.0**	60.4	58.6
High Gamma	**68.9**	60.7	59.4

Table 2. Summary of the overall average classification results for each frequency, separately. The results show the overall percentage of the correctly classified points after bootstrap sampling. The higher values are highlighted for each dataset.

Table 3.: Summary of the correctly classified groups across classification methods

	Random Subspace Ensemble		**K-Nearest Neighbour algorithm- 6 features**		**K-Nearest Neighbour algorithm -18 features**	
	Remembered	Not remembered	Remembered	Not remembered	Remembered	Not remembered
Delta	68.8	68.9	**69.5**	57.2	59.6	58.3
Theta	69.1	69.3	**69.6**	57.1	60.1	58.8
Alpha	69.2	69.4	**69.6**	65.8	60.0	58.8
Beta	69.2	**69.6**	69.1	65.5	59.9	58.7
Gamma	68.8	69.1	**69.9**	57.1	59.3	58.0
High gamma	69.1	69.2	**70.3**	57.3	60.2	58.6

Table 3. Summary of the correctly classified groups across classification methods. The highest values are highlighted for each dataset.

Discussion

Classification Methods

Three classification methods based on K-nearest neighbour (KNN) algorithm have been applied on six EEG frequency power datasets. These were delta, theta, alpha, beta, gamma, and high gamma frequency bands. The three classifiers' primary aim was to be able to reliably classify remembered and not remembered memory items during a working memory task from EEG data.

The overall best classification accuracy was achieved by an ensemble method using random feature selection. This suggests that increasing the feature space generally leads to a lower classification accuracy. Although, KNN is known to be able to generate complex decision boundaries during classification, it seems to have been influenced by the curse of dimensionality and a high noise-to-signal ratio when applied to each dataset. This is also supported by the comparisons of the algorithms that were used. As feature space was reduced, classification accuracies in both classes have always increased.

The overall accuracies ranging from the 58- 60% on each dataset using only KNN has potentially considerable amount of irrelevant information regarding WM status, and therefore is less successful the later discussed classification methods. As frequency power can vary through features within a dataset, with some features holding more relevant information about memory status, further steps were taken towards finding a method that can achieve better classification results.

These steps were KNN with 6 selected features (KNN6) found by neighbourhood component analysis (NCA) and a random subspace ensemble method based on KNN learners (RSE).

KNN6 has been found to achieve better classification accuracies in every dataset compared to the regular KNN classification (using 18 features), which supports the clue that each dataset has large amount of irrelevant information regarding WM memory status.

Interestingly however, KNN6 had the overall best classification accuracies on remembered items. Except beta dataset only, every dataset achieved the best classification accuracies using KNN6 on remembered items (Table3). As these features

were selected using NCA, this suggests that these features might not only reduce the feature space but might have a link to some key components of the remembered items in working memory (WM).

Even more interesting, that the KNN6 on the not remembered memory items resulted in the worst overall accuracies among all tested classifiers. This strongly suggests that the majority of the significant differentiating factors related to not remembered items in WM have been excluded by neighbourhood component analysis (NCA).

By having a closer look at the features found by NCA, there is a clear difference in the spatial distribution between most important features found in different frequency bands. For example, the selected features span around the fronto-temporal regions for the theta and alpha frequency power datasets, while the features selected from high gamma frequency band are located arount the occipital-parietal regions. These areas have been previously linked to WM. An increased frontal-parietal theta and alpha rhythm was shown to be increased during WM load, while phase-synchronization over parieto-occipital regions between theta and high gamma oscillations has also been reported during memory tasks (Gevins *et al.,* 1997, Holz *et al.*, 2010, McEvoy *et al*., 2001). As these areas were previously linked to encoding and retention in WM, this could explain why classification of remembered items are the most successful using these features, despite the features being different throughout the datasets.

The last classification method, RSE, has achieved the best overall classification accuracies throughout the datasets. The classification accuracies increase in every dataset as a result of more successful classification of not remembered items compared to the other classification methods described before. Therefore, this method is the most successful classification method to differentiate between WM success and failure using EEG data. For example, by looking at the most commonly discussed theta band (Table 3), it is clearly visible that however RSE performs worse on remembered items (69.1 % with RSE), a better accuracy is achieved by being able to classify both classes more accurately. KNN6 has the highest classification accuracy, 69.9 % on remembered items, however this could only correctly classify 57.1 % of the not remembered trials. Not remembered items were correctly classified by RSE at 69.3%, which explained the significantly higher overall classification accuracies. The same pattern appears in all datasets. RSE's classification advantage has arisen from being able to classify not remembered items with higher accuracy. In fact, with every dataset, RSE has achieved a better classification accuracy on not remembered trials, showing a slightly higher sensitivity towards this class.

Overall, the classification results suggest that although the data is not easily separable using simple, linear decision planes, but fine, small clustering between classes can be observed. This is supported by the need of a low *k* parameter, regardless of classification method. These small clusters are more prominent between classes in a lower feature space.

The RSE method presented here with the best overall accuracies would probably be the best during live system applications. Such system would allow the live observation and understanding of working memory whilst being able to predict potential task outcome.

Frequency bands and Working Memory

Frequency bands have shown a highly consistent result across frequency bands. This was surprising at first, as it was hypothesized that different frequency power bands might have different significance towards WM memory status, therefore, different datasets would probably result in better or worse classification accuracies.

The results of all three classification methods are very consistent across datasets. This means that with the same classification method, different datasets had highly similar classification accuracies, with differences ranging to a maximum 5 % between datasets. This was most likely the biproduct of the optimized classification methods.

All three classification methods were based on the training sets derived from the same frequency band. As the prediction on any unseen datapoint was based on the distance between the 3 nearest training datapoint, the classification methods do not directly highlight information on the power differences between datasets. This means, the algorithm did not look at differences appearing between and within datasets, rather each dataset was individually classified based on their own training dataset.

The difference between accuracies did not change highly in different neural frequency ranges. The applied classifiers detect the grouping nature of similar datapoints (and classes) for each feature. This may mean that whether an item was remembered or not, have similar tendencies across frequency bands, that these points, regardless of the differences in frequency power, likely to group together in their feature space, in a synchronized manner. It's also possible that different features drive the success of different frequency dataset, which is unknown (Except for those selected by NCA, where features were manually selected).

As mentioned before, the classification results for each frequency bands are too alike to be able to reliably distinguish between specific roles for each frequency band. However, it is worth noting that classification using selected features only on alpha and beta frequency bands resulted a much closer result to RSE than other frequency bands. Selected features classification resulted in 67.5 and 67.1 % overall accuracy for alpha a beta band, while RSE resulted in 69.3 and 69.4 %, respectively. This could mean that the identified features have even better properties that differentiate between WM success and failure in these frequency ranges.

The experiment had limited understanding of the changes of frequency power during working memory load as a result of two main factors; The beginning of each trial, the first three digits presented, was not used for classification, potentially eliminating any incremental power effect within a trial. Also, during this classification, the rank (or difficulty) of the digits (how many numbers were presented before) were not fed into the algorithm. Therefore, the potential power differences between trials, where the tenth digit was presented and where only the fourth digit was shown, was not considered by any of these algorithms.

Although, this generates a more generalised and unbiased classification, but it also removes some key component of understanding WM.

In summary, the above described methods have certainly highlighted the potentials of classification methods applied on EEG data. This experiment also achieved a stable classifier that can distinguish between remembered and not remembered items in WM with near 70% accuracy. The experiment, however, did not show high enough difference between neural frequency bands to obtain reliable conclusions regarding the relationship between WM processes and frequency bands.

References

Abu Alfeilat, H., Hassanat, A., Lasassmeh, O., Tarawneh, A., Alhasanat, M., Eyal Salman, H. and Prasath, V. (2019). Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data*.

Baddeley, A. (2010). Working memory. *Scholarpedia*, 5(2), p.3015.

Brainard, D. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), pp.433-436.

Canolty, R., Edwards, E., Dalal, S., Soltani, M., Nagarajan, S., Kirsch, H., Berger, M., Barbaro, N. and Knight, R. (2006). High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex. *Science*, 313(5793), pp.1626-1628.

Chaieb, L., Leszczynski, M., Axmacher, N., Höhne, M., Elger, C. and Fell, J. (2015). Theta-gamma phase-phase coupling during working memory maintenance in the human hippocampus. *Cognitive Neuroscience*, 6(4), pp.149-157.

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O. and Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), pp.769-786.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), pp.9-21.

Derhab, A., Guerroumi, M., Gumaei, A., Maglaras, L., Ferrag, M., Mukherjee, M. and Khan, F. (2019). Blockchain and Random Subspace Learning-Based IDS for SDN-Enabled Industrial IoT Security. *Sensors*, 19(14), p.3119.

Gevins, A., Smith, M., McEvoy, L. and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7(4), pp.374-385.

Goldberger, J., Hinton, G., Roweis, S. and Salakhutdinov, R. (2005). Neighbourhood Component Analysis. *Advances in Neural Information Processing Systems*, (17), pp.513-520.

Ho, T. (1998). *Nearest Neighbors in Random Subspaces*. [online] 700 Mountain Avenue. 2C-425, Murray Hill, NJ 07974, USA: Bell Laboratories, Lucent Technologies. Available at: https://link.springer.com/content/pdf/10.1007/BFb0033288.pdf [Accessed 17 Aug. 2019].

Holz, E., Glennon, M., Prendergast, K. and Sauseng, P. (2010). Theta–gamma phase synchronization during memory matching in visual working memory. *NeuroImage*, 52(1), pp.326-335.

Howard, M. (2003). Gamma Oscillations Correlate with Working Memory Load in Humans. *Cerebral Cortex*, 13(12), pp.1369-1374.

Jasper, H. (1958). The Ten Twenty Electrode System: International Federation of Societies for Electroencephalography and Clinical Neurophysiology. *American Journal of EEG Technology*, 1(1), pp.13-19.

McEvoy, L., Pellouchoud, E., Smith, M. and Gevins, A. (2001). Neurophysiological signals of working memory in normal aging. *Cognitive Brain Research*, 11(3), pp.363-376.

Meltzer, J., Zaveri, H., Goncharova, I., Distasio, M., Papademetris, X., Spencer, S., Spencer, D. and Constable, R. (2007). Effects of Working Memory Load on Oscillatory Power in Human Intracranial EEG. *Cerebral Cortex*, 18(8), pp.1843-1855.

Palva, J., Monto, S., Kulashekhar, S. and Palva, S. (2010). Neuronal synchrony reveals working memory networks and predicts individual memory capacity. *Proceedings of the National Academy of Sciences*, 107(16), pp.7580-7585.

Pestov, V. (2013). Is the k-NN classifier in high dimensions affected by the curse of dimensionality?. *Computers & Mathematics with Applications*, 65(10), pp.1427-1437.

Roux, F. and Uhlhaas, P. (2014). Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information?. *Trends in Cognitive Sciences*, 18(1), pp.16-25.

Onton, J., Delorme, A. and Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *NeuroImage*, 27(2), pp.341-356.

Sammer, G., Blecker, C., Gebhardt, H., Bischoff, M., Stark, R., Morgen, K. and Vaitl, D. (2007). Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload. *Human Brain Mapping*, 28(8), pp.793-803.

Sarnthein, J., Petsche, H., Rappelsberger, P., Shaw, G. and von Stein, A. (1998). Synchronization between prefrontal and posterior association cortex during human working memory. *Proceedings of the National Academy of Sciences*, 95(12), pp.7092-7096.

Sauseng, P., Griesmayr, B., Freunberger, R. and Klimesch, W. (2010). Control mechanisms in working memory: A possible function of EEG theta oscillations. *Neuroscience & Biobehavioral Reviews*, 34(7), pp.1015-1022.

Sauseng, P., Klimesch, W., Schabus, M. and Doppelmayr, M. (2005). Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory. *International Journal of Psychophysiology*, 57(2), pp.97-103.

Sun, S., Zhang, C. and Zhang, D. (2007). An experimental evaluation of ensemble methods for EEG signal classification. *Pattern Recognition Letters*, 28(15), pp.2157-2163.

Tuladhar, A., Huurne, N., Schoffelen, J., Maris, E., Oostenveld, R. and Jensen, O. (2007). Parieto-occipital sources account for the increase in alpha activity with working memory load. *Human Brain Mapping*, 28(8), pp.785-792.

Yang, W., Wang, K. and Zuo, W. (2012). Neighborhood Component Feature Selection for High-Dimensional Data. *Journal of Computers*, 7(1).

Yao, Z. and Ruzzo, W. (2006). A Regression-based K nearest neighbor algorithm for gene function prediction from heterogeneous data. *BMC Bioinformatics*, 7(S1).