Queen Mary, University of London
School Of Engineering and Material Science

# Multi-Class Classification of EEG Signals Using a Fuzzy Based Rule System for Brain Computer Interfaces

Owais Mohammad Siddiqi

BEng. Biomedical Engineering

Supervisor: Prof Jun Chen

April 2022

School Of Engineering and Materials Science

Engineering/Materials

Third Year Project

DEN318

April 2022

Declaration

This report entitled

MULTI-CLASS CLASSIFICATION OF EEG SIGNALS USING A FUZZY BASED RULE SYSTEM FOR BRAIN COMPUTER INTERFACES

Was composed by me and is based on my own work. Where the work of the others has been used, it is fully acknowledged in the text and in captions to table illustrations. This report has not been submitted for any other qualification.

Name:

Owais Mohammad Siddiqi

Signed:

Owais Mohammad Siddiqi

Date:

22 April 2022

“*Whether you think you can, or you think you can't - you're right.*”

— Henry Ford

# Acknowledgements:

First and foremost, I would like to give a special thanks to my family and friends for their overwhelming love and support and for always reminding me to strive for success.

I would also like to thank all the staff and students at Queen Mary, University of London who have helped me in not only my academic development but my personal one as well.

Thank you to Dr Elham Zareian and Prof Jun Chen, for providing the fully processed data on Subject 1: Alex used in this project, Dr Zareian was also responsible for all the processed data extraction.

Last but not least, I would like to thank my project supervisor Professor Jun Chen for his constant support and guidance and for pushing me out of my comfort zone by broadening my skills as a researcher and helping me develop, improve, and expand upon my knowledge of brain computer interfaces, machine learning, and engineering.

# Abstract

The purpose of this project is to develop a brain computer interface to be able to classify and predict brain activity based on EEG – SSVEP data. This method can potentially be used in assisting neurological condition detection and diagnostics. The main difficulties that BCI's face when interpreting electroencephalogram data are inaccuracies in the model's prediction due to the data's complex and chaotic nature, leading to a great amount of variety in each individual sample. As with any machine learning algorithm, inaccuracies are often due to the algorithm's inability to understand and learn from the data and adapt to new coming information, which is most likely to occur in real life situations. To counteract this problem, the report used ANFIS instead of normal machine learning algorithms such as SVM and K – NN, in hopes that its hybrid composition of both fuzzy modelling and artificial neural networks would overcome this hurdle and achieve greater results. In addition to the use of ANFIS, the bagging ensemble method was also incorporated into the model to improve its predictive capabilities. The ensemble method was found to be effective, beating all baseline models (including a normal multi-class ANFIS classifier) in all evaluation metrics measured, ultimately achieving an accuracy score of 64%.



**Keywords** – Brain Computer Interface, Machine Learning, Neural Networks, ANFIS, Supervised Learning, Multi – Class Classification, Fuzzy Modelling, Evaluation Metrics, EEG - SSVEP

# Contents

I. Introduction ... 1
i. Aims and Objectives ... 2
II. Literature Review ... 2
i. Use of Neural Networks in SSVEP Classification ... 2
ii. Use of Ensemble Method ... 3
iii. Comparing ANFIS To Other Machine Learning Models ... 3
III. Machine learning ... 4
i. Support Vector Machine ... 5
ii. K – Nearest Neighbours ... 5
iii. Fuzzy Modelling ... 6
iv. Neural Networks ... 8
v. ANFIS ... 9
vi. Ensemble Method (Bagging) ... 9
vii. Decomposition Method ... 10
viii. Aggregation Method ... 10
ix. Evaluation Metrics ... 10
IV. Method ... 12
V. Results ... 14
VI. Discussion ... 15
i. Improvements ... 16
VII. Conclusions ... 17
i. Future Work ... 17
VIII. ... 18
IX. Appendix ... 21

## List of Tables

Table 1 - Overview of the Ensemble - ANFIS algorithm ... 13

Table 2 - Evaluation Metrics of All Classifiers Used for All Proposed Models (In Descending Order of Accuracy) | Source: MATLAB & Python ... 14

# Abbreviations

| | |
|---|---|
| ANFIS | **A**daptive **N**euro **F**uzzy **I**nference **S**ystem |
| ANN | **A**rtificial **N**eural **N**etworks |
| BCI | **B**rain **C**omputer **I**nterface |
| DDAG | **D**ecision **D**irected **A**cyclic **G**raph |
| DT | **D**ecision **T**ree |
| EEG | **E**lectroencephalogram |
| FIS | **F**uzzy **I**nference **S**ystem |
| FRBS | **F**uzzy **R**ule-**B**ased **S**ystem |
| FCM | **F**uzzy **C-M**eans |
| K – NN | **K-N**earest **N**eighbours |
| NN | **N**eural **N**etworks |
| OAA | **O**ne – **A**gainst - **A**ll |
| OAO | **O**ne – **A**gainst - **O**ne |
| RMSE | **R**oot **M**ean **S**quare **E**rror |
| SNR | **S**ignal to **N**oise **R**atio |
| SSVEP | **S**teady **S**tate **V**isual **E**voked **P**otential |
| SVM | **S**upport **V**ector **M**achines |

# I. Introduction

To be able to understand the brain is to be able to understand oneself. All human thoughts, feelings and memories are encased in this single organ, and to be able to access this information in its rawest form and understand it clearly is what this project aims to do, by building a foundation that can move on to greater advancements if future research is conducted. This project is aimed towards developing the starting foundations of a brain computer interface, a device that can interpret and understand the electrical charges emitted from human brain activity [1], in the form of Electroencephalogram (EEG) data (figure 1). BCI's can be used for a plethora of tasks, such as the prediction and classification of neurological functions and the restoration of quality of life following a devastating accident or decline [2].

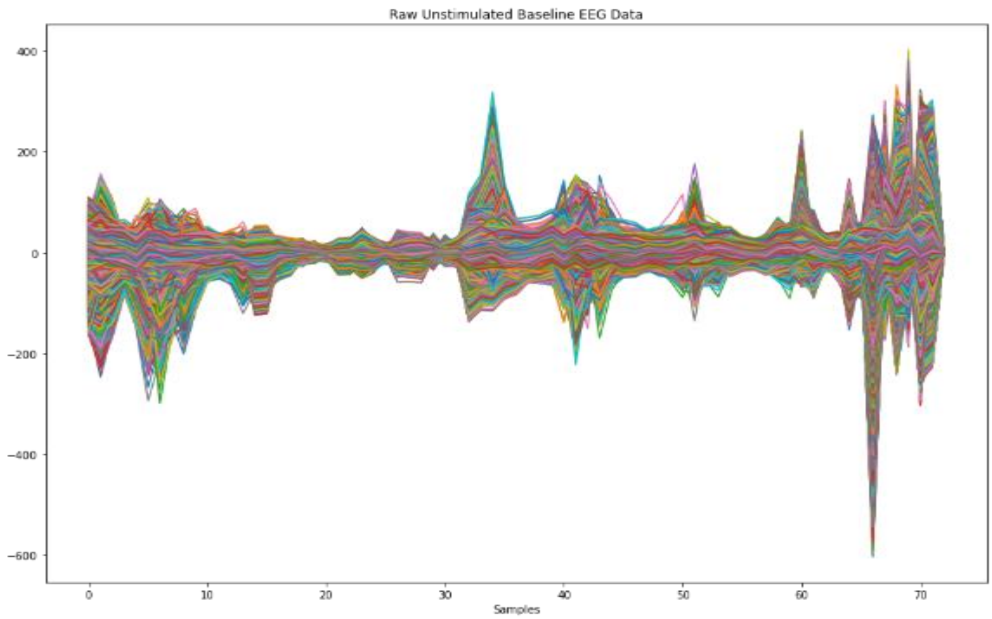


*Figure 1 - Pre - processed EEG data retrieved from subject: Alex during visual stimulation*

EEG is a non – invasive procedure that uses electrodes attached to the head of the patient to measure brain activity, SSVEP data is detected from this brain activity as the natural responses generated from the visual stimulation of the subject at certain frequencies [3]. Rather than using a traditional EEG, which only provides information when the brain is in a state of activity, a Steady State Visual Evoked Potential (SSVEP) uses a flashing light to stimulate the brain in a state of rest. This means that SSVEP is able to provide data regardless of whether the brain is in a state of activity or a state of rest. Using SSVEP data, the model is able to classify labels using ANFIS, SVM, and K-NN algorithms. In this project, the signals received from the brain were recorded as SSVEP data due to its improved SNR and reduced risk of artifacts when compared to EEG data (as EEG data contains a lot of additive noise [4]) and is therefore easier to interpret for the BCI [5].

When the SSVEP data is received, it is grouped into distinct categories based on the brain's activity and level of stimulation at the moment of recording. The main ones of interest in the SSVEP data are the Alpha, Beta and Gamma frequencies [6]. Alpha waves are emitted as times of relaxation, typically found at 8 – 12 Hz [6], Beta waves are emitted at times of awareness, alertness, and anxiousness, found at 12 – 30Hz [6], and lastly Gamma waves are emitted at times of simultaneous information processing, used for learning and memory retention, these waves are found at 30- 40 Hz [6].

### i. Aims and Objectives

The aim of this project is to develop, train and successfully evaluate a multi-Classification model, and improve its prediction capabilities using techniques such as the bagging ensemble method and a fuzzy inference system, resulting in a brain computer interface that can read, classify, and predict new EEG data.

Then evaluate the model's success in its ability to classify EEG data based on trained labels with the use of evaluation metrics to understand its individual performance and performance as a model by comparing it to a baseline and deciding whether ANFIS improves upon conventional machine learning classifiers and its logic.

Referring to the original ROA report, all objectives stated were addressed and achieved. Python was used to complete the SVM and K-NN bagging ensemble algorithms and their performances were judged based on evaluation metrics using sci-kit learn [7], with these values acting as a baseline to compare the ANFIS to. The Bagging ensemble – ANFIS model was successfully created, in MATLAB, using the OAO decomposition approach and Average Voting Strategy aggregation, with its evaluation metrics calculated and compared to the baseline.

## II. Literature Review

The world today is a complex and dynamic place. In this world, machines have been performing a wide variety of tasks that were traditionally the domain of humans. Rapidly developing machine learning algorithms are at the forefront of this change and have been achieving some truly remarkable results. One such algorithm is the Adaptive Neuro-Fuzzy Inference System (ANFIS), which has been demonstrated to be a powerful tool for classification and regression tasks in a variety of domains.

### i. Use of Neural Networks in SSVEP Classification

Neural networks and classical machine learning algorithms have been used in BCI development for years, applying to both regression and classification tasks as NN's are both accurate and provide a fast, efficient solution to a dataset's problem. In this study, different algorithms (including an ANN) were compared by their accuracy in a dataset of diabetic patients, tasked with the goal of predicting diabetes in young patients. The results spoke for themselves, concluding that the ANN outperformed the other machine learning algorithms, as it had the most accurate classification probability (76%) [8]. This implies that ANN work with high accuracy and learning ability and thus should be used when developing classification models. This study supports the use of ANFIS due to its NN composition, the addition of fuzzy logic can only bring hopes in improving upon its classification ability.

In regard to developing classification models for EEG - SSVEP based BCI's, this study found that ANFIS, used in real time classification tasks, successfully identifies frequencies with an accuracy of 90%, and can be used as a viable tool for BCI's and other applications [9]. Using ANFIS resulted in a high classification accuracy and, while there were no baseline models to compare to, its accuracy was high enough to justify its use in this application and in other BCI models.

This study [10] was aimed at presenting a performance analysis on ANFIS, in its ability to classify EEG signals, by comparing it to a linear SVM. ANFIS provided a useful insight between the relationship of input features and class labels [10]. The ANFIS model had also achieved a better classification accuracy than the linear SVM to which it was compared. This study justifies the use of ANFIS instead of traditional machine learning as it improved upon the SVM model even without the application of an ensemble method.

### ii. Use of Ensemble Method (Bagging)

Ensemble based methods have also been incorporated into ANFIS and often generate a greater predictive probability than any individual classifier used [11], such as in this study [12], where an Ensemble – ANFIS, decomposed using the OAO approach and aggregated using the Average Voting Strategy (Bagging Method), was applied in a BCI to classify EEG channels. The study, performing four different BCI sessions, cross validated their data and evaluated the classifiers performance, resulting in a model with a 75% plus average accuracy across all subjects, with the best classifiers achieving above 80% accuracy and F1 scores. Supporting the benefit of ensemble classifiers and their use in multiclass classification tasks.

The method of using fuzzy c – clustering, which is occasionally used in BCI development [13], was considered for this project however literature presents issues with the use of FCM. Subtractive Clustering was compared to FCM in this study [14] revealing that FCM provided inconsistent results and required training to be used whereas Subtractive Clustering did not, as its results were more reliable and consistent in comparison. Thus, the FCM was replaced for the superior Subtractive Clustering Sugeno – type model.

### iii. Comparing ANFIS To Other Machine Learning Models

Research has delved into the comparison of ANFIS to other machine learning algorithms to judge its capabilities, specifically with SVM and K-NN's. In this study [15], where machine learning algorithms were used to predict the risk of COVID – 19 in patients, SVM, K-NN and DT were compared to ANFIS. Research found that SVM had the highest accuracy, Precision, Sensitivity and Specificity, with all measures being 100%, and K-NN achieving a 94.8%

accuracy. Thus, models such as ANFIS can judged on its performance based on its accuracy compared to SVM and K-NN classifiers, due to how powerful those classifiers are. The capabilities of these two algorithms are further supported, where both SVM and K-NN achieved a 100% accuracy score in their ability to classify EEG signals for BCI's [16].

# III. Machine learning

The definition of machine learning is the "capability of a machine to imitate intelligent human behaviour" [17]. Machine learning as a whole can be split into two types of learning, supervised and unsupervised, the main difference between these two are their inclusion of labels, tags that define what the machine is reading so it can learn using context, with supervised predefined and unsupervised not [18]. Supervised learning can then be split into two sub - types of tasks, a regression task, and a classification one.

Regression tasks are ones that predict continuous values, values where there are an unspecified number of possible values between two points, with examples ranging from house prices to age ranges of populations. Classification tasks on the other hand, predict discrete values, values that are finite in nature and fit into specified categories (figure 2).

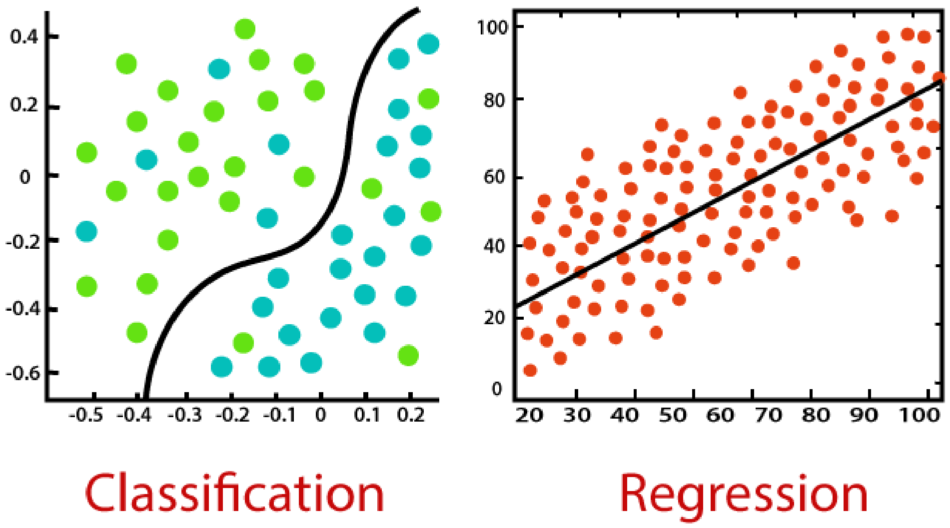


*Figure 2 - Visual representation of a machine learning model used in a classification task and a regression task | Source:* [19]

There are also many different types of classification tasks, the two main types are:

- Binary Classification
- Multi-Class Classification

This project's task is a multi-class classification one, where the model is trained on the EEG – SSSVEP dataset which has three different labels (10 Hz [1], 14 Hz [2], 21 Hz [3]) and based on this training, has to predict, and classify new EEG - SSVEP data within one of these frequencies.

This brain computer interface is a supervised – learning multiclassification model used to classify and predict new EEG - SSVEP data received from volunteers. Typical multiclassification algorithms include SMV and K – NN Classifiers which are both machine learning algorithms, with SVM typically being used in BCI designs [20].

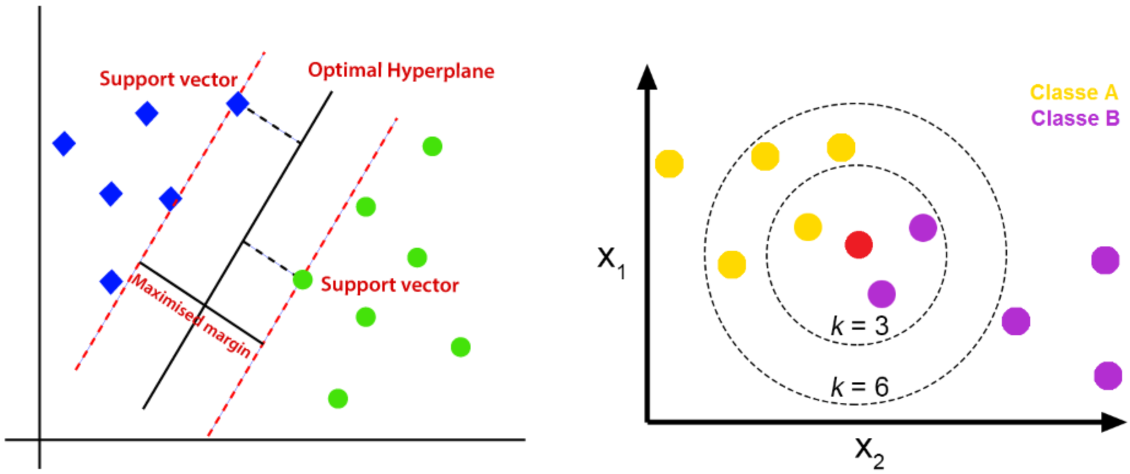


*Figure 3 - (Left) Visual representation of a SVM being used in a classification task, including hyperplane, and maximised margin | Source:* [21] *(Right) Visual representation of a K - NN algorithm in use, demonstrates the effect of changing the k parameter | Source:* [22]

### i. Support Vector Machine

SVM's can be used on both regression tasks and classification ones. When used in regression tasks, this algorithm is designed to find the optimal hyperplane between the two or more classes, to separate them with as great a margin as possible [20] (figure 3). When used for classification purposes, the data is mapped into an additional space, one with a higher dimensionality, and a hyperplane (similar to the regression SVM) is applied to seperate the different labels [20]. One disadvantage of SVM's is that this mapping of data to a higher dimensional place can lead to data loss [4]

### ii. K – Nearest Neighbours

K – NN works in a feature space, where feature vectors are presented (figure 3). The feature vector belonging to the test data that is being predicted is located in this space and is classified based on a *k* parameter, which defines the number of samples closest to the feature vector that will be used to classify the test data. K – NN algorithms are used commonly for EEG data due to "their sensitivity to the local distribution of feature vectors" [23]. One disadvantage to K – NN is difficulty in selecting the optimum value of the *k* parameter [24] without the use of a cross-validation technique.

## iii. Fuzzy Modelling

What makes this project different to current literature is the use of a hybrid classification algorithm called ANFIS to classify the data, combining the human interpretable "IF – THEN" rules generated by Fuzzy logic and the learning capabilities of neural networks to increase the predictive probability of the model [25], [26] than that of traditional machine learning.

The use of fuzzy modelling provides a better predictive probability due to the previously stated "IF – THEN" rules. These rules help the human interpretability of the logic and provide a more generalised approach [27], mimicking human reasoning to more accurately understand and interpret data [27], perfect for varied data such as EEG's. This specific fuzzy logic and rule set is called the Sugeno – model, used typically in non-linear classification systems due to its ability to approximate with linear systems instead, which is currently used in literature [4].

As you can see from figure 4, the FIS starts as a crisp input which is then sent to the fuzzy interface which converts the input to a fuzzy one. The fuzzy input is then sent to the decision – making unit where the inference operation is conducted based on the rule base (figure 5), containing the if – then rules, and the database, which defines the membership functions (figure 6) of the fuzzy sets [28]. Lastly, the defuzzification of the value occurs, where it is converted back into a crisp output.

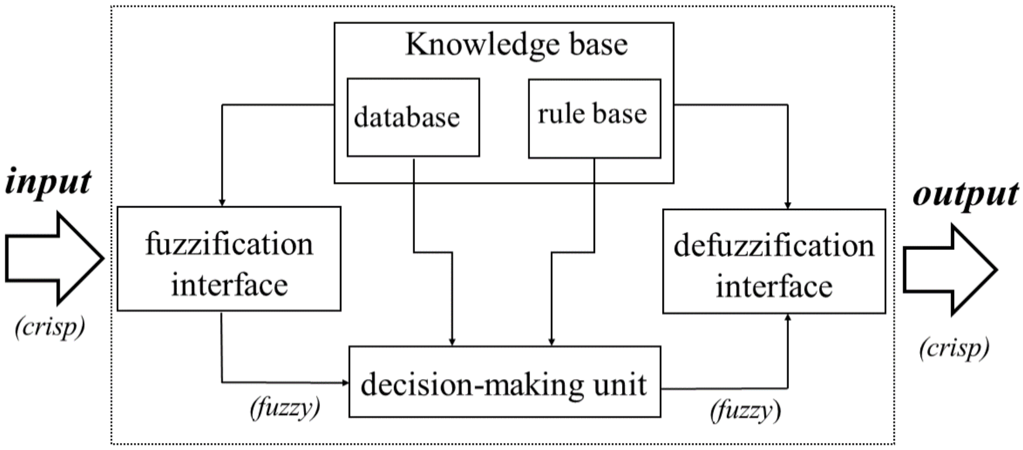


*Figure 4 - Diagram of a Fuzzy Inference System* [4]

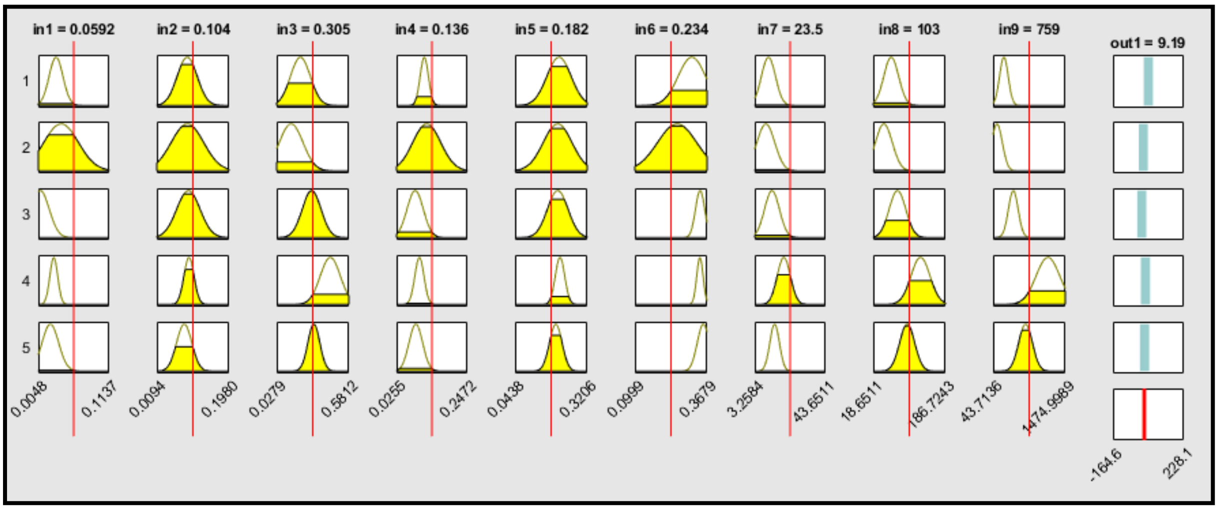


*Figure 5 - ANFIS Rule set derived from one of the binary classifications of SSVEP data | Source: MATLAB*

Membership functions (figure 6) are functions that assign membership values to each input data value, determining which rule set (figure 5) this value belongs to and to what degree of certainty [4], [28]. These fuzzy modelling membership functions increase the smoothness of the rule set [28].

There are different types of FIS's used in recent literature, the main two being the Mamdani [29] and Sugeno [30] types. The Sugeno – type is used in this project due to it being based upon the original Takagi – Sugeno Inference system and having a greater accuracy compared to Mamdani in these studies [31].

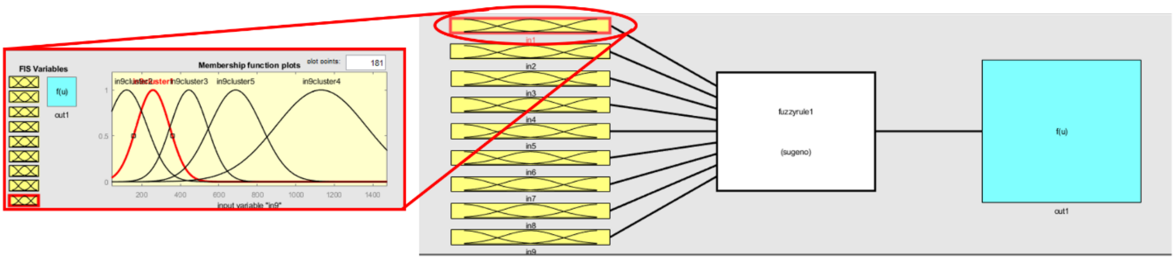


*Figure 6 - Membership functions plot from a Sugeno - Type ANFIS rule set | Source: MATLAB*

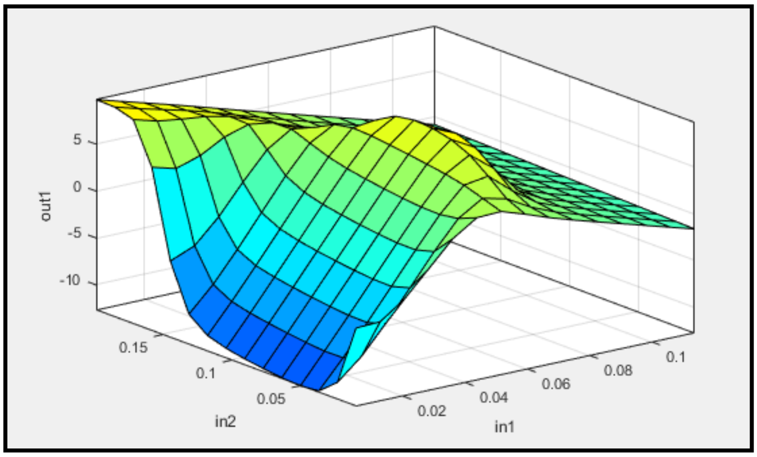


*Figure 7 - Surface profile of a Sugeno - type fuzzy rule set | Source: MATLAB*

Recent literature [32] has used FISs for the classification of EEG signals, with it even shown to outperform the linear classifier used in that study, with a similar accuracy to the SVM classifier also used.

## iv. Neural Networks

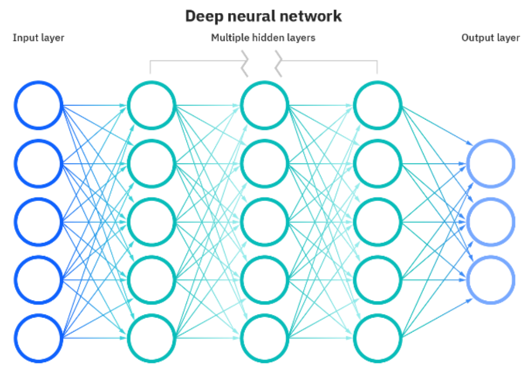


*Figure 8 - Architecture of Neural Networks | Source:* [34]

ANN's are part of a deeper sub – category of machine learning, called deep learning. The difference between the two is that deep learning removes the need for human interaction in data handling. The cleaning, refinement, and construction of feature vectors and subsequent labelling are all completed by the ANN through the use of multiple 'layers' (figure 8). Each layer is comprised of nodes, with the initial 'input' layer receiving the data, and the hidden layers in-between, extracting each and every feature, weight, and threshold from the original data, with each hidden node mostly improving upon the interpretation of the initial data [33]. The architecture was inspired by the human brain, designed to mimic the procedure in which the biological neurons interact and signal one another [34]. Each and every node has a mathematical architecture, called an activation function, to be able to understand and interpret substantial amounts of data quickly, with weights $\omega$, input $x$ and $b$ bias [34]:

$$f(x) = \sum_{i=1}^{m} \omega_i x_i + b$$

At the end of the network lies the output layer, where all complex calculations and functions that have been processed are sent to, leaving a set of outputs. The most important feature is its ability to learn and adapt to convoluted tasks, with it being the forefront of research in a plethora of disciplines and is used in everyday life such as in Google's automatic translator [35]. One of the main drawbacks of using deep learning for multiclassification tasks is that they require requires greater computational resources and data in order to achieve reliable results [36].

## v. ANFIS

ANFIS or (Adaptive Neuro Fuzzy Inference System) is a type of artificial neural network (figure 9), based on the Takagi – Sugeno inference system [25]. It is a unique hybrid system, integrating both neural networks and fuzzy logic into a single machine learning framework, referred to as the "universal estimator" due to its ability at generalisation [37].

ANFIS's main benefit over other neural networks and traditional machine learning is its fuzzification. The first layer is what differs to a typical neural network, as while NN's are mostly used with normalised pre-processed data, the ANFIS does not require this step, as the pre-processing is already completed through the conversion of numerical values to fuzzy ones in the fuzzification layer [38].

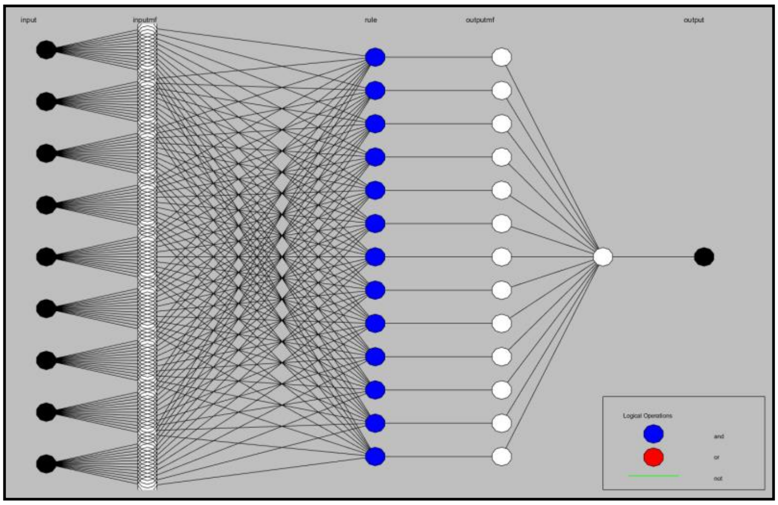


*Figure 9 - Example of the Sugeno – Type ANFIS neural network architecture | Source: MATLAB*

## vi. Ensemble Method (Bagging)

The classification ability of any model used in this project is hindered due to the task being a multiclassification task, instead of a simple binary one. To tackle this hurdle, the multiclass classification task is broken down into its bare parts, classified and predicted separately and then combined together at the end, to improve upon the overall predictive probability of the model [11]. The method to do such a task requires a decomposition method and a subsequent aggregation method. When combined, these two approaches develop into the bagging ensemble method, with studies showing that using the ensemble method "is often more accurate than any of the single classifiers in the ensemble" [11]. This method has been highly successful, leading to its use in several sectors including "optical character recognition, face recognition, scientific image analysis, medical diagnosis" [39].

### vii. Decomposition Method

There are two types of decomposition approaches to break down a multiclass into binary: OAO and OAA. The use of OAA is more computationally complex as it requires one class to be compared to all others and the same again with each other class. However, the OAO approach only requires one class to be compared to another (figure 11), reducing the computation costs and training time required to run this method [40].

### viii. Aggregation Method

The aggregation method is used to build back the separate binary classifiers that have been created from OAO into a singular multiclass classifier. In this project, the 'averaged voting strategy' was used. This method takes the contributions of each binary classifier as a 'vote' and adds them together, once averaged, this vote will be outputted as the final prediction outcome. There are other variations of this type of aggregation methods such as the Voting strategy. With the voting strategy, each binary classifier contributes towards a vote for each class prediction, the class with the maximum vote is put forward as the predicted class for the sample. This study [41] found that overall, the average voting strategy achieved a greater accuracy for most classification tasks, used on both SVM and logistic regression algorithms. It was deemed more effective in a wider range of scenarios and was "recommended for use in practice" [41].

Combining these two processes results in the bagging ensemble method, which takes the singular weak homogenous predictors (the 3 ANFIS models) and re-combines them using an averaging strategy (figure 11). This strategy is used to provide a better prediction compared to each individual model would on its own and reduce the variance within a singular learning algorithm, lowering the risk of overfitting [49].

### ix. Evaluation Metrics

Calculating the evaluation metrics; accuracy, precision, recall, F1 etc, are relatively simple for a binary classification task, however, when additional classes are placed into the mix and the problem turns into a multi-class, the situation becomes difficult as most of the definitions of the metrics relate to one independent class i.e., a 1 in a classification between 1 and 0. While it is certainly possible (figure 10), using a weighted average of each metric instead provides a better idea of how well the model works as it "considers how many of each class there were in its calculation" [42].

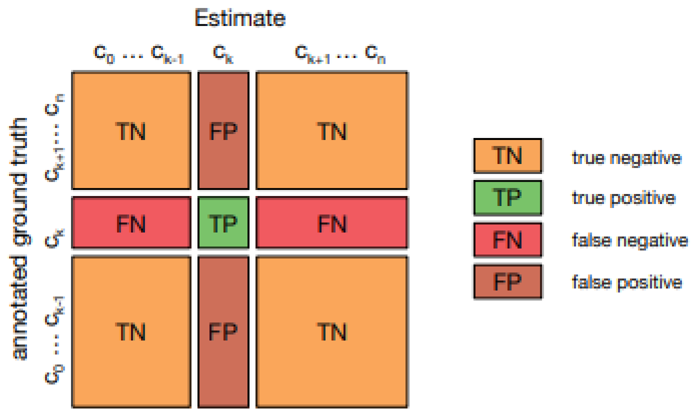


*Figure 10 - A k x k confusion matrix, specific to multiclassification tasks, used in machine learning to view the performance of the model | Source:* [43]

A confusion matrix is an effortless way to view first-hand the performance of a model based on the correct and incorrect predictions made. Figure 10 is a k x k confusion matrix with a k – number of classes, it demonstrates how TP, TN, FP, and FN can be found. From here, all evaluation metrics can be calculated using a combination of these values.

**Accuracy:**

$$\frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

**Precision:**

$$\frac{TP}{TP+FP} \tag{2}$$

**Recall:**

$$\frac{TP}{TP+FN} \tag{3}$$

**F1:**

$$2 * \frac{Precision * Recall}{Precision+Recall} \tag{4}$$

**RMSE:**

$$\sqrt{\frac{\sum_{n=1}^{N}(\hat{r}_n - r_n)^2}{N}} \tag{5}$$

Where $\hat{r}_n$ is the prediction value; $r_n$ is the true rating in testing data set; $N$ is the number of samples in the testing dataset [44].

Accuracy is the measurement of how many classes have been correctly classified, both positive and negative. There is a reason that metrics other than accuracy are used when evaluating a model, accuracy does not consider the balance of data within a data set. If the data set is mostly comprised of one class and the model achieves 100% accuracy under one class only yet a 0% on the rest of the classes, that model would have a still have a high accuracy yet fundamentally it would not be accurate at all. Context is particularly important when understanding these performance measuring metrics, for example if a model has a precision (the prediction probability of only positive values [43]) of 75%, without any context that is a high number. However, if a wrong classification is the difference between life and death then even 75% is not high enough, and in many BCI applications that is the case.

Recall, which is the rate of predicting true positive values [43], has a comparable situation to precision in which the context matters. The F1 score is the "harmonic mean of precision or recall" [45], and while it is a good measure of evaluating a classification model, it has one main drawback, it gives precision and recall equal weighting and depending on the context this may not always be the case, such as if the number of false negatives creates a greater risk than the number of false positives which can be the situation in healthcare applications.

# IV. Method

As discussed previously in this project, using an ensemble classifier for a multiclassification project typically leads to a greater predictive probability [11], or is as good as the best classifier in its ensemble. Thus, ANFIS was used to generate multiple FIS binary classifiers, applying the OAO decomposition method, and the average voting strategy to aggregate these classifiers together into one singular multi-class classifier, to improve upon any singular binary classifier.

The ensemble method works as shown in figure 11, illustrating the main streps taken to develop the ensemble classifier.

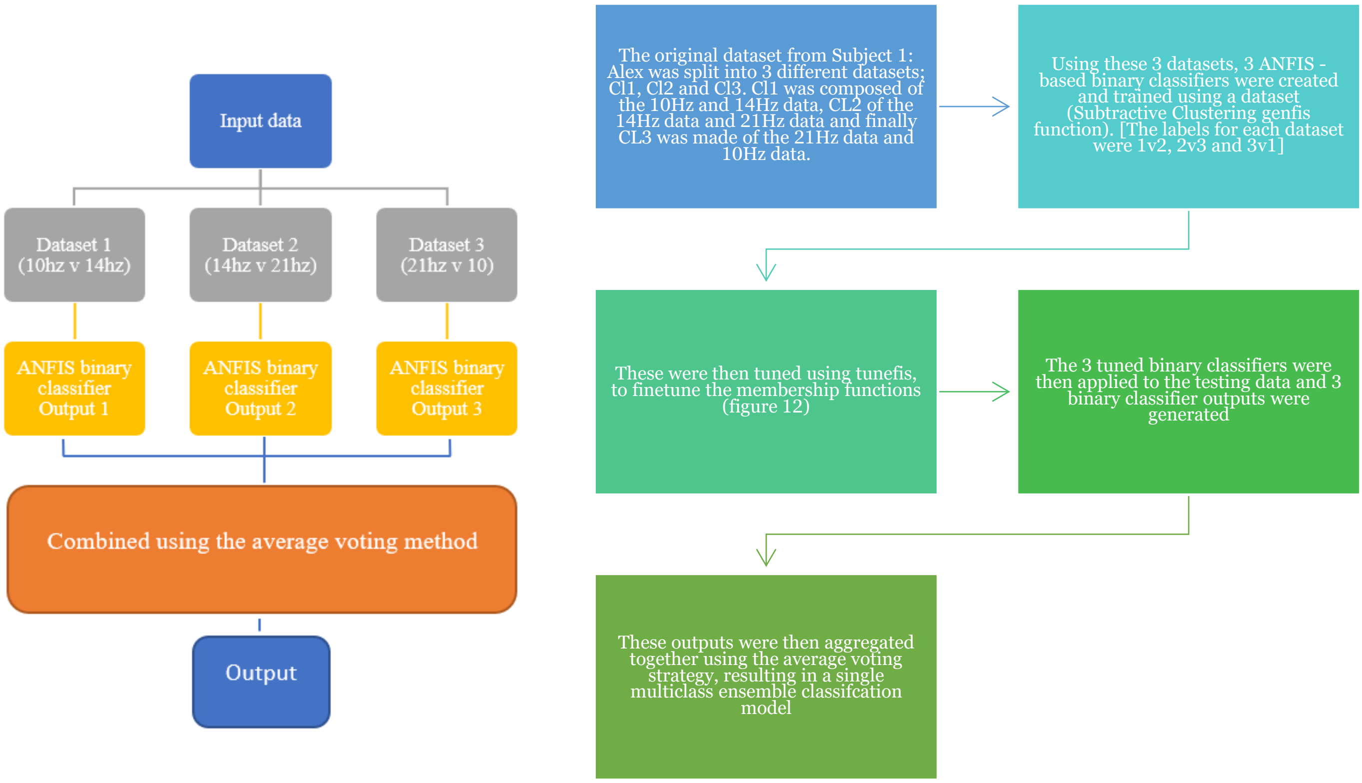


*Figure 11 - The ensemble classification methods (left) and the steps taken to develop an ensemble based multi-classification ANFIS model*

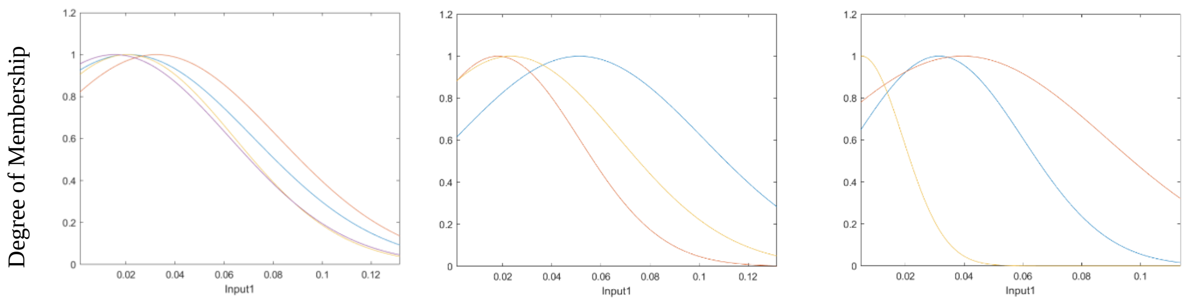


*Figure 12 - The membership functions retrieved from a section of the FRBS (left: Used to classify 1 v 2, centre: Used to classify 2 v 3, right: Used to classify 3 v 1)*

Figure 12 shows an example of the tuned membership functions used to understand, interpret, and classify the SSVEP data.

*Table 1 - Overview of the Ensemble - ANFIS algorithm*

**Bagging Ensemble – ANFIS Multiclass Classification**

---

**Input**:
Training Data, Testing Data (70 – 30 splits, randomised)

**Output**
One Multiclassification Output
Evaluation Metrics
Confusion Matrix

**Procedure**

1: **Load Data**
2: **Split** data into three different datasets of **data** and **labels** (10hz and 14hz [1v2], 14hz with 21hz [2v3], 21hz with 10hz [3v1])
3: Split those three different datasets into **training** and **testing**
4: **Generate FIS rules**: Options include **Subtractive Clustering**, **Sugeno – type** and default **Options** (ClusterInfluenceRange: Radius=1)
5: Generate **three binary FIS's** (1 for each dataset and labels pairing)
6: **Tune** the three binary FIS using the **ANFIS** options
7: **Train** each **FIS** using its corresponding **training data** (50 **epochs**, 0.01 **Step size**, **Step Size Decrease Rate** 0.9, **Step Size Increase Rate** 1.1, **Hybrid Optimisation** method)
8: **Evaluate** each binary classifier using the testing data to **output** three different binary classification arrays for 1v2, 2v3 and 3v1
9: Round the **ANFIS** output to limit its output between its binary labels
10: **Aggregate** all three binary classifier **outputs** using the **Average Voting Strategy** to output a single multi-classifier
11: Find the **RMSE** and **performance metrics** of the outputted classifier

Table 1 demonstrates the algorithm used to create the ensemble classifier, highlighting the splitting of the dataset using the OAO approach, the combining of the classifiers, the output of the single multiclass and the evaluation of it all.

# V. Results

In this section, the three binary classifiers trained on ANFIS, SVM and K – NN are shown in Table 2, aggregated using the average voting strategy, with a regular ANFIS multi – class classifier also recorded. The SVM, K – NN and ANFIS are used as a baseline algorithm to compare their performance metrics, based on equations 1 – 5, against the Ensemble – ANFIS to measure its predictive probability.

*Table 2 - Evaluation Metrics of All Classifiers Used for All Proposed Models (In Descending Order of Accuracy) | Source: MATLAB & Python*

| **Classifier** | **RMSE** | **Accuracy (%)** | **Average Precision (%)** | **Average Recall (%)** | **F1 (%)** |
|---|---|---|---|---|---|
| Ensemble - ANFIS | 0.68 | 64 | 65 | 63 | 64 |
| En - SVM | 1 | 53 | 56 | 51 | 53 |
| En - K – NN | 0.94 | 52 | 53 | 51 | 52 |
| ANFIS | 0.85 | 44 | 53 | 45 | 49 |

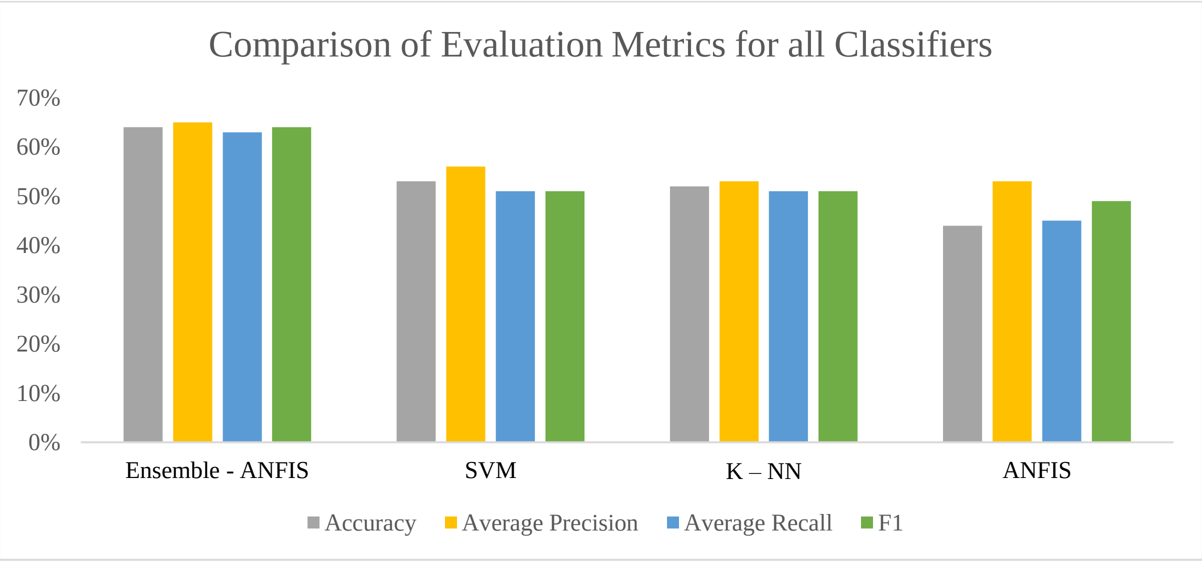


*Figure 13 – Comparison of all Evaluation Metrics for each classifiers used (In Descending order of Accuracy) | Source: Excel*

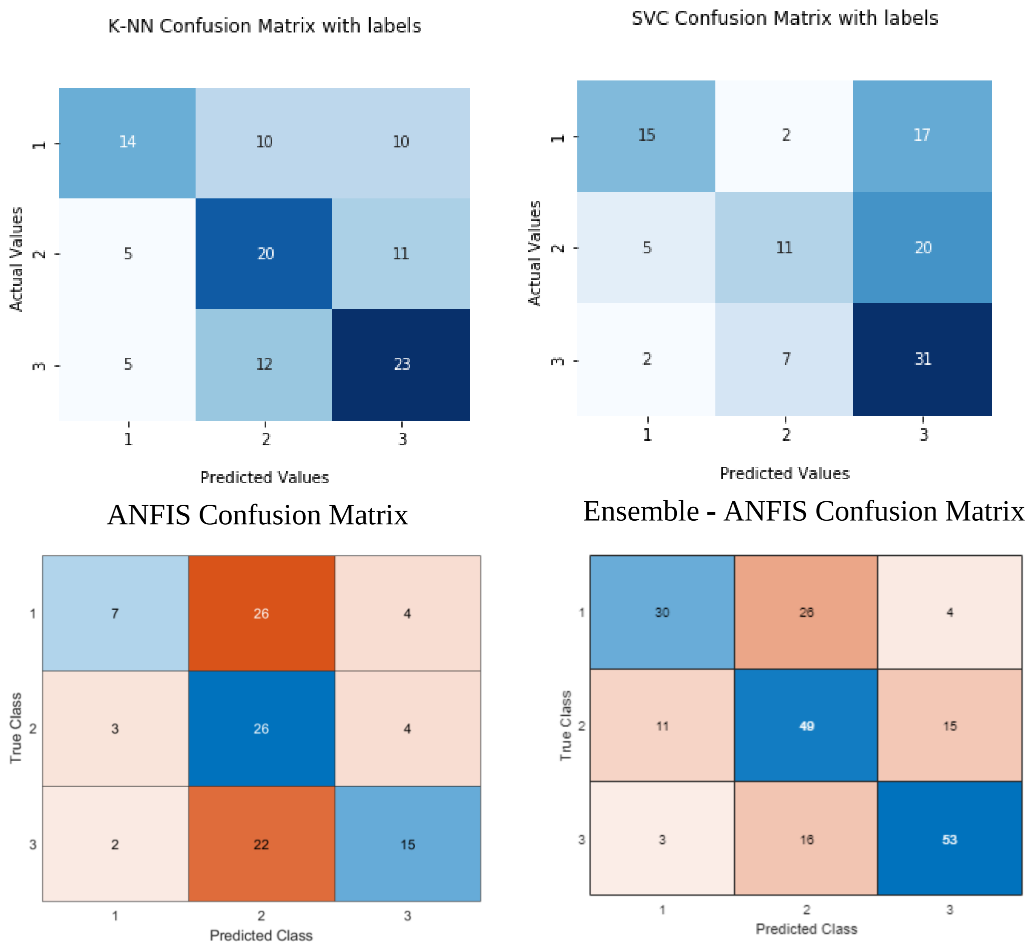


*Figure 14 - Confusion Matrix of all Classifiers | Source: Python (Top Row), MATLAB (Bottom Row)*

# VI. Discussion

Looking at figure 13, it is clearly shown that a normal multiclass ANFIS classifier performs at a much lower success rate, in most performance metrics, compared to all other proposed models at its ability to classify multiple labels. Comparing it to the ensemble – ANFIS, this reinforces the statement that the ensemble method typically leads to a greater, or equal, predictive probability than any other classifier in its ensemble [11]. This observation suggests that even with the use of a neural networks and fuzzy logic, the model found it more difficult to distinguish between three classes then two. With a greater precision, recall and subsequent F1 score, all other models performed better than the stand – alone ANFIS model at classifying positive results and distinguishing them from false negatives and positives.

To highlight the power of ANFIS and its capabilities in predicting labels, the SVM and K – NN models (table 2) were developed and used as a baseline model to evaluate ANFIS, applying the same method of OAO and average voting to create a single multi-class classifier. As seen in table 2, the Ensemble – ANFIS achieved the best results compared to the SVM and K – NN algorithms. ANFIS achieved an accuracy of 64% while SVM and K – NN only achieved 53% and 52% respectively. Indicating that fuzzy logic improves upon traditional machine learning algorithms in this case and the use of neural networks combined with fuzzy logic created more generalised model that could potentially be used for a greater variety of tasks.

Whilst also having the highest accuracy, the ensemble model also achieved the highest F1 score (the combination of recall and position), meaning that it performed the best in achieving both true results and true positive results than any of the baseline models (figure 13). This value is especially important as this determines how accurate the model is in terms of false negative and positives, and in the world of BCI's, these false negatives/positives are damaging to the BCI due to the nature of the tasks they're used for. As stated previously, sectors such as healthcare (where BCI's are commonly used), a wrong false negative/positive could have detrimental effects as it may result in not detecting a problem that may be present resulting in a lack of investigation and solution or detecting a problem that is there leading to unnecessary investigate. Both of which could very well worsen the situation that this failure has caused, possibly developing it into a significant issue.

One potential reason as to why the accuracy of this ensemble ANFIS differs to the one created by Zareian [4], is that Zareian implemented a DDAG Distance algorithm to find the best classifier over a larger set of binary classifiers. Using this algorithm may have given an edge to Zareian's Ensemble – ANFIS model and resulted in their model achieving a 67% accuracy. The lack of DDAG algorithm in this project is discussed in the conclusion.

The best results achieved by both this project and Zareian's [4] were 64% and 67% respectively. Overall, these are poor results, not nearly as high as other BCI models with literature achieving scores as high as 95.71% [3], 90% [9] and 88.7% [46]. This reports, and Zareian's, comparatively poor Subject 1 Multi – Class classification results may indicate that the data used in this experiment may have contained non – task related information that was insufficient in its ability to be used in creating a successful brain computer interface. However, to further validate this claim, more research would be required utilising this data set to give an accurate interpretation of the data and its ability to be used in BCI's.

One last difference between this report and Zareian's, which could be used to explain the discrepancy between our results, would be the implementation of the average voting strategy. Due to difficulties in creating this aggregation method, it may have not been designed and used the way it should have been and consequently, may have reduced the model's accuracy and other performance metrics.

### i. Improvements

Suggested improvements include developing and applying a bespoke DDAG algorithm and implementing it into the ANFIS code to define the best binary classifier for each sub data set and use that as the predictor, resulting in the best aggregated multi class classifier, bettering the one created in this project.

Another suggested improvement would be the standardisation of the average voting strategy, perhaps in the form of a toolbox or package, in hopes that the algorithm used would output using the same method and provide reliable, reproducible results for any ensemble model.

One last suggested improvement would be the use of hyper parameter tuning for the ANFIS model. The application of the Genetic Algorithm [47] to find the optimal parameters for the membership functions generated by ANFIS would provide the best predictive probabilities possible. The algorithm is an “adaptive heuristic search algorithm based off natural selection and the genetic process” [47] Another optimisation algorithm that could be used is particle swarm optimisation [48], an algorithm used to improve upon the performance of ANFIS in a variety of classification tasks by finding the optimal number of rules (figure 5) as well as the optimal quality of interpretation [48].

# VII. Conclusions

The underlying aim of this research project was to develop, train and successfully evaluate a multi classification model and improve its prediction capabilities, compared to a baseline model, using different techniques such as the ensemble method and a fuzzy inference system. Considering all the data presented and results shown in table 2, all aims and objectives that were set out in the ROA were achieved, resulting in the successful development of a Multi – Class Classifier for Brain Computer Interfaces. I was also able to successfully develop baseline models that were used to compare and evaluate the Ensemble - ANFIS model (figure 13). The Ensemble – ANFIS model, created for this project, has successfully been able to read, classify, and predict new EEG - SSVEP data to an accuracy of 64% (Table 2). It also significantly improved upon a normal Multi-class ANFIS classifier (figure 13) in all performance metrics, particularly a 20% increase in accuracy. As well as having a better predictive probability than all baseline models (table 2, figure 13), reinforcing the superiority of neural networks and fuzzy logic to traditional machine learning.

As it did not achieve or surpass the success rate of 67% (the accuracy of the ensemble - ANFIS developed by the study this research is based upon [4]), problems with the ANFIS model were addressed, evaluated, and can be improved upon in the hopes that its future development could potentially do so.

DDAG was difficult to implement in theory and even harder to implement in practise as it required the creation of multiple sets of binary classifiers for each sub – dataset, resulting in a cluster of matrixes that would be difficult to sort through algorithmically. Executing the voting average strategy was also challenging as it required the method of finding the same data sample in each binary dataset, averaging their label values, and outputting the averaged prediction. There was not a great deal of documentation on the idea and its use in MATLAB, as well as MATLAB itself not having a toolbox for decomposition and aggregation methods such as these, so all ensemble processes, including aggregation, were developed from scratch.

## i. Future Work

Further development of this research and future work would involve developing and producing a model with greater performance metrics, based on the addition of all improvements included in the discussion, to be used as a Brain – Computer – Interface to assist in detecting and diagnosing neurological (e.g., epilepsy), and neurodegenerative diseases (e.g., Alzheimer's), based off EEG – SSVEP data.

# VIII. References


[1] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw, "Brain-computer interfaces in medicine," in *Mayo Clinic Proceedings*, 2012, vol. 87, no. 3, pp. 268–279.

[2] R. H. Abiyev, N. Akkaya, E. Aytac, I. Günsel, and A. Çağman, "Brain-computer interface for control of wheelchair using fuzzy neural networks," *Biomed Res Int*, vol. 2016, 2016.

[3] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, and S. Giove, "User adaptive BCIs: SSVEP and P300 based interfaces," *PsychNology Journal*, vol. 1, no. 4, pp. 331–354, 2003.

[4] E. Zareian, "Meta Heuristics based Machine Learning and Neural Mass Modelling Allied to Brain Machine Interface," no. April, 2021.

[5] R. Kuś *et al.*, "On the Quantification of SSVEP Frequency Responses in Human EEG in Realistic BCI Conditions," *PLoS ONE*, vol. 8, no. 10, p. e77536, Oct. 2013, doi: 10.1371/journal.pone.0077536.

[6] A. M. Andrew, "THE HANDBOOK OF BRAIN THEORY AND NEURAL NETWORKS, edited by Michael A. Arbib, MIT Press, Cambridge, Mass. and London, England, 1998, xv+1118 pp, ISBN 0-262-51102-9, paperback, £49.95 (cloth-bound version also available, originally published in 1995, ISBN," *Robotica*, vol. 17, no. 5, pp. 571–572, Sep. 1999, doi: 10.1017/S0263574799211770.

[7] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in {P}ython," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[8] T. Mahboob Alam *et al.*, "A model for early prediction of diabetes," *Informatics in Medicine Unlocked*, vol. 16, no. July, p. 100204, 2019, doi: 10.1016/j.imu.2019.100204.

[9] A. Turnip, A. I. Simbolon, M. F. Amri, and M. A. Suhendra, "Utilization of EEG-SSVEP method and ANFIS classifier for controlling electronic wheelchair," *Proceedings of the 2015 International Conference on Technology, Informatics, Management, Engineering and Environment, TIME-E 2015*, no. September, pp. 143–146, 2016, doi: 10.1109/TIME-E.2015.7389763.

[10] A. Al-Ani and S. Darvishi, "Brain-Computer Interface Analysis Using Continuous Wavelet Transform and Adaptive Neuro-Fuzzy Classifier," 2007.

[11] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1249, 2018.

[12] D. Achanccaray, C. Flores, C. Fonseca, and J. Andreu-Perez, "A P300-based brain computer interface for smart home interaction through an ANFIS ensemble," *IEEE International Conference on Fuzzy Systems*, 2017, doi: 10.1109/FUZZ-IEEE.2017.8015770.

[13] A. Gupta and D. Kumar, "Fuzzy clustering-based feature extraction method for mental task classification," *Brain Inform*, vol. 4, no. 2, pp. 135–145, Jun. 2017, doi: 10.1007/s40708-016-0056-0.

[14] K. M. Bataineh, M. Naji, and M. Saqer, "A comparison study between various fuzzy clustering algorithms," *Jordan Journal of Mechanical and Industrial Engineering*, vol. 5, no. 4, pp. 335–343, 2011.

[15] C. Iwendi, K. Mahboob, Z. Khalid, A. R. Javed, M. Rizwan, and U. Ghosh, "Classification of COVID-19 individuals using adaptive neuro-fuzzy inference system," *Multimedia Systems*, Mar. 2021, doi: 10.1007/s00530-021-00774-w.

[16] M. Rashid, N. Sulaiman, M. Mustafa, S. Khatun, and B. S. Bari, "The Classification of EEG Signal Using Different Machine Learning Techniques for BCI Application," 2019, pp. 207–221. doi: 10.1007/978-981-13-7780-8_17.

[17] "Artificial intelligence Definition & Meaning - Merriam-Webster." https://www.merriam-webster.com/dictionary/artificial intelligence (accessed Apr. 14, 2022).

[18] H. KAWANO, "Knowledge Discovery and Data Mining," *Journal of Japan Society for Fuzzy Theory and Systems*, vol. 9, no. 6, pp. 851–860, 1997, doi: 10.3156/jfuzzy.9.6_851.

[19] "Regression vs Classification in Machine Learning - Javatpoint." https://www.javatpoint.com/regression-vs-classification-in-machine-learning (accessed Apr. 14, 2022).

[20] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, 2007, doi: 10.1088/1741-2560/4/2/R01.

[21] "Support Vector Machine(SVM): A Complete guide for beginners." https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/ (accessed Apr. 14, 2022).

[22] "KNN (K-Nearest Neighbors) #1. How it works? | by Italo José | Towards Data Science." https://towardsdatascience.com/knn-k-nearest-neighbors-1-a4707b24bd1d (accessed Apr. 14, 2022).

[23] T. Kayikcioglu and O. Aydemir, "A polynomial fitting and k-NN based approach for improving classification of motor imagery BCI data," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1207–1215, Aug. 2010, doi: 10.1016/j.patrec.2010.04.009.

[24] K. Gayathri and A. Marimuthu, "Text document pre-processing with the KNN for classification using the SVM," in *2013 7th International Conference on Intelligent Systems and Control (ISCO)*, Jan. 2013, pp. 453–457. doi: 10.1109/ISCO.2013.6481197.

[25] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993, doi: 10.1109/21.256541.

[26] J. Chen and M. Mahfouf, "Improving transparency in approximate fuzzy modeling using multi-objective immune-inspired optimisation," *International Journal of Computational Intelligence Systems*, vol. 5, no. 2, pp. 322–342, 2012.

[27] L. A. Zadeh, "Fuzzy logic," *Computer (Long Beach Calif)*, vol. 21, no. 4, pp. 83–93, 1988.

[28] C. Jun, "Biologically inspired optimisation algorithms for transparent knowledge extraction allied to engineering materials processing." University of Sheffield, 2010.

[29] E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," in *Proceedings of the institution of electrical engineers*, 1974, vol. 121, no. 12, pp. 1585–1588.

[30] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans Syst Man Cybern*, no. 1, pp. 116–132, 1985.

[31] W. E. Sari, O. Wahyunggoro, and S. Fauziati, "A comparative study on fuzzy Mamdani-Sugeno-Tsukamoto for the childhood tuberculosis diagnosis," *AIP Conference Proceedings*, vol. 1755, no. July 2016, 2016, doi: 10.1063/1.4958498.

[32] L. Fabien, L. Anatole, L. Fabrice, and A. Bruno, "Studying the use of fuzzy inference systems for motor imagery classification," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 322–324, 2007.

[33] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 4, no. 4. Springer, 2006.

[34] "What are Neural Networks? | IBM." https://www.ibm.com/cloud/learn/neural-networks (accessed Apr. 16, 2022).

[35] B. Turovsky, "Found in translation: More accurate, fluent sentences in Google Translate," *Google*. https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/ (accessed Apr. 16, 2022).

[36] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225–1231, Nov. 1996, doi: 10.1016/S0895-4356(96)00002-9.

[37] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, "Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]," *IEEE Transactions on automatic control*, vol. 42, no. 10, pp. 1482–1484, 1997.

[38] J.-S. R. Jang, "Self-learning fuzzy controllers based on temporal backpropagation," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 714–723, 1992, doi: 10.1109/72.159060.

[39] Z. H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artificial Intelligence*, vol. 137, no. 1–2, pp. 239–263, May 2002, doi: 10.1016/S0004-3702(02)00190-X.

[40] R. Eichelberger and V. Sheng, "Does one-against-all or one-against-one improve the performance of multiclass classifications?," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2013, vol. 27, no. 1.

[41] K. An and J. Meng, "Voting-averaged combination method for regressor ensemble," in *International Conference on Intelligent Computing*, 2010, pp. 540–546.

[42] "Choosing Performance Metrics. Accuracy, recall, precision, F1 score—... | by S. T. Lanier | Towards Data Science." https://towardsdatascience.com/choosing-performance-metrics-61b40819eae1 (accessed Apr. 18, 2022).

[43] F. Krüger, "Activity, Context, and Plan Recognition with Computational Causal Behaviour Models," *ResearchGate*, no. August, 2018.

[44] W. Wang and Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," *IOP Conference Series: Materials Science and Engineering*, vol. 324, no. 1, 2018, doi: 10.1088/1757-899X/324/1/012049.

[45] Y. Sasaki, "The truth of the F-measure," *Teach Tutor mater*, pp. 1–5, 2007.

[46] B. D. Mensh, J. Werfel, and H. S. Seung, "BCI competition 2003 - Data set Ia: Combining gamma-band power with slow cortical potentials to improve single-trial classification of electrolencephallographic signals," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1052–1056, 2004, doi: 10.1109/TBME.2004.827081.

[47] J. M. Mendel, "Uncertain rule-based fuzzy systems," *Introduction and new directions*, p. 684, 2017.

[48] D. P. Rini, S. M. Shamsuddin, and S. S. Yuhaniz, "Particle swarm optimization for ANFIS interpretability and accuracy," *Soft Computing*, vol. 20, no. 1, pp. 251–262, Jan. 2016, doi: 10.1007/s00500-014-1498-z.

[49] IBM Cloud Education. What is Bagging? [Internet]. Ibm.com. 2021 [cited 2022 Oct 28]. Available from: https://www.ibm.com/cloud/learn/bagging

# IX. Appendix

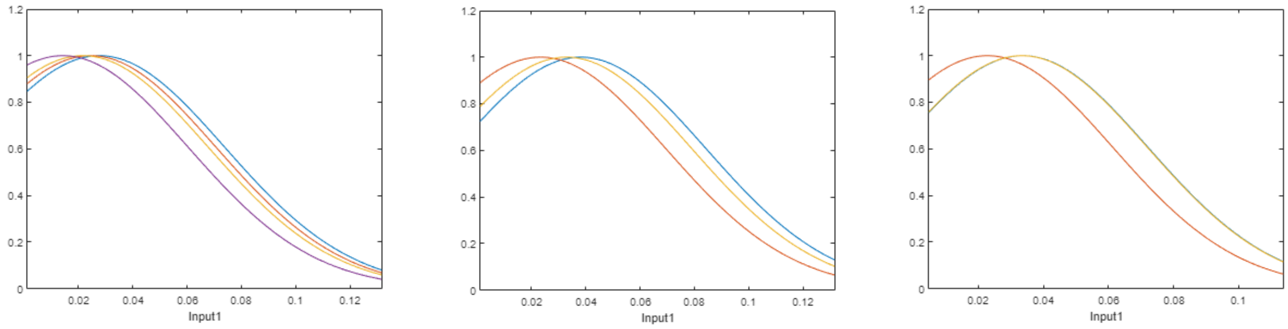


*Figure 15- Untuned membership functions retrieved from a section of figure 12's FRBS (left: Used to classify 1 v 2, centre: Used to classify 2 v 3, right: Used to classify 3 v 1)*

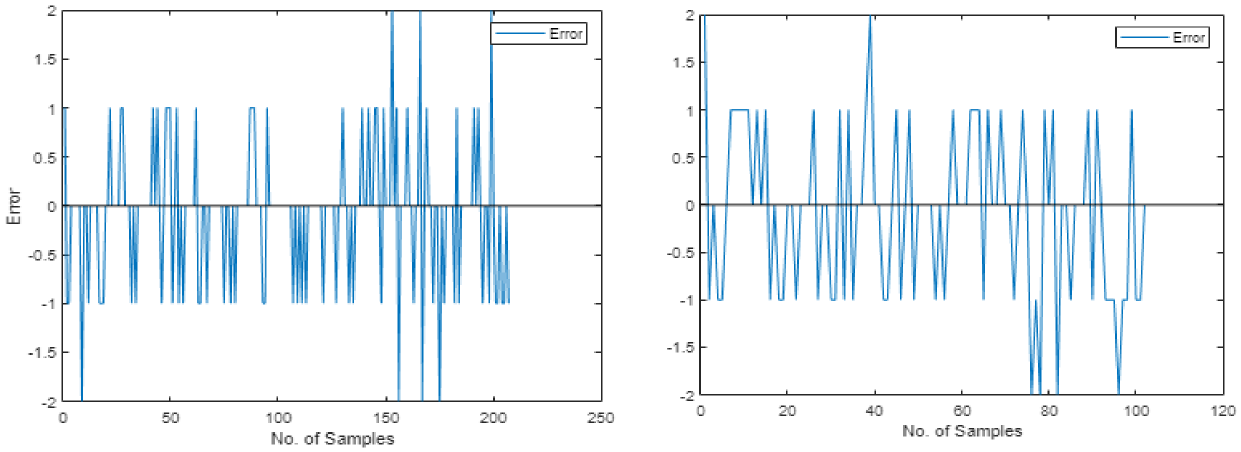


*Figure 16 - Error of testing data for (Left) Ensemble – ANFIS (Right) Multi-Class ANFIS*