Tampere University

Markus Haukipää

SPECTRAL POWER RATIOS IN EPILEPTIC SEIZURE PREDICTION

Bachelor's Thesis
Faculty of Medicine and Health Technology
Narayan Puthanmadam Subramaniyam
May 2024

ABSTRACT

Markus Haukipää: Spectral power ratios in epileptic seizure prediction
Bachelor's Thesis
Tampere University
Biotechnology and Biomedical Engineering
May 2024

1% of the world's population is affected by unpredictable epileptic seizures. These affect the quality of life and are related to increased mortality rates. Prediction of these seizures is vastly researched aiming to improve the quality of life of epileptic patients. Most of the prediction models utilize artificial intelligence. To make these models accurate and precise, there is a need for an efficient feature. The efficient feature results in a prediction model that can accurately predict the start of a seizure. Additionally, the feature should be computationally cheap to process.

The epilepsy seizure can be seen as an abnormal amount of neural activation. This change in neural activation can be used to retrieve a feature. This thesis evaluates the performance of spectral power ratio as a feature. The researched spectral power ratio is formed between different frequency sub-bands. Through electroencephalography data of multiple patients, this thesis aims to represent an efficient feature for these future prediction models. The feature's performance is evaluated through the classification accuracy of different machine learning models. The spectral power ratio's performance in classification is a strong indicator of its possible usefulness in prediction models.

It was concluded that the usage of spectral power ratios implicates the possibility of an effective, useful, and computationally cheap feature. The best-performing models were achieved using theta band ratio to a higher frequency band as a feature. This feature resulted in some of the patients' models achieving over 80% classification accuracy. Concluding a definitive result is affected by artifacts and irregularities across the data. In addition, the noticeable differences between patients in the research affect the result.

The potential use of spectral power ratio as a feature needs to be further researched to be used in the epilepsy prediction models. The research shows the feature's effectiveness may be suited to be part of a larger number of used features in the prediction model. However, it did not implicate a definitive result to be used as the sole feature.

Keywords: epilepsy seizure prediction, power spectral density, machine learning, classification

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Markus Haukipää: Tehospektrien suhteet epilepsiakohtauksen ennustamisessa
Kandidaatintutkielma
Tampereen yliopisto
Bioteknologia ja biolääketieteen tekniikka
Toukokuu 2024

Noin 1 % maailman väestöstä kärsii ennustamattomista epileptisistä kohtauksista. Nämä vaikuttavat ihmisen elämänlaatuun ja ovat yhteydessä lisääntyneeseen kuolleisuuteen. Näiden kohtausten ennustamista tutkitaan paljon epilepsiasta kärsivien elämänlaadun parantamiseksi. Monet näistä ennustusmalleista hyödyntää koneoppimista. Jotta nämä mallit olisivat tarkkoja ja luotettavia, tarvitaan tehokas piirre. Piirre on attribuutti, jota hyödynnetään datan klassifikaatiossa. Tehokas piirre luo ennustusmallin, joka voi ennustaa tarkasti kohtauksen alkamisen. Lisäksi piirteen tulee olla laskennallisesti halpa käsitellä.

Tässä opinnäytetyössä arvioidaan spektritehosuhteen suorituskykyä piirteenä. Tutkittu spektritehosuhde muodostetaan eri taajuusosakaistojen välille. Tämä opinnäytetyö hyödyntää useiden potilaiden elektroenkefalografiadataa piirteen muodostamiseen. Tavoitteena on esittää tehokas piirre, jota tulevaisuuden ennustusmallit voivat hyödyntää. Piirteen suorituskykyä arvioidaan erilaisten koneoppimismallien klassifikaatiotarkkuuden avulla. Piirteen klassifikaatiotarkkuus on vahva indikaattori sen mahdollisesta käyttökelpoisuudesta tulevaisuuden ennustusmalleissa.

Tutkimus osoitti, että spektritehosuhteet ovat mahdollisesti hyödynnettäviä piirteitä ennustusmalleissa. Spektritehosuhde olisi mahdollisesti tehokas, hyödyllinen ja laskennallisesti halpa piirre. Parhaiten suoriutuvat mallit saavutettiin käyttämällä piirteenä theetakaistan suhdetta korkeampiin taajuuskaistoihin. Tätä piirrettä hyödyntämällä osassa potilaiden malleista saavuttiin yli 80 %:n klassifikaatiotarkkuus. Definitiivisen tuloksen saavuttamiseen vaikutti artefaktit ja datan epäsäännöllisyydet. Lisäksi tutkimuksessa havaitaan potilaiden välisten erojen vaikuttavan tulokseen.

Spektritehosuhteen mahdollista käyttöä piirteenä on edelleen tutkittava, jotta sitä voidaan hyödyntää epilepsiakohtauksien ennustusmalleissa. Tutkimus osoitti, että piirteen tehokkuus voi osoittautua yhdeksi hyödylliseksi osaksi suurempaa määrää ennustemallien käytettyjä piirteitä. Tutkimus ei kuitenkaan suoraan tukenut spektritehosuhteen käyttöä ainoana piirteenä ennustusmalleissa.

Avainsanat: epilepsiakohtauksien ennustus, tehospektritiheys, koneoppiminen, klassifikaatio

CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

BUT	Butterworth filter
EEG	Electroencephalography
FIR	Finite impulse response
ICA	Individual class accuracy
iEEG	Intracranial electroencephalography
KNN	K-nearest neighbour
ML	Machine learning
OA	Overall accuracy
PSD	Power spectral density
RF	Random forest
sEEG	Scalp electroencephalography
SVM	Support vector machine
SWEC-ETH	Sleep-Wake-Epilepsy-Center - ETH Zurich (dataset)
f	frequency
s	second
V	voltage
ω	radian frequency

1. INTRODUCTION

Epilepsy is one of the most common neurological diseases. It affects around 1% of the world's population. Epilepsy is a neurological disease with a wide array of etiology suffering from epileptic seizures. These epileptic seizures are the result of imbalanced electrical activity of the central nervous system. Epileptic seizure describes the resulting immoderate activation of neurons. [1], [2] The unpredictable and recurrent nature of epileptic seizures [3] may impose difficulties on the patient's life. The epileptic seizures have been proposed to be a major cause in the increased mortality rate of epileptic patients. Patient suffering from epileptic seizures have a two to three times more increased mortality rate compared to the general population. [4] Making epileptic seizures predictable could improve epileptic patients' quality of life.

Epilepsy seizures can be analyzed with electroencephalography (EEG). EEG measures voltage differences in the brain caused by neural activation. The data is simultaneously captured from multiple sensors leading the data to be multi-channel. [5, pp. 1–30] EEG signal can be measured non-invasively with a scalp EEG (sEEG) where the electrodes are placed on top of the head. The downside of sEEG is its rather high risk of artifacts and noise in the signal due to the long distance between the neurons and the sensors in addition to the interfering electrical activity of the medium. [6] Alternatively, intracranial electroencephalography (iEGG) is an invasive method to retrieve EEG data. The electrode placement differs from sEEG. In iEEG, through surgery, the electrodes are placed directly in contact with the brain. Therefore, iEEG data has fewer artifacts and needs less signal processing. In other words, iEEG has a high signal-to-noise ratio compared to sEEG resulting in a more detailed signal compared to sEEG. [7], [8]

The explosive growth of artificial intelligence has compelled research to increasingly implement machine learning (ML) technology. ML can be utilized to create seizure detection programs using the EEG epileptic seizure data points as the training data. Research has been conducted using different machine learning models, datasets, and signal processing techniques, including various combinations of them. A number of them have been proven to formulate working seizure detection programs. [6]

The purpose for this thesis is to formulate a working model of epileptic seizure prediction. Furthermore, it gives more insight to making epileptic seizure prediction models step by step starting from the open access epileptic seizure data. The aim is to utilize different EEG signal's frequency bands' power ratios. The model is based on the open-access EEG dataset from the Sleep-Wake-Epilepsy-Center (SWEC) of the University Department of Neurology at the Inselspital Bern and the Integrated Systems Laboratory of the ETH Zurich (SWEC-ETH dataset) [9].

The formulation of the thesis will be as follows. There will be a theoretical literature review and an experimental part. Furthermore, the literature review will be divided into biological and computational theory parts. The biological part including the theory and information of EEG data. Computational theory includes data transformation, filtering, and machine learning methods. The experimental part includes the methodology and the results of the machine learning models using different features. After there will be a performance rating and conclusion of the methodology, models' performance, and the results.

2. THEORETICAL BACKROUND

2.1 Electroencephalography

Electroencephalography is a direct measure of neural activity. EEG utilizes the voltage difference between two electrodes to measure neuroelectric potentials caused by a large population of neurons. The use of electrical potentials as the means of measure results in an excellent temporal resolution. The measurement is therefore considered to be recorded in real-time. Using multiple electrodes, EEG outputs multiple channels of real-time voltage data of neuroelectric activity. Depending on the procedure, the total number of electrodes used may differ. [5], [10]

EEG data can be divided by the channels and furthermore by the channels' frequency bands. The channel data can be categorized into 5 different sub-bands according to the frequency. These subsets of frequency bands are delta (δ, < 4 Hz), theta (θ, 4 - 8 Hz), alpha (α, 8 - 13 Hz), beta (β, 14 - 30 Hz), and gamma (γ, > 30 Hz) (Figure 1). [5], [11], [12] For the wide bandwidth of the gamma band, it can be further divided into low gamma (30 - 80 Hz) and high gamma frequencies (> 80 Hz). [13] The amplitude of a non-ictal EEG signal ranges approximately from 10 μV to 100 μV, where V is the voltage [6].

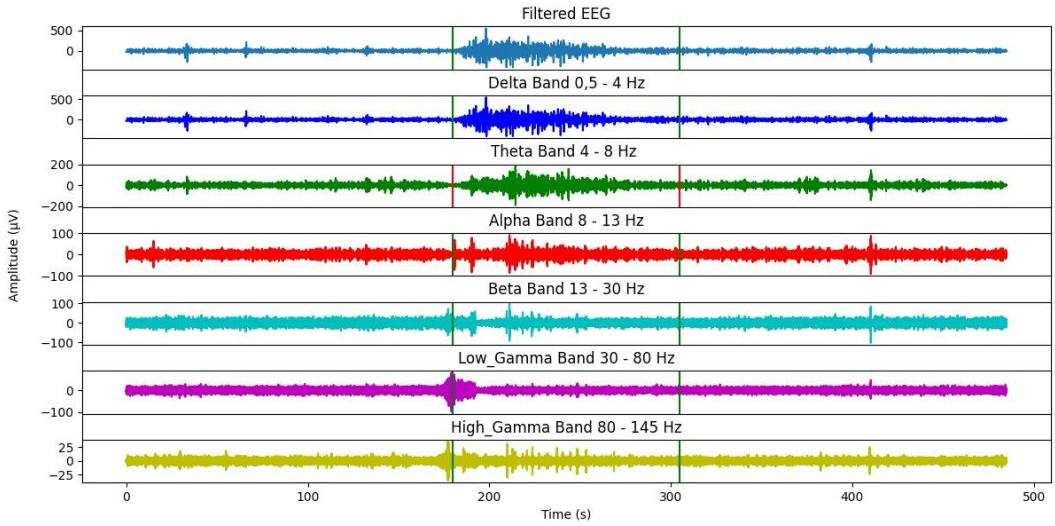

Figure 1. *Data from different sub-bands around a seizure. Vertical lines represent the start and the end of the seizure. The data used in the figure is from SWEC-ETH [9].*

Dividing the data into subsets contributes to a more precise analysis of the brain's neural activity. Important indicators of seizure could be neglected in the full-spectrum analysis of the signal. Smaller amplitude sub-bands' substantial percentual increase may be concealed by the overall increase of the absolute amplitude value. [14] It needs to be noted that the frequencies bandwidths may differ between sources.

2.1.1 Noise and artifacts

Intracranial EEG data has a high signal-to-noise ratio [7], [8]. In contrast, sEEG signal has a lower signal-to-noise ratio. This is due to the high amplification of the low voltage signal from the source, partly due to the distance between electrodes and the neurons, resulting in amplification of the wanted neuron signal **in addition** to the unwanted electrical activity or movement of the medium [5, p. 20], [11].

In the iEEG the effect of the medium is negligible compared to sEEG. In addition, the close contact of the electrodes and the source allows a lower amplification. Nevertheless, iEEG contains artifacts from skeletal muscle activation and myocardium activation. [5, p. 20], [8] In addition, the environment may cause artifacts. These include near electrical devices. From these, the most significant noticeable interference in iEEG is caused by the line current (50 Hz). [5, pp. 100–120]

2.2 Dataset

The open-access dataset is from the SWEC-ETC. The data used is preselected and annotated by professionals in seizure segments having ictal (seizure), 3 minutes of pre-ictal (before seizure), and 3 minutes of post-ictal (after seizure) data with minimal artifacts. The ictal phase length ranges from 10 seconds (s) to 1002 s. The electrode count ranges from 42 to 100 electrodes averaging at 64 electrodes. [9], [15]

The data is iEEG data gathered from 16 separate patients. The age of the patients ranges from 20 to 59 years old averaging at 33. [15] The data is converted to digital with a 16-bit converter with a rate of 512 Hz. The data is pre-processed with a fourth-order Butterworth filter (BUT) to a bandwidth of 0.5 -150 Hz and with a forward and backward filter. [9] There is a total of 99 seizure files for one of the 100 files is an artifact. [9]

2.3 Power spectral density

Power spectral density (PSD) is a presentation of the signal's power in relation to its frequency. PSD value is real-valued and nonnegative. The PSD can be presented in several ways. For a voltage signal, PSD can be formulated in the following way:

$$PSD = \frac{V^2}{f}, \tag{1}$$

where f is the frequency and *V* is the voltage of the signal. PSD gives an insight into the power distribution over the frequencies in a finite sample size. This can be utilized in discrete signal analysis to calculate the PSD value of individual samples. [16, p. 352], [17] The PSD value is directly proportional to the voltage amplitude of power of 2 and inversely proportional to the frequency.

EEG frequency sub-bands have differing PSD value compared to each other. These different PSD ratios have been used in formulating working seizure detection models. The power spectrum of the sub-band is calculated by

$$p_i = \sum PSD_i(x_i) \ , \tag{2}$$

where p_i is the spectral power of the sub-band k, x_i is the windowed EEG signal of the sub-band and PSD_i is the power spectral density of the EEG signal in the frequency band. This gives the absolute value of the spectral power. [17] For continuous signal x(n), the PSD of a signal can be defined discrete-time Fourier transform of the covariance sequence as follows:

$$PSD(\omega) = \sum_{k=-\infty}^{\infty} r(k)\, e^{-i\omega k}, \tag{3}$$

where PSD(ω) is the power spectral density of radian frequency ω, r(k) is the auto-covariance sequence of the signal x(k). [18]

To minimize the calculation time, the Welch method is often used to estimate the PSD value. In addition, the Welch method results in a smoother representation of the power spectrum. The Welch's method divides the data into smaller segments and the PSD is calculated for each windowed segment. Welch method utilizes the overlapping of the segments. The segments are divided mathematically with

$$\mathrm{X_K(j)} \ = \ \mathrm{X(j + (K - 1) * D)}\,, \tag{4}$$

where $X_K(j)$ is the Kth segment, and D represents the length of the segment. These segments are subsequently averaged resulting in the PSD estimation. Mathematically Welch PSD estimation is as follows:

$$\widetilde{PSD}(\omega) = \frac{1}{K}\sum_{k=1}^{K} \widetilde{PSD}_k(\omega)\,, \tag{5}$$

where $\widetilde{\mathrm{PSD}}(\omega)$ is the PSD estimate of radian frequency ω, S is the number of segments and $\widetilde{\mathrm{PSD}}_k(\omega)$ is the segments PSD at estimate. [19]

Different frequency bands' amplitude fluctuates individually between the seizure phases. These differences in expression in pre-ictal, ictal, and post-ictal phases can used to recognize the phase. The frequency bands' individual expressions during the change of a phase can be seen in figure 2.

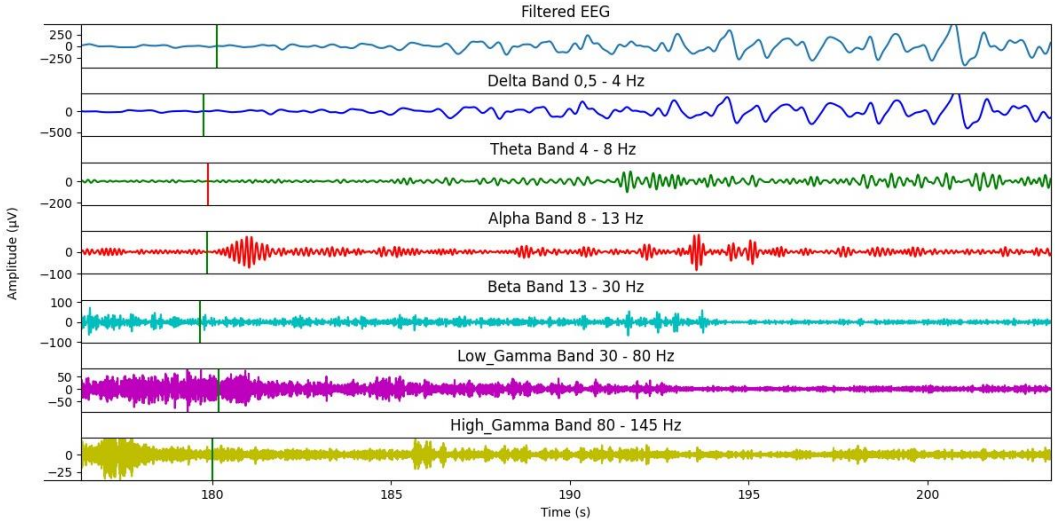

Figure 2. *EEG data and its different frequency sub-bands at the start of ictal phase (vertical line at 180 s).*

The EEG data can be observed to mostly be constructed by the lower frequencies and more resamples delta band. During the start of the ictal phase, commonly the higher frequencies amplitude, therefore the spectral power, increases. As seen, commonly the lower frequencies have smaller relative change during the shift from pre-ictal to ictal phase. [14] Therefore, the spectral power ratio change between the sub-bands could be implemented as an efficient indicator of the classification of the ictal phases.

2.4 Machine learning

ML can be divided into reinforced, unsupervised, and supervised learning. Reinforced learning utilizes feedback from the environment through rewards to formulate the best rewarding classification model. Unsupervised learning applies its own classification to the unlabelled learning data by finding different patterns. In contrast, supervised learning inputs use labelled learning data. New presented unlabelled data is classified according to the classifications made from the learning data. [6] Supervised machine learning algorithms used in EEG data classification include support vector machine (SVM), K-nearest neighbours (KNN) and random forest (RF).

SVM is a frequently used ML algorithm used for EEG data classification [6], [20]. SVM uses a vector line(s) to divide the data points between the classes defined (Figure 3). SVM can be used to classify the data in numerous classes. [21]

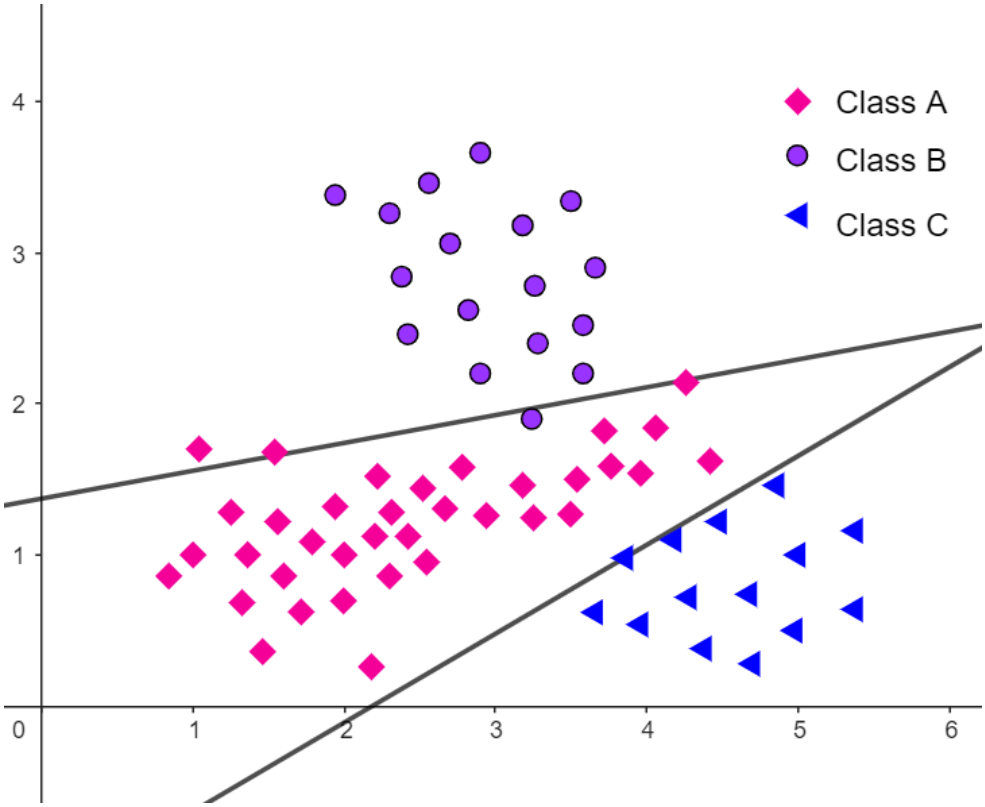

Figure 3. *A figurative visualization of linear SVM algorithm on learning data A, B and C.*

SVM has been used in various EEG data classification models with high-performance results (> 90% sensitivity) [6]. SVM can utilize different kernels to classify the data. Deciding the kernel function affects the dividing vectors shape and therefore the classification of the data. Linear and complete polynomial of degree of n are examples of commonly used kernels. [21]

KNN algorithm classifies the data by connecting the datapoints. KNN forms a set number of nearest neighbours to each datapoint. The distance to different classes' neighbours is used to determine the classification of a new datapoint. The classification line dividing the different classes is made accordingly (Figure 4) [22]. KNN, like SVM, is a used EEG analysis method with high-performance results (> 90% sensitivity) [6]. The classification is affected by the chosen number of neighbours per datapoint.

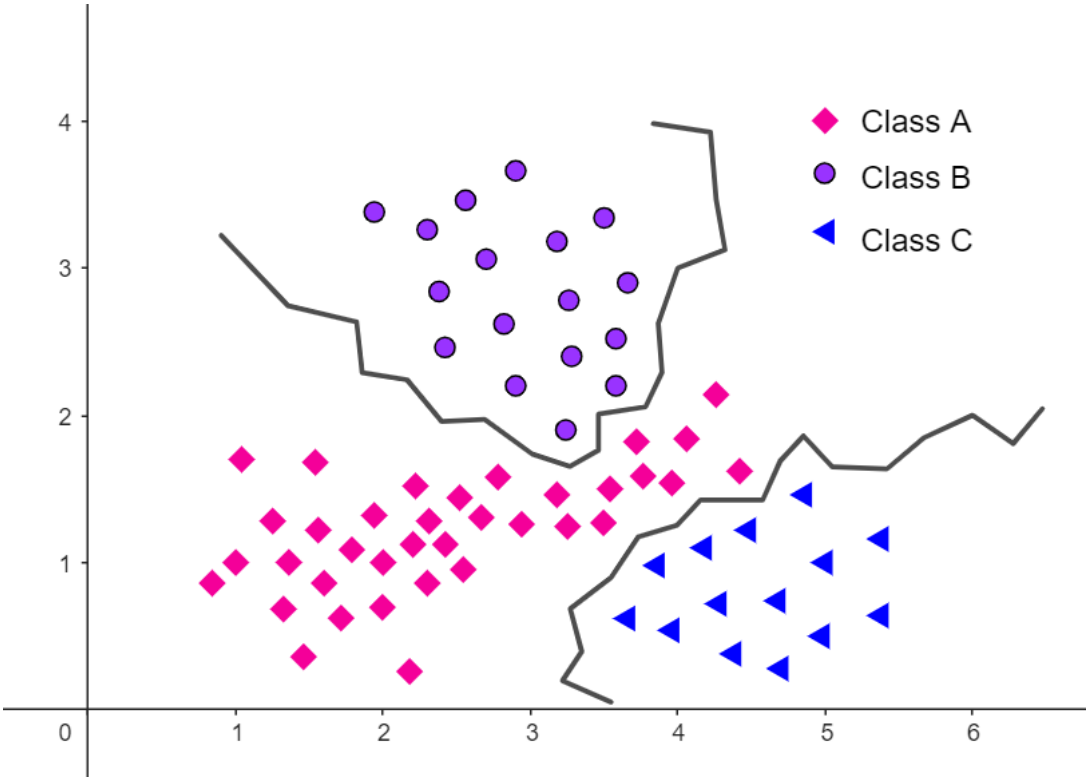

Figure 4. *A figurative visualization of the KNN algorithm on data A, B and C.*

RF model uses multiple randomly chosen decision trees to classify the data (Figure 5). The decision tree is a classifier on its own which uses the data's characteristics. The characteristics and parameters differ between different decision trees. The aggregate of the decision trees is used to compute the whole RF model classification areas. [23]

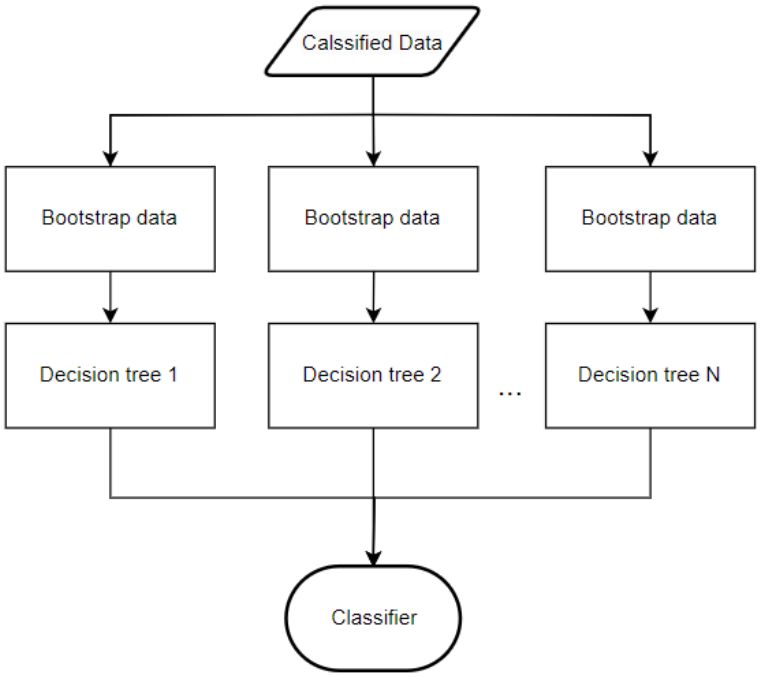

Figure 5. *A figurative visualization of RF classifier.*

Overall ML methods succeed in handling big data that would take humans an unnecessarily large amount of time. For biosignals, such as EEG data, patterns are important to be recognized from the continuous dataflow. ML is used for its capability to go through this large data, recognize patterns, and make real-time predictions of the data.

3. METHODOLOGY

The research pipeline is depicted in the following block chart (Figure 6). The overall process involves downloading the data from the database, preprocessing, feature extraction and classification and ultimately learning and testing the machine learning model. The result is the classification accuracy of the known testing data.

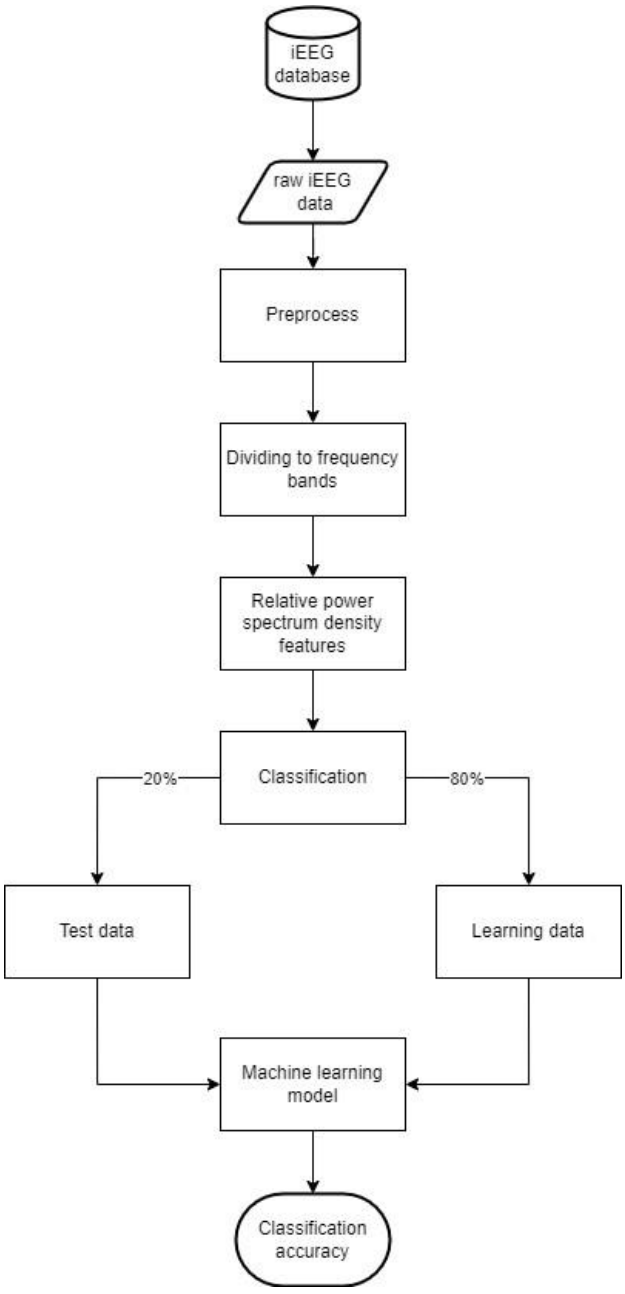

Figure 6. *Pipeline of the process.*

The research program is done using Python 3.10. MNE library is used to 1.6.1 preprocess the EEG signals [24]. Open source python library SciPy 1.12.0 the signal module is used in filtering and calculating signals power spectral density [25]. Open source Python library Scikit-learn 1.4.0 works as the machine learning library [26]. It is utilized in making and analyzing the machine learning model. The code used in the thesis is attached[27].

3.1 Preprocess

The iEEG data was handpicked by certified epileptologists to have low amounts of artifacts [9]. In addition, the iEEG has a natively high signal-to-noise ratio. The proving of the hypothesis can be determined with minimal preprocessing.

The preprocess involved filtering the line current's noise and its harmonics. Correlating to filtering the frequencies 50 Hz and 100 Hz. The EEG data is low pass filtered with 145 Hz cut frequency, for frequencies above 128 Hz have increasingly lower amplitude approaching 0 [17]. Therefore, there is no need to filter the harmonic noise at 150 Hz. The filter is a digital finite impulse response (FIR) notch filter from the MNE library [24]. The notch filter uses hamming window, a sampling frequency of 512 Hz, and a notch filtering width of 1/200 part of the filtered frequency.

3.2 Feature extraction

The data was divided by the channels and by time to classes: pre-ictal, ictal, and post-ictal. Finally, it is divided by the frequency bands. Subsequently, the PSD is made by SciPy's welch method. The segmenting is done at 1024 Hz at 512 Hz sampling frequency meaning samples every 0.25 Hz. The frequency bands' mean spectral power values are calculated. This is done for every channel, class, and frequency band.

The mean spectral power value ratio between the different frequency bands is used as the feature for the ML. To minimizing the computational time, the ratios are normalized using the biggest ratio of the learning set's specific frequency pair. A specific frequency band pair is used to formulate an ML classification model.

3.3 Machine learning

The ML is made with SVM, KNN, and with RF models. The models are done for every frequency band ratio patient specifically for patients who had at least 7 seizure files

recorded. These included a total of 7 patients with IDs 1, 4, 5, 9, 12, 13, and 14. In addition, models are done by utilizing all 16 of the patient files. The learning and test data is pre-classified and divided into approximately 80% of learning data and 20% of testing data.

SVM classifier is done with a linear kernel with a “one over rest” decision boundary. KNN is done with 4 neighbours and uniform weights. RF is done with 200 estimators. Other parameters are according to the set values by the documentation [26]. These parameters resulted in the highest classification rates with a reasonable computational time of the finite sets tested.

3.4 Classification

The classification is done using the ictal phases. From those, 2 classification models were made. A 3-class classification model and a binary classification model. 3 class classification model classifies the data into pre-ictal, ictal, and post-ictal. The binary model classifies the model in only 2 classes, pre-ictal and ictal. Classification accuracy is calculated by the correctly classified data in ratio to all data according to the following formula:

$$Accuracy(\%) = \frac{RC}{RC + WC} * 100\%, \tag{7}$$

where RC is the right classification, WC is the wrong classification and accuracy is the percentage of correctly classified data. The accuracy is used for overall classification accuracy (OA) and individual class accuracy (ICA).

The 3-class classification model is used to test the efficiency of the feature’s precise detection of the phase overall in the EEG data. The model theoretically could imply the start and end of the seizure. The binary classification model is done to imply a more definitive effectiveness of the features’ capabilities as a seizure start indicator. Using the pre-ictal and ictal data prevents the data from mixing between post- and pre-ictal having an effect on the overall accuracy. On the other hand, the efficiency of the binary classification does not directly correlate to the efficiency of the possible model made from all the data.

The EEG data classified as seizure and non-seizure data may overlap. The classification is between pre-ictal, ictal, and post-ictal, can vastly differ between experts’ classification. This affects the models detection accuracy. [11]

4. RESULTS AND DISCUSSION

Results consist of the ratios ICA percentages per model. The model is made per patient ID, including ID0 where the model uses all patients' data. Results consist of the binary model, a 3-class classification model, and a cherry-picked model.

4.1 2-class classification model

2-class classification model results from all the models imply a difference in the success rate of model accuracy between the patients as seen in figure 7.

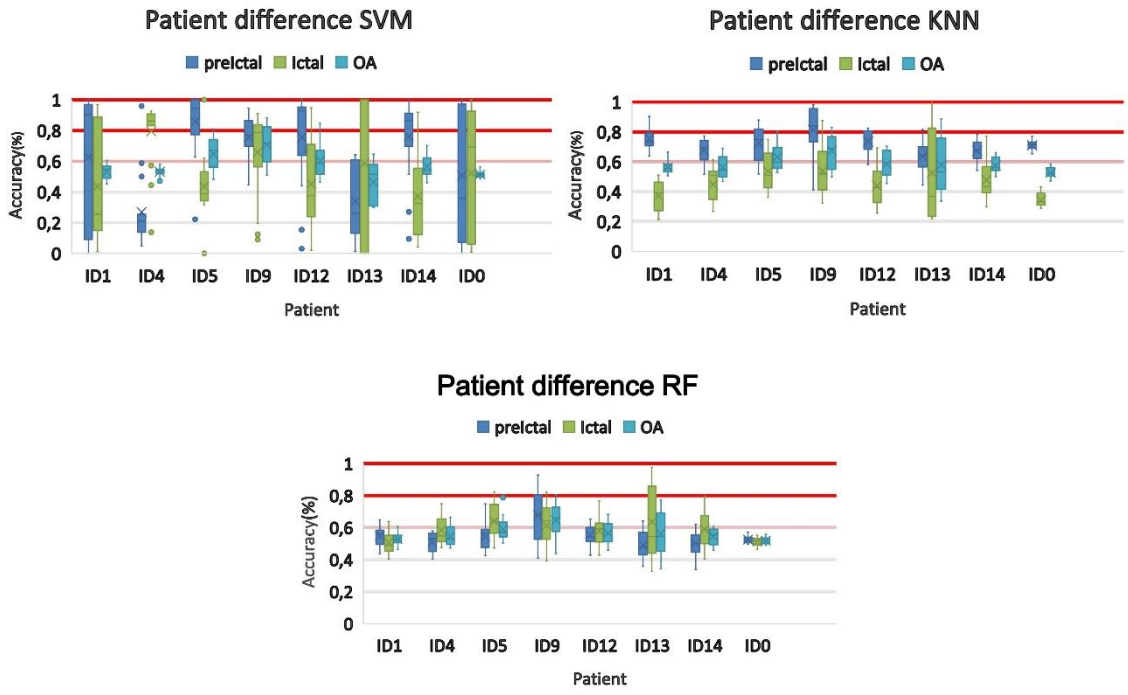

Figure 7. *Accuracy between patients and ML models. The red gradient implicates more significant results starting from 60%.*

Overall model results, patients ID5, ID9, ID12, and ID13 have significant results with more than 60% OA results. ID0 seems to have insignificant results with low variance emphasizing more the patient differences between the EEG seizure expression. The low performance accuracy in different models is affected by multiple things. One is

artifactual channels as seen in figure 8. Therefore, we focus on these patients in the rest of the thesis.

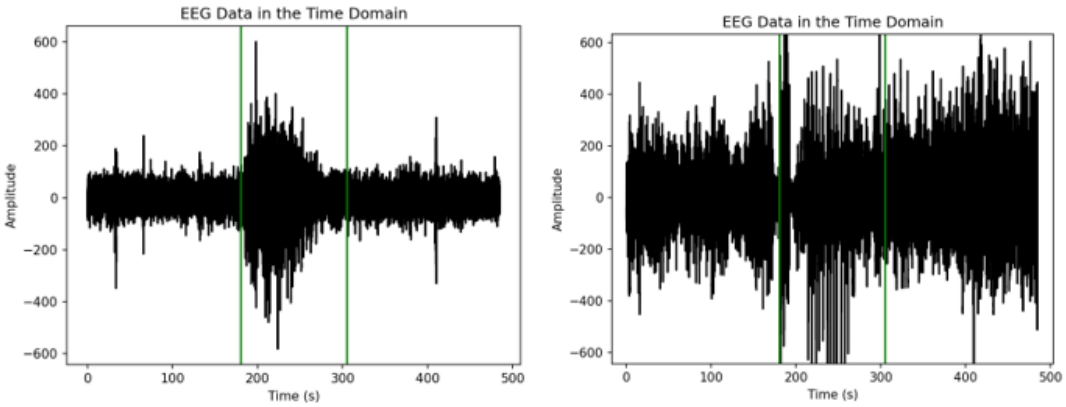

Figure 8. *ID1, Seizure 1, the difference between channels. Channels 1 and 34, sequentially. The annotated ictal phase is represented with green lines.*

The ratio performance is evaluated using the patient ID 5, 9, 12, and 13. The OA and the ictal accuracy imply most about the performance. The performance can be seen in figure 9, 10 and 11, where the notation “all” references all other ratios average, same as “others”.

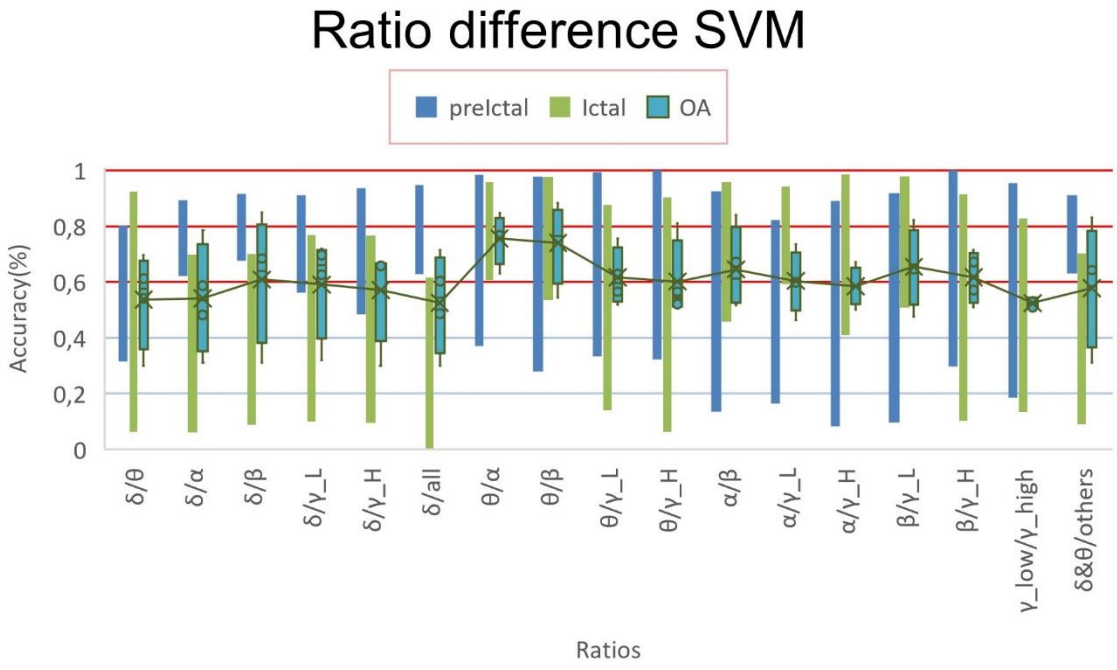

Figure 9. *Representation of classification accuracy done with SVM using patients ID 5, 9, 12, and 13.*

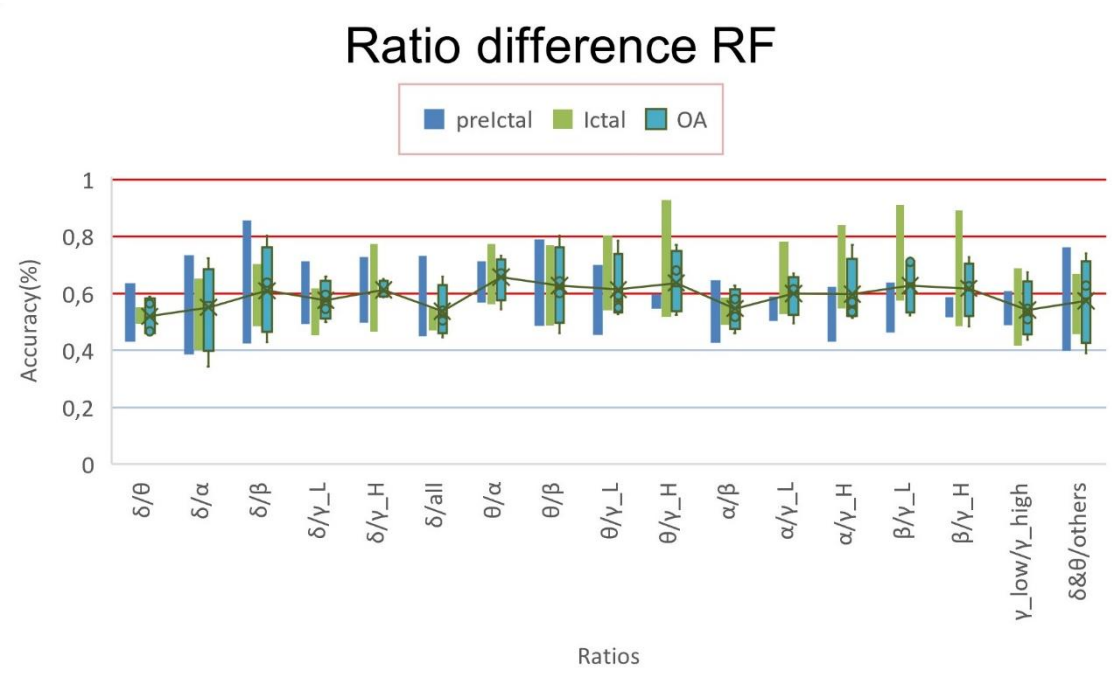

***Figure 10**. Representation of classification accuracy done with RF using patients ID 5, 9, 12, and 13.*

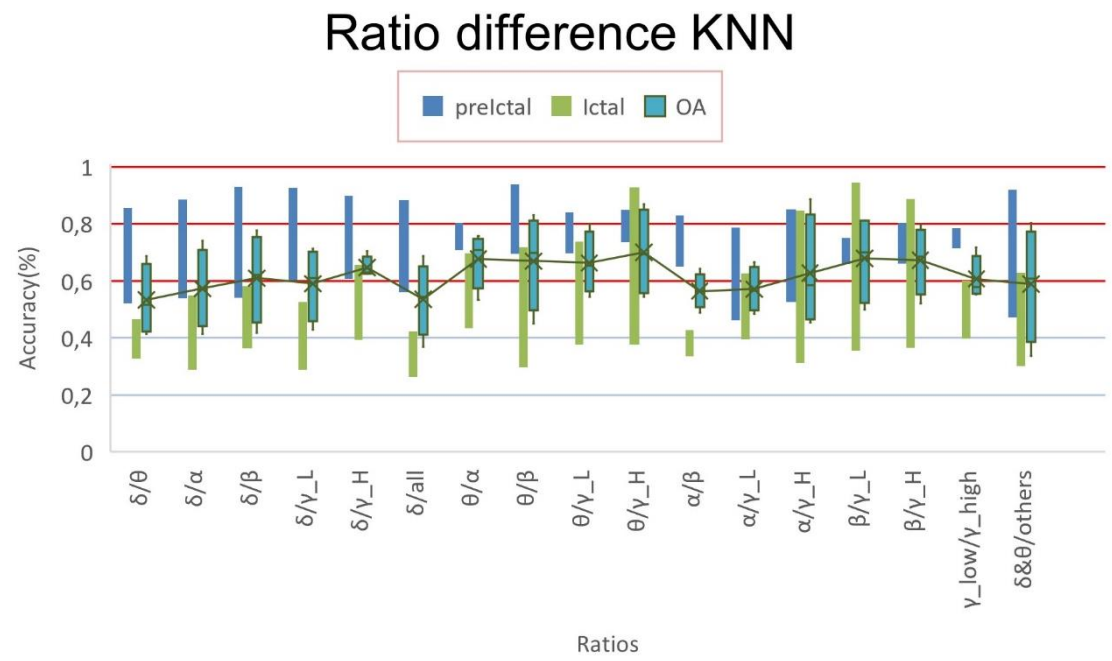

***Figure 11.** Representation of classification accuracy done with KNN using patients ID 5, 9, 12, and 13.*

The ratios between two lower frequencies seem to have generally lower performance with low ictal classification accuracy. In contrast, the ratio between a higher frequency seemed to perform better in general. Supporting this, the combined delta and theta PSD's ratio to other bands' PSD had a noticeable performance.

To be added, theta ratios seem to have the highest resulting classification overall. Theta sub-band has been used to make working epileptic seizure prediction models by itself as a feature [28]. It is left to be debated if the use of bands' ratios increases the classification accuracy or if could it be done using just the PSD of a band. This could explain the higher performance of lower frequency to higher frequency PSD ratios higher performance for higher frequencies have generally lower PSD increasing the effectiveness of a singular band's PSD.

Most of the high-performing ratios (OA > 75%) were theta ratios. Overall, best-performing models and cases are seen in table 1.

Table 1. *Best-performing models by patient*

ML	Ratio	Patient	Preictal	Ictal	OA
KNN	α/γ_H	**ID13**	0,815217	0,956522	0,88587
SVM	θ/β	**ID9**	0,857143	0,910714	0,883929
	θ/α	**ID12**	0,938776	0,755102	0,846939
KNN	θ/β	**ID5**	0,851852	0,75	0,800926

KNN had the most models with over 75% OA (11 models). SVM had most models with over 80% OA (7 models). RF was the lowest performing.

4.2 3-class classification model

The same best-performing patients are used for 3-class classification model visualization as concluded in 2-class classification model. 3-class classification seems to have low performance as seen in figure 12.

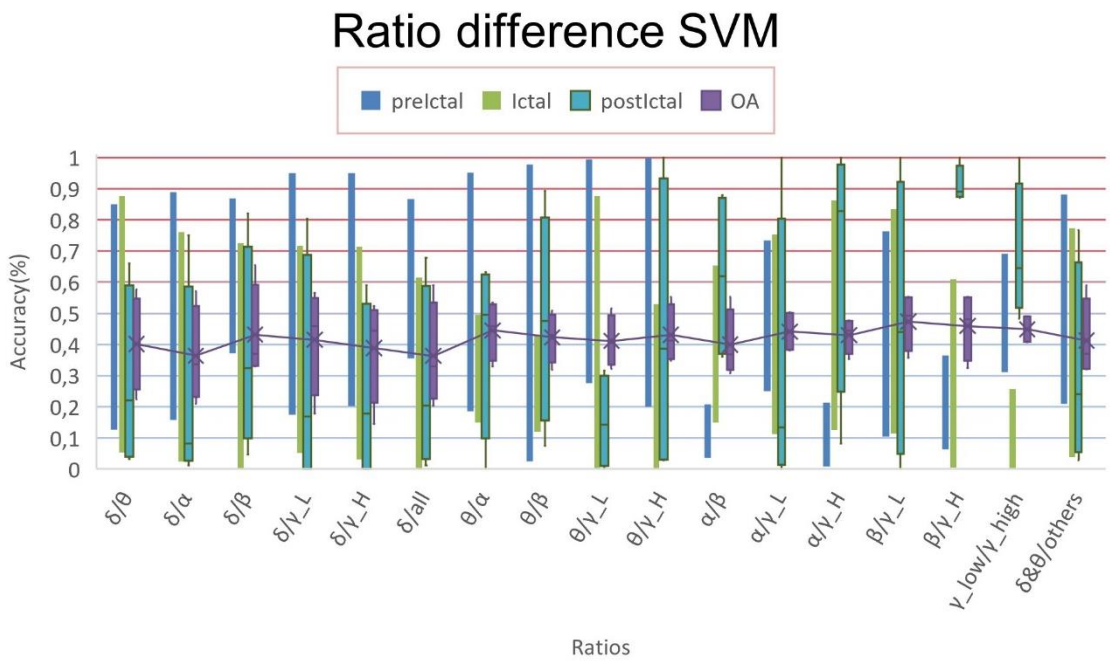

Figure 12. *3-class classification model's performance across patient ID 5, 9, 12, and 13*

This is due to the classification mixing up pre- and postictal data, or in contrast, disregarding ictal classification to fit pre- and postictal data. This can be viewed in figure 13 where "preictal + postictal" preictal classification is done to pre- and postictal data. Likewise, "postictal + preictal" is done to postictal classification.

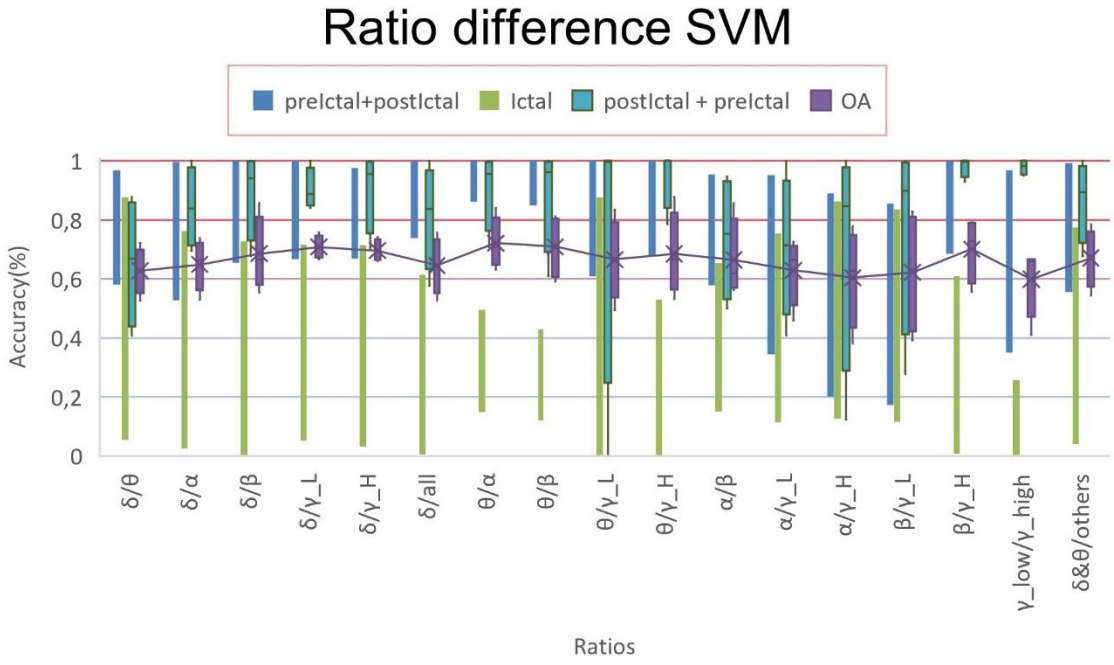

Figure 13. *3-class classification model's ratios performance allowing pre- and postictal mislabeling*

All models made with 3-class classification are low performing regardless of the ML model used.

4.3 Cherry-picked model

One model was made to a lower-performance patient by cherry-picking good channels. The aim was to showcase the feature's functionality on low-performing patients who highly affected by the artefactual channels. Continuingly, deducting the effects caused by the different natures of seizures between the patients implies a more universal application of the feature.

The model was done to ID1. From every seizure, there were 2 - 3 channels handpicked by plotting the EEG channels. This totalled to 35 channels of which 6, from 2 seizures, were left as the testing data. The results were increased from the 2-class model performance across all ratios (Figure 14).

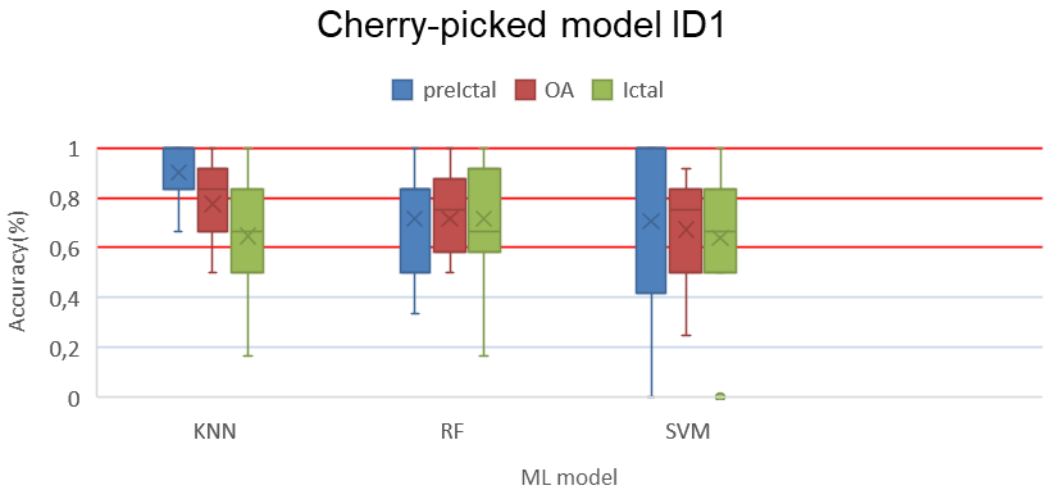

Figure 14. *Cherry-picked model of ID1 accuracy between different ML models.*

The cherry-picked model is only looked upon as an overall indicator. The performance of individual ratios is left unchecked for the limited learning and testing data.

4.4 Discussion

The definitive performance result is limited by the dataset's quality. Additionally, the minimal signal processing allows artifacts to remain in the signal that further affect the results. Nevertheless, the result suggests that PSD ratio could be utilized as a feature

in epilepsy seizure prediction. The result does not clearly support PSD ratio usability as a sole feature in epilepsy seizure prediction.

The neural activation balances during the seizure until the seizure ultimately ends. This decrease of neural activation may lower the classification accuracy for the signal increasingly resembles non-ictal data. Therefore, the classification may be impacted by the usage of the whole ictal data rather than just the beginning.

Epilepsy seizure prediction and classification models need a computationally low cost and a quick feature. The PSD ratio offers this for it utilizes only the frequency domain in feature extraction requiring low processing. In contrast, competitive features and models have high performance with the cost of increased computational or durational cost or they require too much from the hardware [6]. For example, using solely the sub bands normalized spectral power value as a feature is shown to perform worse than the relative power feature [29]. Using the relation of sub bands spectral power normalizes the value inheritably lowering the calculation complexity. The usage of multiple ratios as a feature could increase the performance.

5. CONCLUSION AND OUTLOOK

Using the relation of different sub-bands SPD indicates a possibility of being used as a feature in epileptic seizure predictions. SPD ratio is a computationally cheap and fast way of indicating the difference between the ictal and non-ictal phases. Results show a considerable performance in a few patients up to 88% OA.

The artefactual channels and the disputed definition of the seizure phases greatly affect the performance of the models. In iEEG, where there is a comparatively high signal-to-noise ratio to sEEG, several artefactual channels can be found. These ultimately affect the performance of certain patients and the overall model made across all the patients. The effect of artefactual channels can be seen in the cherry-picked model.

The 3-class classification model didn't exceed notable performance accuracy for the latter reasons mentioned. In addition, the division between pre-ictal and post-ictal data was not managed to be made for the similarities of data or without the cost of ictal classification accuracy.

The 2-class classification model resulted in higher performance on features of low-frequency band's PSD in ratio to a higher frequency band's PSD. The most high-performance models were achieved with theta ratios. In addition, PSD of theta and delta in ratio to other bands was noted to have relatively high performance.

PSD ratio features performance should be further studied. The feature's performance should be tested to recognize more distinct parts of the of the signal rather than the whole ictal phase. The classification accuracy may improve without losing the functionality as a feature in epilepsy prediction models. Continuing, the utilization of multiple PSD ratios as a feature may increase performance and may be worthy of implementing.

REFERENCES

[1] A. Vera-González, 'Pathophysiological Mechanisms Underlying the Etiologies of Seizures and Epilepsy', in *Epilepsy*, S. J. Czuczwar, Ed., Brisbane (AU): Exon Publications, 2022. Accessed: Jan. 17, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK580618/

[2] P. Chauhan, S. E. Philip, G. Chauhan, and S. Mehra, 'The Anatomical Basis of Seizures', in *Epilepsy*, S. J. Czuczwar, Ed., Brisbane (AU): Exon Publications, 2022. Accessed: Jan. 17, 2024. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK580614/

[3] T. Tomson, 'Mortality in epilepsy', *J. Neurol.*, vol. 247, no. 1, pp. 15–21, Jan. 2000, doi: 10.1007/s004150050004.

[4] R. S. Fisher *et al.*, 'ILAE Official Report: A practical clinical definition of epilepsy', *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014, doi: 10.1111/epi.12550.

[5] K. E. Misulis and K. E. Misulis, *Atlas of EEG, Seizure Semiology, and Management*. Oxford, UNITED STATES: Oxford University Press, Incorporated, 2013. Accessed: Dec. 04, 2023. [Online]. Available: http://ebookcentral.proquest.com/lib/tampere/detail.action?docID=1426695

[6] K. Rasheed *et al.*, 'Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review', *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 139–155, 2021, doi: 10.1109/RBME.2020.3008792.

[7] J. Parvizi and S. Kastner, 'Human Intracranial EEG: Promises and Limitations', *Nat. Neurosci.*, vol. 21, no. 4, p. 474, Apr. 2018, doi: 10.1038/s41593-018-0108-2.

[8] 'Advances in human intracranial electroencephalography research, guidelines and good practices - ScienceDirect'. Accessed: Feb. 01, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811922005559

[9] 'The SWEC-ETHZ iEEG Database and Algorithms'. Accessed: Jan. 19, 2024. [Online]. Available: http://ieeg-swez.ethz.ch/

[10] R. P. Lystad and H. Pollard, 'Functional neuroimaging: a brief overview and feasibility for use in chiropractic research', *J. Can. Chiropr. Assoc.*, vol. 53, no. 1, pp. 59–72, Mar. 2009.

[11] A. Shoeb and J. Guttag, 'Application of Machine Learning To Epileptic Seizure Detection'.

[12] G. Strobbe, 'Advanced forward models for EEG source imaging'.

[13] J. Winawer, K. N. Kay, B. L. Foster, A. M. Rauschecker, J. Parvizi, and B. A. Wandell, 'Asynchronous Broadband Signals Are the Principal Source of the BOLD Response in Human Visual Cortex', *Curr. Biol.*, vol. 23, no. 13, pp. 1145–1153, Jul. 2013, doi: 10.1016/j.cub.2013.05.001.

[14] N. Panigrahi and S. P. Mohanty, *Brain Computer Interface: EEG Signal Processing*. Milton, UNITED KINGDOM: Taylor & Francis Group, 2022. Accessed: Mar. 04, 2024. [Online]. Available: http://ebookcentral.proquest.com/lib/tampere/detail.action?docID=7001104

[15] A. Burrello, K. Schindler, L. Benini, and A. Rahimi, 'One-shot Learning for iEEG Seizure Detection Using End-to-end Binary Operations: Local Binary Patterns with Hyperdimensional Computing', in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Cleveland, OH: IEEE, Oct. 2018, pp. 1–4. doi: 10.1109/BIOCAS.2018.8584751.

[16] M. P. Norton and D. G. Karczub, *Fundamentals of Noise and Vibration Analysis for Engineers*. Cambridge University Press, 2003.
[17] 'Epileptic seizure prediction using relative spectral power features - ScienceDirect'. Accessed: Feb. 21, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245714002971
[18] P. G. Stoica, R. Moses, P. Stoica, and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson, Prentice Hall, 2005.
[19] P. Welch, 'The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms', *IEEE Trans. Audio Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.
[20] S. M. Usman, S. Khalid, R. Akhtar, Z. Bortolotto, Z. Bashir, and H. Qiu, 'Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: Review of available methodologies', *Seizure*, vol. 71, pp. 258–269, Oct. 2019, doi: 10.1016/j.seizure.2019.08.006.
[21] V. Kecman, 'Support Vector Machines – An Introduction', in *Support Vector Machines: Theory and Applications*, L. Wang, Ed., in Studies in Fuzziness and Soft Computing. , Berlin, Heidelberg: Springer, 2005, pp. 1–47. doi: 10.1007/10984697_1.
[22] O. Kramer, 'K-Nearest Neighbors', in *Dimensionality Reduction with Unsupervised Nearest Neighbors*, O. Kramer, Ed., Berlin, Heidelberg: Springer, 2013, pp. 13–23. doi: 10.1007/978-3-642-38652-7_2.
[23] G. Biau and E. Scornet, 'A random forest guided tour', *TEST*, vol. 25, no. 2, pp. 197–227, Jun. 2016, doi: 10.1007/s11749-016-0481-7.
[24] 'MNE — MNE 1.6.1 documentation'. Accessed: Mar. 07, 2024. [Online]. Available: https://mne.tools/stable/index.html
[25] 'SciPy -'. Accessed: Mar. 07, 2024. [Online]. Available: https://scipy.org/
[26] 'scikit-learn: machine learning in Python — scikit-learn 1.4.1 documentation'. Accessed: Mar. 07, 2024. [Online]. Available: https://scikit-learn.org/stable/
[27] 'BachelorThesis2024/main.py at main · MarkusHaukipaa/BachelorThesis2024 · GitHub'. Accessed: Apr. 19, 2024. [Online]. Available: https://github.com/MarkusHaukipaa/BachelorThesis2024/blob/main/main.py
[28] N. Jiwani, K. Gupta, and N. Afreen, 'Automated Seizure Detection using Theta Band', in *2022 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2022, pp. 1–4. doi: 10.1109/ESCI53509.2022.9758331.
[29] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, 'Epileptic seizure prediction using relative spectral power features', *Clin. Neurophysiol.*, vol. 126, no. 2, pp. 237–248, Feb. 2015, doi: 10.1016/j.clinph.2014.05.022.