Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: September 30, 2022

Citation: Tsolakis AC, Timplalexis C, et al., 2022. Electroencephalography classification of healthy, mild cognitive impairment and probable Alzheimer's disease through linear and non-linear biomarkers, Medical Research Archives, [online] 10(9). https://doi.org/10.18103/mra.v10i9.3064

DOI
https://doi.org/10.18103/mra.v10i9.3064

ISSN: 2375-1924

RESEARCH ARTICLE

Electroencephalography Classification of Healthy, Mild Cognitive Impairment and Probable Alzheimer's Disease Through Linear and Non-Linear Biomarkers

Apostolos C. Tsolakis*[a], Christos Timplalexis[b], Magda Tsolaki[c,e], Elias C. Aifantis[d,e]

[a]Laboratory of Materials for Electrotechnics, School of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Greece
[b]School of Science and Technology International Hellenic University Thessaloniki, Greece
[c]1st Department Neurology, School of Medicine, Faculty of Health Sciences, Aristotle University of Thessaloniki, Greece
[d]Laboratory of Mechanics and Materials, School of Civil Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Greece
[e]Laboratory of Neurodegenerative Diseases, Center for Interdisciplinary Research and Innovation (CIRI - AUTh), Thessaloniki, Greece

*aptsolak@gmail.com

ABSTRACT
Alzheimer's disease is one of the main challenges of modern medicine since no cure has been found yet, the scientific community still does not fully understand the reasoning behind it, and any interventions found can delay the progress for only a limited amount of time. Over the years, research has shifted from attempts for curing the disease to efforts towards understanding the mechanisms behind it as well as finding tools that will speed up diagnosis many years before its clinical manifestations, when the brain deterioration begins. One of the many promising tools towards this direction is electroencephalography. Electroencephalography employs a variety of different measures that can be used as biomarkers for early diagnosis and differentiation of Alzheimer's disease from other neurodegenerative disorders. Literature has produced a number of methods that have established reliable correlation between electroencephalography signals and structural abnormalities in Alzheimer's disease. To that end, the present work proposes the combination of Tsallis Entropy and Higuchi Fractal Dimension within a common classification framework using machine learning techniques for classification among healthy, Mild Cognitive Impairment, and probable Alzheimer's disease. The proposed methodology is applied on 75 subjects with different feature utilisation scenarios, reaching to an accuracy of 98.03% when classifying a signal epoch, following a 10-fold cross validation, as compared with other similar studies. Nevertheless, in a leave-one-out scenario with the same approach, the average accuracy drops significantly, suggesting that this method could complement other diagnosis approaches but cannot be used on each own.

Keywords: Alzheimer's disease, EEG, Tsallis Entropy, Higuchi Fractal Dimension, Signal Processing

1. Introduction

According to the International Alzheimer's Association,[1] Alzheimer's Disease (AD) is the most common form of dementia. With more than 1 in 9 people (11.3%) age 65 and older having AD, the decease remains as one of the most severe factors that affect brain dysfunctioning, mainly in elderly people. It is expected to affect roughly 12.7 million 65+ people by 2050. Although modern medicine and advanced technology breakthroughs are continuously applied to moderate the situation, the incidence and mortality due to AD keeps rising. From 2000 to 2019 an increase of 145.2% has been observed. To that end an enormous amount of resources have been employed not only to postpone the effects of AD (which is still currently the only successful course of action) but to also understand and fully analyse the underlying physiological processes responsible for the brain degradation, both in terms of tissue quality and volume shrinkage (Figure 1).

Figure 1 From left to right: a healthy brain, an AD brain, and the comparison between the two[2].

Given the fact, that the scientific community has failed so far to develop a cure for AD and that the solutions available only target symptoms but not the cause of the disease,[3] efforts have shifted towards better understating of the initial mechanisms that cause cognitive decline that could lead to an early AD diagnosis, especially at the Mild Cognitive Impairment (MCI) level, as this is considered a precursory stage of AD.[4] In this connection, it is also pointed out that to this day, diagnosis is mainly based only on clinical criteria, a fact that introduces additional challenges.[4-5] An early diagnosis may contribute not only to the development of more effective interventions that could delay the progress or even inhibit it entirely, but it could also prevent some of the symptoms to evolve when dealt with at an early stage. Towards the direction of accomplishing an early diagnosis, several different methodologies have been proposed; some of them are invasive (e.g., blood and Cerebrospinal Fluid – CSF), others are expensive (i.e., MRI, SPECT, or PET), with only a number of them eluding significant results.[6] In contrast with these, a non-invasive, low-cost and high-resolution method in terms of brain activity is the electroencephalography (EEG).

With research going back a few decades,[7-8] a plethora of studies have been focused on the use of EEG in AD, revealing certain commonly agreed features and some other rather controversial[9]. The most interesting features that are commonly agreed upon regarding EEG and AD are summarized below[10-13]:

- **Overall retardation of specific rhythms;** in particular, the observations so far present an increase in delta (0.1 - 4 Hz) and theta (4 – 8 Hz) activities and decrease in alpha (8 – 13 Hz) and beta (13 – 30 Hz) activities. Earliest changes are characterised by an increase in theta and a decrease in beta activities, followed by a decrease in alpha, while delta increases later during the progress of the disease. This is supported by the fact that patients with severe dementia exhibit a decrease in alpha and an increase in delta activity, whereas patients with mild dementia show a decrease in beta and an increase in theta activity;
- **Decreased complexity;**
- **Decreased coherence** in general and among different brain regions;
- **Overall topography changes;** in particular, observations indicate that slow activity is prominent in the left temporal area of AD patients, whereas differences between pre-senile patients and normal controls are detected in the right posterior temporal area. Largest differences between senile patients and the controls are found in the midfrontal and anterior frontal lobes bilaterally.

When evaluating complexity, significant effort has been focused on non-linear dynamics,[14] under the assumption that EEG signals are generated by nonlinear deterministic processes with nonlinear coupling interactions between neurons. Studies employing such measures have found that AD patients have reduced values of the correlation dimension (D2) in the occipital EEG compared with those of healthy subjects, and with probable AD subjects.[14-18] In addition, it has been highlighted that AD patients exhibit reduced spatio-temporal brain activity in comparison with that in normal controls,[19] and in some cases the former subjects are

characterized by specific patterns of dysfunction in dementia.[20] Investing in the analysis of EEG complexity, a lot of novel biomarkers have been extracted from non-linear approaches (e.g., entropies, fractality, lacunarity) towards providing the necessary methods for accurate and early diagnosis of AD.

Such biomarkers have gained an even greater importance with the emerging trend of artificial intelligence based solutions, such as machine learning. The proper feature selection in order to train models that can successfully classify an epoch (most commonly followed in the literature along with a k-fold cross validation) or an entire subject (leave one out methodology) is consider of utmost importance. Interestingly enough, going through recent literature, although quite promising findings are reported (e.g., Yang et al.[21] reported a classification accuracy of 98%), there are still significant challenges that need to be addressed, as stated by Tanveer et al.,[22] such as the classification of MCI, which hasn't been researched a lot, as well as the size of the datasets explored, and the noise/artefacts introduced in AD, which if removed (as seen in most studies), remove a significant portion of the signal.

As this is a vast field of research nowadays, the present study focuses mainly on two specific non-linear biomarkers (i.e., Tsallis Entropy and Higutchi Fractal Dimension) that have been found through the literature to have promising potential and combines them for the intended purpose. The methodology designed and the methods selected are presented in the following section.

1.1. Entropy and the Tsallis approach

Entropy is a measure of the uncertainty associated with a random variable or otherwise it's defined as a measure of uncertainty of information in a statistical description of a system, and has been applied initially to thermodynamics, and later on to statistical mechanics. Up to this day, over forty (40) different types of specific, generalized, extended, etc. entropies have been introduced. For Statistical Mechanics one of the entropies proposed over the years was the Tsallis entropy.[23] The Tsallis entropy (or Tsallis statistics), was introduced as a generalized version of the Boltzmann-Gibbs statistics and from that point onward have been applied in various domains, one of which is statistical bio-mechanics and particularly brain-related cases where the ways that the "system" can be arranged are limitless as far as we know.

To the knowledge of the authors, the first application of the Tsallis entropy on EEG signals was performed by Gamero et al.[24] in 1997. They used wavelet transforms in multiple resolutions, extracting probabilities from the wavelets coefficients and calculating the Shannon and Tsallis entropies of EEG data focusing on spike–wave paroxysms. Within this context, they proved the robustness of the Tsallis over the Shannon entropy for the detection of these waves. Taking it a step further, Capurro et al.[25] continue working on human brain dynamics using the Tsallis measure on EEG data. By examining again Shannon and Tsallis entropies over different levels of wavelet analysis on epilepsy patients' EEG data, they produced similar results to those reported in Gamero et al.,[24] but for sharp waves. A year later, Martin et al.[26] used Tsallis analysis on EEG focusing also on epileptic seizures that provided valuable insight on signal analysis by revealing a particular degree of sensitivity as a detector of changes in the parameters of dynamical systems.

Following their work, Thakor et al.[27] in 2001 used both Shannon and Tsallis entropies to quantitatively access brain rhythms; in particular normal versus injured brain signals. A reduction in the entropy of the brain rhythm has been observed after calculating the mean and standard deviation of the Tsallis entropy. As a result, they proposed Tsallis entropy as a non-redundant information measure of brain dynamics with promising applications to various brain-related research areas. Three years later, they expanded their research by reviewing in general quantitative EEG analysis methods, including Tsallis entropy, highlighting the nonlinear properties of the EEG source. They provided in particular, an initial link between Tsallis entropy, EEG and AD, thus establishing a new robust method for assessing EEG complexity which seems to be affected by the neurodegeneration caused by the disease.[28]

Sneddon[29] demonstrated an estimator of the Tsallis entropy in order to assess the information in the EEG of subjects that were trying to recall objects or faces, under the hypothesis that neurodegeneration is linked/caused by faulty information processing. Some of the subjects were diagnosed with probable AD, whereas the others were considered healthy. In order to compare results, a ratio between frontal to posterior brain information was utilised with the threshold of 1 being the separator between the two (normal >1 and AD < 1). The overall results suggested that this decision criterion exhibits a very high accuracy for AD detection. However, the sample that this method used was too small (30 healthy and 16 AD) and moreover, it was not accompanied with neurophysiologic information about the subjects. After more screening was performed, better results were established but,

again, the sample consisted of 10 normal and 10 AD subjects only. Nevertheless, interesting findings were obtained regarding Tsallis entropy monitoring of AD treatment.

In the same year, Zhao et al.[30-31] presented a normalised Tsallis entropy for EEG to be used as a biomarker for AD. In their studies two small subject databases were examined with a high diversity and only between normal and AD (most of which probable AD). The experimentally extracted values for the Tsallis Entropy characterisation parameters were: N=5120 and q=0.5. They applied the normalized Tsallis entropy (NTSE) calculation to all 21 EEG channels and found that the AD group had clearly lower NTSE than the normal group. An interesting outcome was their threshold value of p=0.22, according to which subjects with NTSE lower than that are AD whereas higher than that they can be considered as normal. However, their analysis between datasets produced low sensitivity results and highlighted the need for efficient pre-processing due to high levels of noise.

Given the fact that MCI diagnosis is more useful than early AD diagnosis, De Bock et al.[32] following the work of Shennon[29] presented two ratios which were calculated using the Tsallis entropy: prefrontal cortex to posterior parietal lobe, and prefrontal cortex to occipital lobe. The study did not provide any definite results and, once again, the sample examined was considered too small (15 normal and 11 MCI). In the same context, McBride et al.[33] explored a multiscale entropy theory that combined Tsallis, "approximate" and "sample" entropies, and through support vector machine (SVM) models discriminated AD, MCI and normal subjects (43 subjects). Their findings suggest high discrimination accuracy, through promising ratios for AD vs. Normal discrimination clustered in the parietal and frontotemporal (frontal and temporal channels) regions of the head; especially for responses to matching target held in working memory. Differences between MCI and the other groups appeared to be widespread across the head, including the occipital, parietal, and frontal regions. Garn et al.[34] conducted one of the most extensive studies of quantitative EEG (QEEG) markers in order to identify which ones (either individually or combined) can best correlate to AD severity. For complexity measures, the Tsallis entropy was once again employed, examining only on the band of 2–15 Hz. Although the complexity measures obtained are not the optimal ones for AD severity, when combined with other factors, they can provide added value to the overall correlation. An interesting highlight of their work is that left sided indices in temporal and parietal regions consistently showed most significant results in their subjects.

Al-nuaimi et al.[35] continued the work from Zhao et al.[30-31] by taking the normalised Tsallis entropy which was applied to each EEG channel and for each subject from their dataset. They established two reference feature vectors (AD and normal) and they used K-means clustering to compare a new dataset to their reference vectors. Their approach provided improved results in terms of *sensitivity*, *specificity*, *accuracy* and *precision*. However, the dataset was the same as the one used in the previous work; i.e. small enough to derive reliable results. Later on, in 2018 they presented[36] a more complete work, investigating once more the complexity measures, in order to quantify changes in EEG for AD. They focused again on the most promising methods that have been used so far: according to them, the Tsallis Entropy, the Higuchi Fractal Dimension (see next section) and the Lempel-Ziv complexity. Their results indicate that all three measures are significantly lower for AD subjects for specific EEG bands and channels, and that we can now start detecting AD with a sensitivity and specificity of more than 90%.

Finally, in a most recent extensive analysis, Tzimourta et al.[37] explored 38 linear and non-linear features on EEG signals, consisting of multiple entropies including Tsallis, in an effort to identify the most promising ones, through their correlation to the Mini-Mental State Examination (MMSE) scores. By applying a multi-regression linear analysis on a dataset consisting of 24 subjects (again a very limited dataset with only 5 moderate AD and 9 mild AD), they presented a high correlation of MMSE score variation with Permutation Entropy. Interestingly, however, the Tsallis Entropy has not been found to have any correlation with the MMSE score.

1.2. Fractality and the Higuchi Approach

In general, the fractal dimension can be used, in particular, as an index of irregularity in signals and patterns to evaluate time series with non-periodic and turbulent behaviour,[38] thus making it a very suitable tool for EEG waveforms. Two of the first researchers to apply this complexity measure were Woyshville and Calabrese.[14] By employing the original Hausdorff[39] fractal dimension it was established that the normal subjects group had the largest fractal dimension, whereas lowest scores were obtained through the (autopsy confirmed) AD patients.

Following a different path, Besthorn et al.[40] estimated the dimensional complexity (fractal dimension) using the approach proposed by

Pritchard and Duke.[41] They pointed out that it is possible (given that the same algorithm and parameters are used to all subjects) to investigate group effects without being biased by the heterogeneity of inter-individual variability. They also concluded that AD subjects have significantly lower values of dimensional complexity, which can be correlated with known dementia scales (i.e., MMSE) and EEG band power scores.

Two years later, Accardo et al.[42] identified the Higuchi Fractal Dimension (HFD)[43] as a fast and efficient computational method that is able to successfully and accurately estimate the fractal dimension also for segments shorter than 250 ms, thus enabling the study of brief EEG events and the identification of behavioural variations with a good temporal resolution. A decade afterwards, Henderson et al.[44] further assessed EEG signals as a tool for detecting dementia. They concluded that the fractal dimension can be a good candidate to fulfil the need for a low-cost, easy to administer and reasonably accurate method for detecting dementia.

Staudinger and Polikar[45] employed the HFD as an EEG based biomarker for AD (among others). Their finding agrees with the overall literature where AD subject have lower values of HFD, while they also attributed the relevant information carried by HFD to the parietal and temporal areas. Finally, an interesting suggestion made by the authors is the combination of different non-linear dynamics measures into a feature vector to significantly improve classification accuracy by more than 10%.

Smits et al.[46] examined the HFD between healthy and AD subjects, to assess the sensitivity of brain activity changes for the two groups. They also calculated another measure based on HFD to study symmetry between the two hemispheres. They named this new measure as Homologous Areas Interhemispheric Symmetry (HArS), according to which, left-higher-than-right HFD asymmetries correspond to positive values and right-higher-than-left HFD asymmetries correspond to negative values. Their results indicate that HFD depends on age in healthy subjects but is reduced in AD ones, whereas focusing on a regional correlation among HFD, age and AD, healthy subjects' HFD depends most strongly on age in parietal and central brain whereas for AD subjects there was a strong dependency in temporal and occipital areas. The last one is a bit controversial with other research findings that suggest significant changes on the frontal areas. Finally, based on the new metric introduced they found a loss of HArS of parietal HFD that depends on age.

Al-nuaimi et al.[47] also elaborated on the HFD in their most recent work.[36] They demonstrated that HFD is indeed a promising biomarker that is able to capture the areas of the brain that are considered to be affected first in the early stages of AD. Their HFD values are lower on AD subjects, while (just like Tsallis Entropy) the analysis of separate channels holds more promise than that of the whole EEG record.

Summarising, it is evident that there are several controversial findings in the literature when employing complexity measures such as entropy and fractality towards the ultimate goal of differentiating and quantitatively characterising the various stages that govern the evolution of AD. Furthermore, the most interesting findings retain to research that examines (probable) AD vs. healthy samples without investigating the stages in between (i.e., MCI). On top of that, the vast majority of the related studies are presenting results based on a very limited sample of subjects, a fact that considerably limits their validity. Finally, to the authors' knowledge, entropy and fractality have not been combined together, although they have delivered promising results when used individually. To address these shortcomings, the present study combines the abovementioned tools of entropy and fractality on a larger number of subjects (from the majority of the relevant literature mentioned above), within a unifying framework which, in addition, can differentiate among the three main stages of brain deterioration: i.e., Healthy controls, MCI, and probable AD. This study extends the initial high level analysis from a previous work of the authors.[48]

The manuscript is organised as follows: Section 2 describes the overall methodology and the features employed in the analysis. Section 3 presents the obtained results based on several scenarios employed for this study, whereas Section 4 includes the discussion on the results. The conclusions along and future direction are included in Section 5, also including limitations of this study.

2. Methods

2.1. EEG Dataset

Within this study, the EEG signals of 75 anonymised subjects were collected from the Greek Association of Alzheimer's Disease and Related Disorders (25 Healthy, 25 MCI, and 25 probable AD) in the context of a retrospective cohort study. Their cognitive status was assessed by the Association's medical experts (neurologists, psychiatrists, and psychologists) through an extensive battery of neuro-psychometric tests (i.e. the Mini-Mental State

Examination – MMSE[49], and the Functional Cognitive Assessment Scale – FUCAS[50] tests).

Regarding the sample size, the selection was performed based on availability of samples at the time of the investigation and not in a pre-hoc power analysis. However, comparing with the literature findings analysed in Section 1, most of previous studies contacted had around 50 or less participants, whereas only 4 had over 100, with the highest sample being 161. Hence, the authors believe that the sample size of this study is of a similar order of magnitude of the literature, thus it can be considered comparable in terms results.

The EEG signals were collected through a set of 21 electrodes Nihon Kohden Neurofax J_921A following the 10-20 international reference system[51] at 500Hz. The input impedance was set to Z<10kΩ. The signals were digitised with the Neurofax EEG-12200 Ver. 01-93. The protocol used for the acquisition of the EEG signals refers to resting stage and lasts for 10 minutes with 5 minutes *eyes closed* and 5 minutes *eyes open*, while being seated in an upright position. For the analysis performed within this study, only the first part was explored (5 min closed eyes), resulting in signals of approximately 300 seconds each (some had slightly less). The reason for selecting only the *eyes-closed* part has to do with the fact that with eyes closed, the subjects have less irregularities to their waveforms due to external stimulation.

For all of the signals, upon acquisition, a 50Hz filter is applied, to remove any noise due to electromagnetic disturbances from surrounding equipment and cables.

2.2. Methodology

In previous work, both non-linear measures (TsEn and HFD) have been tried individually for the analysis of EEG signals for (probable) AD diagnosis, providing promising results (sensitivity and specificity above 90% per epoch). Building on this, the purpose of the present study is to develop an approach that will focus mainly on the MCI stage, thus exploring not only the effect of these complexity measures individually but also combining them in formulating a new biomarker for more informed and accurate diagnosis. A similar approach, but with Sample entropy instead of Tsallis, and towards detecting depression and not MCI or AD, has been presented by Cucic et al.[52] with very high accuracy, ranging from 90.24% to 97.56%.

In accordance with literature findings that indicate alteration on the EEG signals between stages on the frontal and temporal brain regions, initially only specific electrode channels have been examined. Furthermore, since it has also been suggested that both MCI and AD have different effect on the four main rhythms (delta, theta, alpha, and beta), these are also evaluated separately to identify any distinguished alterations on the proposed metrics.

Time-series, data-driven statistical analysis heavily depends on the availability of data. Hence, to increase the number of epochs to be analysed, each signal was further divided in non-overlapping epochs of 10 seconds, as proposed by other studies as well (Tzimourta et al.37), thus resulting in a total of 22,340 seconds (or 2,234 epochs). A more clear presentation of the sample signals is depicted in Table 1 As it can be understood from the numbers shown, it was not possible in some cases to include and analyse all 5 minutes.

Table 1 Sample Pool.

	Number of Total Duration Samples/Signals	Total Duration (seconds)
Healthy	25	7490
MCI	25	7370
AD	25	7480

In order to extract the desired features not only on the whole signal but also in each frequency sub-band of interest, four band-pass filters (0.5 – 4 Hz, 4 – 8 Hz, 8 – 13 Hz, and 13 – 30 Hz) are applied. Then, for each epoch, for each of the five bands (Whole, Alpha, Beta, Theta, and Delta), and for each of the nineteen electrodes included in the analysis (Fp1, Fp2, T3, T5, C3, T4, T6, C4, F3, F7, F4, F8, P3, P4, O1, O2, Fz, Cz, and Pz), six linear and non-linear features were extracted towards training a classification model that could allow the differentiation and identification among Healthy, MCI, and probable AD stages. To ensure the validity of the results a 10-fold cross validation is employed during the evaluation process.

To address the classification problem, three machine learning techniques have been explored, the Support Vector Machines (SVM),[53] the Gradient Boosted Trees (GBT),[54] and the Light Gradient Boosting Machine (LGBM).[55] Since these techniques

are well known to the machine learning community no further elaboration is provided on reviewing them. For the analysis, a custom software was developed using Python and the libraries MNE, Scikit-learn, XGBoost, and LightGBM.

At this point, it has to be mentioned that in order to facilitate the analysis (even from non-experts), there was not any artefact clearance procedure used in this study. Hence, the signals were analysed including potential artefacts towards exploring the possibility of obtaining positive results without the cumbersome task of artefact removal that requires both extra time and expertise.

2.3. Linear and Non-Linear Features

By utilising combined information from the literature, six features are extracted from each epoch, for every band and electrode examined, as follows:

Mean

$$\acute{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

Variance

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \acute{x})^2 \tag{2}$$

Skewness

$$\gamma_1 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{x_i - \acute{x}}{\sigma}\right)^3 \tag{3}$$

where σ denotes, as usual, the standard deviation (or the square root of variance).

Kurtosis

$$\gamma_2 = \frac{1}{N}\sum_{i=1}^{N} \left(\frac{x_i - \acute{x}}{\sigma}\right)^4 \tag{4}$$

Tsallis Entropy (TsEn)

Given a discrete set of probabilities p_i with the condition that $\sum p_i = 1$, and q any real number (in this study q=2), then the Tsallis Entropy is defined as:

$$TsEn = \frac{\sum_{i=1}^{N}(p_i - p_i^q)}{q-1} \tag{5}$$

Higuchi Fractal Dimension (HFD)

For an *N-sample* EEG data sequence *(1), (2), . . . , x(N)*, the data is first divided into a *k-length* subdata set as:

$$x_k^m : x(m), x(m+k), x(m+2k), \ldots, x\left(m + \left[\frac{N-m}{k}\right]k\right) \tag{6}$$

where [.] is Gauss' notation, *k* is constant, and $m = 1, 2, \ldots, k$. The length *L_m(k)* for each subdata set is then computed as:

$$L_m(k) = \tag{7}$$

The mean of *L_m(k)* is then computed to find the HFD for the data as:

$$HDF = \frac{1}{k}\sum_{m=1}^{k} L_m(k) \tag{8}$$

2.4. Evaluation Metrics

For the evaluation and validation of the approach proposed, a 10-fold cross validation is performed to obtain various metrics, as is usually done in the machine learning (ML) domain. Following the assessment presented by Sokolova et al.[56] and most recently by Tharwat[57] for multi-class classification problems, four distinct metrics are adopted. Besides the overall *accuracy* of the classification results, which is rather easy to comprehend, the *precision*, *recall* and *f* or *f1-score* are calculated.

3. Results & Discussion

3.1. ML Technique and Feature Combination Selection

As mentioned in the previous sections three ML techniques have been used towards classifying the EEG signal epochs in one of the three clinical stages: the SVM, GBT, and LGBM techniques. The configuration for the latter, that provided the highest accuracy in the optimal scenario, was 64 leaves in one tree, 50 estimators, and -1 depth for the tree model. On the other hand the configuration for the GBT, which was the second best, was 400 estimators and 5 layers (depth) for the size of the decision trees.

It is evident that LGBM outperforms both SVM and GBT in almost all examined cases, with a sole exception of *Delta-O2*, in which GBT seems to be slightly better. Nevertheless, that particular combination is one of the least accurate combinations. Interestingly, the *Whole* band seems to outperform all other besides Theta, which holds the highest accuracy, for electrode *T4*, with 82.99%.

Through the experimental procedure (see Table 2) it was clear that the LGBM outperforms the other two in providing higher accuracy results. Hence, for the detailed results to be presented in the following sections, only the data acquired through the LGBM technique are given. Nevertheless, to support this selection some indicative results are included in the table given below, covering both the highest and lowest accuracy results in each band.

Table 2 Accuracy comparison between SVM, GBT, and LGBM techniques.

Band-Electrode	SVM	GBT	LGBM
Whole-T6	0.43912	0.65443	0.81289
Alpha-O1	0.48702	0.58057	0.58236
Beta-O1	0.50940	0.57833	0.58729
Delta-Cz	0.45345	0.75112	0.76231
Theta-T4	0.47314	0.82095	**0.82990**
Whole-T3	0.38406	0.50134	0.50806
Alpha-F4	0.44047	0.48747	0.48791
Beta-F8	0.42704	0.47225	0.48568
Delta-O2	0.39928	0.68174	0.68084
Theta-F8	0.49418	0.63250	0.64235

Accordingly, even though literature findings indicate that both the *TsEn* and *HFD* provide high accuracy results individually, this is not fully supported by the approach presented in this study (see Table 3). Therefore, their combination is preferred, along with the other four features proposed. The following table include indicative examples that support this claim.

Table 3 Accuracy comparison when using TsEn or HFD as single features or their combination over the LGBM technique.

Band-Electrode	TsEn	HFD	TsEn & HFD	Six Combined
Whole-T6	0.42480	0.41495	0.54297	0.81289
Alpha-O1	0.46106	0.41003	0.55416	0.58236
Beta-O1	0.49642	0.35542	0.56938	0.58729
Delta-Cz	0.35497	0.67816	0.68218	0.76231
Theta-T4	0.64727	0.62086	0.77798	**0.82990**

3.2. Scenario A: Per Band and Per Electrode

After validating the fact that the use of all six features with the use of the LGBM technique provides the optimal setup for the analysis and classification of the examined EEG epochs, a thorough step by step procedure, consisting of four experimental scenarios, is followed to ensure the validity of the presented methodology. As this analysis is based on each epoch, and not the entire signal, a leave one out scenario is also covered, after the most accurate combination of features is presented.

Within this very first experimental scenario all six features are examined per band and electrode. For example, only the six features for the epochs from the Alpha band and the Fp1 electrode are used towards training and testing the methodology presented above. In the following tables, results from each individual scenario are presented: starting with the overall accuracy results per case and then followed by the *precision*, *recall* and *f1-score* metrics for each stage (i.e., Healthy, MCI, AD).

Table 4 Accuracy results for Scenario A

Electrode	Whole	Alpha	Beta	Delta	Theta
Fp1	0.57923	0.50224	0.52059	0.72963	0.67502
Fp2	0.62086	0.49463	0.50269	0.73053	0.67681
T3	0.50806	0.53536	0.51209	0.69114	0.79051
T4	0.62399	0.50224	0.53715	0.73366	**0.82990**
T5	0.60788	0.54342	0.52775	0.72874	0.79678
T6	0.66294	0.55157	0.53626	0.75783	0.81289
C3	0.57833	0.50582	0.52507	0.73187	0.67592
C4	0.59982	0.50895	0.51522	0.72739	0.67860
F3	0.64011	0.50716	0.55953	0.75828	0.68666
F4	0.63474	0.48791	0.51164	0.74530	0.69024
F7	0.57252	0.49687	0.49373	0.71844	0.65533
F8	0.54611	0.49239	0.48568	0.69517	0.64235
P3	0.64056	0.56580	0.53089	0.72874	0.72739
P4	0.58908	0.54879	0.54342	0.71665	0.66607
O1	0.62220	0.58236	0.58729	0.73187	0.70859
O2	0.59624	0.55282	0.57610	0.68084	0.68353
Fz	0.59042	0.49642	0.53268	0.75560	0.68218
Cz	0.62578	0.50716	0.54252	0.76231	0.66831
Pz	0.61817	0.53850	0.54745	0.75380	0.71979

As depicted by the results in Table 4 and Table 5, the higher accuracy is achieved for the *Theta* band and electrode *T4*, reaching a value of 82.99% (as also presented in Table 3. However, *Theta* doesn't hold the highest accuracy for all other electrodes. Interestingly enough, *Delta* is more accurate than *Theta* in 14 electrodes, whereas Theta just in 5, including T4 which has the highest percentage. Another interesting finding is that when examining the *Whole* band, we can get better results, than examining *Alpha* or *Beta* bands. From a regional perspective, it can be observed that the temporal electrodes have in general higher values than other regions (as also supported by other research on the field analysed in Section 1, with the right-temporal side having slightly better results.

Table 5 Detailed classification results for Scenario A

Electrode	Metric	Whole	Alpha	Beta	Delta	Theta
Fp1	Precision	0.5801	0.5034	0.5207	0.7308	0.6763
	Recall	0.5792	0.5022	0.5206	0.7296	0.6750
	F1-score	0.5789	0.5025	0.5206	0.7299	0.6753
Fp2	Precision	0.6220	0.4942	0.5025	0.7325	0.6778
	Recall	0.6209	0.4946	0.5027	0.7305	0.6768
	F1-score	0.6205	0.4936	0.5026	0.7303	0.6765
T3	Precision	0.5086	0.5357	0.5115	0.6929	0.7920
	Recall	0.5081	0.5354	0.5121	0.6911	0.7905
	F1-score	0.5080	0.5353	0.5112	0.6914	0.7906
T4	Precision	0.6240	0.5028	0.5369	0.7337	**0.8299**
	Recall	0.6240	0.5022	0.5372	0.7337	**0.8299**
	F1-score	0.6240	0.5018	0.5370	0.7337	**0.8299**
T5	Precision	0.6077	0.5435	0.5269	0.7293	0.7974
	Recall	0.6079	0.5434	0.5278	0.7287	0.7968
	F1-score	0.6077	0.5430	0.5270	0.7288	0.7968
T6	Precision	0.6628	0.5279	0.5345	0.7577	0.8129
	Recall	0.6629	0.5278	0.5363	0.7578	0.8129
	F1-score	0.6628	0.5274	0.5346	0.7575	0.8129
C3	Precision	0.5781	0.5056	0.5252	0.7331	0.6767
	Recall	0.5783	0.5058	0.5251	0.7319	0.6759
	F1-score	0.5774	0.5054	0.5248	0.7319	0.6758
C4	Precision	0.5997	0.5092	0.5149	0.7284	0.6790
	Recall	0.5998	0.5090	0.5152	0.7274	0.6786
	F1-score	0.5992	0.5091	0.5150	0.7275	0.6785
F3	Precision	0.6408	0.5075	0.5593	0.7587	0.6893
	Recall	0.6401	0.5072	0.5595	0.7583	0.6867
	F1-score	0.6398	0.5071	0.5594	0.7582	0.6871
F4	Precision	0.6363	0.4878	0.5116	0.7455	0.6911
	Recall	0.6347	0.4879	0.5116	0.7453	0.6902
	F1-score	0.6348	0.4875	0.5116	0.7452	0.6902
F7	Precision	0.5724	0.4970	0.4939	0.7189	0.6549
	Recall	0.5725	0.4969	0.4937	0.7184	0.6553
	F1-score	0.5720	0.4969	0.4936	0.7184	0.6550
F8	Precision	0.5461	0.4916	0.4855	0.6965	0.6418
	Recall	0.5461	0.4924	0.4857	0.6952	0.6423
	F1-score	0.5457	0.4918	0.4854	0.6948	0.6420
P3	Precision	0.6403	0.5660	0.5309	0.7289	0.7276
	Recall	0.6406	0.5658	0.5309	0.7287	0.7274
	F1-score	0.6402	0.5654	0.5308	0.7285	0.7274
P4	Precision	0.5887	0.5489	0.5438	0.7171	0.6682
	Recall	0.5891	0.5488	0.5434	0.7167	0.6661
	F1-score	0.5882	0.5487	0.5435	0.7165	0.6665
O1	Precision	0.6223	0.5833	0.5870	0.7320	0.7087
	Recall	0.6222	0.5824	0.5873	0.7319	0.7086
	F1-score	0.6221	0.5826	0.5870	0.7318	0.7083
O2	Precision	0.5956	0.5548	0.5773	0.6808	0.6836
	Recall	0.5962	0.5528	0.5761	0.6808	0.6835
	F1-score	0.5959	0.5532	0.5761	0.6806	0.6831
Fz	Precision	0.5909	0.4968	0.5320	0.7559	0.6828
	Recall	0.5904	0.4964	0.5327	0.7556	0.6822
	F1-score	0.5902	0.4962	0.5319	0.7557	0.6820
Cz	Precision	0.6264	0.5069	0.5425	0.7625	0.6675
	Recall	0.6258	0.5072	0.5425	0.7623	0.6683
	F1-score	0.6252	0.5061	0.5423	0.7623	0.6675
Pz	Precision	0.6170	0.5384	0.5499	0.7539	0.7199
	Recall	0.6182	0.5385	0.5474	0.7538	0.7198
	F1-score	0.6174	0.5383	0.5484	0.7538	0.7197

3.3. Scenario B: Per Electrode and Joint Bands

In an effort to improve the results of the classification methodology proposed by following common ML practices, it was decided to increase the features per epoch. Hence, as a second step, this scenario examines for each electrode all bands together presenting in total 30 features per epoch and per electrode. The accuracy results in general are provided in Table 6, followed by the detailed classification results in Table 7.

Table 6 Accuracy results for Scenario B

Electrode	All Bands	Electrode	All Bands
Fp1	0.8272	F7	0.8165
Fp2	0.8209	F8	0.7945
T3	0.8026	P3	0.8290
T4	0.8209	P4	0.8178
T5	0.7950	O1	0.8250
T6	0.8192	O2	0.8142
C3	0.8089	Fz	0.8491
C4	0.8115	Cz	0.8339
F3	**0.8523**	Pz	0.8509
F4	0.8357		

Table 7 Detailed classification results for Scenario B

Electrode	Metric	All Bands	Electrode	Metric	All Bands
Fp1	Precision	0.5801	F7	Precision	0.7308
	Recall	0.5792		Recall	0.7296
	F1-score	0.5789		F1-score	0.7299
Fp2	Precision	0.8207	F8	Precision	0.7945
	Recall	0.8214		Recall	0.7947
	F1-score	0.8209		F1-score	0.7945
T3	Precision	0.8028	P3	Precision	0.8290
	Recall	0.8040		Recall	0.8295
	F1-score	0.8026		F1-score	0.8290
T4	Precision	0.8208	P4	Precision	0.8178
	Recall	0.8210		Recall	0.8183
	F1-score	0.8209		F1-score	0.8178
T5	Precision	0.7949	O1	Precision	0.8248
	Recall	0.7952		Recall	0.8251
	F1-score	0.7950		F1-score	0.8250
T6	Precision	0.8191	O2	Precision	0.8142
	Recall	0.8191		Recall	0.8144
	F1-score	0.8192		F1-score	0.8142
C3	Precision	0.8089	Fz	Precision	0.8490
	Recall	0.8095		Recall	0.8492
	F1-score	0.8089		F1-score	0.8492
C4	Precision	0.8116	Cz	Precision	0.8338
	Recall	0.8130		Recall	0.8344
	F1-score	0.8115		F1-score	0.8339
F3	Precision	**0.8523**	Pz	Precision	0.8510
	Recall	**0.8529**		Recall	0.8511
	F1-score	**0.8523**		F1-score	0.8509
F4	Precision	0.8358			
	Recall	0.8360			
	F1-score	0.8357			

It is easy to observe that combination of the extracted features can increase significantly the accuracy of the classification process. Compared with Scenario A, results across all electrodes have over 80% (except T5 that is slightly below at 79.5%). Furthermore, it can now been seen that the

most accurate outcome is identified in F3 (85.23%) and not in T4 (82.99%) as seen previously. This is quite interesting, taken into account that this is opposite both in terms of left and right but also front and back, and it does not agree with the best results of previous studies. Nevertheless, since the estimates obtained are quite close to each other, it implies that in the approach followed the results will be more or less similar no matter which type of electrode is examined. These findings support previous studies, where differences have been identified to be widespread across the head, and not entirely focused on one specific region or electrode.

3.4. Scenario C: Per Band and Joint Electrodes

In an effort to increase the features even more for the third scenario the feature selection was performed in reverse order from Scenario B. For each band all the electrodes were merged together presenting in total 114 (19 electrodes x 6 measures) features per epoch and per band. The results for this scenario are all provided in Table 8.

Table 8 Classification results for Scenario C

Electrode	Metric	*Whole*	*Alpha*	*Beta*	*Delta*	*Theta*
All	Accuracy	0.9651	0.8885	0.9409	**0.9776**	0.9570
	Precision	0.9651	0.8893	0.9410	**0.9776**	0.9571
	Recall	0.9651	0.8885	0.9409	**0.9776**	0.9570
	F1-score	0.9651	0.8886	0.9409	**0.9776**	0.9570

In the results of this scenario, and in contrast to Scenario A, we can observe that the best accuracy is on the *Delta* band (97.76%) and not on *Theta* (even though quite close with 95.70%). Interestingly enough, the second best performance is observed for the *Whole* band (96.51%), which may lead to the assumption, that through such combination of features, the exploration on individual bands may not be needed. The results obtained through this scenario, are significantly higher than the previous two, at least by 10-12%.

Even though this specific scenario has not be found in previous studies (which also applies for scenario B), the results obtained are quite close or better from most similar findings, leading to the assumption, that the combination of electrodes (and bands) in the form of features for training an ML model, can introduce better results, than exploring each one of them individually, when trying to diagnose a subject. However, the drawback of these approaches (as well as the next Scenario) is that they do not allow extracting meaningful information towards better understanding the pathology behind the decease.

3.5. Scenario D: Joint Bands and Electrodes

Subsequently to the previous scenarios, as one would expect, the final step would be to join all bands and electrodes in one complete set of features, providing for each epoch a total of 570 features. The obtained results are very promising with the classification accuracy observed reaching the remarkable level of 98.03%, as can be seen in Table 4. These are extremely promising findings, which extend previous research results by differentiating successfully not only Healthy versus (probable) AD epochs, but also MCI simultaneously.

Table 9 Classification results for Scenario D

Electrode	Metric	All Bands
All	Accuracy	0.9803
	Precision	0.9803
	Recall	0.9803
	F1-score	0.9803

3.6. Scenario E: Joint Bands and Electrodes - Leave one out

Even though the above evaluation approach is most commonly found in literature, it can be considered non-practical (or performing over-training), as it cannot be applied to a new unclassified subject and its epochs. Hence, this scenario extends a bit the methodology explored in this work, following the fact that the most promising results have been observed when combining all features together (i.e. Scenario D), and performs a *leave-one-out* evaluation. This means that for 75 iterations (the total number of our subjects) the LGBM model is trained utilizing 74 subjects each time and then tested on the 1 subject left outside the training set. In each iteration, a different subject is left out, so

that all possible combinations are tested. The average results acquired are depicted in Table 10.

Table 10 Classification results for Scenario E

Electrode	Metric	All Bands
All	Accuracy	0.4319
	Precision	0.4365
	Recall	0.4329
	F1-score	0.4342

It is evident that the accuracy drops drastically to just 43.19%, which means that in average the model cannot distinguish among the three classes. However, as can be seen in Table 2 there is wide distribution for each class, with the majority of subjects not being able to be classified accurately. Nevertheless, there are also quite a few that have over 75% accuracy, which leads to the assumption that there is potential in the methodology presented, but perhaps a different approach (either in terms of features or modelling) is required. Interestingly enough, 10 of the subjects have 0% accuracy (which means that none of the epochs were classified correctly), whereas 8 of them have 100% accuracy (which means that all of the epochs were classified correctly). Both these cases are found distributed among all three classes. Such diversity in the results, leads to the conclusion that more thorough research is needed for delivering concrete evidence on whether such models can effectively work on a completely new subject.

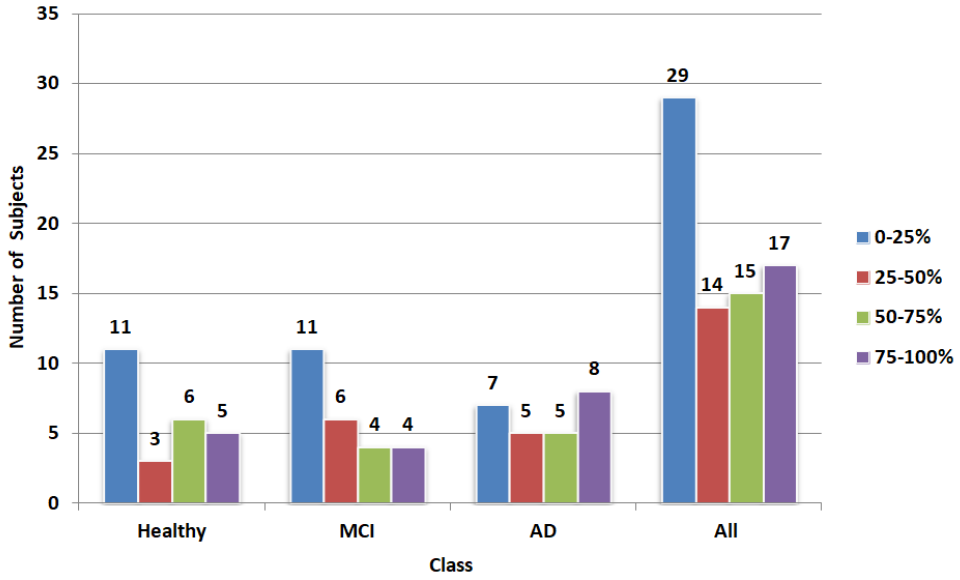

Figure 2. Distribution of Accuracy results per Subject in the Leave-one-out Scenario. The accuracy range refers to the percentage of accurately classified epochs of each subject (e.g. classifying correctly 21 out 30 epochs of one subject means 70% accuracy for that subject

3.7. Scenario F: Joint Bands and Electrodes - Leave one out – Importance selection

In an effort to improve the above findings for the *leave-out-out* scenario, the importance of the 570 features has been explored, using once more the inherent functionalities of the LGBM library. Using a trial-and-error approach, the authors identified that the best accuracy is achieved if features that participate with over 0.6% are included in the model (the full list of the selected features is presented in Table 11. Using only these features, the accuracy rises to approximately 52.5%, which is better than using all 570 features. As shown in Table 12, if we go lower (0.5%) or higher (0.7%) for the importance threshold, the accuracy drops once again below 50%.

Interestingly enough, in terms of the initial features, most of the selected ones are based on the *Mean* value, followed by *HFD* and *Var*. Hence, it seems that *HFD* (which also holds the highest score in T3) is generally more important than *TsEn*, whereas for such models, it may be sufficient to work with simpler features. In terms of bands, once again, the majority is identified on the *Whole* band, which

enforces even more the assumption, that in ML approaches, the analysis of the entire signal could be sufficient for classifying towards the correct direction. On top of that, we can see that even though the *Delta* band, which has been supported by the literature as the most information-rich for non-linear analysis, has been prominent in the previous scenarios of this study, here there aren't any features originating from that band. Finally, we can see that the frontal electrodes play a more important role than other regions.

Table 11 Features with importance over 0.6%

Feature	Band	Electrode	Average Importance (%)
Mean	Whole	Fp1	0.8271
Mean	Whole	F4	0.6267
Var	Whole	Pz	0.6355
HFD	Whole	F7	0.6559
Var	Beta	O2	0.6764
Var	Beta	Fp2	0.6892
HFD	Beta	Fp1	0.6973
Var	Theta	Fp1	0.7039
Mean	Whole	F8	0.721
Mean	Whole	Pz	0.7797
Mean	Whole	Fz	0.8333
HFD	Whole	C3	0.8631
HFD	Whole	F4	0.8746
Mean	Whole	T4	0.8853
Mean	Whole	Cz	0.956
Mean	Whole	F7	1.0173
Mean	Whole	T6	1.0684
Mean	Whole	F3	1.2263
HFD	Alpha	T3	1.2921

Table 12 Classification Results for joint bands and electrodes - Leave One Out - Most important Features

Electrode	Metrics	All Bands (0.5%)	All Bands (0.6%)	All Bands (0.7%)
All	Accuracy	0.4946	**0.5249**	0.4673
	Precision	0.4968	0.5263	0.4702
	Recall	0.4960	0.5260	0.4687
	F1 Score	0.4961	0.5261	0.4688

Figure 3 shows the updated distribution of the 75 subjects in terms of accuracy range. As observed, the results are quite improved, with 9 less subjects in the range between 0-25%, and 7 more subjects in the range between 75-100%. Out of all this, only 4 of them remain with 0% accuracy (as described above), whereas 7 (one less) remain at 100%.

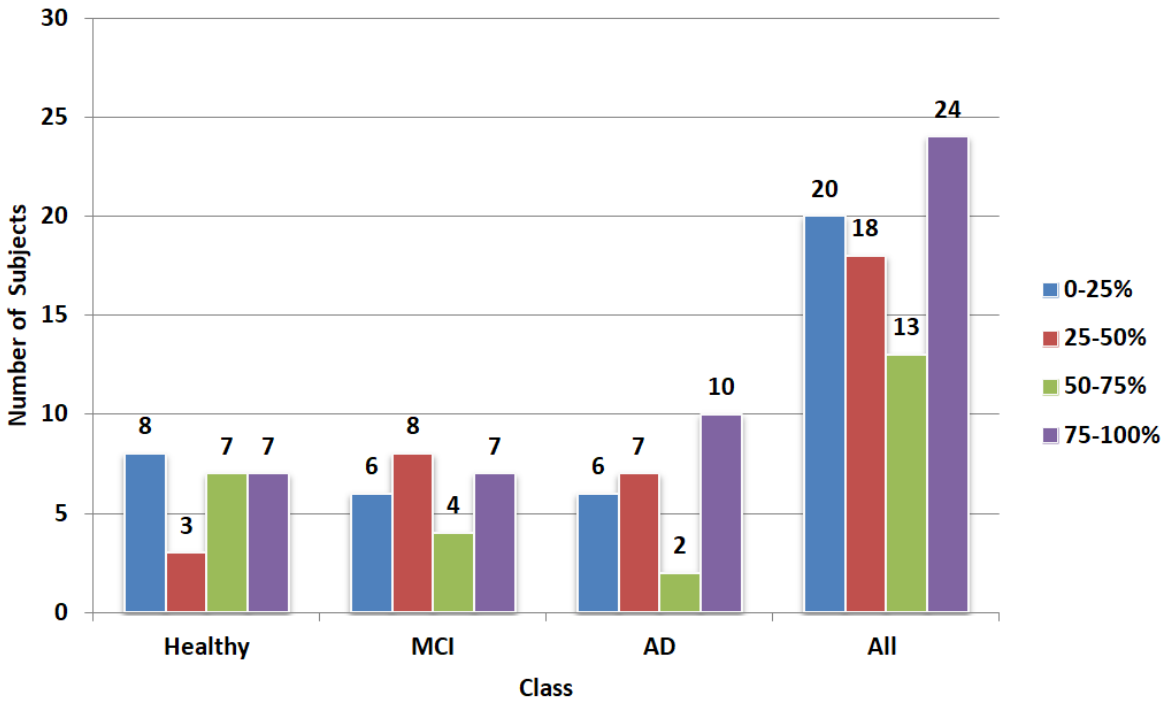

Figure 3. Distribution of Accuracy results per Subject in the Leave-one-out Scenario only with features over 0.6% importance.

4. Discussion

As has been discussed in the state-of-the-art in the introduction, there is an increasing interest in studying the potential of non-linear features of EEG signals for supporting the classification of healthy, MCI, and AD subjects. The most recent technological focus is on the rapidly emerging AI-driven techniques, and more specifically ML-based ones, which have been found to hold immense potential for the health domain, including signal processing in AD. In this article, the focus is on two specific non-linear features, i.e., the Tsallis Entropy and the Higuchi Fractal Dimension, along with other more common linear and non-linear variables such as mean, variance, skewness and kurtosis. Their combination has been examined over three well-known ML tools (i.e., SVM, BBT, and LGMB), and the overall performance in properly classifying subjects as healthy, MCI, and AD has been assessed under several scenarios. To the authors' knowledge, this is one of the first ML-based approaches that combines these specific features for addressing the challenge of classifying EEG signals (either at epoch or complete signal level) in one of three stages, without performing artefact removal. Regarding the main results obtained, a higher performance (>98%) from the respective literature (in terms of the specific features examined) has been observed when performing a 10-fold cross validation on 10 seconds epochs, on the entire sample pool, when combining all of the selected features across all 19 electrodes. Nevertheless, when examining a *leave-one-out* scenario, the accuracy of the model drops significantly, creating room for improvements and future research. A more detailed discussion, aligned with the results follows.

4.1. ML Technique

In the results presented in Section 3.1, it is evident that by employing different ML techniques and tools, a better performance can be achieved, using the same features and modelling methods. LGBM outperforms both SVM and GBT in almost all examined cases, with a sole exception of *Delta-O2*, in which GBT seems to be slightly better. Nevertheless, that particular combination is one of the least accurate combinations. Interestingly, the *Whole* band seems to outperform all other besides Theta, which holds the highest accuracy, for electrode *T4*, with 82.99%.

4.2. Feature Selection

It can be seen that the *TsEn* provides better results when used individually than the *HFD*. This could potentially mean that the entropy gives more information in terms of classifying AD than the fractality, an assumption that would require further and more in-depth research by also employing other types of fractality and entropy measures, as well as a larger dataset. Another interesting

outcome of this comparison is the slight difference in between bands. When investigating only *TsEn* in the separate bands, the one with the highest accuracy is *Theta*, whereas for *HFD* the band that offers the highest accuracy is *Delta*. What's even more interesting is that their combination exhibits the highest accuracy in *Delta*. These results suggest that complexity measures seem to provide the richest information for classification purposes in different bands. Nevertheless, when combined with the other four selected features, the accuracy improves significantly, reaching in two scenarios a value over 80%, leading to the assumption that both *TsEn* and *HFD* cannot capture optimally the differences among the three stages, using the proposed approach. This is something that contradicts previous findings in the literature, in which only one of them was enough to present highly accurate results.

Moreover, from the abovementioned comparison, it is observed that when exploring either *TsEn* or *HFD* individually, there is a large gap between individual bands and the entire signal (*Whole*), whereas when all six features are combined, it is clear that even if the analysis doesn't go into individual bands, but remains at the *Whole* level (entire signal), the results are quite close to their best accuracy. This is quite promising, as it can further reduce pre-processing, computation time and make the diagnosis faster, making it a more practical tool to healthcare professionals.

4.3. Classification towards accurate supported diagnosis

In the classification scenarios that have been presented in this manuscript, several interesting insights can be extracted and discussed. Starting with the first scenario (i.e., Scenario A) it can be observed that when investigating each electrode per band, the results are quite accurate (i.e., reaching up to 82.99%), but not as accurate as when combining multiple electrodes or bands, or even both electrodes and bands. From a band perspective, *Delta* and *Theta* seem to have the highest accuracy in the examined scenario, whereas considering the regional characteristics; the temporal electrodes offer in general higher values than other regions. Both of these findings are aligned with the literature analysed.

Getting in more detail, it is also interesting to observe the consistency of the results. Even though *Theta* holds the highest accuracy in this scenario, this is not observed for all electrodes. In fact, *Delta* has been found more accurate than *Theta* in 14 electrodes, whereas *Theta* just in 5. This implies that there is diversity in the bands and electrodes that offer valuable information, and research shouldn't focus on one or another independently but taking into account their interrelation. This extends to the other scenarios as well, within which it is easy to observe that combination of the extracted features (over both bands and electrodes) can increase significantly the accuracy of the classification process. This is even more evident through Scenario C, where the joint analysis of all electrodes leads to an increase of at least 10% in the accuracy, compared to the previous two scenarios examined, reaching up to 98.03% in Scenario D.

Following closely the scenarios and their results, the findings support previous studies, where differences have been identified to be widespread across the head, and not entirely focused on one specific region or electrode. Hence, it is imperative to identify the optimal combination of both the electrodes and bands to provide meaningful data towards effectively modelling an accurate classification process, as also clearly demonstrated in Scenario F.

Even though the above results are quite promising and follow a similar approach to previous studies in terms of epoch-level classification, it can be argued that including epochs of the same signal in the training and testing can affect the performance of the models. Nevertheless, even in this case, such as tool could support clinicians in their diagnosis, by allowing them to label epochs that are clearly belonging to one stage or another, and then seek "advice" for the remaining ones that are not so easy to decipher (always leaving the final decision to the clinician).

Going however one step ahead, and in an effort to avoid the abovementioned shortcoming of the analysis explored by the previous scenarios, the leave-one-out scenario has been explored, in which a whole signal is left outside the training process, and its epochs are tested independently. There, a significant drop in the accuracy leads to the conclusion that classifying a completely unlabelled signal remains quite challenging and requires more elaborate approaches. By evaluating the importance of the features explored, it is possible to improve the performance of the model, but again only to slightly go above 52%, which cannot be used by healthcare professionals on the field.

When taking a closer look to the signals and the performance of the model on them on the last scenario (i.e., Scenario F) it is interesting to observe the fact that there are signals that cannot be classified correctly at all (0%) and there are also signals that can be perfectly classified (100%). Hence, a closer look is required to these signals to better identify and understand the factors that

affect this behaviour, which may reveal additional modelling parameters that can increase the performance of the described approach.

5. Conclusions

This work presented a novel approach employing ML techniques (i.e., LGBM) for diagnosing Healthy, MCI, and probable AD subjects through EEG signals. The main merits of the presented methodology are five-fold: (i) slightly above 98% accuracy (with precision and recall/Sensitivity also slightly above 98%) is achieved when classifying an epoch in one of these three stages/classes in a balanced sample pool; (ii) the high accuracy achieved was on epochs originating from a very simple EEG acquisition protocol (5 minutes eyes closed in resting stage), (iii) the approach dispenses with laborious artefact-related pre-processing (completely damaged signals were not included at all in the analysis); and finally, (iv) different types of complexity features perform better when combined as features in the same classification algorithm. The results acquired are on par with other findings in the literature, however with quite a simpler methodology for the recording and the pre-processing procedures.

Furthermore, identifying the limitation of over training, a comparison with a *leave-one-out* approach has shown that the best model created, can only reach an average accuracy of 52.5%, which makes clear the fact that there is level of uniqueness in EEG recordings, which cannot be easily captured by ML models, at least not with so few samples and the features examined in the present manuscript. A lot more work is required, in order to effectively classify epochs from a new subject, without any previous knowledge, besides the signal itself. A promising yet not fully exploited approach would be the use of Deep Learning engineering, along with an automatic feature extraction, as presented recently Ieracitano et al.,[58] which however requires a lot more sophisticated data science.

6. Limitations

Finally, even though very interesting findings were obtained, within this study, there are several other extra parameters that could aid towards a better understanding and further improvement of the approach. By adding extra complexity features in the analysis and cross validating with other additional information such as demographics (age, sex, etc.) and neuro-psychometric tests (MMSE, FUCAS, etc.) more reliable results may be extracted, while increasing the accuracy of the models employed. Another limitation of this study, which also characterised most other related literature findings, is the sample size. Towards that direction, effort is already denoted in order to be able to include additional samples and data in the future work, leading to a statistically significant sample size, taking into consideration the population targeted.

Conflicts of Interest Statement

None

Funding Statement

This work has received funding by the European Union and the Greek National Strategic Reference Framework (NSRF) 2014-2020, under the call Researchers' Support with Emphasis on Young Researchers "ΕΔΒΜ34".

Acknowledgment

We would also like to show our gratitude to the Greek Association of Alzheimer's Disease and Related Disorders (Alzheimer Hellas) for providing us with the diagnosed anonymised EEG signals.

References

1. 2022 Alzheimer's Disease facts and figures. Alzheimer's Association. https://www.alz.org/media/documents/alzheimers-facts-and-figures.pdf. Accessed July 25, 2022.
2. Inside the brain. Alzheimer's Association. https://www.alz.org/alzheimers-dementia/what-is-alzheimers/brain_tour_part_2. Accessed July 25, 2022
3. Kumar A, Singh A, Ekavali. A review on Alzheimer's disease pathophysiology and its management: an update. *Pharmacological reports*. 2015;*67*(2):195-203.
4. Petersen RC, Lopez O, Armstrong MJ, Getchius TSD, Ganguli M, Gloss D, Gronseth GS, Marson D, Pringsheim T, Day GS, Sager M, Stevens J, Rae-Grant A. Practice guideline update summary: Mild cognitive impairment: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology. *Neurology*, 2018;90(3):126-135.
5. McKhann GM, Knopman DS, Chertkow H, Hyman, BT, Jack Jr, CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 2011;*7*(3):263-269.
6. Blennow K, Zetterberg H. Biomarkers for Alzheimer's disease: current status and prospects for the future. *Journal of Internal Medicine*, 2018;284(6):643-663.
7. Weiner H, Schuster DB. The electroencephalogram in dementia. - Some preliminary observations and correlations. *Electroencephalography and clinical neurophysiology,* 1956;8(3):479-488.
8. Berger H. On the electroencephalogram of man. Twelfth report. *Electroencephalography and clinical neurophysiology*, 1969;267+.
9. Jeong J. EEG dynamics in patients with Alzheimer's disease. *Clinical neurophysiology*, 2004;*115*(7):1490-1505.
10. Dauwels J, Vialatte F, Cichocki A. Diagnosis of Alzheimer's disease from EEG signals: where are we standing?. *Current Alzheimer Research*, 2010;*7*(6):487-505.
11. Alberdi A, Aztiria A, Basarab A. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial intelligence in medicine*, 2016;*71*:1-29.
12. Houmani N, Vialatte F, Gallego-Jutglà E, Dreyfus G, Nguyen-Michel VH, Mariani J, Kinugawaet K. Diagnosis of Alzheimer's disease with Electroencephalography in a differential framework. *PloS one*, 2018;13(3):e0193607.
13. Tsolaki A, Kazis D, Kompatsiaris I, Kosmidou V, Tsolaki M. Electroencephalogram and Alzheimer's disease: clinical and research approaches. *International journal of Alzheimer's disease*, *2014*.
14. Jeong J. Nonlinear dynamics of EEG in Alzheimer's disease. *Drug development research*, 2002;*56*(2):57-66.
15. Woyshville MJ, Calabrese JR. Quantification of occipital EEG changes in Alzheimer's disease utilizing a new metric: the fractal dimension. *Biological psychiatry*, 1994;*35*(6):381-387.
16. Besthorn C, Zerfass R, Geiger-Kabisch C, Sattel H, Daniel S, Schreiter-Gasser U, Förstl H. Discrimination of Alzheimer's disease and normal aging by EEG data. *Electroencephalography and Clinical Neurophysiology*, 1997;*103*(2):241-248.
17. Jeong J, Kim SY, Han SH. Non-linear dynamical analysis of the eeg in alzheimer's disease with optimal embedding dimension, *Electroencephalography and clinical Neurophysiology,* 1998;106:220–228.
18. Jelles B, Van Birgelen JH, Slaets JPJ, Hekster REM, Jonkman EJ, Stam CJ. Decrease of non-linear structure in the EEG of Alzheimer patients compared to healthy controls. *Clinical Neurophysiology*, 1999;*110*(7):1159-1167.
19. Yagyu T, Wackermann J, Shigeta M, Jelic V, Kinoshita T, Kochi K, Julin P, Almkvist O, Wahlund LO, Kondakor I, Lehmann D. Global dimensional complexity of multichannel EEG in mild Alzheimer's disease and age-matched cohorts. *Dementia and geriatric cognitive disorders*, 1997;8(6):343-347.
20. Stam CJ, Jelles B, Achtereekte HAM, Rombouts, SARB, Slaets, JPJ, Keunen RWM. Investigation of EEG non-linearity in dementia and Parkinson's disease. *Electroencephalography and clinical neurophysiology*, 1995;*95*(5):309-317.
21. Yang S, Bornot JMS, Wong-Lin K, Prasad G. M/EEG-based bio-markers to predict the MCI and Alzheimer's disease: a review from the ML perspective. *IEEE Transactions on Biomedical Engineering*, 2019;66(10):2924-2935.
22. Tanveer M, Richhariya B, Khan RU, Rashid AH, Khanna P, Prasad M, Lin CT. Machine learning

techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2020;*16*(1s):1-35.
23. Tsallis C. Possible generalization of Boltzmann-Gibbs statistics. *Journal of statistical physics*, 1988;*52*(1):479-487.
24. Gamero LG, Plastino A, Torres ME. Wavelet analysis and nonlinear dynamics in a nonextensive setting. *Physica A: Statistical Mechanics and its Applications*, 1997;*246*(3-4):487-509.
25. Capurro A, Diambra L, Lorenzo D, Macadar O, Martin MT, Mostaccio C, Plastino A, Pérez J, Rofman E, Torres ME, Velluti J. Human brain dynamics: the analysis of EEG signals with Tsallis information measure. *Physica A: Statistical Mechanics and its Applications*, 1999;265(1-2):235-254.
26. Martin MT, Plastino AR, Plastino A. Tsallis-like information measures and the analysis of complex signals. *Physica A: Statistical Mechanics and its Applications*, 2000;*275*(1-2):262-271.
27. Thakor NV, Paul J, Tong S, Zhu Y, Bezerianos A. Entropy of brain rhythms: normal versus injury EEG. In *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing (Cat. No. 01TH8563)*, 2001:261-264.
28. Thakor NV, Tong S. Advances in quantitative electroencephalogram analysis methods. *Annu. Rev. Biomed. Eng.*, 2004;*6*:453-495.
29. Sneddon R. The Tsallis entropy of natural information. *Physica A: Statistical Mechanics and its Applications*, 2007;*386*(1):101-118.
30. Zhao P, Van-Eetvelt P, Goh C, Hudson N, Wimalaratna S, Ifeachor E. Characterization of EEGs in Alzheimer's disease using information theoretic methods. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007:5127-5131.
31. Zhao P, Van Eetvelt P, Goh C, Hudon N, Wimalaratna S, Ifeachor E. C. EEG markers of alzheimer's disease using tsallis entropy. In *Proceedings of the 3rd International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007), Plymouth, UK*, 2007;27.
32. De Bock TJ, Das S, Mohsin M, Munro NB, Hively LM, Jiang Y, Smith CD, Wekstein DR, Jicha GA, Lawson A, Lianekhammy J, Walsh E, Kiser S, Black C. Early detection of Alzheimer's disease using nonlinear analysis of EEG via Tsallis entropy. In *2010 Biomedical Sciences and Engineering Conference*, 2010:1-4.
33. McBride J, Zhao X, Munro N, Jicha G, Smith C, Jiang Y. EEG multiscale entropy dynamics in mild cognitive impairment and early Alzheimer's disease. In *Proceedings of the 2014 Biomedical Sciences and Engineering Conference*, 2014:1-4.
34. Garn H, Waser M, Deistler M, Benke T, Dal-Bianco P, Ransmayr G, Schmidt H, Sanin G, Santer P, Caravias G, Seiler S, Grossegger D, Fruehwirt W, Schmidt R. Quantitative EEG markers relate to Alzheimer's disease severity in the Prospective Dementia Registry Austria (PRODEM). *Clinical Neurophysiology*, 2015;126(3):505-513.
35. Al-Nuaimi AH, Jammeh E, Sun L, Ifeachor E. Tsallis entropy as a biomarker for detection of Alzheimer's disease. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2015:4166-4169.
36. Al-Nuaimi AHH, Jammeh E, Sun L, Ifeachor E. Complexity measures for quantifying changes in electroencephalogram in Alzheimer's disease. *Complexity*, 2018.
37. Tzimourta KD, Afrantou T, Ioannidis P, Karatzikoy M, Tzallas AT, Giannakeas N, Astrakas LG, Angelidis P, Glavas E, Grigoriadis N, Tsalikakis DG, Tsipouras MG. Analysis of electroencephalographic signals complexity regarding Alzheimer's Disease. *Computers & Electrical Engineering*, 2019;*76*:198-212.
38. Campbell P, Abhyankar S. Fractals, form, chance and dimension, 1978.
39. Mandelbrot BB. The fractal geometry of nature/Revised and enlarged edition, New York, WH Free. Co, 1983;*495*:1.
40. Besthorn C, Sattel H, Geiger-Kabisch C, Zerfass R, Förstl H. Parameters of EEG dimensional complexity in Alzheimer's disease. *Electroencephalography and clinical* disease. In *2011 Annual International Conference of the IEEE Engineering in neurophysiology*, 1995;*95*(2):84-89.
41. Pritchard WS, Duke DW. Measuring chaos in the brain-a tutorial review of EEG dimension estimation. *Brain and cognition*, 1995;*27*(3):353-397.
42. Accardo A, Affinito M, Carrozzi M, Bouquet F. Use of the fractal dimension for the analysis of electroencephalographic time series. *Biological cybernetics*, 1997;*77*(5):339-350.
43. Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, 1988;*31*(2):277-283.
44. Henderson G, Ifeachor E, Hudson N, Goh C, Outram N, Wimalaratna S, Del Percio C,

Vecchio F. Development and assessment of methods for detecting dementia using the human electroencephalogram. *IEEE Transactions on Biomedical Engineering*, 2006;53(8):1557-1568.

45. Staudinger T, Polikar R. Analysis of complexity based EEG features for the diagnosis of Alzheimer's. *Medicine and Biology Society*, 2011:2033-2036).
46. Smits FM, Porcaro C, Cottone C, Cancelli A, Rossini PM, Tecchio F. Electroencephalographic fractal dimension in healthy ageing and Alzheimer's disease. *PloS one*, 2016;*11*(2):e0149587.
47. Al-Nuaimi AH, Jammeh E, Sun L, Ifeachor E. Higuchi fractal dimension of the electroencephalogram as a biomarker for early detection of Alzheimer's disease. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017:2320-2324.
48. Tsolakis AC, Petsos G, Kapetanou O, Nikolaidis IN, Aifantis, EC. Model analogies between pattern formation in deforming engineering materials & morphogenesis in ageing human brains. *Journal of the Mechanical Behavior of Materials*, 2019;*28*(1):95-106.
49. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 1975;*12*(3):189-198.
50. Kounti F, Tsolaki M, Kiosseoglou G. Functional cognitive assessment scale (FUCAS): a new scale to assess executive cognitive function in daily life activities in patients with dementia and mild cognitive impairment. *Human Psychopharmacology: Clinical and Experimental*, 2006;*21*(5):305-311.
51. Jasper HH. The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, 1958;*10*:370-375.
52. Cukic M, Pokrajac D, Stokic M, Radivojevic V, Ljubisavljevic M. EEG machine learning with Higuchi fractal dimension and Sample Entropy as features for successful detection of depression. *arXiv preprint arXiv:1803.05985*, 2018.
53. Tian Y, Shi Y, Liu X. Recent advances on support vector machines research. *Technological and economic development of Economy*, 2012;*18*(1):5-33.
54. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016:785-794.
55. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. *Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems*, 2017;30.
56. Sokolova M, Japkowicz N, Szpakowicz S. Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation. *In Australasian joint conference on artificial intelligence*. 2006:1015-1021.
57. Tharwat A. Classification assessment methods. *Applied Computing and Informatics*, 2021;*17*(1):168-192.
58. Ieracitano C, Mammone N, Hussain A, Morabito FC. A novel multi-modal machine learning based approach for automatic classification of EEG recordings in dementia. *Neural Networks*, 2020;123:176-190.