Machine Learning Approach for Dementia Classification using EEG Signal Analysis

Rageshri Bakare[1], Dr. Virendra Shete[2], Dr Magda Tsolaki[3], Dr Ioannis Kompatsiaris[4]

[1] MIT SOES, MITADT University, Pune, India
[2] Director, MIT SOES, MITADT University, Pune, India
[3] Professor of Neurology, MD, PhD, GAADRD, Thessaloniki, Greece
[4] Researcher A, CERTH -ITI, Thessaloniki, Greece

***Abstract* -Alzheimer's disease belongs to the most quintessential types of dementia and it counts a huge percentage of dementia cases as more than 60 percent. Diagnosing dementia definitively and early is a huge challenge. This study's objective is to identify Alzheimer's Disease during its earliest stages like subjective cognitive decline and Mild cognitive impairment using EEG signal analysis.**

Methods: Data is provided by CRETH, includes the sample EEG signal recordings of 48 AD-patients, 79 MCI-patients, 34 SCD-patients, and 33 HC's. The EEG signals are analyzed by extracting features using FFT, CWT and PSD techniques and Machine learning algorithms like, k-Nearest Neighbor, Neural networks, Support vector machine and Random Forest , are used as classifiers.

Results: Receiver-operating characteristic analysis of the EEG results 90%for HC ,81% for SCD, 90% for MCI and 89% for AD when CWT feature extraction for Beta band using KNN classifier is used corresponding to a sensitivity of 95%, specificity of 84%, and F1-score of 84%. Using PSD feature extraction of beta band with KNN classifier yielded 94 % of specificity, but sensitivity of 80% and F1-score of 81% is observed.

Keywords-*EEG, Dementia, Alzheimer's Disease, FFT, CWT, PSD, Machine learning*

I. INTRODUCTION AND RELATED LITERATURE

World Health Organization (WHO) anticipates that 1.2 billion individuals worldwide have dementia, with more than sixty percent coming from lower- and middle-income countries. 78 million in 2030 and 139 million in 2050 can be affected by this disorder. [1] One of the main causes of incapacitated and dependent older individuals globally, it is currently the seventh most prevalent killer among all diseases. The condition is a slowly developing ailment. Physical alterations to the brain, including protein buildup and nerve damage, are what cause it to happen. [11] Hence, the research in area of neurological disorders like dementia is of great worth. The dementia associated with Alzheimer's has a propensity to take years to develop and progressively get worse. Alzheimer's disease (AD) progressively impacts the majority of brain regions causing. It causes effect on mobility, temperament, reasoning, thinking, language, real concern, and remembrance and many such activities of a being which increases progressively.[1]

Although the precise pathophysiology of AD is still unknown, related pathology ideas suggest that amyloid plaques, neurofibrillary tangles, and synapse loss may all play a role [3]. It is estimated that the pathophysiological processes of AD begin for as long as twenty years before clinical symptoms might be identified [4]. There have been no viable treatment options to prevent the advancement of AD and repair this ailment. [4].

Subjective Cognitive Decline (SCD), which was devised to encompass an earlier stage of AD before Mild Cognitive Impairment (MCI) and was defined as conscience cognitive decline before the abnormalities are apparent on cognitive tests. [8,12] Between the predicted cognitive loss associated with normal ageing and the more severe cognitive decline associated with dementia, there is a stage called MCI. It entails issues with memory, language, reasoning, and judgement in addition to the normal aging. The onset of AD may be delayed if MCI is picked up early. The first lineament of the AD triad SCD, which is a conscience decrease in cognitive function without evidence of actual cognitive impairment [9, 12]. A conceptual framework for research on subjective cognitive decline in preclinical AD is discussed by the

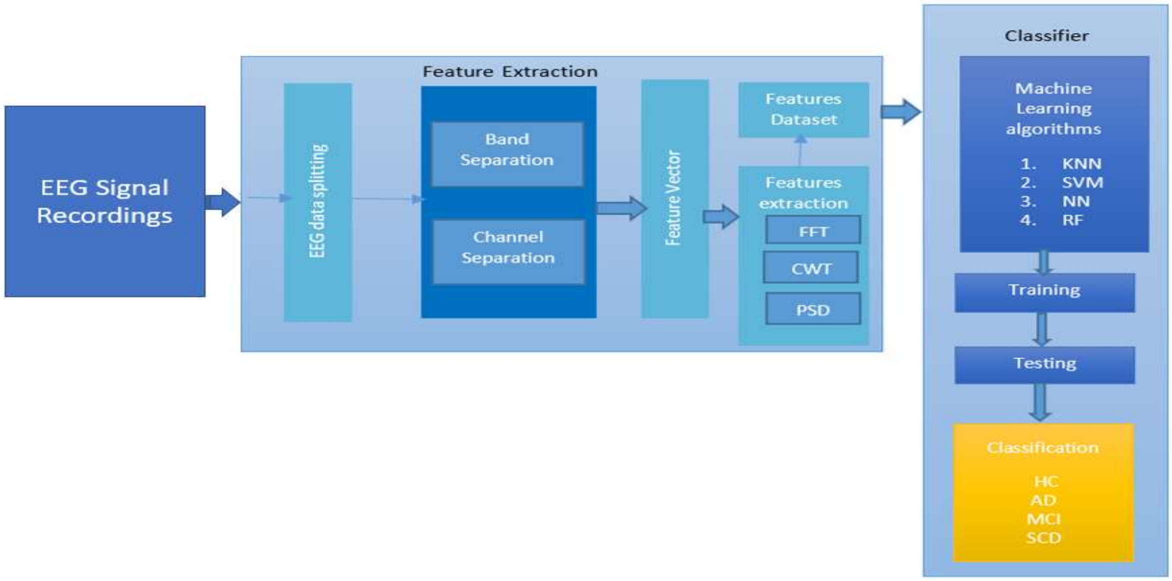

Fig 1: Block diagram of proposed system for dementia early-stage classification using ML techniques

researchers. Studies conducted give testimony that the likelihood for mild cognitive impairment and dementia is greater among those with SCD. [10].

Positron Emission Tomography-(PET) and Single Photon Emission Computed Tomography-(SPECT) are modalities for detecting alterations in brain function and physiology that have been used in some clinical testing accompanying blood biomarkers [1]. In some cases, structural and functional Magnetic Resonance Imaging-(MRI), and to some extent, Computed Tomography-(CT) is used too. [11]

With its capacity to quantify changes in brain atrophy in dementia, Electroencephalography (EEG) is becoming more and more prevalent as a means of determining the presence of alterations caused by brain activities for the clinical assessment. [13] The EEG recordings is an upcoming explored research area for diagnosing dementia. EEG recording abnormalities are studied to determine various types of dementia.[6] Similar to other methods, the goal is to achieve earlier diagnosis with EEG, but one of the advantages that can be considered for dementia is non-invasive and less expensive compared to various other modalities.

It is possible to perform time domain, frequency domain, and time-frequency domain analysis using a multitude of feature extraction techniques. In this work, the FFT, CWT, and PSD techniques are experimented.

Machine Learning-(ML) is a discipline of artificial intelligence that integrates algorithms and the ability to comprehend hidden patterns and behaviors in enormous quantities of information. [6].

I. Proposed Research Methodology:

The research paper aims to analyze the EEG signal recordings using time and time-frequency feature extraction techniques followed by various ML classifiers to classify into categories of dementia namely AD, MCI, SCD and Healthy Controls. The proposed system is represented in Fig 1. The next subsections address the specifics of this proposed system.

A. EEG dataset:

The dementia clinic of the Greek Association of Alzheimer's Disease and Associated Disorders (GAADRD) provided the participants and the set required for the diagnosis. And the overall dataset was provided by Centre for Research and Technology Hellas (CERTH), Information Technologies Institute (ITI).

GAADRD contributed the dataset of resting-state EEG recordings for 48 recordings of Alzheimer's Disease

(AD), 79 recordings of Moderate Cognitive Impairment (MCI), 34 recordings of Subjective Cognitive Decline (SCD), and 33 recordings of Healthy Controls (HC) using the HD-EEG EGI GES 300. (GAADRD). The EEG recording data is pre-collected for the patients in Resting-State High-Density EEG using EGI GES 300 with 256 Channels (GES 300, CERTH-ITI, Thessaloniki, Greece). Standard procedure is adopted by the experts. Sampling rate of 250 Hz, With AFz serving as the ground electrode and the electrode impedance being less than 50 k, the signals were recorded in relation to a vertex reference electrode (Cz). Using the software Net Station 4.3, the HD-EEG data were pre-processed (filtered, segmented, and replaced with faulty channels) in order to detect any artefacts (EGI). A 5th-order bandpass Butterworth IIR filter with a 0.3–30 Hz frequency range was initially used to filter HD-EEG data.[2]

Delta ranging from 1 – 4 Hz, theta ranging from 4–8 Hz, Alpha ranging from 8–13 Hz, Beta ranging from 13–30 Hz and Gamma >30 Hz are the five sub-frequency bands that make up the brainwave. These frequency bands are analyzed for the detection of early stages of dementia.

The standard 20 channels are selected as per 10-20 channel mapping system for the analysis. [5,7]

B. Feature extraction

The process for obtaining elements from EEG signals is referred to as feature extraction. To make content categorization smoother, attributes should be unique and autonomous.[2] The primary technique approach for one-dimensional signals in the frequency or time-frequency domain are the Fourier transform (FT) and Wavelets. The nature of EEG signals is stochastic and heterogeneous. A good alternative to remedying this issue is the wavelet transform (WT), which collapses a signal into a collection of functions (wavelets) with distinct frequencies and bounded durations. [15]

A highly efficient version of the Fast Fourier Transform (FFT) is Discrete Fourier Transform (DFT). By employing the DFT transform, the frequency content of a succession of sampled data (a signal) can be estimated.

$$FFT(x) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{\left(-\frac{2\pi ik}{N}\right)}$$

Hence, frequency domain representation of the signal is obtained. But FFT coefficients do not carry time information.

Wavelet are used for both time and frequency information. Since the translation and dilation factors (or coefficients) contribute extra details about the signal, the digital signal is regarded as the aggregate of substantially identical wavelets. Thus, wavelet coefficients can suffice as the depiction of a signal. These coefficients offer crucial temporal and frequency variables which can be utilized to assess a signal. Wavelets provide more reliable information by deriving the modest changing coefficients that users may be able to comprehend. Wavelets are the transforms used in the area of digital signal processing regularly.

$$CWT(\tau, s) = \frac{1}{\sqrt{|s|}}\int_{-\infty}^{\infty} x(t)\varphi^{*}\frac{t-\tau}{s}\,dt$$

Where, $\varphi\,(\tau, s)$ is the wavelet function and τ is the dilation parameter and s are the scaling parameter.

The estimated autocorrelation sequence produced by nonparametric procedures is given a Fourier transform, and this Fourier transform is used to compute the PSD. Welch's method is one of the generally used method. When the data sequence is used with data windowing, improved periodograms result [16].

Power spectral density (PSD) gives the details of the power levels of the frequency components present in a signal. It specifies the power of various frequencies present in the signal. Essentially, the PSD profile is a plot of the power over frequency.

C. Classifier

There are lots of different classification methods available. Some of the most popular techniques are K-Nearest-Neighbours-(KNN), Artificial Neural Networks-(ANNs), Support Vector Machines-(SVMs), and ensembles of classification trees like Random Forest-(RF) [17]

Machine learning algorithms-(MLAs)have drawn more attention as computer power has grown, and the calibre of signal processing algorithms has risen in tandem. Hence, the primary classifiers yielding high accuracy are reported in the majority of classification research to be RF, KNN, and SVM. [17]

Even so, the effectiveness of various classification techniques still heavily depends on the common characteristics of the entries to be classified.[17]. It is still unclear how exactly the facts that need to be categorised relate to various categorization systems' overall effectiveness. There is currently no classification technique that effectively solves any given issue. There have been a lot of problems with classification techniques used nowadays. Consequently, a trial-and-error strategy is utilised to determine the first-class classification approach for a certain dataset in order to determine the best performance.

1) Random Forest: A machine learning technique called ensemble techniques combines multiple base models to create a single, optimum forecasting model. Random Forest is a type of ensemble method. The key premise behind the random forest approach includes establishing lots of decision trees during training and applying a majority vote across them for classification.[14] Providing scores that earlier forecasters incorrectly anticipated more weight is known as boosting. The forecast is eventually decided by a weighted vote. By utilising bagging, every tree is separately constructed using a bootstrap sample of the data set; succeeding trees are independent of earlier trees. A simple majority vote ultimately determines the outcome of the forecast.[17].

Using the RF involves establishing the quantity of trees (ntree) as well as the number of attributes in each split. [18].

For best performance, more predictors necessitate more trees. To determine the number of trees needed, compare predictions from a forest to estimate from a portion of a forest. There are enough trees when the subgroups function just as effectively as the entire forest. [19]

2) k-Nearest Neighbour: Ever since early 1970s, the k-NN, non-parametric approach, has been used in statistical applications. [2]. The basic principle of k-NN is to pick the subset of k samples in the standardization dataset that are closest to the specimens. A response variables' average is derived (i.e., the class attributes of the k nearest neighbours), With these k samples, it is achievable to identify the label (class) of specimens. As the fundamental tuning parameter for this classifier, the k has quite a considerable influence on the k-NN's effectiveness.[2]

3) Support Vector Machine: The SVM classifier, a kernel-based supervised learning technique created for binary classification, separates data into two or more groups and is not advised for large numbers of training instances. A mapping approach used on a training collection to boost its similarity to a dataset that can be linearly segregated is alluded as a core function. By employing kernel functions, it is possible to effectively increase the density of the dataset using mapping. Quadratic, multilayer perceptron kernel, linear, RBF, and polynomial kernel are some instances of widely used kernel functions. Linear kernel functions are suitable for linearly separable data sets, while RBF kernel functions are suitable for non-linear data sets. Linear kernels contrast to RBF kernels, an SVM consumes shorter time to train. The linear kernel function also tends to be less vulnerable to overfitting than the RBF kernel function. [17]

4) Neural Networks: In the field of machine learning, neural networks, frequently known to as artificial neural networks (ANN), are one of the most commonly employed techniques for regression and classification modelling. [19] The use of neural networks for categorization has become increasingly significant. They can adapt to the data without any explicit specification of the workable or distributive form for the model parameters, neural networks are data-driven self-adaptive methodologies. As neural networks are non - linear models, they can easily represent the complicated interactions that occur in actual situations. [20] A way to think of a NN classifier is as a very huge number of connected basic processors that operate simultaneously [17].

One type of neural network that is often used is the multi-layered feed-forward perceptron, which has several layers of neurons coupled to one another. Nonlinear data can be segregated using the multi-layered perceptron, which often comprises three or more distinct layers. [17]

II. Experiment results:

Fig 1. shows the block schematic representation for the work proposed in this paper. The EEG signals database obtained is split into 4 seconds so as to achieve larger number of samples. Delta, theta, alpha and beta are the four frequency bands across which the data is segmented. The average of the 20 channels coefficients is taken for every band. This feature vector is applied to feature extraction techniques FFT, CWT and PSD respectively.

Features thus extracted are applied to each classifier k-NN, SVM, RF and NN individually. For each set, the performance metrics Sensitivity, Specificity, Precision, Loss, and F1-score are investigated.

Some of the sample observations are demonstrated in the Figures 3 (a-d) through 5 (a-d).

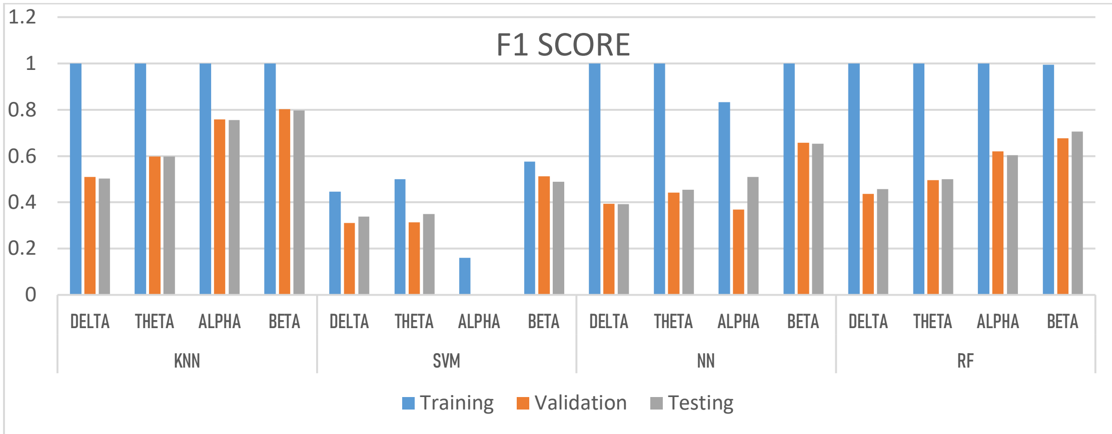

Fig. 2 Sample performance parameter F1-SCORE derived with FFT feature extraction, using KNN, SVM, NN and RF classifier for 4 Delta, Theta, Alpha and Beta frequency band

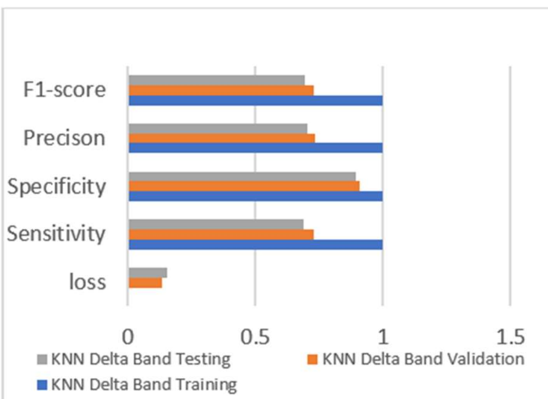

Fig 4.a F1-score, Precision, Specificity, Sensitivity and loss performance measures of Delta band, using CWT Feature extraction and KNN classifier

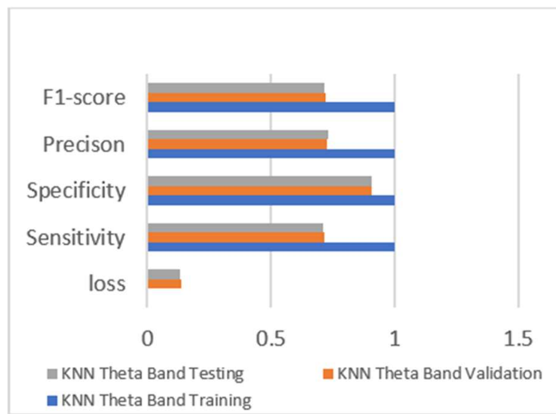

Fig 4.b F1-score, Precision, Specificity, Sensitivity and loss performance measures of Theta band, using CWT Feature extraction and KNN classifier

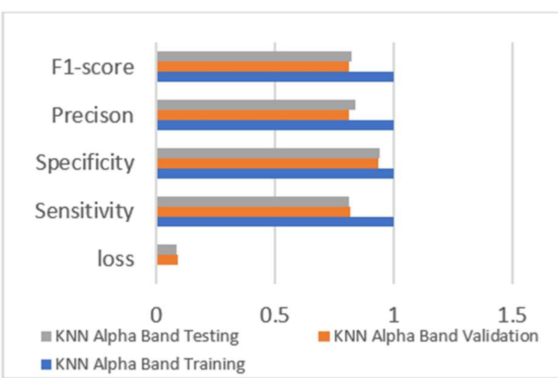

Fig 4.c F1-score, Precision, Specificity, Sensitivity and loss performance measures of Alpha band, using CWT Feature extraction and KNN classifier

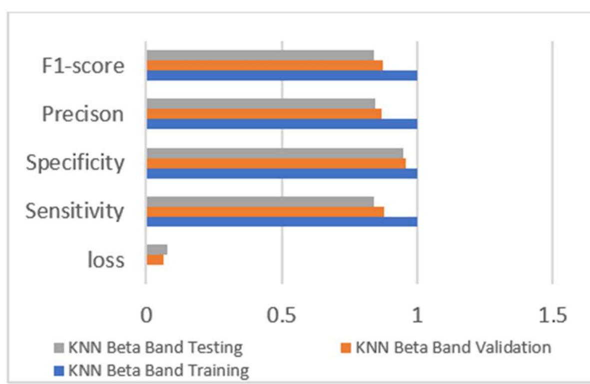

Fig 4.d F1-score, Precision, Specificity, Sensitivity and loss performance measures of Beta band, using CWT Feature extraction and KNN classifier

The Fig 2 shows the F1-score, given by each classifier namely k-NN, SVM, NN and RF, for each band of frequency respectively, with FFT feature extraction as a sample. Compared to all other combinations of feature extraction approaches used and all other classifier used in this work, k-NN classifier is seen to give F1-score of approximately 84 % for validation as well as testing stage of beta band, when CWT feature extraction is used.

The sample bar graphs for Sensitivity, Specificity, Precision, Loss and F1-score performance parameters, observed for the given dataset are shown in the Fig 3 to Fig 5. The bar graphs demonstrate the parameters derived by applying the k-NN classifier, to the features each extracted from FFT, CWT and PSD feature extraction technique. The dataset generated only includes data for the distinct frequency ranges delta, theta, alpha, and beta; gamma is not included.

It is observed that k-NN algorithm gives better specificity over other performance parameters. Also, 94% sensitivity is observed when CWT feature extraction with k-NN classifier is used.

Compared to FFT and PSD feature extraction technique, it is observed that sensitivity of 94 % and F1-score of 84.12% for beta band is given by k-NN classifier, when CWT feature extraction technique is used. It is also observed that specificity of 93% was given by Random Forest classifier, when PSD is applied to the beta band signals.

Receiver-operating characteristic analysis of the EEG results 90% for HC ,81% for SCD, 90% for MCI and 89% for AD when CWT feature extraction for Beta band using KNN classifier is used corresponding to a sensitivity of 95%, specificity of 84%, and F1-score of 84%. Using PSD feature extraction of beta band with KNN classifier

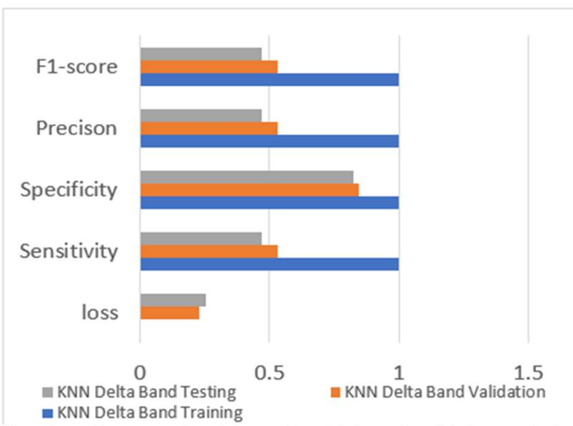

Fig 5.a F1-score, Precision, Specificity, Sensitivity and loss performance measures of Delta band, using PSD Feature extraction and KNN classifier

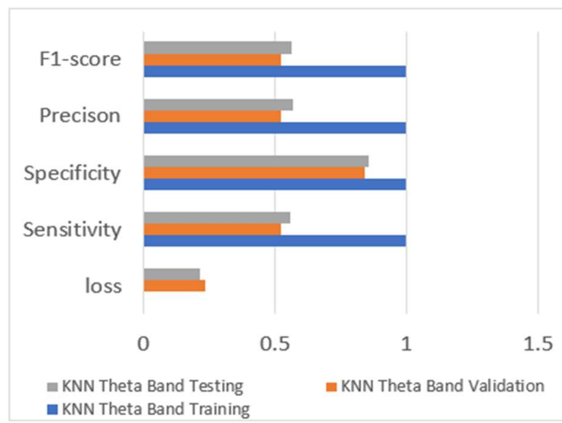

Fig 5.b F1-score, Precision, Specificity, Sensitivity and loss performance measures of Theta band, using PSD Feature extraction and KNN classifier

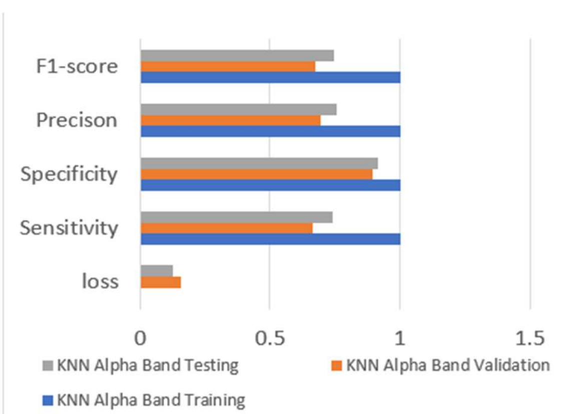

Fig 5.c F1-score, Precision, Specificity, Sensitivity and loss performance measures of Alpha band, using PSD Feature extraction and KNN classifier

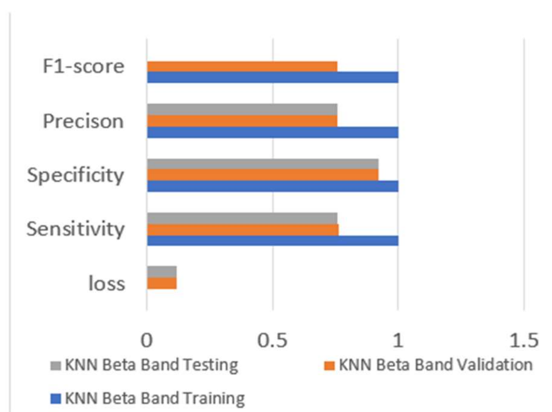

Fig 5.d F1-score, Precision, Specificity, Sensitivity and loss performance measures of Beta band, using PSD Feature extraction and KNN classifier

yielded 94 % of specificity, but sensitivity of 80% and F1-score of 81% is observed.

III. Discussion:

The rigorous experimentation was done for early-stage detection for dementia stages namely AD, MCI, SCD with respect to HC; by analysing EEG signals of these subjects, deriving various features by applying frequency domain and time-frequency domain techniques FFT, CWT and PSD feature extraction techniques individually. The extracted features are applied to well-known classifiers in Machine learning namely k-NN, NN, RF and SVM.

The observations indicate that each technique has unique benefits and drawbacks. The paper demonstrates the samples for results in terms of FI-score, Precision, Specificity, Sensitivity and loss when each of feature extraction technique is applied to individual frequency bands and KNN classifier is used. The sensitivity of 94 % and F1-score of 84.12% for beta band is given by k-NN classifier, when CWT feature extraction technique is used. It is also observed that specificity of 93% was given by Random Forest classifier, when PSD feature extraction is used for beta band signals.

Our further research aims to explore other feature extraction techniques like determining correlation coefficients, determining the coherence between various channels used for EEG signal recording. Also, various Deep learning algorithms supported by explainable AI methods to possibly improve the performance of the system.

Acknowledgment

The authors are grateful to Centre for Research and Technology Hellas - Information Technologies Institute (CERTH - ITI), Thessaloniki Greece and Greek Association of Alzheimer's Disease and Related Disorders (GAADRD), Thessaloniki Greece, for providing access to the dataset.

References

[1] http://www.who.int/news-room/factsheets/detail/dementia , 01 Jan 2023

[2] Lazarou, I., Georgiadis, K., Nikolopoulos, S., Oikonomou, V. P.Tsolaki, A., Kompatsiaris, I., Tsolaki, M., & Kugiumtzis, D. (2020*). A Novel Connectome-Based Electrophysiological Study of Subjective Cognitive Decline Related to Alzheimer's Disease by Using Resting-State High-Density EEG EGI GES 300. Brain sciences*, 10(6), 392. https://doi.org/10.3390/brainsci10060392

[3] Colom-Cadena, M., Spires-Jones, T., Synaptic Health Endpoints Working Group (2020) et al. *The clinical promise of biomarkers of synapse damage or loss in Alzheimer's disease. Alzheimer's research & therapy*, 12(1), 21. https://doi.org/10.1186/s13195-020-00588-4

[4] Sperling, R., Mormino, E., & Johnson, K. (2014). *The evolution of preclinical Alzheimer's disease: implications for prevention trials. Neuron,* 84(3), 608–622. https://doi.org/10.1016/j.neuron.2014.10.038

[5] Luu, Phan & Ferree, Thomas. (2000). *Determination of the Geodesic Sensor Nets' Average Electrode Positions and Their 10 – 10 International Equivalents*. Technical Note.

[6] Lazarou, I., Georgiadis, K., Nikolopoulos, S., Oikonomou, V. P., Stavropoulos, T. G., Tsolaki, A., Kompatsiaris, I., Tsolaki, M., & RADAR-AD Consortium (2022). *Exploring Network Properties Across Preclinical Stages of Alzheimer's Disease Using a Visual Short-Term Memory and Attention Task with High-Density Electroencephalography: A Brain-Connectome Neurophysiological Study*. Journal of Alzheimer's disease: JAD, 87(2), 643–664. https://doi.org/10.3233/JAD-215421

[7] Acharya, J. N., Hani, A. J., Thirumala, P. D., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 3: *A Proposal for Standard Montages to Be Used in Clinical EEG*. Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society, 33(4), 312–316. https://doi.org/10.1097/WNP.0000000000000317

[8] Cheng YW, Chen TF, Chiu MJ. *From mild cognitive impairment to subjective cognitive decline: conceptual and methodological evolution.* Neuropsychiatry Dis Treat. 2017; 13:491-498 https://doi.org/10.2147/NDT.S123428

[9] Rabin LA, Smart CM, Amariglio RE. *Subjective Cognitive Decline in Preclinical Alzheimer's Disease*. Annu Rev Clin Psychol. 2017;13:369–96

[10] Jessen F, Amariglio RE, Buckley RF, van der Flier WM, Han Y, Molinuevo JL, Rabin L, Rentz DM, Rodriguez-Gomez O, Saykin AJ, et al. *The characterisation of subjective cognitive decline*. Lancet Neurol. 2020;19: 271–8.

[11] Turner R. Scott, Stubbs Terry, Davies Don A., Albensi Benedict C., *Potential New Approaches for Diagnosis of Alzheimer's Disease and Related Dementias,* Frontiers in Neurology, VOLUME=11, year=2020, 10.3389/fneur.2020.00496, ISSN=1664-2295

[12] The Stages of Dementia: How Dementia Progresses (healthline.com)

[13] Al-Qazzaz, N. K., Ali, S. H., Ahmad, S. A., Chellappan, K., Islam, M. S., & Escudero, J. (2014). *Role of EEG as biomarker in the early detection and classification of dementia.* The Scientific World Journal, 2014, 906038. https://doi.org/10.1155/2014/906038

[14] Tanishq Abraham, Austin Todd, Daniel A. Orringer, Richard Levenson, Chapter 7 - *Applications of artificial intelligence for image enhancement in pathology*, Editor(s): Stanley Cohen, Artificial Intelligence and Deep Learning in Pathology, Elsevier, 2021, Pages 119-148, ISBN 9780323675383, https://doi.org/10.1016/B978-0-323-67538-3.00007-5.

[15] Faust O, Acharya RU, Allen AR, Lin CM. *Analysis of EEG signals during epileptic and alcoholic states using AR modelling techniques. IRBM.* 2008;29(1):44–52

[16] Al-Fahoum AS, Al-Fraihat AA. *Methods of EEG signal features extraction using linear analysis in frequency and time-frequency domains.* ISRN Neurosci. 2014 Feb 13; 2014:730218. Doi: 10.1155/2014/730218. PMID: 24967316; PMCID: PMC4045570.]

[17] Chen, W., Xie, X.S., Wang, J.L., Pradhan, B., Hong, H.Y., Bui, D.T., Duan, Z. and Ma, J.Q. (2017), *A Comparative Study of Logistic Model tree,*

Random Forest, and Classification and Regression Tree Models for Spatial Prediction of Landslide Susceptibility. Catena, 151, 147-160. https://opus.lib.uts.edu.au/handle/10453/119788#https://doi.org/10.1016/j.catena.2016.11.032

[18] Breiman L. *Random forests*. *Mach. Learn.* 2001; **45:5**–32. Doi: 10.1023/A:1010933404324.

[19] Krizhevsky, A., I. Sutskever, and G.E. Hinton. 2012. *ImageNet Classification with Deep Convolutional Neural Networks.* In Advances in Neural Information Processing Systems 25, edited by F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, 1097–1105. Curran Associates, Inc.

[20] K. Hornik, *Approximation capabilities of multilayer feedforward networks,* Neural Networks, vol. 4, pp. 251–257, 1991