Big Data-Machine Learning Processing of Recorded Radiofrequency Physiological and Pathological Measurements to Predict the Progression of Alzheimer's Disease

Rahmat Ullah
School of Engineering
University of Edinburgh
Edinburgh, UK
rahmat.orakzai@ed.ac.uk

Imran Saied
School of Engineering
University of Edinburgh
Edinburgh, UK
Imran.Saied@ed.ac.uk

Tughrul Arslan
School of Engineering
University of Edinburgh)
Edinburgh, UK
tughrul.arslan@ed.ac.uk

***Abstract*—Alzheimer's disease is expected to be the largest growing disease around the world with the increase in the ageing population. It affects the livelihood of not just patients, but those who take care of them. Currently, MRI and PET scans are used to monitor its progression. However, these systems are inconvenient for patients to use. There is a need for more intelligent and efficient methods to predict the current stage of the disease along with strategies on how to slow down its progress over time. In this paper, RF measurements were obtained from previous studies which looked at physiological and pathological changes in the brain due to Alzheimer's disease. Big data techniques were implemented to scale the dataset to simulate measurements for 100 patients. Data was then processed in several machine learning algorithms in order to validate how well each algorithm could classify the stage of Alzheimer's disease. Results showed that classical machine learning algorithms were accurate in classifying the different stages of Alzheimer's disease using the combined RF dataset.**

***Index Terms*—Machine learning, RF, Medical Microwave, Medical diagnostics, Classification**

I. Introduction

The rapidly growing ageing population has led to Alzheimer's disease (AD) becoming one of the fastest growing diseases in the world. In addition, AD poses an even larger and costlier burden and socioeconomic threat to society for the next 30 to 40 years [1]. The disruption that the ongoing pandemic has caused to ongoing care and research for AD is likely to impact and increase these numbers further [2].

Therefore, it is of utmost importance to investigate and adopt solutions that can intelligently, quickly, and noninvasively detect and monitor the progression of AD in patients. This will allow clinicians to predict the progression of AD and determine timely treatment plans. Machine learning (ML) techniques, combined with advanced sensing technology, has the potential to fill in this gap and provide timely contributions in early AD progression monitoring.

ML has been used extensively in the past decade to detect certain biomarkers in MRI scans for AD [3]. Such methods are used to improve the prediction of AD by using personalized classifiers and support vector machine (SVM) in order to distinguish between AD cases and cognitively normal cases, and between stable forms and progressive forms of MCI [3] [4] [5] [6].

Microwave medical sensing and imaging has turned into a transformative area of research for several decades, due to its non-ionising technology and capacity of translating to portable and wearable systems. This technology has been used extensively in the detection of breast cancer, stroke, and most recently, neurodegenerative diseases [7] [8] [9]. Recently, ML has been utilised to efficiently process the captured RF signals from such devices and classify different diseases in the heart and breast using received RF signals [10]. The authors have also recently proved the use of ML to classify different stages of AD using recorded RF measurements on pathological changes from 9 different head models [11].

While the amount of research interest in ML for classifying AD is increasing over the past decade, there has not been too much work done on raw signals or datasets, such as electroencephalogram (EEG) signals or radiofrequency (RF) signals. This paper aims to build upon the previous work done by authors in [11] [12] [13] by applying Big Data and ML techniques to scale measured RF data for physiological and pathological changes in the brain due to AD. This work aims to replicate an environment where several RF measurements from 100 patients are generated and processed in ML algorithms in order to determine how efficient ML is in classifying and predicting the stage of AD.

II. Recorded RF Data for AD

For this study, RF data was obtained from previous studies that investigated the use of antennas to detect physiological and pathological changes in the brain that were investigated in [11]–[13]. For these cases, a rectangular planar monopole antenna was used to capture reflected signals from the head that is shown in Fig. 1, and having the following dimensions (in mm) : W=35, L=75, W_R=30, L_R=50, W_{feed}=17, L_{feed}=12, W_{QWTL}=7, and L_{QWTL}=13.

This antenna was fabricated using silicone rubber as the substrate and a conductive textile material as the radiating patch. The thickness of the substrate was 3 mm, while the thickness of the radiating patch was 0.75 mm. The radiating patch was embedded inside the silicone rubber substrate in order to ensure that the conducting material does not get damaged, while also enhancing the antenna's flexibility. 6 antennas were placed around the head models in CST in order to capture the reflection coefficients (S11) at different regions of the brain.

A. Physiological Data

The recorded RF data for physiological changes in the brain due to AD is based on S11 measurements that were captured by the antennas for brain atrophy [12] [13]. For

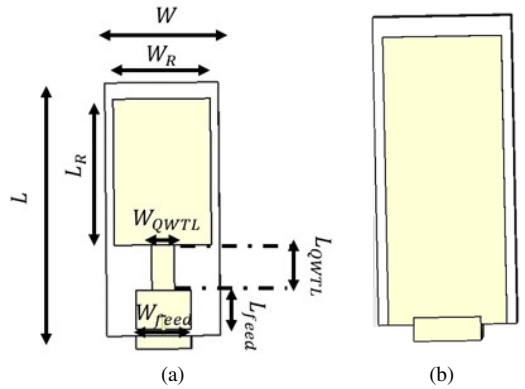

Fig. 1: (a) Front and (b) back of the rectangular planar monopole antenna used to capture RF measurements

this study, 3 cases were considered: 1) Healthy (0% brain atrophy), 2) Mild AD – 10% brain atrophy and 3) Severe AD – 25% brain atrophy. Fig. 2 shows the CST head models that were used to represent the different cases of brain atrophy.

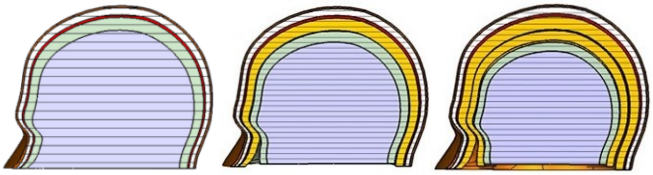

Fig. 2: CST head models used to simulate the physiological changes in the brain (i.e., brain atrophy) for a) Healthy, b) Mild AD (10% brain atrophy), and c) Severe AD (25% brain atrophy).

B. Pathological Data

An important change that occurs in the brain during AD is the presence of beta-amyloid plaques and tau tangles that affect the transport of nutrients within the brain cells [14]. This leads to the pathological changes in the brain tissue composition as AD progresses. Based on the progression of plaques and tangles in the brain, which were presented by Braak [14], we recreated CST voxel models that accurately represented the regions of the brain that were affected by the spread of AD pathology. Fig. 3 shows the CST head models that were used to represent: 1) Healthy, 2) Mild AD, and 3) Severe AD.

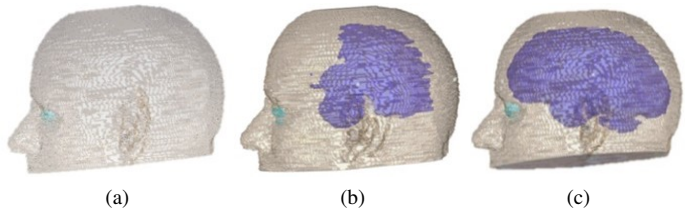

Fig. 3: CST head models used to simulate the pathological changes in the brain for a) Healthy, b) Mild AD, and c) Severe AD.

III. Big Data Implementation

Current machine and deep learning methods are using images from different imaging modalities such as MRI and PET scan. The algorithms are trained using different features such as brain volume, gray matter densities and cerebral amyloid accumulation in the hippocampus [15]. Since ML algorithms required a large data for better results in term of accuracy, therefore using images to train model is easy to find. However, getting microwave scattered data is a challenge, since the technology is still in the testing and evaluation phase.

In this work, synthetic data was used to investigate different ML algorithms to classify stages of AD based on both pathological and physiological changes in the brain. Simulations were performed to obtain data for 10 different cases. The values of the reflected signal were statistically analysed to find the approximate difference between different level of atrophy for physiological changes. A similar technique was employed for pathological data. These differences are then used to replicate the original data for 100 patients and fed to the ML algorithms for training and testing.

IV. Machine Learning Implementation

A Python environment was used in this study along with SciPy, which is an ecosystem of Python libraries such as NumPy (work with data in arrays), Matplotlib (create 2D plots), and Pandas (tools to organize and analyze data) to run the machine learning algorithms on the dataset. Additionally, the scikit-learn library was declared to evaluate ML models. The following ML algorithms were used: 1) Logistic Regression (LR), 2) Linear Discriminant Analysis (LDA), 3) K-Nearest Neighbours (KNN), 4) Classification and Regression Trees (CART), 5) Naïve Bayes (NB), and 6) Support Vector Machines (SVM).

Each ML algorithm was evaluated on the RF dataset. Later, statistical methods were used to estimate the accuracy of the models on unseen data (or validation data). Specifically, some data from the RF dataset was held back so that the algorithms will not get biased during its training phase. This method will provide a second and independent idea of how accurate the best model might actually be.

To accomplish this, first, the RF dataset was split into two sets; approximately 80% of the RF dataset was used to train the models (e.g. training dataset) and the remaining 20% was used as the validation dataset. This split was done randomly by using a random number generator in Python Next, a test harness was developed to use k-fold cross-validation to estimate the accuracy of the models on the training dataset. An evaluation score is then generated and used to summarise the skill of the model. The resulting accuracy metric is then used to evaluate the models. The final step is to validate the performance of the best ML algorithm on the validation dataset that was held back to determine how well the selected model is able to classify AD based on unseen data.

V. Results

After running each machine algorithm on the test dataset, the following table was generated to show the results of each ML algorithm and its accuracy. A visual representation of the accuracy of each ML algorithm is also shown in Fig. 1 as box plots.

Based on the results, it is shown that CART has the highest accuracy of approximately 81.5%, while LR follows up with

TABLE I: AVERAGE MEASURED ACCURACY FOR EACH ALGORITHM

Algorithm	Accuracy (%)	Std. Dev.
LR	77.29	0.0811
LDA	61.04	0.1135
KNN	74.17	0.0761
CART	81.46	0.0553
NB	70	0.0987
SVM	71.25	0.0817

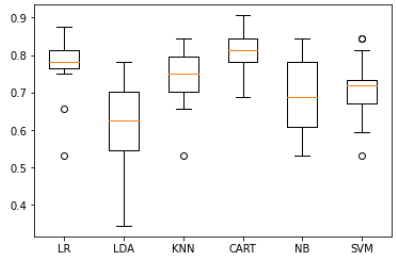

Fig. 4: Box plot comparing the distribution of estimated accuracy values for each algorithm.

an accuracy of 77.3%. Given the large dataset and multi-class nature of the data, CART looks to be the obvious choice in classifying AD from the dataset, as it has the ability to segment different variables thoroughly in order to come up with a decision model that it can rely on for classification. We now take the CART algorithm and apply it to the validation dataset in order to determine how well the CART model can classify the cases. After running the CART model on the validation dataset, the following confusion matrix and classification report was provided in Table.II and Table.III, respectively.

TABLE II: CONFUSION MATRIX FOR CART ALGORITHM

	Healthy	Mild AD	Severe AD
Healthy	29	1	6
Mild AD	3	31	5
Severe AD	3	6	36

TABLE III: CLASSIFICATION REPORT FOR CART ALGORITHM

	Precision (%)	Recall (%)	F1 Scores (%)	Support (#)
Healthy	76	81	78	36
Mild AD	89	79	84	39
Severe AD	77	80	78	45

It was found that CART had an overall accuracy of 83.2% on the validation dataset. Precision, which is defined as the percent of predictions that were correct, was less than 90% for all cases. Additionally, it can be noted that there were more misclassifications for Healthy and Severe AD cases as opposed to the Mild AD case. This could be due to the CART algorithm not being able to properly differentiate or extract features that can separate the cases efficiently. While the classification report shows that the CART's performance on the validation dataset has some limitations, the overall accuracy of 83.2% is promising and gives the incentive to process the data further so that it can differentiate features more easily. In addition, the precision of Mild AD detection is relatively high, which is promising as our goal is to use ML on the RF dataset to determine early-stage detection of AD.

VI. Conclusion

This paper investigates the use of big data and machine learning-based techniques to monitor the progression of AD. An integrated dataset was first generated based on both pathological and physiological changes in the brain caused by AD. The dataset was used to train different classification algorithms. The results of different classification algorithms show that classical ML algorithms can detect the progression with sensible accuracy. Future work will focus on more realistic data and neural network-based approaches to investigate it further.

References

[1] A. Association, "2020 alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 16, pp. 391–460, 2020.

[2] "The impact of covid-19 and the global pandemic on alzheimer's research, long-term care and the brain — aaic 2020 — alzheimer's association," http://alz.org/aaic/releases-2020/covid-19-cognition-media-panel.asp, (Accessed on 07/27/2021).

[3] J. Escudero, E. Ifeachor, J. P. Zajicek, C. Green, J. Shearer, S. Pearson, A. D. N. Initiative *et al.*, "Machine learning-based method for personalized and cost-effective detection of alzheimer's disease," *IEEE transactions on biomedical engineering*, vol. 60, no. 1, pp. 164–168, 2012.

[4] S. M. Landau, D. Harvey, C. M. Madison, R. A. Koeppe, E. M. Reiman, N. L. Foster, M. W. Weiner, W. J. Jagust, A. D. N. Initiative *et al.*, "Associations between cognitive, functional, and fdg-pet measures of decline in ad and mci," *Neurobiology of aging*, vol. 32, no. 7, pp. 1207–1218, 2011.

[5] S. Haller, D. Nguyen, C. Rodriguez, J. Emch, G. Gold, A. Bartsch, K. O. Lovblad, and P. Giannakopoulos, "Individual prediction of cognitive decline in mild cognitive impairment using support vector machine-based analysis of diffusion tensor imaging data," *Journal of Alzheimer's disease*, vol. 22, no. 1, pp. 315–327, 2010.

[6] S. Al-Shoukry, T. H. Rassem, and N. M. Makbol, "Alzheimer's diseases detection by using deep learning algorithms: a mini-review," *IEEE Access*, vol. 8, pp. 77 131–77 141, 2020.

[7] E. Fear and M. Stuchly, "Microwave detection of breast cancer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 11, pp. 1854–1863, 2000.

[8] M. Persson, A. Fhager, H. D. Trefná, Y. Yu, T. McKelvey, G. Pegenius, J.-E. Karlsson, and M. Elam, "Microwave-based stroke diagnosis making global prehospital thrombolytic treatment possible," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 11, pp. 2806–2817, 2014.

[9] M. Manoufali, A. T. Mobashsher, B. Mohammed, K. Bialkowski, P. C. Mills, and A. Abbosh, "Implantable sensor for detecting changes in the loss tangent of cerebrospinal fluid," *IEEE transactions on biomedical circuits and systems*, vol. 14, no. 3, pp. 452–462, 2020.

[10] J. Saluja, J. Casanova, and J. Lin, "A supervised machine learning algorithm for heart-rate detection using doppler motion-sensing radar," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 1, pp. 45–51, 2019.

[11] I. Saied, T. Arslan, and S. Chandran, "Classification of alzheimers disease using rf signals and machine learning," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, 2021.

[12] I. M. Saied and T. Arslan, "Noninvasive wearable rf device towards monitoring brain atrophy and lateral ventricle enlargement," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 4, no. 1, pp. 61–68, 2019.

[13] I. M. Saied, S. Chandran, and T. Arslan, "Integrated flexible hybrid silicone-textile dual-resonant sensors and switching circuit for wearable neurodegeneration monitoring systems," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1304–1312, 2019.

[14] H. Braak and E. Braak, "Neuropathological stageing of alzheimer-related changes," *Acta Neuropathologica*, vol. 82, no. 4, pp. 239–259, 2015.

[15] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers in aging neuroscience*, vol. 11, p. 220, 2019.