A microwave sensing system enhanced by a machine learning algorithm for Alzheimer’s disease early detection

A Microwave Sensing System Enhanced by a Machine Learning Algorithm for Alzheimer's Disease Early Detection

Leonardo Cardinali
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
leonardo.cardinali@polito.it

Mattia Spano
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
s292254@studenti.polito.it

Martina Gugliermino
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
martina.gugliermino@polito.it

David O. Rodriguez-Duarte
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
david.rodriguez@polito.it

Marco Ricci
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
marco.ricci@polito.it

Jorge A. Tobon-Vasquez
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
jorge.tobonvasquez@polito.it

Rosa Scapaticci
National Research Council of Italy
IREA
Napoli, Italy
scapaticci.r@irea.cnr.it

Roberta Palmeri
National Research Council of Italy
IREA
Napoli, Italy
palmeri.r@irea.cnr.it

Lorenzo Crocco
National Research Council of Italy
IREA
Napoli, Italy
crocco.l@irea.cnr.it

Francesca Vipiana
dept. of electronics and telecommunications
Politecnico di Torino
Torino, Italy
francesca.vipiana@polito.it

***Abstract*—Diagnostic techniques able to detect Alzheimer's disease in its early stage are costly, invasive, and lack portability. Early detection is fundamental to give patients the best treatment and the possibility to plan their future life accordingly, and benefits caregivers and medical professionals. To address the drawbacks of current detection technologies, microwave imaging and sensing are emerging as a non-intrusive, relatively low-cost alternative methods to differentiate between normal and pathological states. The physiological basis behind this difference is the change in permittivity within the cerebrospinal fluid of Alzheimer's patients. In this study, we investigate machine learning applied to microwave sensing data using a multilayer perceptron classifier. Data collection is made through experiments involving a realistic multi-tissue head phantom filled with liquids mimicking different disease severities. We explore various architectural approaches for the classification algorithm, evaluating the performances of binary classification for different combinations of hyperparameters and using different subsets of the total data pool as training data. Our preliminary findings affirm the potential of the proposed technique, reaching high correct classification scores on test data not used during training.**

Keywords—Alzheimer's disease, microwave antennas, machine learning, realistic phantom.

I. Introduction

Every year nearly 10 million new cases of Alzheimer's disease (AD) arise. This pathology causes major disabilities and dependency and is the seventh leading cause of death worldwide [1]. AD diagnosis is a crucial moment for patients, detecting AD in its early stage is fundamental to give patients better treatments, to avoid medication errors [2] and to plan their life decisions accordingly. Knowing in advance about AD also allows the early development of coping strategies, important to improve life quality when suffering cognitive impairment, also helping caregivers and healthcare professionals [3]. In most cases the diagnosis is given by analyzing the individual's behavioral and cognitive changes and the family history. This evaluation happens when the illness is already in a developed state since cognitive symptoms arise years after the appearance of physiological changes correlated with the pathology [4]. The current techniques to diagnose AD in its early stage include computed tomography (CT) scans, functional magnetic resonance imaging (fMRI) and positron-emission tomography (PET) to highlight the presence of amyloid plaques, neurofibrillary tangles, and brain atrophy [5]. Another method is performing laboratory tests on blood and cerebrospinal fluid (CSF) since abnormal concentrations of tau, tau-phosphorylated and amyloid-beta ($A\beta_{1\text{-}40}$, $A\beta_{1\text{-}42}$) proteins are CSF biomarkers of AD [6].

PET and CT scans are invasive imaging technologies, since they use ionizing radiation to produce a result, and invasiveness is also a problem for laboratory tests since they require CSF extraction through a lumbar puncture to search for AD biomarkers [7]. fMRI, while being non-invasive, has the disadvantage of being expensive and not easily portable, when more than 60 % of people affected by dementia live in low-and middle-income countries [1], and patients may have metal implants not compatible with fMRI.

In this context, microwave imaging is a valid alternative technology since it does not use ionizing radiation, is relatively not expensive and the required hardware is easily portable. This technique allows to locate and estimate the extension of a hidden region with different dielectric properties respect to its surrounding material

through a suitable elaboration of backscattered fields produced by a set of antennas surrounding the region. Numerous applications of microwave imaging on the medical field are being studied, as brain stroke monitoring [8] [9] [10] [11], breast tumor detection [12] [13] [14], bone health monitoring [15] and also AD detection and monitoring [16] [17]. While previous applications of microwave imaging to AD focus on the detection of plaques and tangles in the brain [16] or on brain atrophy evaluation [17], the novelty of our approach consists in the use of microwaves to non-invasively detect AD in its early stage monitoring the CSF dielectric properties change related to abnormal proteins concentration due to the illness [18] [19]. In this case we are performing a sensing of the CSF dielectric properties instead of trying to locate and estimate the extension of a region with different permittivity, and we talk about microwave sensing.

Obtaining the dielectric properties from the backscattered fields is called inverse scattering problem: it is an ill-posed and non-linear problem, and a difficult task. Our objective is not a precise measurement of the CSF properties, but we want to assess whether changes in CSF dielectric properties associated with AD can be detected using microwave sensing, and if a CSF classification as healthy or affected by AD is possible. For this reason, we apply machine learning to scattering parameters collected in controlled experiments and create a binary classifier to distinguish between data from a healthy-like condition and data from different severity degrees of the illness, meaning a reduced permittivity.

II. Methods

Data collection is performed on a realistic multi-tissue phantom representing a human head specifically built for microwave imaging and sensing applications [20]. Human tissues are replicated using a mixture of graphite and rubber in percentages such that each tissue mimics the specific material dielectric properties. Bone, skin, fat, white and grey matter, ventricles, and cerebellum are solid and built this way, while CSF is liquid and is realized using Triton X-100 mixed with water and salt [21]. CSF can easily be inserted by pouring it in the dedicated space and removed opening a valve in the bottom of the phantom. Five different liquids mimicking CSF are produced, one representing the healthy case and four representing various AD severity degrees, where a higher degree corresponds to a lower permittivity. The reference values for dielectric properties come from the IFAC-CNR database [22]. The compounds mimicking the pathological CSF have permittivity lowered by 7 % respect to the healthy case for the lowest severity, then 11 %, 19 % and 24 % for the highest severity. The permittivity and conductivity of each liquid are measured using a coaxial probe and the Keysight software suite [23].

The instrument used to generate and collect the electromagnetic fields is a Keysight 4-ports network analyzer (VNA), combination of two 2-ports VNAs: P9375A and P9371A [24]. The four antennas connected to the VNA are flexible circuit-printed monopoles and have already been used in a microwave imaging application on a human head phantom [10]. The investigated frequency range spans from 0.5 GHz to 6.5 GHz. This range was chosen to reach balance between resolution and the penetration depth. Penetration is severely limited beyond

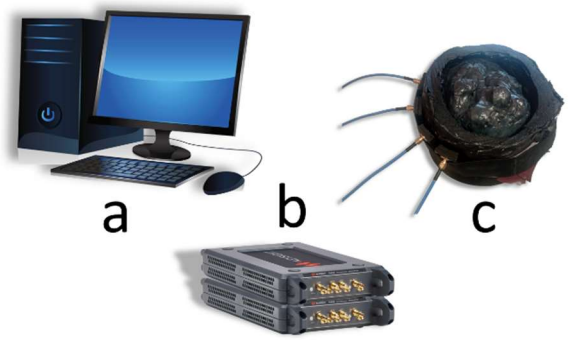

Fig. 1: Schematic representation of the measurement setup; (a) central processing unit; (b) vector network analyzer; (c) phantom representing human head, filled with fluids mimicking human CSF dielectric properties. 4 antennas connected with the VNA are placed on one side.

6.5 GHz, while below 0.5 GHz the subarachnoid region's electrical dimension becomes extremely small. Our goal for data collection is to create a suitable dataset to train a binary classification algorithm, which could differentiate between healthy and AD condition. The experimental setup is displayed in Fig. 1.

We conduct 28 measurement subsets over four different days. In each subset, we fill a phantom with artificial CSF representing the healthy case. We then perform 10 consecutive measurements of the scattering parameters. Afterward, we empty the phantom, refill it with different CSF to simulate the AD-affected case, and conduct another 10 measurements. We repeat this process for all CSF-mimicking fluids. Since there are four different mixtures representing AD-affected CSF compared to one representing healthy CSF, we add 29 additional subsets of the healthy case. Consequently, the final dataset comprises 570 measurements of CSF representing the healthy condition and 1120 measurements of CSF representing the various AD severity degrees, with 280 measurements for each severity level.

From this dataset, we build four distinct training sets. This approach allows a deeper understanding of how various subsets of the whole dataset impact the learning process and to evaluate the influence of different data distributions on the system's resilience and ability to generalize. The first training set contains measurements from all four days, in such a way that the measurements are widely spread. The second set contains measurements from the first and the last measurements of each day. The third set consists of all measurements except the ones from day three. The fourth set consists of measurements from the first and the second day. The measurement subsets excluded from training in each training set scenario are utilized as test set to assess the classification performance of the algorithm trained with the corresponding training set.

Because of the non-linearity of the problem, multilayer perceptron (MLP) [25] [26] is the chosen algorithm, being able to approximate complex non-linear functions [27]. We employ three distinct approaches to explore the optimal network configuration for our classification task. We start from the method with the fewest degrees of freedom, escalate to the second method with more degrees of freedom, and then use the third method that has the highest number of degrees of freedom. We employ these three methods across the four training sets. We initiate our

investigation by employing a two-layer feed-forward neural network for classification tasks, utilizing the MATLAB ML toolbox [28]. We employ random data partitioning, scaled conjugate gradient algorithm as the training function, and cross-entropy error as the loss metric for assessing performance. The training set is divided into three portions: 70 % for training, 15 % for validation, and the remaining 15 % for testing. During training the only degree of freedom is the number of neurons in the hidden layer, that varies from 1 to 100, employing a sigmoid activation function. The output layer consists of a single neuron with softmax activation function. Notably, this method offers the lowest degree of freedom and serves as the initial stage in our exploration of hyperparameters.

For the second method, utilizing the identical toolbox [28], we conduct an exploration of the design space for our MLP model that encompasses more degrees of freedom, exploring combinations of the following variables:

- the number of hidden layers, ranging from 1 to 20;
- the number of neurons within each layer, spanning from 1 to 40 with consistent neuron counts across all hidden layers;
- the learning rate parameter;
- the choice of training function for updating network parameters;
- the performance evaluation function, or loss function, employed during learning.

The training functions examined included resilient backpropagation, scaled conjugate gradient, conjugate gradient with Powell/Beale restarts, one-step secant, and gradient descent. For performance evaluation, we explored both cross-entropy and mean squared error functions, typically employed in binary classification scenarios. It is important to note that the output layer remained consistent with the first method throughout this experimentation, and random data partitioning was consistently applied.

Some parameters, such as the activation functions for neurons in the hidden layers, are fixed in earlier methods. To optimize these parameters, we employ Python's machine learning development tools from the scikit-learn library [29]. We utilize the GridSearchCV tool for a systematic search across hyperparameters to identify the combination yielding the best performance on test data, that is data excluded from the given training set. The parameters under investigation include hidden layer sizes as in the second method, maximum iteration count, activation functions for hidden layer neurons, training functions for weight optimization, and the alpha parameter, that is the strength of L2 regularization. Activation functions explored are: logistic sigmoid function, hyperbolic tangent function, and rectified linear unit function. We implement three training functions: limited-memory BFGS [30], stochastic gradient descent, and the Adam solver [31]. For stochastic gradient descent, we consider various learning rate types: constant, inverse scaling (decreasing at each time step), and adaptive (reduced when training loss remains unchanged for two consecutive epochs). The output layer consists of a single neuron using the logistic sigmoid activation function.

For all methods, default settings are maintained for parameters that were not specifically referenced. For each method and for each training set, the combination of hyperparameters that gives the best results in terms of correct classification is listed in Table 1.

TABLE 1
RESULTS OF HYPERPARAMETERS OPTIMIZATION

METHOD 1				
	Training set 1	*Training set 2*	*Training set 3*	*Training set 4*
Number of neurons	20	20	14	17
METHOD 2				
	Training set 1	*Training set 2*	*Training set 3*	*Training set 4*
Training function	One step secant	Scaled conjugate gradient	Scaled conjugate gradient	Scaled conjugate gradient
Perform function	Cross-entropy	Mean squared error	Mean squared error	Cross-entropy
Number of hidden layers	2	3	3	1
Neurons per hidden layer	30	19	17	10
Learning rate value	3×10^{-4}	4.2×10^{-2}	4×10^{-2}	1×10^{-3}
METHOD 3				
	Training set 1	*Training set 2*	*Training set 3*	*Training set 4*
Training function	Limited-memory BFGS	Limited-memory BFGS	SGD (adaptive learning rate)	SGD (adaptive learning rate)
Activation function	Hyperbolic tangent	Hyperbolic tangent	Logistic sigmoid	Logistic sigmoid
Alpha parameter	7.5×10^{-3}	1×10^{-2}	1.5×10^{-4}	1×10^{-3}
Number of hidden layers	20	10	1	1
Neurons per hidden layer	40	20	19	19
Maximum iterations	1000	1000	1000	1000

III. PRELIMINARY RESULTS AND DISCUSSION

The classification performances of the best network for each method are displayed in Table 2. All applied methodologies across all datasets have consistently yielded high accuracy scores, always reaching over 83 % correct classification. The best accuracy is achieved when

employing training set 4, where all methods attain an accuracy exceeding 90 % when tested with previously unseen data belonging to the test set.

TABLE 2
CORRECT CLASSIFICATION PERCENTAGES ON TEST DATA

	Training set 1	*Training set 2*	*Training set 3*	*Training set 4*
Method 1	85.1 %	85.1 %	85.9 %	90.8 %
Method 2	88.0 %	86.3 %	89.6 %	90.6 %
Method 3	90.2 %	88.1 %	83.9 %	90.1 %

It's noteworthy that training set 4, comprising only 2 out of 4 days and featuring the fewest number of healthy subjects, outperforms in terms of classification on data that was not part of its training. Additionally, all testing data stems from days separate from those employed for training. Intriguingly, the optimal network architecture for this specific training set consistently features a single hidden layer. Method 3, when applied to training set 1, produces results of comparable quality, albeit employing a dissimilar hidden layer configuration and a distinct activation function for the hidden layer neurons. This observation suggests that the problem's decision boundaries may possess a degree of smoothness.

IV. CONCLUSION AND PERSPECTIVES

In this paper we presented a new approach to early, non-invasive AD detection, which leverages machine learning to classify microwave sensing data. Specifically, our binary classification, distinguishing the presence or absence of the pathology, relied on detecting variations in measured scattering parameters attributed to the reduced permittivity of pathological CSF. To compile our dataset, we conducted measurements using a realistic phantom, and subsequently, we trained a variety of classifiers. The most successful network in terms of accuracy when classifying test data was a single-hidden-layer multilayer perceptron. This model was trained on data collected during two out of the four days of measurement, achieving a high classification accuracy of 90.8 % when provided with all data from the remaining two days.

One of our immediate objectives is to enhance the quality of measurements. This enhancement can be accomplished through various means, such as expanding the number of antennas and increasing the quantity of measurement sets. Furthermore, with a larger dataset at our disposal, we can explore multiclass classification to differentiate between different AD severities. To further enhance our machine learning approach, we can consider implementing alternative algorithms, and expanding the search space of hyperparameters during the tuning phase.

ACKNOWLEDGMENTS

This work was supported in part by the project PON Research and Innovation “Microwave Imaging and Detection powered by Artificial Intelligence for Medical and Industrial Applications (DM 1062/21)”, funded by MUR, in part by the project “INSIGHT – An innovative microwave sensing system for the evaluation and monitoring of food quality and safety”, funded by MAECI and in part by the project “THERAD - Microwave Theranostics for Alzheimer’s Disease”, funded by Compagnia di San Paolo. It was carried out partially within the Agritech National Research Center, funded by the European Union Next-Generation EU (Piano Nazionale di Ripresa e Resilienza (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022).

V. REFERENCES

[1] C. Greenblat, "World Health Organization - Dementia," World Health Organization, 15 march 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dementia. [Accessed 5 september 2023].

[2] I.G. McKeith et al., "Diagnosis and management of dementia with Lewy bodies: Third report of the DLB consortium," *Neurology,* vol. 65, no. 12, p. 1863–1872, 2005.

[3] S. Gauthier, C. Webster, S. Servaes, J. Morais and P. Rosa-Neto, "World Alzherimer Report 2022: Life after diagnosis: Navigating tratment, care and support," Alzheimer's Disease International, London, England, 2022.

[4] K. A. Jellinger, B. Janetzky, J. Attems and E. Kienzl, "Biomarkers for early diagnosis of Alzheimer disease: 'ALZheimer ASsociated gene'–a new blood biomarker?," *Journal of molecular and cellular medicine,* vol. 12, no. 4, pp. 1094-1117, 2008.

[5] K. Blennow and H. Zetterberg, "Biomarkers for Alzheimer's disease: current status and prospects for the future," *J Intern Med,* vol. 284, no. 6, pp. 643-663, 2018.

[6] J. Dumurgier, S. Schraen, A. Gabelle et al., "Cerebrospinal fluid amyloid-β 42/40 ratio in clinical setting of memory centers: a multicentric study," *Alzheimer's research & therapy,* vol. 7, no. 30, pp. 1-9, 2015.

[7] B. Dubois et al., "Research criteria for the diagnosis of Alzheimer’s disease: revising the NINCDS–ADRDA criteria," *The Lancet Neurology,* vol. 6, no. 8, pp. 734-746, 2007.

[8] R. Scapaticci, O. M. Bucci, I. Catapano and L. Crocco, "Differential microwave imaging for brain stroke followup," *International Journal of Antennas and Propagation,* vol. 2014, 2014.

[9] J. A. Tobon Vasquez, R. Scapaticci, G. Turvani, G. Bellizzi, D. O. Rodriguez-Duarte, N. Joachimowicz, B. Duchêne, E. Tedeschi, M. R. Casu, L. Crocco and F. Vipiana, "A Prototype Microwave System for 3D Brain Stroke Imaging," *Sensors,* vol. 20, no. 9, p. 2607, 2020.

[10] D. O. Rodriguez-Duarte, C. Origlia, J. A. Tobon Vasquez, R. Scapaticci, L. Crocco, F. Vipiana, "Experimental assessment of real-time brain stroke monitoring via a microwave imaging scanner," *IEEE Open Journal of Antennas and Propagation,* vol. 3, pp. 824-835, 2022.

[11] D. O. Rodriguez-Duarte, J. A. Tobon Vasquez, R. Scapaticci, G. Turvani, M. Cavagnaro, M. Casu, L. Crocco, F. Vipiana, "Experimental Validation of a Microwave System for Brain Stroke 3-D Imaging," *Diagnostics, Special Issue on Electromagnetic Imaging for a Novel Generation of Medical Devices,* 2021.

[12] E. C. Fear, J. Sill and M. A. Stuchly, "Experimental feasibility study of confocal microwave imaging for breast tumor detection," *IEEE Transactions on Microwave Theory and Techniques,* vol. 51, no. 3, pp. 887-892, 2003.

[13] R. Scapaticci, G. Bellizzi, I. Catapano, L. Crocco and O. M. Bucci, "An effective procedure for MNP-enhanced breast cancer microwave imaging," *IEEE Transactions on Biomedical Engineering,* vol. 61, no. 4, pp. 1071-1079, 2013.

[14] M. T. Bevacqua, S. Di Meo, L. Crocco, T. Isernia and M. Pasian, "Millimeter-waves breast cancer imaging via inverse scattering techniques," *IEEE Journal of Electromagnetics,* vol. 5, no. 3, pp. 246-253, 2021.

[15] B. Amin, A. Shahzad, M. O'halloran, B. Mcdermott and A. Elahi, "Experimental Validation of Microwave Imaging Prototype and DBIM-IMATCS Algorithm for Bone Health Monitoring," *IEEE Access,* vol. 10, pp. 42589-42600, 2022.

[16] R. Ullah and T. Arslan, "Detecting Pathological Changes in the Brain Due to Alzheimer Disease Using Numerical Microwave Signal Analysis," *2020 IEEE International RF and Microwave Conference (RFM),* pp. 1-4, 2020.

[17] R. Ullah, I. Saied and T. Arslan, "Measurement of whole-brain atrophy progression using microwave signal analysis," *Biomedical Signal Processing and Control,* vol. 71, p. 103083, 2022.

[18] T. Yoshikawa, Z. Zhang, K. Yamashita, M. Noda, "Biosensing of interaction between phospholipid membrane of liposome as model cell membrane and amyloid-beta protein in human serum by dielectric dispersion analysis," *Sensors and Actuators B: Chemical,* vol. 236, pp. 1028-1035, 2016.

[19] M. Manoufali, A. T. Mobashsher, B. Mohammed, K. Bialkowski, P. C. Mills, 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.

[20] C. Origlia, M. Gugliermino, D. O. Rodriguez-duarte, J. A. Tobon Vasquez, F. Vipiana, "Anthropomorphic Multi-tissue Head Phantom for Microwave Imaging Devices Testing," in *17th European Conference on Antennas and Propagation (EuCAP)*, Florence, 2023.

[21] N. Joachimowicz, B. Duchêne, C. Conessa, O. Meyer, "Anthropomorphic breast and head phantoms for microwave imaging," *Diagnostics,* vol. 8, no. 4, p. 85, 2018.

[22] D. Andreuccetti, R. Fossi and C. Petrucci, "An Internet resource for the calculation of the dielectric properties of body tissues in the frequency range 10 Hz - 100 GHz," IFAC-CNR, 1997. [Online]. Available: http://niremf.ifac.cnr.it/tissprop/. [Accessed 6 september 2023].

[23] Keysight Technol., "N1500A Materials Measurement Suite," 2022. [Online]. Available: https://www.keysight.com/it/en/product/N1500A/materials-measurement-suite.html. [Accessed 6 september 2023].

[24] Keysight Technol., "Keysight streamline series USB vector network analyzer P937XA 2-port, up to 26.5 GHz," 2018. [Online]. Available: https://www.keysight.com/it/en/product/N1500A/materials-measurement-suite.html. [Accessed 7 september 2023].

[25] P. J. Werbos, "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. dissertation, Harvard University, Cambridge, MA, USA, 1974.

[26] D. E. Rumelhart, G. E. Hinton, R. J. Williams, "Learning representations by back-propagating errors," *Nature,* vol. 323, no. 6088, pp. 533-536, 1986.

[27] K. Hornik, M. Stinchcombe and H. White, "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks,* vol. 2, pp. 359-366, 1989.

[28] The MathWorks Inc., "Deep Learning Toolbox Documentation," 2023. [Online]. Available: https://www.mathworks.com/help/deeplearning/. [Accessed 7 september 2023].

[29] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *The Journal of Machine Learning Research,* vol. 12, no. 85, pp. 2825-2830, 2011.

[30] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on scientific computing,* vol. 16, no. 5, pp. 1190-1208, 1995.

[31] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference for Learning Representations*, San Diego, 2015.