Model-based data generation for support vector machine stroke classification

Model-Based Data Generation for Support Vector Machine Stroke Classification

Valeria Mariano, Jorge A. Tobon Vasquez, Mario R. Casu, Francesca Vipiana
Dept. of Electronics and Telecommunication
Politecnico di Torino
Torino, Italy
{valeria_mariano, jorge.tobon, mario.casu, francesca.vipiana}@polito.it

***Abstract*—This paper presents a new and efficient method to generate a dataset for brain stroke classification. Exploiting the Born approximation, it derives scattering parameters at antennas locations in a 3-D scenario through a linear integral operator. This technique allows to create a large amount of data in a short time, if compared with the full-wave simulations or measurements. Then, the support vector machine is used to create the classifier model, based on training set data with a supervised method and to classify the test set. The dataset is composed by 9 classes, differentiated for presence, typology and position of the stroke. The algorithm is able to classify the test set with a high accuracy.**

I. Introduction

In the last years, microwaves entered the world of medical diagnostic. This technology allows the creation of a portable, low cost and low power intensity device. These particular characteristics make the device very versatile and suitable for brain stroke detection and monitoring [1].

For what concerns the use of microwaves in brain stroke detection and classification, there are two main approaches: the deterministic one, based on the solution of the electromagnetic scattering problem [2], and a machine learning (ML) approach as in [3]–[5]. ML is a great alternative because it does not need specific theoretical requirements. However, it involves a large amount of data, and for this specific application, the collection of measurement data and full-wave simulations requires significant time. In order to overcome this limit, here we propose to synthesize the dataset through high-speed simulations, based on a linear integral operator that allows the creation of a large dataset in a short time. Then, the support vector machine (SVM) is the algorithm used to classify the data with a supervised method [6].

II. Dataset Generation

This section contains a description of the method used to generate the dataset. The considered 3-D scenario is constituted by a human head phantom wearing a helmet formed by an antenna array, where each antenna acts as transmitter and receiver as described in [7], [8]. The features of the SVM algorithm will be the scattering parameters S at the antennas ports. The imaging domain corresponds to the head, with a complex relative permittivity equal to the average of dielectric properties of brain tissues. The dielectric contrast is defined as

$$\Delta\chi(\underline{r}) \triangleq \frac{\epsilon_r(\underline{r}) - \epsilon_b(\underline{r})}{\epsilon_b(\underline{r})}, \tag{1}$$

where ϵ_b and ϵ_r are the dielectric complex permittivity of the average brain and stroke area, respectively. Instead, $\underline{r}$ is the position vector in the domain of interest (DOI). Considering S^{tot} and S^{inc} as the scattering matrix at antennas ports with and without the stroke in the DOI, the differential scattering matrix is

$$\Delta S = S^{tot} - S^{inc}. \tag{2}$$

Exploiting the Born approximation, for each pair of antennas m and n, $\Delta\chi$ and ΔS are related by means of a linear integral operator:

$$\Delta S_{m,n} = -\frac{j\omega\epsilon_b}{2a_m a_n}\iiint_V \underline{E}_{b,m}(\underline{r}) \cdot \underline{E}_{b,n}(\underline{r})\Delta\chi(\underline{r})\, d\underline{r} \tag{3}$$

where, $\omega = 2\pi f$ is the angular frequency, a_m and a_n are the power waves, at antenna port m and n, respectively. The symbol " $\cdot$ " identifies a dot product between the background fields $\underline{E}_{b,m}$ and $\underline{E}_{b,n}$ radiated by antennas m and n.

The evaluation of ΔS via (3), allows to obtain results close to a full-wave approach (e.g. the finite element method described in [9]), with a significant reduction in time of around 3 orders of magnitude. The first step for dataset creation is the definition of $\Delta\chi$ that is different from zero only in the stroke. In the results reported here, the stroke is a sphere with a fixed radius of 1.5 cm, and it assumes a different value of contrast depending on the type of stroke (ischemic or hemorrhagic). Different types of scenario are created by moving the target randomly in the brain. Once created the contrast, we sum white noise with an order of magnitude equal to 1/10 of ischemic dielectric contrast (the smaller one). The noise allows to have data variety also in the case without the target. Now, we calculate ΔS parameters through the integral operator in (3), summing the S^{inc} in order to obtain S^{tot}. The S parameters matrix is symmetrical, so we consider only the superior triangular matrix. The machine learning does not work with complex number, for this reason for each S parameter there are two features: the real and the imaginary part. The final dataset has 9 classes based on presence, typology (ischemic or hemorrhagic) and position of stroke. In Fig. 1 there are the 4 regions of the head used for classes division: front-left, front-right, back-left and back-right. Each class contains almost 500 records, with a total number of 4500 records in the dataset. The time to generate the whole dataset with a not yet parallelized code, on a standard laptop is around 1 hour.

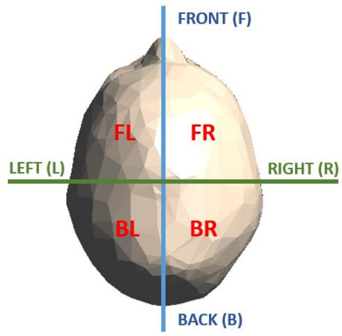

Fig. 1. Transverse plane view of human head. Subdivision in 4 regions: front left (FL), front right (FR), back left (BL), back right (BR).

III. Numerical Results

The dataset is divided into training set and test set with a percentage of 80 and 20, respectively. The first step, is the selection of SVM hyperparameters that are the regularization parameter C and the kernel used in the algorithm [10]. They are selected with the grid search method, which chooses the combination of parameters that minimize a metric, in our case the accuracy of the model. Fig. 2 shows the normalized confusion matrix obtained with SVM. The diagonal values are very high, moreover looking at the 3 macro-classes (no target, ischemic and hemorrhagic stroke) they are completely differentiated. Analysing the positions classification in some cases the algorithm assigns an unexpected class: this happens when the target is located in an ambiguous position, i.e. the center of the sphere is very close to the axis that divided two regions. Three important metrics for machine learning performance evaluation are *recall*, *precision*, and *accuracy* [11]. For each class they assume values very close to 1, in particular $recall > 0.93$, $precision > 0.89$ and $accuracy > 0.98$.

IV. Conclusion and Perspectives

An important drawback of applying machine learning in brain stroke classification is the need for a large amount of data to train the algorithm. This paper presents a novel and fast way to generate a dataset for brain stroke classification. The dataset, composed of S parameters at the antennas ports, is created using a linear integral operator, obtained applying the Born approximation. The operator allows passing from the space of dielectric contrast in the domain of imaging to the space of S parameters. The results obtained using the SVM algorithm have shown that the three macro-classes (no target, ischemic and hemorrhagic stroke) are easily distinguishable. Moreover,

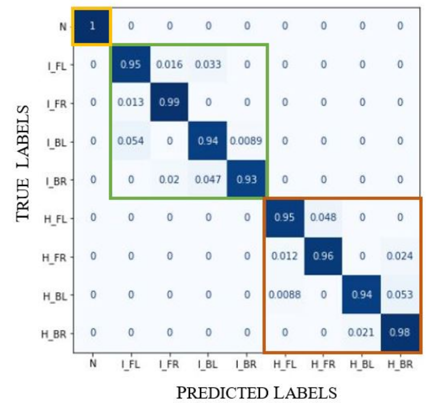

Fig. 2. Confusion matrix. The 3 macro-classes are identified by the colored squares: yellow for N (no target), green for I (ischemic stroke) and red for H (hemorrhagic stroke). The acronyms FL, FR, BL, BR identify the 4 spatial regions (front-left, front-right, back-left and back-right).

the algorithm always found the stroke position, except when it is located in an ambiguous position. Future work will deal with exploiting this method also with measurements data obtained with the system described in [7].

Acknowledgement

This work was supported by the MIUR under the PRIN project "MiBraScan", and by the European Union's Horizon 2020 Research and Innovation Program under the EMERALD project, grant agreement No. 764479.

References

[1] L. Crocco, I. Karanasiou, M. James, and R. C. Conceicao (Eds.), *Emerging Electromagnetic Technologies for Brain Diseases Diagnostics, Monitoring and Therapy*. Springer int. pub., 2018.

[2] M. Pastorino, *Microwave Imaging*. John Wiley & Sons, Ltd, 2010.

[3] G. Zhu, A. Bialkowski, L. Guo, B. Mohammed, and A. Abbosh, "Stroke classification in simulated electromagnetic imaging using graph approaches," *IEEE J. Electromagn., RF, Microw. Med. Biol.*, vol. 5, no. 1, pp. 46–53, 2021.

[4] M. Salucci, A. Polo, and J. Vrba, "Multi-step learning-by-examples strategy for real-time brain stroke microwave scattering data inversion," *Electronics*, vol. 10, no. 1, 2021.

[5] 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.

[6] R. P. Lakshmi, M. S. Babu, and V. Vijayalakshmi, "Voxel based lesion segmentation through svm classifier for effective brain stroke detection," in *2017 International Conference WiSPNET*, pp. 1064–1067, 2017.

[7] J. A. Tobon Vasquez *et al.*, "A prototype microwave system for 3D brain stroke imaging," *Sensors*, vol. 20, May 2020.

[8] D. O. Rodriguez-Duarte, J. A. T. Vasquez, R. Scapaticci, L. Crocco, and F. Vipiana, "Brick-shaped antenna module for microwave brain imaging systems," *IEEE Antennas Wireless Propag. Lett*, vol. 19, no. 12, pp. 2057–2061, 2020.

[9] D. O. Rodriguez-Duarte, J. A. Tobon Vasquez, R. Scapaticci, L. Crocco, and F. Vipiana, "Assessing a microwave imaging system for brain stroke monitoring via high fidelity numerical modelling," *IEEE J. Electromagn., RF, Microw. Med. Biol.*, pp. 1–1, 2021.

[10] C.-w. Hsu, C.-c. Chang, and C.-J. Lin, "A practical guide to support vector classification," 11 2003.

[11] J. Davis and M. Goadrich, "The relationship between precision-recall and roc curves," in *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, (New York, NY, USA), p. 233–240, Association for Computing Machinery, 2006.