POLITECNICO DI TORINO

Master's Degree in Biomedical Engineering
Biomedical Instrumentation

Master Thesis

Machine learning techniques for microwave brain stroke detection and classification

Candidate
Giulia FOLLARI
s274491

Supervisor
Prof. Francesca VIPIANA

Co-Supervisors
Prof. Mario R. CASU
PhD Jorge A. TOBON V.
M. Sc. Valeria MARIANO

Academic year 2021/2022

Ai miei genitori

e

a chi da lassù,

veglia sempre su di me.

Acknowledgements

I would like to thank my supervisor Francesca Vipiana for giving me the opportunity to develop my degree thesis on such an interesting project.

I also thank the whole research group of the DET of the Politecnico di Torino, for having welcomed me in these months: I met some brilliant people, from whom I learned a lot.

I thank Jorge Alberto Tobon Vasquez, David Orlando Rodriguez Duarte, and Cristina Origlia for their immense availability.

A special recognition goes to my co-consultant Valeria Mariano who has constantly supported me in the work and in the drafting of this thesis.

Summary

The innovative microwave brain imaging system (MWI) allows to carry out a pre-diagnosis of the stroke in an ambulance and a continuous monitoring of bedridden patients, thanks to a non-invasive, easy-to-use, portable and low-cost device. The operating principle of the system exploits the dielectric contrast between healthy and pathological tissues at microwaves frequencies.
The combination of artificial intelligence techniques and the proposed imaging method can effectively assist the clinician in making decisions about the therapeutic treatment of potential stroke patients.

In this regard, my thesis project consists in the development of algorithms capable of solving classification problems. The aim of the work is therefore to identify the presence and the location within the head of the cerebral stroke, distinguishing the cases of ischemia from those of hemorrhage. Classes are detected via supervised Machine Learning Algorithms (ML) as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and K-Nearest Neighbors (k-NN).
The S parameters measured at the antennas ports of the MWI system represent the features that are given as input to the ML algorithms. They are provided in the form of amplitude and of real and imaginary part.

Data collection and processing are two key aspects in the learning process: algorithms need thousands of known examples to identify patterns useful to build a model that is then able to correctly recognize the class of an unknown case. However, carrying out a sufficiently large number of measurements requires a great effort in terms of time. For this reason, the first step was to create a series of synthetic training data, using the Born approximation and performing a linearization of the scattering operator. This method allowed to generate 10000 examples in a very short time. The relative permittivity and conductivity values adopted for the creation of the synthetic training set refer to the dielectric characteristics of the brain tissues at the considered frequencies. Ad hoc mixtures that mimic the dielectric characteristics of both ischemic and hemorrhagic stroke and healthy brain tissue, intended as a homogeneous medium, were created. At this point the tuning of the hyper-parameters, the model construction and the training of ML algorithms were performed.

The second part of the work involved the creation of a testing-set used to evaluate the performance of the previously trained algorithms. It consists of examples much more similar to reality, obtained through full-wave Finite Element Method (FEM) simulations. It came out that all the classifiers can identify the presence or not of the stroke and among the algorithms used, the MLP proved to be the most performing. From the results achieved it is evident that the linearization of the scattering operator is a reasonable approximation.

Future developments will consist in testing ML algorithms on a series of experimental measurements performed with the MWI system and the 3D human head phantom.

Contents

Acronyms

AI	Artificial Intelligence
ASPECTS	Alberta Stroke Program Early CT Score
CAD	Computer-aided-design
CT	Computed tomography
DOI	Domain of Interest
DUT	Device Under Test
EM	Electromagnetic
FEM	Finite Element Method
FN	False Negative
FP	False Positive
GND	Ground
HU	Hounsfield Unit
H_BL	Hemorrhagic Back-Left
H_BR	Hemorrhagic Back-Right
H_FL	Hemorrhagic Front-Left
H_FR	Hemorrhagic Front-Right
I_BL	Ischemic Back-Left
I_BR	Ischemic Back-Right
I_FL	Ischemic Front-Left
I_FR	Ischemic Front-Right
K-NN	K-Nearest Neighbors
ML	Machine Learning
MLP	Multi-layer Perceptron
MRI	Magnetic resonance imaging
MWI	Microwave imaging
NIHSS	National Institutes of Health Stroke Scale
N	No target
NN	Neural Network
PEC	Perfect electric conductor
ReLU	Rectified Linear Unit
RF	Radio Frequency
rTPA	Recombinant tissue plasminogen activator
RX	Receiver
S	Scattering
SVM	Support Vector Machine
TIA	Transient Ischemic Attack
TN	True Negative
TP	True Positive
TSVD	Truncated singular value decomposition
TX	Transmitter
VNA	Vector Network Analyzer

Chapter 1

Introduction

1.1 Brain stroke

Brain stroke is a pathology affecting the cerebrovascular system. It is the third leading cause of death in industrialized countries after tumors and cardiovascular diseases, and the first leading cause of long-term disability [1].
The World Health Organization (WHO) has defined stroke as "*a syndrome characterized by the sudden and rapid development of symptoms and signs referable to focal and/or global deficit of brain functions, which persist for more than 24 hours or lead to death, not attributable to any other apparent cause than the vascular one* "[2].
There are different types of stroke (Fig.1.1):

- **Ischemic stroke**: occurs when an artery supplying the brain is blocked by the formation of an atherosclerotic plaque (*thrombotic stroke*), or by a blood clot coming from the heart or other vascular area (*thrombo-embolic stroke*). About 80% of all strokes are ischemic [3].
  During the ischemic event, cellular energy (adenosine triphosphate) is lost due to a reduction in glucose and oxygen.
  The functionality of the Na^+/K^+ ATPase pumps is compromised: the accumulation of potassium ions outside the cell causes a depolarization of the plasma membrane of neurons. At this point, calcium ions enter the cell through the membrane channels, stimulating the release of glutamate into the extracellular space and promoting excitotoxicity. The cell activates the calcium dependent enzyme to extract excess calcium. The metabolic products damage the membrane wall and the cytoskeleton of the neuron, leading the cell itself to apoptosis [4].
  Ischemia can be prolonged or transient, complete, or incomplete. When the ischemia is transient and incomplete (<15 min.), the brain tissue does not suffer any permanent damage. The *transient ischemic attack* (TIA) differs from ischemic stroke for the shorter duration of symptoms (less than 24 hours) [5].

- **Hemorrhagic stroke**: occurs when an artery located in the brain ruptures, thus causing non-traumatic intracerebral hemorrhage (this form represents 15-20% of all strokes) or in the sub-arachnoid space (this form accounts for approximately 3-5% of all strokes). This can be caused by high blood pressure and weak arteries [3].

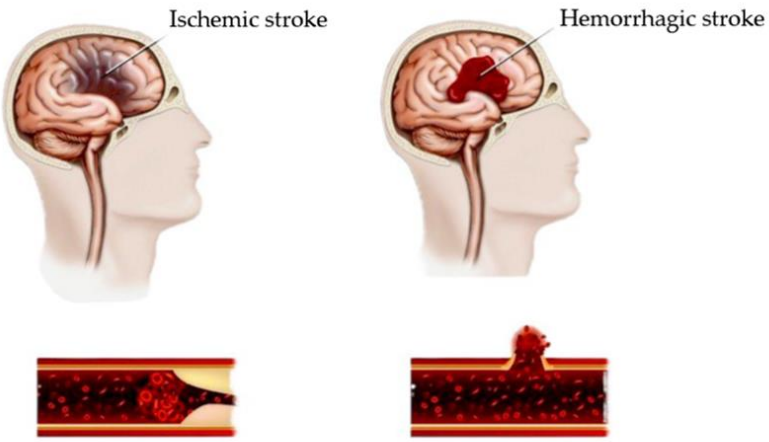


Figure 1.1. Distinction between ischemic and hemorrhagic stroke: the first occurs when a brain vessel is clogged; the second happens when there is a bleeding.

A stroke causes the death of nerve cells and consequently the neurological functions controlled by the affected area are compromised.
It has been estimated that about 1.9 million neurons and 14 billion synapses are lost every minute since stroke onset. For this reason, the expression "*time is brain*" is often used. In most hemispheric strokes due to atherothromboembolism of the great vessels, the damage is completed in about 10 hours. It is believed that, except for exceptional cases, there is no more salvage brain tissue after about 8 hours. The table 1.1 below shows some estimates that correlate time with associated damage [6].

	Neurons Lost	Synapses Lost	Myelinated Fibers Lost	Accelerated Aging
Per Stroke	1.2 billion	8.3 trillion	7140 km/4470 miles	36 y
Per Hour	120 million	830 billion	714 km/447 miles	3.6 y
Per Minute	1.9 million	14 billion	12 km/7.5 miles	3.1 wk
Per Second	32 000	230 million	200 meters/218 yards	8.7 h

Table 1.1 Estimated pace of neural circuitry loss in typical large vessel due acute ischemic stroke [6].

Unfortunately, stroke is a very common condition. The American Heart Association, in collaboration with the National Institutes of Health, annually reports the most updated statistics relating to diseases affecting the cardiovascular system. It is estimated that 15 million people suffer from strokes every year worldwide: 1/3 of them die, while the 50% who survive will need long-term care [6].
According to reports from the Ministry of Health in Italy, 196,000 stroke cases would occur every year, of which 80% are new episodes and the remaining 20% relapses. 20-30% of people affected by stroke die within one month of the event and 40-50% within the first year. Only 25% of stroke survivors recover completely, while 75% survive with some form of disability. 50% of them completely lose self-

sufficiency [7]. Because of its high incidence, stroke is a welfare, social and economic problem. The cases of stroke are in fact constantly increasing due to the aging of the population. Improving the effectiveness of preventive, therapeutic and care measures can significantly reduce the incidence and mortality of cerebrovascular events [8].

1.2 Risk factors and symptoms

Ischemic and hemorrhagic stroke share several risk factors: hypertension and atrial fibrillation are the most common. It was estimated that the first affects the development of stroke for 30-40%, while the second for 5% [9].

The probability increases for men, especially if they have first-degree relatives also suffering from a stroke, or TIA, or with other vascular problems. Age is also an important risk factor to be considered: elderly people are more prone to developing a stroke than young people, because of their medical conditions. 95% of stroke cases involves people aged 45 or over, while 2/3 are over 65 years old [10]. The use of contraceptives for women and sickle cell anemia for young patients (under the age of 20) are also risk factors. Among the most common diseases, diabetes mellitus can lead to stroke, considering the associated high cholesterol levels and obesity. Cigarette smoke irreversibly damages the walls of blood vessels, promoting the formation of atherosclerotic plaques and favoring the aggregation of platelets. Alcohol also has negative effects on the cardiovascular system by decreasing the level of vitamin B1, the deficiency of which can cause damage to both the cardiovascular and nervous systems. Finally, sedentary lifestyle, stress, and lack of physical activity related to incorrect nutrition can negatively affect the patient's general clinical conditions, favoring the onset of stroke [8].

For hemorrhagic stroke risk factors include vascular malformations such as aneurysms, angioma, and thrombophilic disorders. [11].

The symptomatology of the stroke patient varies according to the damaged area.

While for the ischemic stroke, whether thrombotic and/or embolic, it produces a focal type symptom, the hemorrhagic stroke produces a wider and more severe effects, due to the expansion of the blood content in other areas of the brain through the subarachnoid pathway.

A stroke patient often experiences motor aphasia, personality problems, numbness of the contralateral leg and paresis or facial paralysis, with asymmetry of the mouth. Strokes can cause sudden loss of coordination of movements and a sudden and severe headache. Visual disturbances in one or both eyes and partial or total reduction of the visual field may occur. In the most serious cases there may be an alteration of the state of consciousness [9]. It is important to note the time of onset of the first symptoms, because some treatments can only be done within a certain time frame.

1.3 Therapeutic treatment

The patient who manifests the symptoms of stroke is rushed to centers that, presumably, have adequate equipment and highly specialized personnel (Unit stroke). There is a standardized procedure that involves stabilization of the airways, breathing and circulation (ABC). Thereafter, anamnesis and a neurological evaluation are made to first assess the patient's level of consciousness. Assessment is usually done with the so-called National Institutes of Health stroke scale (NIHSS) [12]. A score ranging from 0 to 4 is assigned for each of the stroke-related symptoms: the higher the score, the more severe the situation. To check for dysarthria and ataxia, for example, the patient is asked to smile, to touch the tip of the nose, to raise a limb. The scale is also used in the monitoring phase to keep track of any improvements or worsening.
Following the physical examination, blood tests and instrumental tests are performed in order to trace the causes, classify the type of stroke.
Imaging techniques allow to identify the *core* and the tissue in *penumbra:* the first is the area irreversibly affected by the injury, the second is the area damaged reversibly, and therefore potentially savable [13].
Stroke treatment differs depending on whether it is ischemic and hemorrhagic stroke.
In the case of ischemic stroke, perfusion to the brain tissues must be re-established by removing the clot that causes the obstruction and preventing it from spreading. This can happen in two ways. The first is called fibrinolysis and consists in removing the clot by administering antiplatelet and anticoagulant drugs.
The administration of the drug, called tissue plasminogen activator rTPA, is a possible option only if no more than 4.5-6 hours have elapsed from the onset of symptoms. After this time window, it is not possible to receive the rTPA because the risks outweigh the benefits.
The rtPA is prepared during the imaging phase and, once the absence of bleeding is ascertained, it is injected by intravenous (IV) administration.
The second - the most innovative technique - involves the surgical removal of the thrombus, through an operation called thrombectomy. The minimally invasive surgery consists of inserting a catheter into the femoral artery that reaches the area of the brain where the obstruction is present.

The treatment of hemorrhagic stroke consists in stopping the bleeding by administering drugs with a coagulating action. Blood loss should be limited and controlled to reduce pressure which can damage the brain tissues surrounding the ruptured area of the vessel. If the bleeding is small, it can spontaneously reabsorb within a certain time, otherwise if the blood loss has been significant, the leaking blood must be surgically removed [14].

Hemorrhagic stroke may require surgery such as classic craniotomy, aneurysm clipping and removal of the arteriovenous malformation.

Once the therapeutic treatment has been carried out, the patient must be kept under observation. After 24 hours, the NIHSS exam is repeated, and it would be advisable to repeat the instrumental tests to verify that the core area has not extended and that the treatments begin to give the desired results.

Physical and psychological rehabilitation for a stroke patient is a fundamental and obligatory step. The rehabilitation process can allow the recovery of some motor, coordination, and language faculties.

It is desirable that the prognosis of these patients not only depends on the timing of the intervention, but also on the brain area or areas involved in the vascular event, the age of the patient and the general condition of the patient himself before the stroke [15].

1.4 Thesis focus

In recent years, more and more devices, both in the medical and non-medical fields, incorporate features that have to do with artificial intelligence (AI). Machine Learning (ML) techniques, for example, have been used for diagnostic systems aiming to recognize early breast tumors, or to detect the presence of contaminants within food [16-17].

The *Wavision Research Group* of the Department of Electronics and Telecommunications (DET) at the Politecnico di Torino has developed an innovative microwave imaging system (MWI) for stroke diagnosis and monitoring. Thanks to a helmet of 24 receiving and transmitting antennas, it is possible to reconstruct the images of the head and identify the injured portion of the brain, exploiting the dielectric inhomogeneities of the tissues at microwave frequencies [18]. The first task is to recreate in the laboratory a mixture that mimics the dielectric characteristics of the healthy brain and that fills the 3D head anthropomorphic phantom, used for the measurements.

The goal of the thesis project is to combine AI with the new MWI, in view of its future use on patient, reducing the intervention time significantly.

To be more specific, ML algorithms should be able to identify the region affected by ischemia or bleeding, simply by "looking" at the parameters recorded by the measurement imaging system, (Fig.1.2).

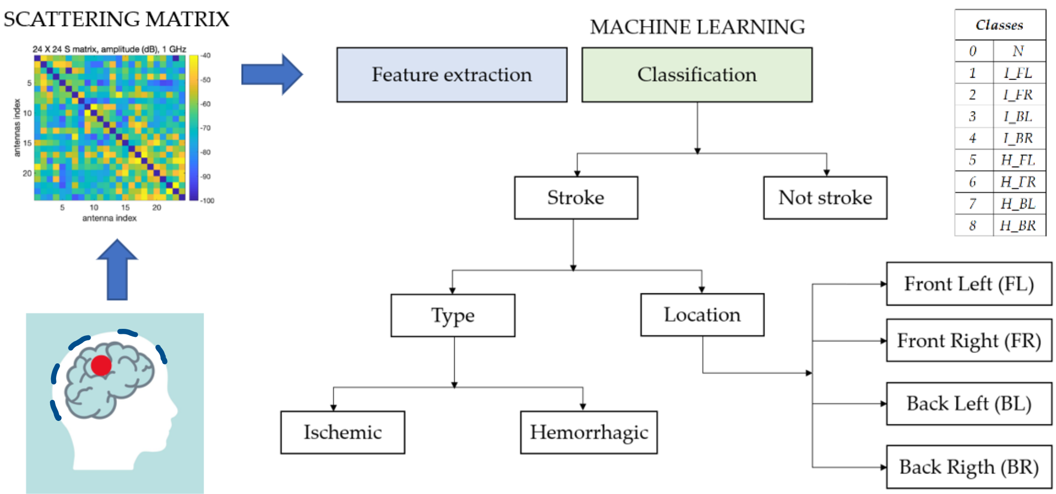


Figure 1.2 Brain stroke classification problem whit Machine Learning technique.

However, ML algorithms need a very large training dataset to learn patterns useful for classifying and detecting stroke. Clinical data, laboratory measurements with anthropomorphic phantoms or electromagnetic (EM) simulations do not represent a feasible solution because they require great efforts in terms of time.

Therefore, a method for rapidly generating the training set will be validated.

1.5 Thesis outline

The thesis is organized as follows.
Chapter 2 describes the traditional imaging technologies used to diagnose stroke: Computed Tomography and Magnetic Resonance Imaging. It highlights pros and cons and briefly explains their principles of operation.
Chapter 3 focus on Microwave brain Imaging system. In the first part it describes the basis of electromagnetism and then it discusses each element of the entire prototyping tool and illustrates the image reconstruction algorithm.
Chapter 4 is about the 3D head anthropomorphic phantom. It describes the process that led to the creation of the mixtures that mimics the dielectric characteristics of the healthy brain tissue.
Chapter 5 is an introduction to the world of Machine Learning. It briefly describes the three models used as classifiers in this project: Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and K-Nearest Neighbors (K-NN). Finally, it illustrates the problem of overfitting and performance evaluation metrics.
Chapter 6 details the steps of the implemented method to generate the training set.
Chapter 7 schematically shows the flowchart of the code used for the training set creation.
Chapter 8 deals with the steps followed for the generation of the synthetic testing set via Full-Wave Finite Element Method simulations.
Chapter 9 is about the implementation of the three ML algorithms in Python, on the Google Colab platform. It illustrates the steps taken to solve the classification problem: feature scaling, tuning, training, and test phases.
Chapter 10 shows the results obtained by the classifiers.
Finally, Chapter 11 summarizes the thesis, discusses the relevant results, and anticipates the possible future developments.
Chapter 12 contains bibliographic references and sitography.

Chapter 2

Traditional Imaging Techniques

The traditional biomedical imaging techniques, used in case of a suspected stroke, are computed tomography (CT) and magnetic resonance imaging (MRI). CT and MRI base their operating principle on the interaction that occurs between the type of electromagnetic wave and the irradiated tissue: in the first case an X-ray source is adopted, while in the second are used magnetic fields and radio waves.
The high cost and large size are two aspects shared by both imaging methodologies. CT exposes the patient to ionizing radiation while MRI is harmless. Due to its design, however, the MR device is not suitable for all patients as it can be claustrophobic. CT is able to perform much faster scan and it is more readily available in the hospital. Both techniques may involve the use of contrast media, which can cause potential kidney toxicity or allergies [19]. Table 2.1 summarizes the strengths and weaknesses of CT and MRI techniques.

CT		MRI	
high cost	✗	high cost	✗
harmful (ionizing radiation)	✗	harmless	✓
not portable	✗	not portable	✗
large size	✗	large size	✗
not available for bedside monitoring	✗	not available for bedside monitoring	✗
		not always available in the hospital	✗
fast acquisition	✓	slow acquisition	✗
		good spatial resolution	✓
		claustrophobic	✗

Table 2.1. Comparison of CT and MRI: pros (✓) and cons (✗).

The diagnostic images obtained by CT or MRI allow clinicians to:
1. Distinguish the ischemic case from the hemorrhagic one: understand, therefore if the patient is a candidate for the tissue plasminogen activator (rtPA);
2. Know the position of the stroke within the brain;
3. Identify stroke etiology;
4. Estimate the relative volume of penumbral regions that can be saved with timely reperfusion therapy [15].
The dynamic nature of the stroke requires continuous monitoring. In the case of ischemic stroke, if left untreated in time, occlusion of the vessel, due to the pressure exerted by the blood, can cause its rupture causing severe bleeding. Neuroimaging also makes it possible to predict outcomes, plan rehabilitation and prevent early secondary stroke.

2.1 Computed Tomography (CT)

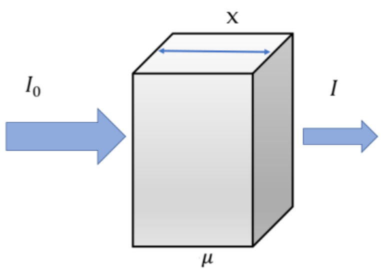
Figure 2.1. Computed Tomography (CT) device [20]

Nowadays Computed tomography (CT) is recognized as the first-instance neuroradiological method, because it allows to discriminate the ischemic event from the hemorrhagic one in a short time.

CT is a diagnostic tool that bases its operating principle on the Lambert-Beer law. The following equation (2.1) shows the relationship between the intensity of the monochromatic incident radiation (I_0), and the (attenuated) intensity of the outgoing radiation (I) [21]:

$$I = I_0 \cdot e^{-\mu x} \tag{2.1}$$

where x is the thickness of the medium and μ the attenuation coefficient (Fig.2.2).



Figures 2.2. Schematic representation of the Lambert-Beer law.

An x-ray source rotates around the subject's head emitting a beam or more beams that pass through the tissues and collimate on sensors located on the opposite side (Fig.2.1). There are at least two acquisition modes: stop & shut and spiral CT. The latter mode is feasible only if the CT device is equipped with slip-ring technology.

X-ray detectors collect photons that are not absorbed by the tissues and obtain the attenuation coefficient in each point of the head with sufficient precision. Water is regarded as a reference material.

In computed tomography a normalized version of the attenuation coefficient called the Hounsfield Unit (HU) is represented (2.2):

$$\mu\,(HU) = 1000\,\frac{\mu - \mu_{H_2O}}{\mu_{H_2O}} \tag{2.2}$$

The equivalent HU of a tissue that has a linear attenuation coefficient μ, is equal to the product between 1000 and the difference between μ of that tissue and μ of water, divided by the attenuation coefficient of water.

All tissues that have a HU lower than water are materials in which there is a significant percentage of fat, or which have a predominant gas content.

The following Table 2.2. shows the ranges of attenuation coefficients in HU for some materials of interest.

Materials		HU
Water		0
Air		−1000
Skin		-770 to -200
Fat		−120 to −90
Soft tissue on contrast CT		+100 to +300
Bone	Cancellous	+300 to +400
	Cortical	+500 to +1900
Subdural hematoma	First hours	+75 to +100
	After 3 days	+65 to +85
	After 10–14 days	+35 to +40
Other blood	Unclotted	+13 to +50
	Clotted	+50 to +75
CSF		15
White matter		+20 to +30
Grey matter		+37 to +45
Skin		-770 to -200

Table 2.2 Range of the attenuation coefficient expressed in HU for some materials of interest [22].

The coding of the images in HU allows to have a superior contrast between tissues that also have similar μ. The numerical output is an image that associates the gray tones with HU units. The filtered rear projection algorithm allows to reconstruct the 3D volume of the head by superimposing all the slices.

CT has good specificity ranging from 56% to 100% and low sensitivity ranging from 20% to 75% in detecting early ischemic changes. In complete neurological deficits, the radiological signs become evident after a few hours and depend on the nature and severity of the vascular obstruction. The first signs of ischemic infarction cause low densitometric attenuation values due to cytotoxic edema [23].

The Alberta Stroke Program Early CT Score (ASPECTS) allows clinicians to quantify ischemic changes in the anterior circulation. This approach consists of dividing the middle cerebral artery (MCA) into 10 regions: if the area has ischemic signs, such as focal swelling or parenchymal hypoattenuation, one point is subtracted. If the score is equal to 10 the CT is normal, if it is equal to 0 it indicates diffuse ischemic [15].

According to the American Heart Association guidelines on the management of acute stroke, endovascular therapy is recommended in patients with ASPECT greater than 6 [24].

In the case of a hemorrhagic stroke, extravasated blood appears as a clear hyperdensity with a HU value of around 50, interpreted as the effect of the hemoglobin concentration [22].

Although CT allows for rapid acquisitions, it exposes the patient to ionizing radiation, and therefore it is not a particularly suitable technique for continuous monitoring. Figure 2.3. shows, by way of example, two images obtained with CT of a patient with ischemic stroke (left) and hemorrhagic stroke (right) [25,26].

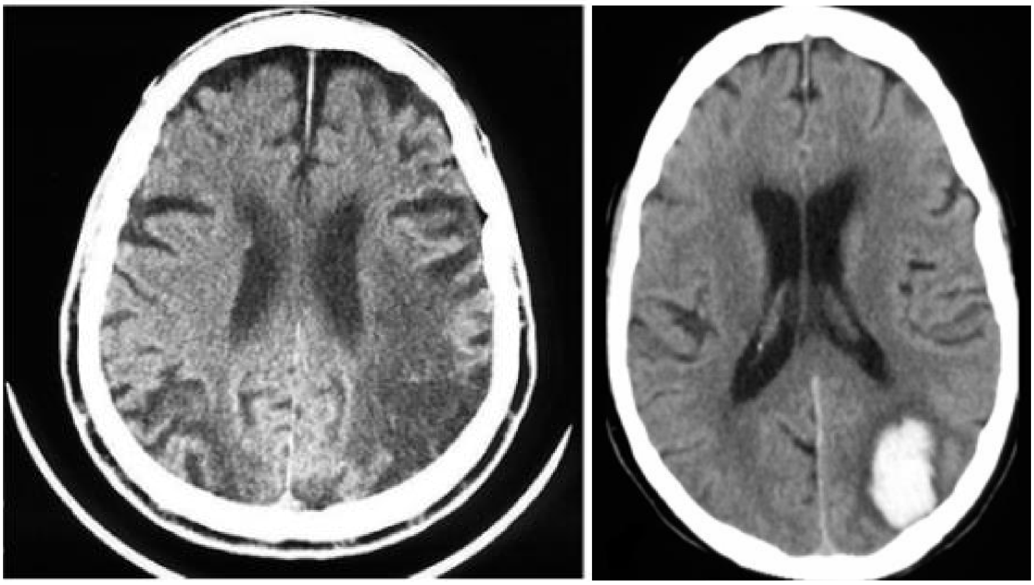


Figure 2.3. Example of CT images: acute ischemic infarction on the left and hemorrhagic stroke on the right [25,26].

2.2 Magnetic Resonance Imaging (MRI)

Figure 2.4. Magnetic Resonance Imaging (MRI) device [27]

The magnetic resonance (MR) device (Fig.2.4) bases its operating principle on the use of an intense, homogeneous, and static magnetic field produced by a large magnet (0.2 - 7 T). The addition of a second radiofrequency (RF) magnetic field, variable in time and space, modifies the orientation of the hydrogen atoms present in the single cells of the anatomical of the anatomical area under examination. Once the modification has taken place, the gradient fields are deactivated. The hydrogen atoms begin to move to restore the original arrangement: during this phase the protons release their excess energy at different times, depending on the type of tissues in which they are immersed. Therefore, the (proton) density of a tissue will be proportional to the percentage of aqueous content. The released energy is captured by special detectors and the signals are processed to reconstruct a volume[28].

MRI produces very high-resolution images and has a high sensitivity and specificity in the diagnosis of acute ischemic infarction in the first hours after onset [29]. When combined in various sequences such as Diffusion Weighted Imaging (DWI) Fluid Attenuated Inversion Recovery (FLAIR) and Gradient echo (GRE), MRI allows clinicians to acquire a large amount of information about stroke.

Based on the images acquired in DWI it is possible to obtain the so-called ADC, which is the apparent diffusion coefficient of water. The ADC value (opposite to that of DWI) reflects the diffusivity of the tissue: the values are high if the diffusivity is high (water/liquor), while they are low when physical obstacles prevent water from moving freely (white/gray matter) [30].

In the healthy brain, the white and gray matter ADC is similar, but that of the latter is slightly higher. During an ischemia, the cell undergoes a swelling which causes a reduction in volume and an increase in the tortuosity of the extracellular space, which hinders the diffusivity of the water molecules. In DWI images a hyperintense zone followed by a reduced ADC map represents irreversible

ischemia [15]. FLAIR and T2-weighted images show a hyperintense lesion in the case of ischemic infarction. Usually, these sequences can show the lesion if imaging occurs within 3-8 hours of the start of the stroke (Fig.2.6) [28].
Finally, the mismatch between DWI and FLAIR can provide an estimate of the time to onset of stroke. Lesions visible on a DWI image, but not evident on the FLAIR scan, indicate that the onset of the stroke was less than 4.5 hours ago [30,31].
Although MRI allows for very detailed images to be acquired, it is extremely slow and not always available in the hospital.
Figure 2.5. shows, by way of example, two images obtained with MRI of a patient with ischemic stroke (left) and hemorrhagic stroke (right) [32].

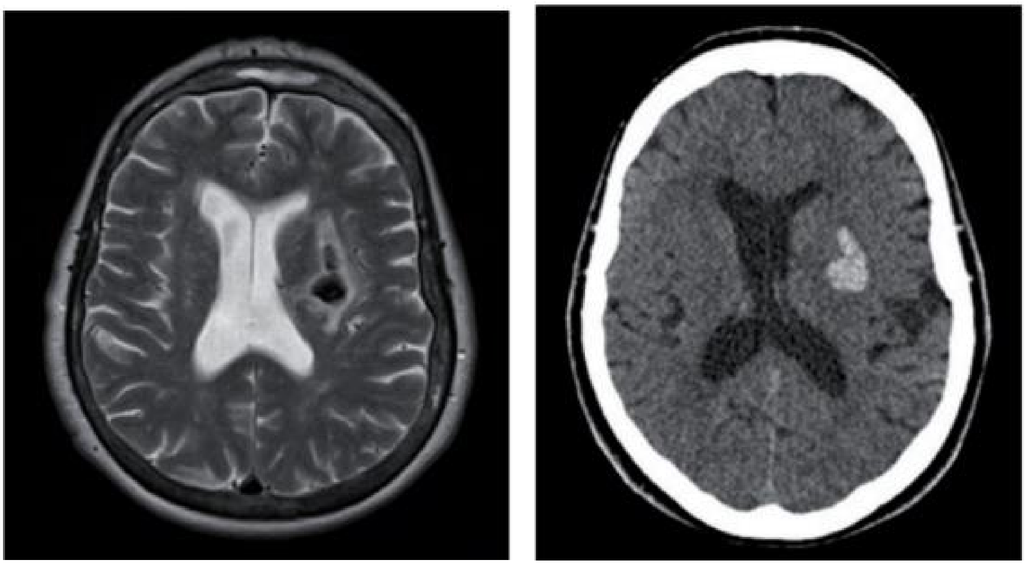


Figure 2.5. Example of MR images: ischemic infarction on the left and hemorrhagic stroke on the right [32].

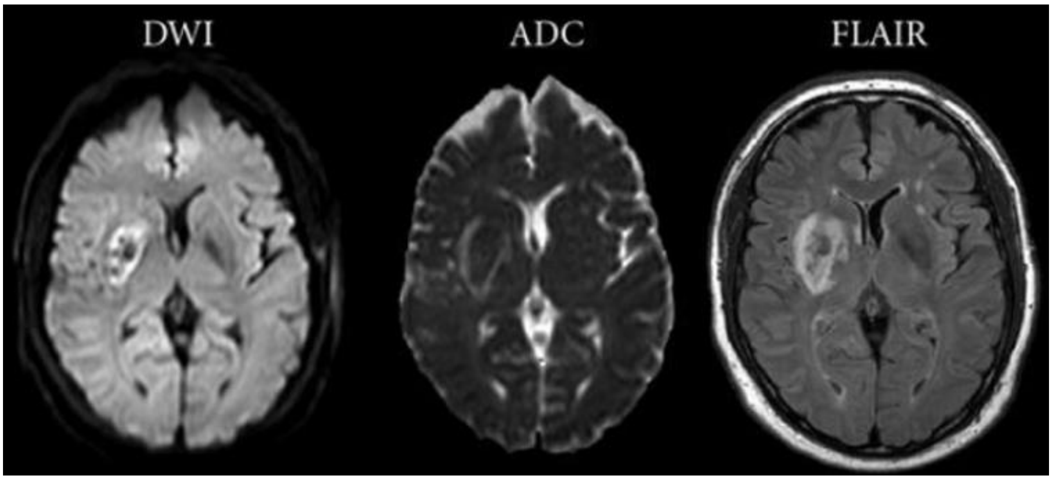


Figure 2.6. Examples of images obtained through different sequences combined with MR: DWI, ADC and FLAIR. The clearest area in DWI and FLAIR images highlights the core of the ischemic stroke [31].

Chapter 3

Microwave brain imaging system

The previous chapter highlighted the pros and cons of traditional imaging techniques. As seen, CT and MRI can effectively identify the location and type of stroke, but they are not suitable to be used for continuous monitoring and they do not allow to carry out a pre-diagnosis directly in the ambulance or at the accident site.
The innovative Microwave Brain Imaging System (MWI) prototype, developed at the DET of the Politecnico di Torino, aims to overcome these limits [18].
In this perspective, microwave technology offers completely new possibilities for optimizing the treatment of stroke patients.
A pre-hospital stroke diagnosis reduces the time from injury to therapy and it helps to predict which patients need acute intervention and which patients might be safely transported to a non-specialized center.
The device (Fig.3.1) is non-invasive, easy to use, portable, low-cost and it is proposed as a complementary tool to existing imaging techniques.

Figure 3.1 A Prototype Microwave System for 3D Brain Stroke Imaging developed at the DET of the Politecnico di Torino: the photo shows the helmet of antennas surrounding the 3D human head phantom and the switching matrix.

MWI system consists of a helmet made of plastic material where 24 antennas are hosted. The antennas can work as both transmitters (TX) and receivers (RX) of electromagnetic (EM) waves at a frequency of 1GHz. The switching matrix has the task of selecting an antenna as a transmitter, leaving the other 23 as receivers. The Vector Network Analyzer (VNA) derives the scattering parameters (S parameters) in the form of complex numbers with a real part and an imaginary part. The attenuation and phase shift of the wave passing through the head, are obtained from the ratio between the transmitted and received signals [33].
The measured scattering parameters due to a full scan are the starting point of the image reconstruction algorithm. The mechanism is based on transmission through the tissues and it exploits the variation in dielectric properties due to the presence of stroke.
The main components of the microwave imaging system are shown in Figure 3.2.

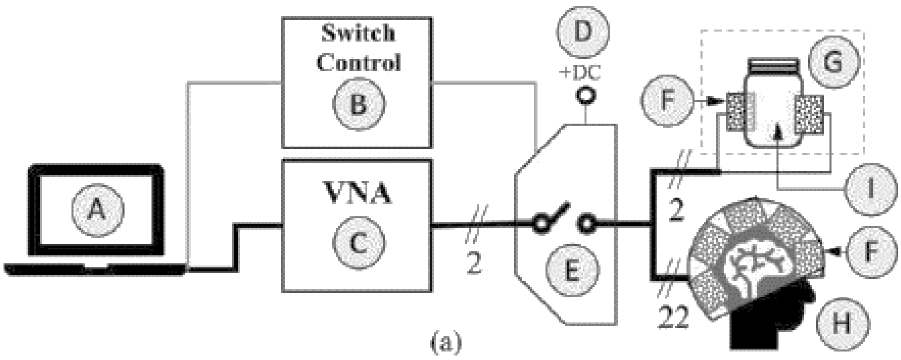


Figure 3.2. Block diagram of the main components of the MWI system [34].

Before going into the details of each individual component of the MWI system, a brief review of the theory of electromagnetism is made [35,36].

3.1 Electromagnetism overview

Electromagnetism is branch of physics which studies the interactions between electric and magnetic fields [35].
The properties of the electromagnetic (EM) fields are described by physical quantities which have the nature of vector fields. These time-variant/frequency-dependent quantities are listed below:

$\vec{E}$ is the electric field measured in $\left(\frac{V}{m}\right)$
$\vec{H}$ is the magnetic field measured in $\left(\frac{A}{m}\right)$
$\vec{D}$ is the electrical induction or electrical displacement, measured in $\left(\frac{C}{m^2}\right)$
$\vec{B}$ is the magnetic induction or magnetic displacement, measured in $\left(\frac{Wb}{m^2}\right)$
$\vec{J}$ is the density of electric current, measured in$\left(\frac{A}{m^2}\right)$
$\vec{\rho}$ is the electric charge density, measured in $\left(\frac{C}{m^3}\right)$

The electromagnetic problem can be described through the Maxwell equations, reported below in differential form (3.3-3.6):

$$\nabla \cdot \vec{D} \tag{3.3}$$

$$\nabla \cdot \vec{B} = 0 \tag{3.4}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{3.5}$$

$$\nabla \times \vec{H} = \vec{J} + \frac{\partial \vec{D}}{\partial t} \tag{3.6}$$

These four equations show that electric field and magnetic field are closely related. In fact, a variable magnetic field induces a variable electric field and vice versa, and these variable fields propagate as an EM wave through space, mutually perpendicular to each other. An electromagnetic wave is characterized by frequency, wavelength, speed, and a direction of propagation [35,36].
The electromagnetic spectrum shown in Figure 3.3. represents the set of all possible frequencies of electromagnetic radiation.
The whole spectrum is conventionally divided into bands: the microwave one is in the frequency range between 10^{10} Hz and 10^{12} Hz (300 MHz – 300 GHz).

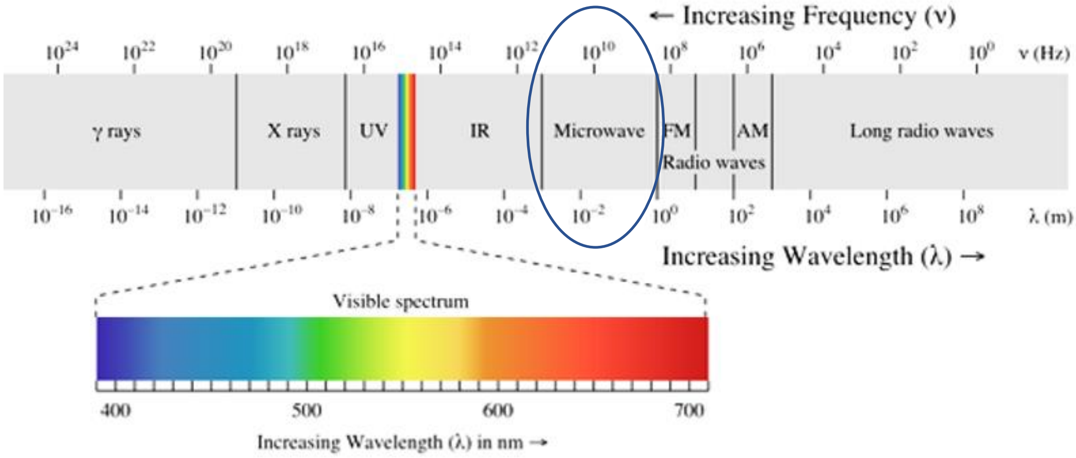


Figure 3.3. Electromagnetic spectrum [37].

Materials are conventionally divided into conductors and dielectrics. The subdivision is based on the behavior that the materials have when they are exposed to an EM field. The response to an external excitation is determined by the atomic structure of an element: the arrangement of free external electrons for the electric field and the atomic moments for the magnetic one. In the conductors the charges are free to move, while in the dielectrics they encounter a certain resistance.

Dielectric properties of a material in a vacuum space are described by (3.7-3.9):

$$\varepsilon_0 = 8.854187 \cdot 10^{-12} = \frac{10^{-9}}{36\pi} \quad \left(\frac{F}{m}\right) \tag{3.7}$$

$$\mu_0 = 4\pi \cdot 10^{-7} \quad \left(\frac{\mathrm{H}}{\mathrm{m}}\right) \tag{3.8}$$

$$c_0 = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \approx 3 \cdot 10^8 \quad \left(\frac{m}{s}\right) \tag{3.9}$$

The electrical permittivity ε_0 defines the capacitance of a material exposed to an electrical field to store electrical energy. The magnetic permeability μ_0 expresses the ability of the material to become magnetized in the presence of a magnetic field. c_0 is the light speed.

Dielectric properties of a material vary as a function of the frequency of the EM field applied. They are usually expressed as $\tilde{\varepsilon}$ or $\tilde{\sigma}$: the first is the relative complex permittivity (3.10) and the second is the complex conductivity (3.11):

$$\tilde{\varepsilon} = \varepsilon + \frac{\sigma}{j\omega} = \varepsilon' - j\,\varepsilon'' \tag{3.10}$$

$$\tilde{\sigma} = \sigma + j\,\omega\,\varepsilon = \sigma' - j\,\sigma'' \tag{3.11}$$

where

$\varepsilon^{'} = \varepsilon_{rR}$ ($\tilde{\varepsilon}$ real part: dielectric constant);

$\varepsilon^{''} = \varepsilon_{rI}$ ($\tilde{\varepsilon}$ imaginary part: loss factor);

$\omega = 2\pi f$

ω is the angular frequency $\left(\frac{rad}{s}\right)$. The imaginary part of $\tilde{\varepsilon}$ ($\varepsilon^{''} = \varepsilon_{rI}$) is an equivalent conductivity, and thus data are usually reported as (3.12):

$$\varepsilon_r = \frac{\varepsilon'}{\varepsilon_0} \tag{3.12}$$

Sometimes the loss-tangent $\frac{\varepsilon''}{\varepsilon'}$ is also considered. For the sake of clarity, the previous expressions can be written as (3.13-3.14):

$$\varepsilon = \varepsilon_0 \varepsilon_r = \varepsilon_0(\varepsilon_{rR} - j\varepsilon_{rI}) \tag{3.13}$$

$$\sigma_{eq} = \sigma' = \sigma - \omega\,\mathrm{Im}(\varepsilon) = \sigma + \omega\,\varepsilon_0 \varepsilon_{rI} \quad \text{(S/m)} \tag{3.14}$$

To the Maxwell equations mentioned previously (3.3-3.6), the following constitutive relations of the materials are added (3.15-3.17):

$$\vec{D} = \varepsilon \vec{E} \tag{3.15}$$

$$\vec{B} = \mu \vec{H} \tag{3.16}$$

$$\vec{J} = \sigma \vec{E} \tag{3.17}$$

The table below summarizes the values of relative dielectric permittivity and conductivity at the frequency of 1 GHz for each tissue in the human head [38]

Name tissue	ϵ_r / σ[S/m]
Cerebellum	48.86 / 1.308
CSF/Ventricle	68.44 / 2.455
Fat	5.47 / 0.053
Grey matter	52.28 / 0.985
Skin	40.94 / 0.089
Skull	12.36 / 0.156
White matter	38.58 / 0.622

Table 3.1. Dielectric properties of all the head tissues at 1 GHz [38].

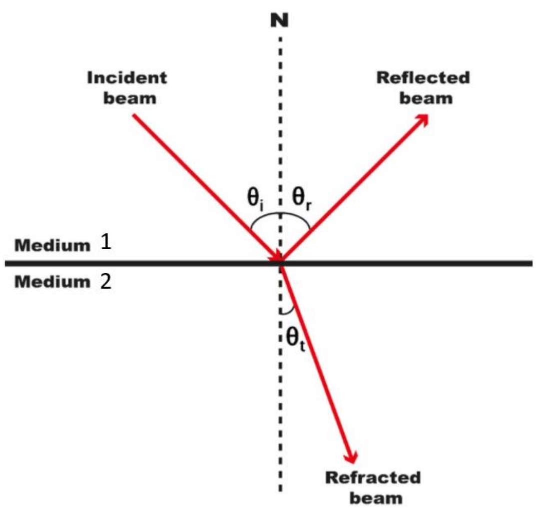


Figure 3.4. Snell representation. Wave reflection and transmission [39]

Figure 3.4. schematically shows the behavior of an EM wave that propagates from a medium 1 to a medium 2 with different dielectric properties. The incident wave that hits the surface of a medium 2 is partly reflected and partly absorbed. The angle of reflection and the angle of incidence are the same (3.18):

$$\theta_r = \theta_i \tag{3.18}$$

Snell's law relates the transmission θ_t and reflection θ_i angles (measured from the normal of the boundary N) to the refractive indices and properties of the materials (3.19):

$$\sin\theta_t = \sin\theta_i \frac{n_1}{n_2} = \sin\theta_i \frac{c_2}{c_1} = \sin\theta_i \frac{\sqrt{\varepsilon_2\,\mu_2}}{\sqrt{\varepsilon_1\,\mu_1}} \tag{3.19}$$

where *c* is the speed of the wave in the medium and *n* the refractive index of materials. The so-called reflection coefficient Γ and transmission coefficient T are also defined:

$$\Gamma = \frac{I_r}{I_i} = \frac{(n_1 - n_2)^2}{(n_1 + n_2)^2} \tag{3.20}$$

$$|T| = 1 - |\Gamma|^2 = \frac{I_t}{I_i} \tag{3.21}$$

Where I_i is the incident field, I_r is the reflected field and I_t the transmitted one.

3.2 Antennas' System

The MWI system is characterized by a very simple and compact hardware, consisting of a system of monopolar antennas mounted on a support that adapts to the patient's head [18].
Antennas are electrical devices capable of transmitting and/or receiving electromagnetic waves and in the case in exam, are immersed in a dielectric matching medium to address the signal, reducing losses. A monopolar antenna consists of a perfect electrical conductor (PEC), mounted perpendicular to a type of conductive surface called ground plane (GND) (Fig.3.5a-b). On PEC faces the only necessary boundary condition (BC) is the vanishing of the tangential electric field. The constitutive relationships, to be considered together with Maxwell's equations(3.3-3.6), are listed below (3.22-3.25):

$$\hat{n}\, x\, \overrightarrow{H_1}\,|_{\Sigma} \;=\; \vec{J_s} \tag{3.22}$$

$$\hat{n}\, x\, \overrightarrow{H_2}\,|_{\Sigma} \;=\; 0 \tag{3.23}$$

$$\hat{n} \cdot \overrightarrow{D_1}\,|_{\Sigma} \;=\; \rho_s \tag{3.24}$$

$$\rho_s \;=-\frac{1}{j\omega}\,\nabla_s \cdot \vec{J} \tag{3.25}$$

where 1 is the external PEC face and 2 is the internal face [35].
One side of the antenna feed line is attached to the lower end of the monopole and the other side is attached to the GND [34]. Each antenna is manufactured with printed circuit technology, ensuring low cost and high repeatability. Monopolar antennas are printed on a standard FR4 slab, an insulating plastic laminate made with a fiberglass fabric and an epoxy resin matrix. The FR4 laminate has a rectangular extension (48x30 mm) and a thickness of 1.55 mm. The relative permittivity is equal to 4.4 and the conductivity value is 0.012 S/m (Fig.3.5c-Fig.3.6c) [40].

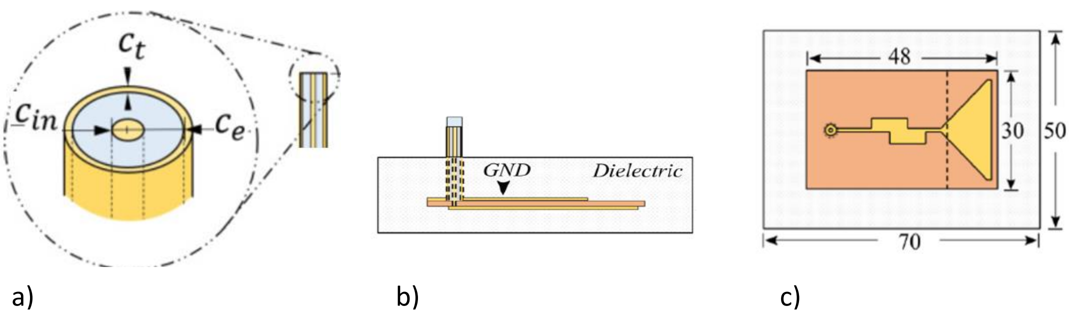


Figure 3.5. Schematic representation of the monopolar antenna used for the MWI system [40][48].

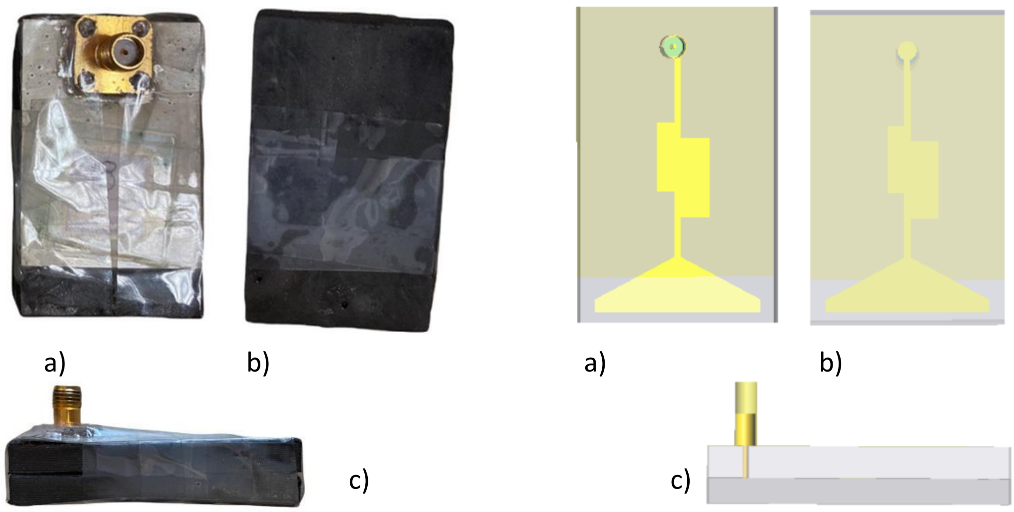


Figure 3.6. Example of MWI system antenna: (a) front, b) back, and c) lateral view); on the left the photos and on the right the 3D CAD models.

The antenna design includes a triangular-shaped radiating element, a trimmed back-placed GND plane, and a line, with two stubs, fed by a rigid coaxial cable on the back [40]**.** The characteristic impedance of the antennas used has been measured and it is equal to 47.3 Ω.

The operating frequency, the coupling medium, the antenna number, and their configuration are the result of a study aimed at respecting the following specifications:

- maximize the spatial resolution;
- maximize the power of the incident radiation field (E_{inc}) in the head, in order to maximize the retro-scattered signal used for the image reconstruction.

The spatial resolution depends on the EM wavelength used (and therefore the frequency): it must have dimensions such as to be comparable with those of the object under test. According to this, it would seem convenient to choose a small wavelength and therefore a very high frequency. However, by increasing the frequency an increase in the conductivity of the tissue is obtained which results in greater losses within the brain. For this reason, it is necessary to find a compromise. A frequency range of 0.5 to 4 GHz was considered as a starting point.

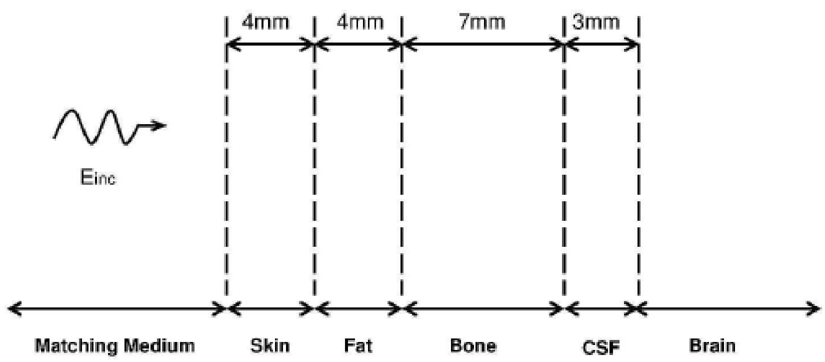


Figure 3.7. The planar layered model of the human head [41].

A layered planar model simulates the interaction that occurs between the incident wave and the tissues present within the head (skin, fat, bone, Cerebrospinal fluid (CSF) and brain) (Fig.3.7). The layered model was then considered as an equivalent transmission line (Fig.3.8).

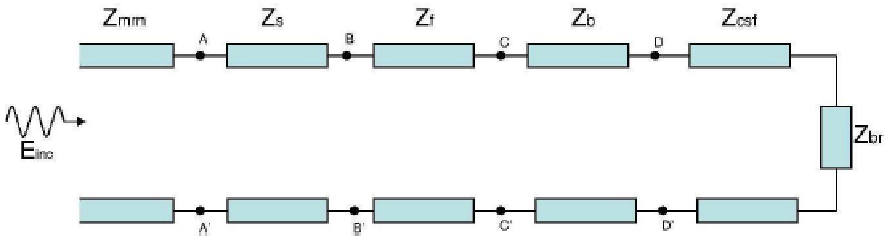


Figure 3.8 Equivalent transmission lines of the multilayer planar model.

The values of the length of the layers and the associated impedance were obtained from the literature. To take into account the frequency dispersive behavior of biological tissues, the calculated impedances were evaluated according to the Cole-Cole model [41].

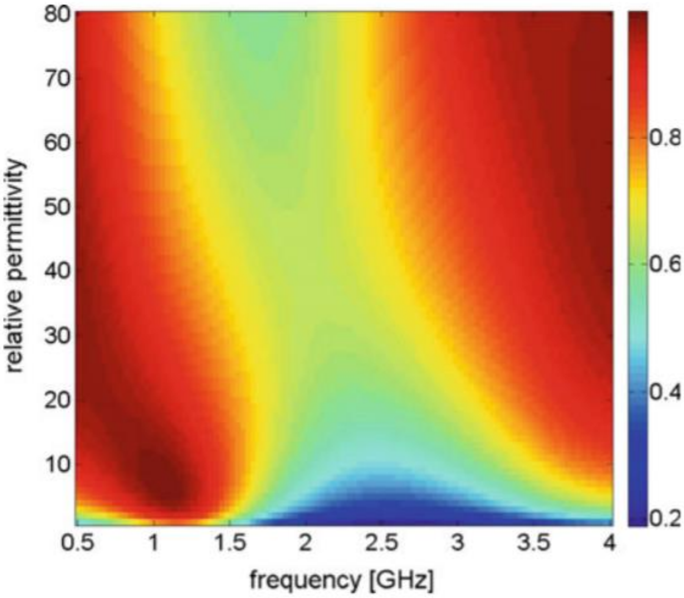


Figure 3.9. Transmittance as a function of the frequency and the permittivity of the matching medium [41].

The graph shows the transmission coefficient as a function of the frequency and relative permittivity of the coupling medium.
The red areas highlighted in Figure 3.9. indicate the frequency ranges for which the transmission is maximum: one is around 1 GHz and the other one, on the right, exceeds 4 GHz. Based on these results, a working frequency of 1GHz and a coupling medium with relative permittivity of about 20 were chosen [18].
The dielectric material was made with a semi-flexible mixture composed of 65% urethane rubber and 35% graphite powder [40]. The compound initially has a liquid appearance, as it cools it tends to solidify (Fig.3.10).

The final material has a relative permittivity of 18,425 and a conductivity of 0.204 S/m [40].

Figure 3.10 Coupling material for the antennas.

The optimal number of antennas that allows to have good quality images is twenty-four. To obtain this result, several full wave simulations have been made. The antennas are distributed over the entire upper surface of the helmet to provide uniform irradiation to the patient's head (Fig.3.11) [33]. The position of the antennas was chosen after having carried out an analysis of the spectral properties of the scattering operator [18].

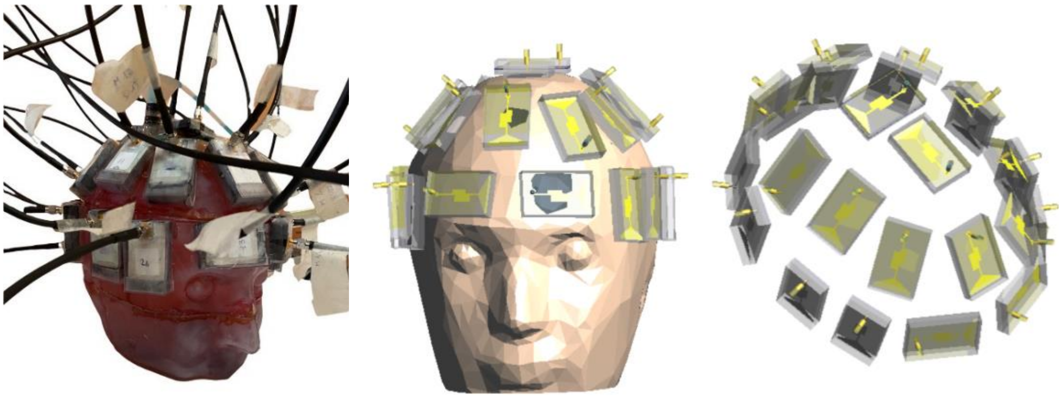


Figure 3.11. Conformal distribution of the antenna array around the head: on the left the developed prototype, in the middle and on the right the CAD model.

3.3 Scattering parameters matrix

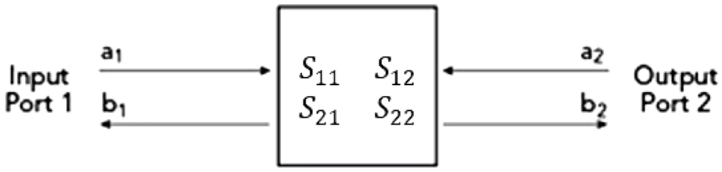


Figure 3.12. Example of a 2- ports microwave circuit and S parameters.

The scattering matrix is a mathematical construct that quantifies the way in which the energy transmitted by the electromagnetic wave (EM) propagates through a multi-ports network [36].

The S matrix accurately describes the properties of even very complex circuits: the device under test (DUT) is considered as a "black box" (Fig.3.12).

For a circuit with N input ports, the matrix contains N^2 coefficients called S parameters. If a circuit is composed of linear elements, it can be described with a set of linear equations. The matrix algebraic representation of the 2-port S parameters is as follows [42] (3.26):

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \times \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{3.26}$$

which in linear form can be written with these two equations (3.27-3.28):

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{3.27}$$

$$b_2 = S_{21}a_1 + S_{22}a_2 \tag{3.28}$$

Variable a_1 represents a wave incident at port i and variable b_j represents a wave reflected from port j. The magnitude of a_i and b_j variables can be considered as voltage, normalized using a specified reference impedance [43].

As seen in paragraph 3.1, an incident EM wave that hits a port is partly reflected outwards, and partly dispersed inside the network. The port from which the wave is irradiated, measures the reflected signal, the other ports instead measure the scattered signal within the network. Amplitude and phase of the incident signal are respectively attenuated and distorted. Each parameter represents a possible input-output path.

The parameters S_{ij} are complex numbers with real and imaginary parts and are defined for a given frequency and impedance Z_0 of the system.

The first number in the subscript (i) refers to the receiving port, while the second number (j) refers to the transmitting port. Therefore S_{12} indicates the response to port 1 due to a signal sent from port 2.

The parameters along the diagonal of the S matrix are called reflection coefficients, while those outside the diagonal are called transmission coefficients [42]. The four S-parameters of a 2-door system are defined as (3.29-3.32):

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2=0} \tag{3.29}$$

$$S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1=0} \tag{3.30}$$

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2=0} \tag{3.31}$$

$$S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1=0} \tag{3.32}$$

S_{11} is the reflected signal measured at port 1 , when port 1 works as a transmitter: no signals are sent from port 2, so $a_2 = 0$ (3.29). To quantify S_{21}, a signal is injected at port 1 and the resultant signal power at port 2 is measured ($a_2 = 0$)(3.30).
For S_{12} calculation must be considered that port 2 acts as a transmitter and port 1 works as a receiver ($a_1 = 0$) (3.31) and finally, for S_{22} a signal is injected to port 2 and its reflected signal is measured ($a_1 = 0$) (3.32) [43].
The S matrix is by construction symmetric (Fig.3.13), and the S parameter can be expressed either as a linear quantity or in logarithmic decibel (dB) by applying the following formula (3.33):

$$S_{ij}\,(\mathrm{dB}) = 20\ \log_{10}(S_{ij}) \tag{3.33}$$

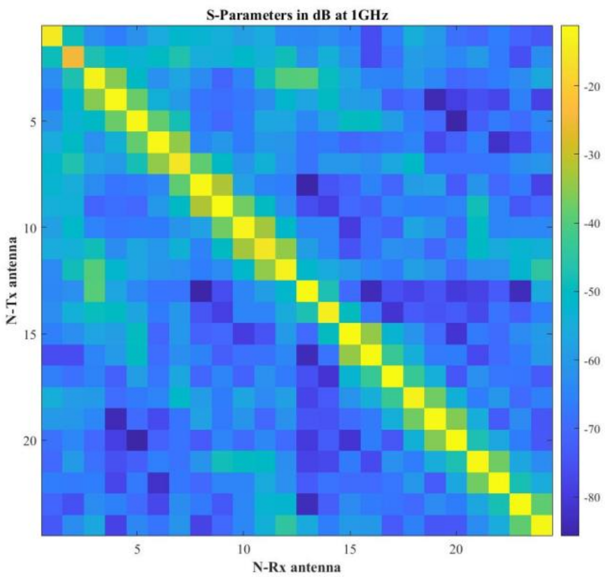


Figure.3.13 Example of a 24x24 scattering matrix in dB.

Furthermore, the S parameters can be obtained from the reflection and transmission coefficients with these two equations (3.34-3.35):

$$S_{11} = \frac{\Gamma(1-T^2)}{1-\Gamma^2 T^2} \tag{3.34}$$

$$S_{21} = \frac{T(1-\Gamma^2)}{1-\Gamma^2 T^2} \tag{3.35}$$

3.4 Vector Network Analyzer (VNA)

The Vector Network Analyzer (VNA) is the device that allows to measure the scattering parameters. The latter, as explained in the previous paragraph, describe the relationship between magnitude and phase of incident and reflected waves at antennas ports and represent the input of the image reconstruction algorithm.
Since the VNA has two ports (one for transmitting and one for receiving), a controller configures the switching matrix in such a way that only one transmitter and one receiver can be simultaneously active and connected to the device [33].
The VNA disposable in the LACE laboratory at Politecnico di Torino is P9375A Keysight Streamline USB Vector Network Analyzer (Fig.3.14) [45].

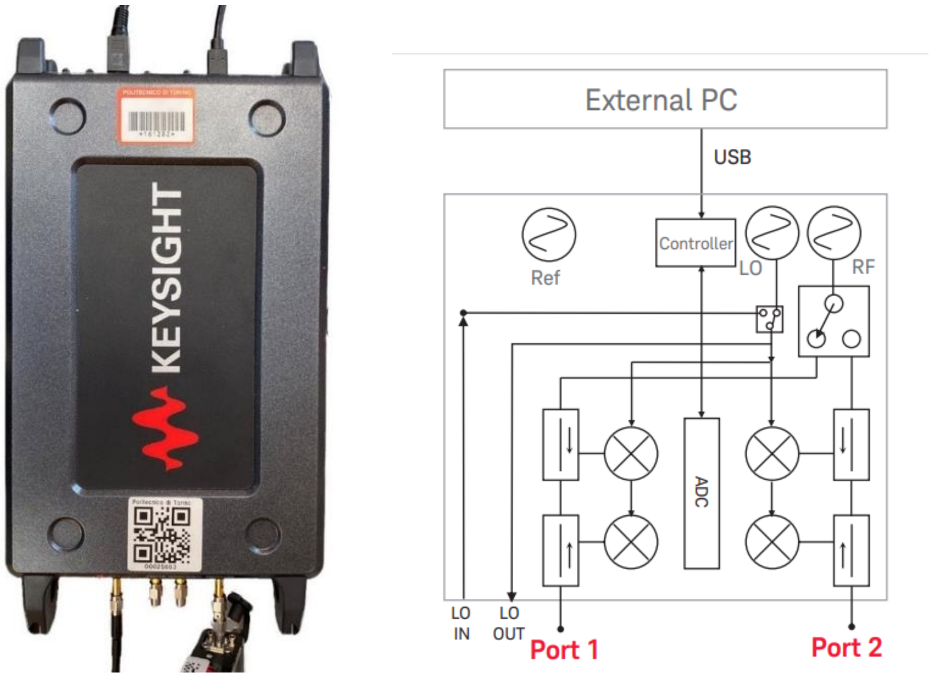


Figure 3.14 P9375A Keysight Streamline USB Vector Network Analyzer: a photo of the device on the left and the block diagram on the right [45].

Its key features are:
- compactness, portability, and easy connections;

- wide frequencies range from 300 kHz up to 26.5 GHz;
- support of Electronic Calibration Modules (Fig.3.16);
- dynamic range up to 115 dB;
- stability;
- fast in making measurements.

The VNA is packaged in a compact chassis and controlled by an external laptop. The visualization of the signals is carried out directly on the computer, thanks to a dedicated software (PNA Network Analyzer)[45].

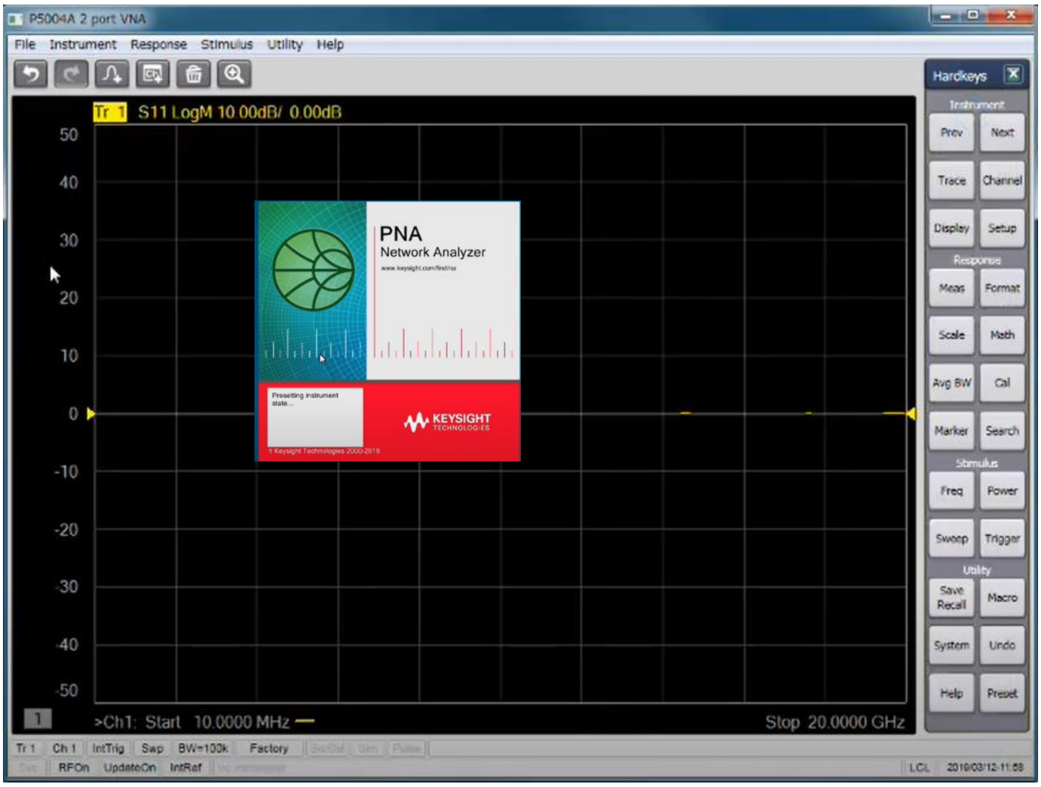


Figure 3.15 P9375A Keysight Vector Network Analyzer Guided User Interface.

The measurement parameters are set directly from the Guided User Interface (GUI) (Fig.3.15).

Below are the most important ones, adopted for the real measurements used to test the image reconstruction algorithm:

- Frequency range: 0.5 - 2.5 GHz
- Sweep: 201 points
- Power: 0 or -5dB
- Filter: 100 Hz

It is also possible to apply the averaging technique: the VNA performs the measurement on several points (for example 5) and returns the mean value between them. The higher the power, the more accurate the measurement. The narrower filter bandwidth, the less noise, the slower the measurement and the better the quality of the reconstructed image.
Before taking the measurements, it is necessary to calibrate the VNA: connect ports 1 and 2 of the VNA to ports A and B of the electronic calibration module (Fig.3.16). From the GUI select: Cal → Other Cal → ECal.

Figure 3.16 Electronic Calibration Modules of the P9375A Keysight Streamline USB Vector Network Analyzer.

The calibration phase is essential because it corrects systematic errors. Once the measurement has been performed, the scattering parameters are provided in the form of a list in a file with the ".s2p" or ".s1p" extension.

3.5 Switching matrix

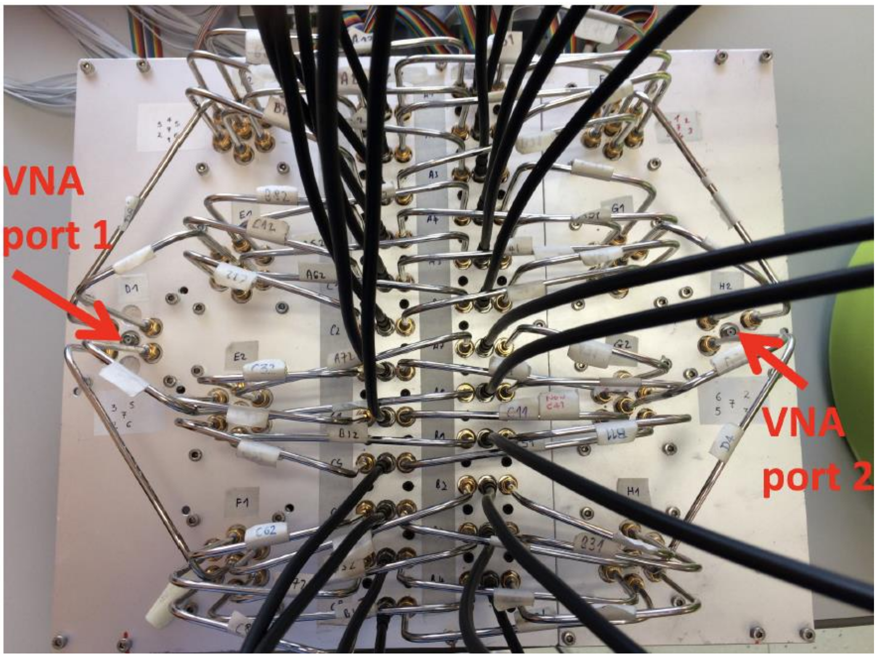


Figure 3.17. Switching matrix [46].

It has been seen that the VNA has only two ports, so it is possible to connect together only two antennas out of twenty-four. The switching matrix (Fig.3.17) manages this limit, selecting two antennas at a time: one is used as a source and the other is used for reception. Considering all pairs of antennas there are 576 combinations.

The switching matrix is realized with different high-quality electromechanical coaxial switches and two electronic control boards (Fig.3.19) [34].

The switching matrix combines two single-pole-four-throw (SP4T) (Keysight 8762B Coaxial Switches), eight single-pole-six-throw (SP6T) (Keysight 87206B Multiport Coaxial Switches)(Fig.3.18) and twenty-four single-pole-double-throw (SPDT) [46]. The main characteristics of these electromechanical switches are:

- small package size and portability;
- insertion loss less than 0.5 dB @ 10 GHz;
- isolation between ports greater than 90 dB @ 10 GHz;
- maximum switching speed of 30 ms for model 8762B and 15 ms for model 87206B

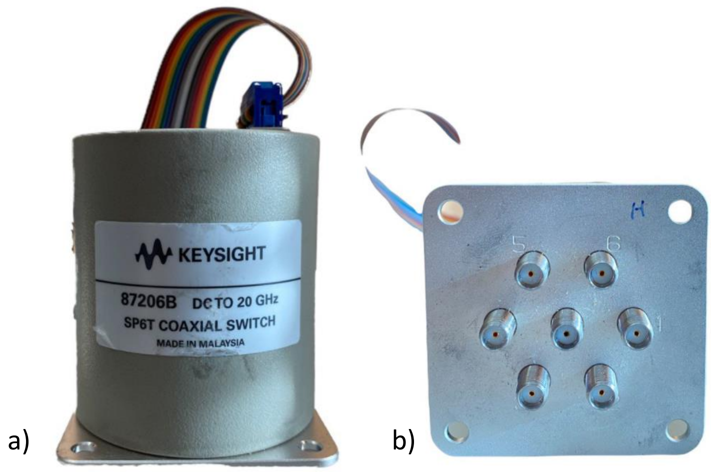


Figure 3.18. Single-pole-six-throw (SP6T) Keysight 87206B Multiport Coaxial Switches. a) front and b)bottom view.

To create the proper connections between the antennas and the VNA, all the switches are connected to an electronic control board. The Ethernet connectivity enables the connection to the network [46].

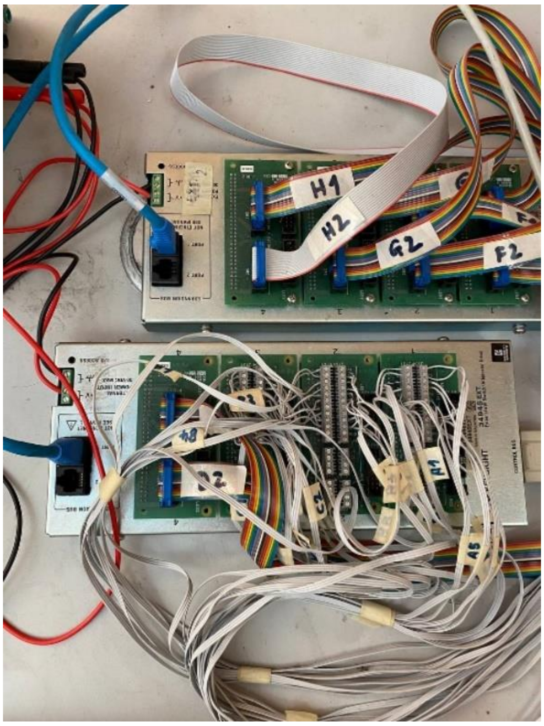


Figure 3.19. Electronic boards (powered 24V DC) that manage the switching of the switches.

The 24 antennas positioned on the helmet are not directly connected to the VNA but are connected via flexible coaxial cables to the switching matrix [34]. Connections between switches were made with semi-rigid coaxial cables with a black plastic coating (Fig.3.20) to maximize the isolation and minimize the insertion losses [18].

Figure 3.20. Semi-rigid coaxial cables connected to the switching matrix and flexible coaxial cables connected to the antennas.

There are 24 paths from VNA port 1 to the corresponding 24 antennas and 24 paths back to VNA port 2: all the paths were designed to have the same length [18]. The switches are controlled through the VISA/TCPIP(VXI-11) protocol by using standard scripting languages Matlab [46].

The instrument control session comprises:

1. Instrument Connection
2. Instrument Configuration and Control
3. Disconnect and Clean Up

3.6 Image reconstruction algorithm

The MWI system exploits the dielectric contrast $\Delta\chi$ between healthy and pathological tissues, at microwave frequencies. However, the output from the measurement system is not the dielectric contrast but the scattering parameters [34]. Differential dielectric contrast and differential scattering matrix ΔS are related to each other by the following non-linear equation (3.36):

$$\Delta S\left(\overline{r_p},\overline{r_q}\right) = \frac{-j\omega\varepsilon_b}{4}\int_{DOI}\overline{E}_b\left(\overline{r_p},\overline{r}\right)\cdot\overline{E}\left(\overline{r},\overline{r_q}\right)\Delta\chi(\overline{r})\,dr \tag{3.36}$$

where:

- DOI is the domain of interest, that is the volume of the object under test (the head);
- $\overline{r_p},\overline{r_q}$ are the positions of the transmitting and receiving p and q antennas, respectively;
- $\overline{E}_b\left(\overline{r_p},\overline{r}\right)$ is the "background" electric field radiated in each point r of the DOI by the antenna in position r_p, when the volume has no target;
- $\overline{E}\left(\overline{r},\overline{r_q}\right)$ is the total field measured by antenna q given by the sum of the incident field and the scattered field;
- ε_b is the "background" dielectric constant (the brain without stroke);
- ω is the angular frequency used by the antennas;
- $\cdot$ is a dot product.

The image reconstruction algorithm programmed ad hoc and implemented on Matlab, allows to obtain images starting from the measured scattering parameters (Fig.3.21).

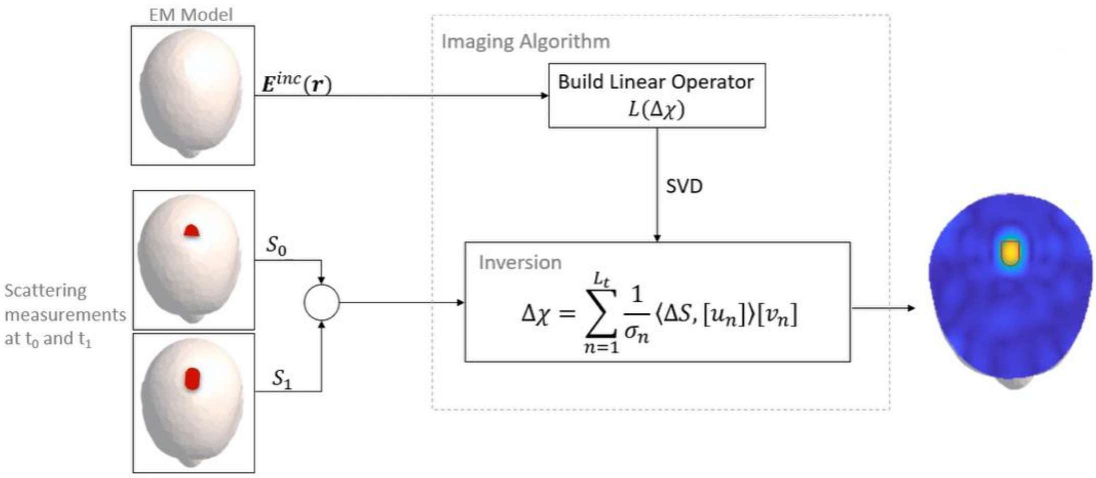


Figure 3.21 Block diagram: operating principle of the algorithm implemented for image reconstruction [34].

The software section manages highly computational steps which consist of the resolution of an inverse scattering problem, not linear and wrongly posed.
The resulting image is obtained from the difference between two images. To diagnose the presence of stroke, the reference image is a background scenario without target. To monitor the evolution of the stroke at instant t, the reference image relates to the same scenario but at instant t_{-1}.
The difference between the two scenarios is in a portion of the head and it should be very small: this allows to linearize the model and use the Born approximation [46].
The parameters are collected within a 24 x 24 matrix in two different instants. It is assumed that the only change between the two scenarios is due to a dielectric contrast variation. The difference is then calculated by obtaining the so-called differential scattering matrix (ΔS).
The variations and therefore ΔS are only related to the unknown variation of the contrast through a linear operator, defined by the following Kernel (3.37) [46]:

$$\frac{-j\omega\varepsilon_b}{4} * \overline{E}_b(r_p, r_m) \cdot \overline{E}_b(r_m, r_q) \tag{3.37}$$

The relationship between differential dielectric contrast and differential scattering matrix can therefore be linearized as follows (3.38):

$$\Delta S\ (\mathrm{r_p}, \mathrm{r_q}) \ = \ \mathcal{L}(\Delta\chi) \tag{3.38}$$

The linear operator is computed off-line for all combination of antennas p and q and for all the positions inside the DOI with Finite Element Method (FEM) simulations [48].
The decomposition of the singular value (SVD) of the scattering operator relates the data of the problem to the unknown contrast function. The decomposition of the singular value (SVD) is applied to the known linear operator (3.39):

$$\mathcal{L} = USV \tag{3.39}$$

SVD is a particular factorization of a matrix based on the use of eigenvalues and eigenvectors: U is a unitary matrix (right singular vectors), S a diagonal matrix (singular values) and V the conjugate transpose [46].
At this point it is possible to obtain the unknown variation of the contrast with a simple linear combination of the right singular vector of the operator's SVD (3.40):

$$\Delta\chi \ = \ \sum_{\mathrm{n}=1}^{\mathrm{L_t}} \frac{1}{\sigma_\mathrm{n}} \langle \Delta \mathrm{S}, [u_n] \rangle [v_n] \tag{3.40}$$

Where:

σ_n : singular values of the discretized scattering operator

u_n : left singular of vector the discretized scattering operator

v_n : right singular of vector the discretized scattering operator

L_t : truncation factor.

The truncation factor works as a regularization parameter, and it defines the level of information to retain after the TSVD. It can be considered the threshold and its choice is a trade-off between the stability against the noise affecting the measured scattering parameters and the accuracy of the reconstructed image [47,48].
To obtain the reconstructed image, $\Delta\chi$ (r) is simply plot in three-dimensional. The operation, once the S parameters have been obtained, is quite fast. Below is an example with ischemic ictus related to a measurement carried out on a phantom in laboratory (Fig.3.22).

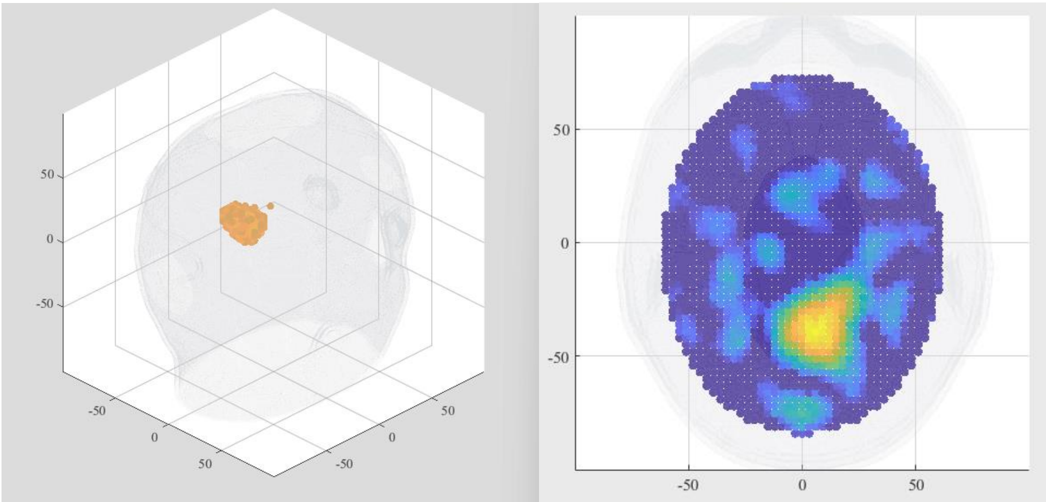


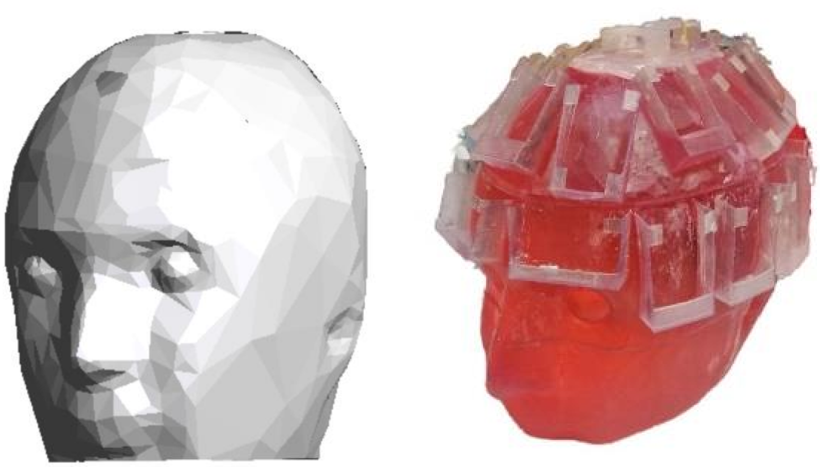
Figure 3.22 Example of a tomographic image obtained via MWI system: the stroke in the back of the head is visible.

Chapter 4

Anthropomorphic Head Phantom

4.1 3D model

Before moving on to clinical trials, the operation of the MWI system was tested using head models that reproduce the dielectric properties of the tissues. The different characteristics of the materials have been obtained from mixtures suitably studied.
A phantom is a physical model that simulates the behavior of different tissues exposed to electromagnetic waves. They should have high durability over time, easy availability, reproducibility, and low production costs.
The 3D anthropomorphic head phantom used to validate the MWI system was made from polyester casting resin. The 3D model was created with a computer-aided design (CAD) software by editing a stereolithography (STL) file derived from an MRI scan of the head of a healthy subject [18]. The different blocks of the phantom were made through 3D printing, and then they were fixed together with a sticky material. In the realized model, there are also the housings for the antennas. It has an ellipsoidal section with a minor axis of 20 cm and a major axis of about 26 cm, and a wall thickness of 3 mm.
The phantom is hollow: in the upper part of the head there is a slot, which is then sealed once the liquid has been added.
This head model represents the background scenario for the image reconstruction algorithm and the condition without target (healthy case).



a) b)

Figure 4.1. Human head phantom: a) CAD model, b)3D anthropomorphic phantom filled with a mixture created ad hoc that mimics the dielectric characteristics of the healthy brain.

4.2 Dielectric properties

The human head is made up, starting from the outside inwards, from different types of tissue: skin, fat, bone, cerebrospinal fluid (CSF), gray matter, white matter, and cerebellum.
However, for practical reasons, an ad hoc mixture was created that mimics the dielectric characteristics of the brain intended as a homogeneous medium.
A homogeneous material is an object with a single constituent. The constitutive parameters that characterize it, are the same in the entire volume. Since white matter and gray matter make up a good percentage of the brain, the recreated mixture has a permittivity and dielectric conductivity values, equal to the mean of the permittivity and conductivity values of these two tissues (Tab.3.1).
The first liquid blend created was composed of 38% Triton X-100, 62% water and salt. [18]. The graph below represents the dielectric characteristics in the frequency range from 0.5 to 1.5 GHz (Fig.4.2).

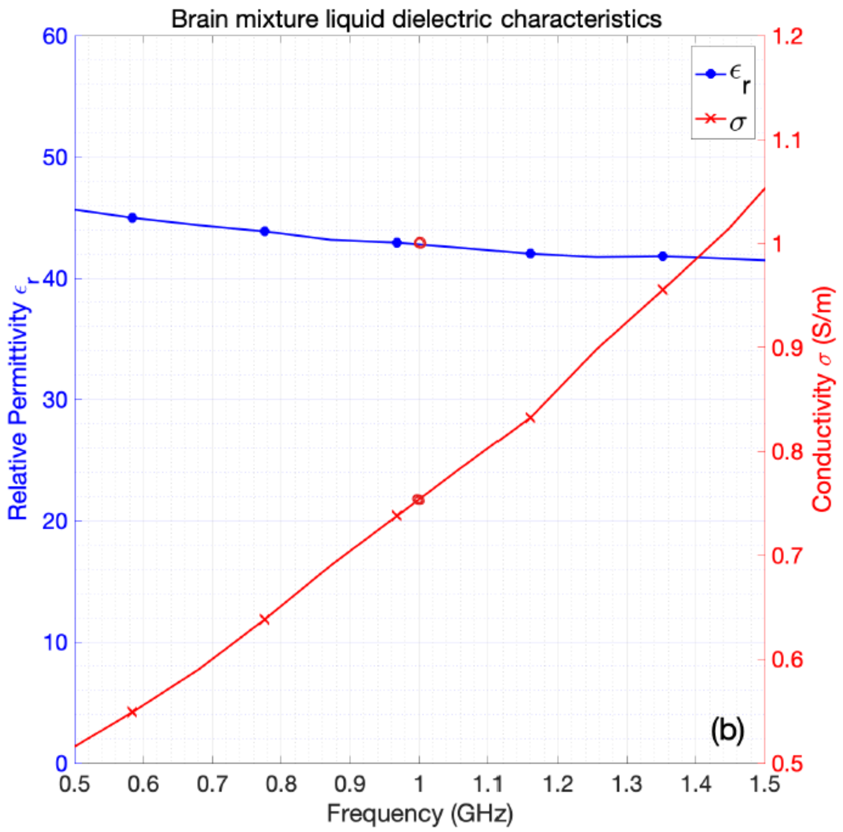


Figure 4.2. Brain mixture liquid dielectric characteristics in the frequency range from 0.5 to 1.5 GHz [18].

The relative permittivity (4.1) and conductivity (4.2) values at the 1 GHz working frequency were:

$$\varepsilon_r = 45.37 \tag{4.1}$$

$$\sigma = 0.7729 \text{ S/m} \tag{4.2}$$

It has been seen that the gelatinous substance changes its properties. Decreasing the temperature, the compound tends to solidify, making it difficult to move and place the balloon that mimics the stroke inside the phantom.
For these reasons, it was decided to use another "recipe" that is simpler and cheaper to make. At first the ingredients were: demineralized water, isopropyl alcohol, and table salt (NaCl).
Measurements were carried out daily to verify that the evaporation of the alcohol did not significantly change the dielectric properties of the mixture.
It was therefore decided to replace isopropyl alcohol with ethyl alcohol, and it has seen that the properties remain almost stable over time. The final mixture is made up of 32.53% demineralized water (1816.3 g), 67.18% ethyl alcohol (3751 g) and 0.29% salt (15.75g).
Before carrying out any measurements, it is however advisable to take a sample of liquid to measure its dielectric properties and, if necessary, correct the mixture by adding small quantities of alcohol.
The graph below represents the dielectric characteristics of the so-called *average brain*, in the frequency range from 0.5 to 2.5 GHz (Fig.4.3).

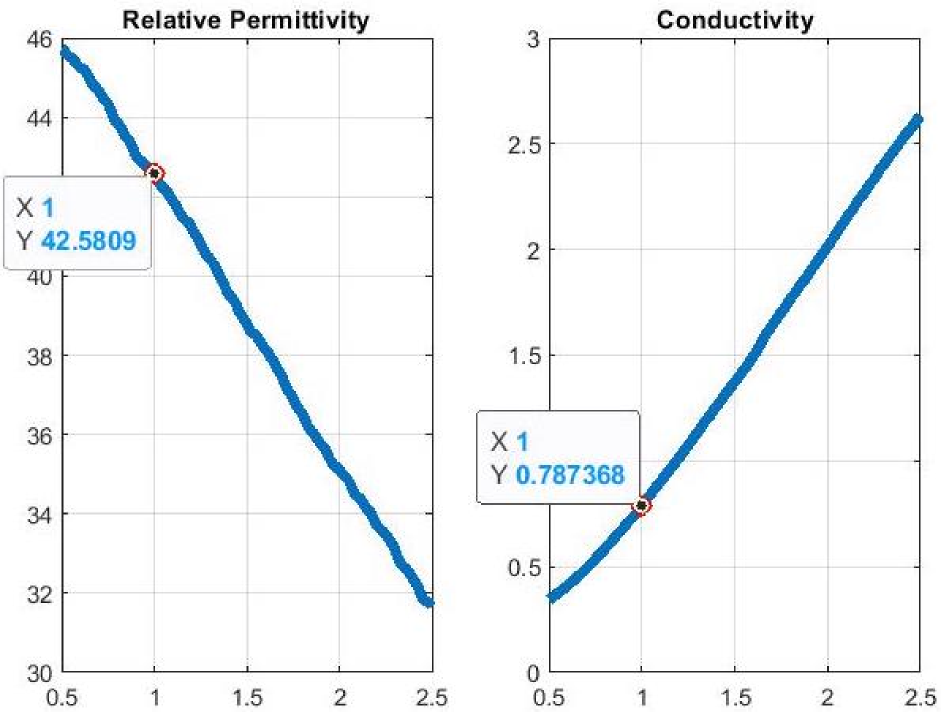


Figure 4.3. Average brain dielectric characteristics in the frequency range from 0.5 to 2.5 GHz.

The relative permittivity (4.3) and conductivity (4.4) values at the 1 GHz working frequency are:

$$\varepsilon_r = 42.58 \tag{4.3}$$

$$\sigma = 0.7874 \text{ S/m} \tag{4.4}$$

These values are taken as a reference for the creation of the model of the healthy case, used as a background scenario for the image reconstruction algorithm. The same process was followed to create the blends that mimic the condition of ischemic and hemorrhagic stroke. The stroke is simulated in the laboratory with balloons anchored to wooden sticks and filled with ad hoc liquids.
A dye was added to the ischemic stroke mixture to differentiate it from the hemorrhagic one. The relative permittivity (4.5) and conductivity (4.6) values at the working frequency of 1 GHz are, for the hemorrhagic blood:

$$\varepsilon_r = 63.41 \tag{4.5}$$

$$\sigma = 1.576 \text{ S/m} \tag{4.6}$$

and for ischemic blood (4.7)(4.8):

$$\varepsilon_r = 36.00 \tag{4.7}$$

$$\sigma = 0.7200 \text{ S/m} \tag{4.8}$$

The following paragraph briefly illustrates the method used to measure the dielectric properties of the mixtures.

4.3 Coaxial probe method

The dielectric properties of the materials depend on the working frequency and on their internal molecular structure. The liquids that fill the phantom were electromagnetically characterized thanks to an open coaxial cable probe.
The coaxial probe method (Fig.4.4) consists in putting the terminal region of the probe in contact with the sample whose properties are to be known. This technique is simple, convenient, non-destructive and it is very well suited for liquid and semi-solid materials.
The probe transmits an electromagnetic field in the liquid and determines the scattering parameter S_{11} from which the value of the complex dielectric constant is obtained.
A typical measurement system therefore consists of a network analyzer (VNA), the coaxial probe and an ad hoc software for real time calculation of the dielectric parameters [49].

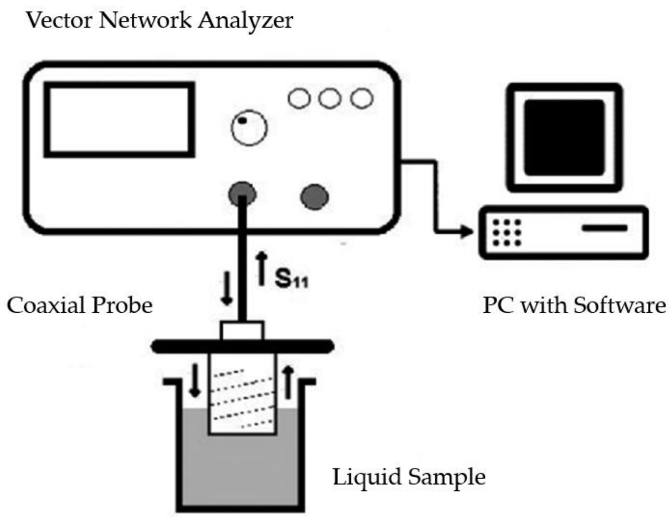


Figure 4.4. Open-ended coaxial-probe method [50].

The probe used in the laboratory is the 85070E dielectric probe from Keysight Technologies (Fig.4.5).
It offers high performance: it can work in a very wide range of frequencies (from 200 MHz to 50 GHz) and temperature (from -40ºC to + 200ºC). At one end there is a 2.4 mm male connector [49].

Figure 4.5. Keysight 85070E Dielectric Probe Kit 200 MHz to 50 GH [49]

The probe must be calibrated to avoid errors due to the different impedances and to reduce the ripple on the signal (Fig.4.6).

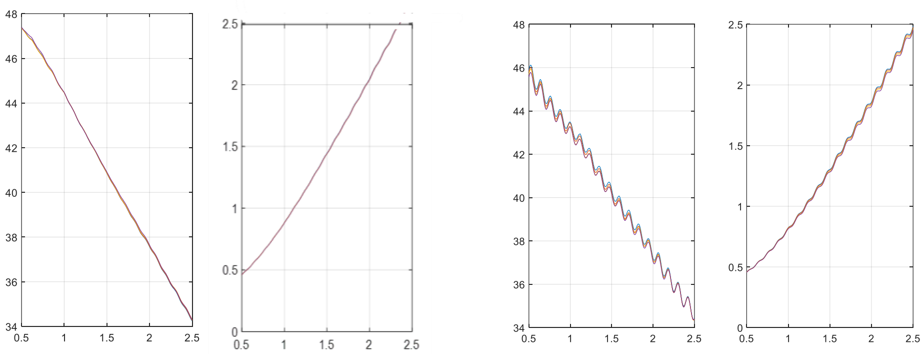


Figure 4.6. Example of measurement of dielectric properties: a) with calibration and b) without calibration.

Three measurements are performed under known conditions to calibrate the probe:

- open circuit measurement (air)
- short circuit measurement (the probe is closed with a very high impedance resistance (Γ= -1))
- polar liquid (water) measurement

This calibration is transferred to the ECal module and is performed automatically before each measurement [49]. The prepared liquids were mixed thoroughly inside a normal plastic cup. To measure the dielectric properties of the prepared liquid it is sufficient to put it in contact with the calibrated probe (Fig.4.7-Fig.4.8).

Figure 4.7. Coaxial probe calibration: a) short measurement, b) water measurement.

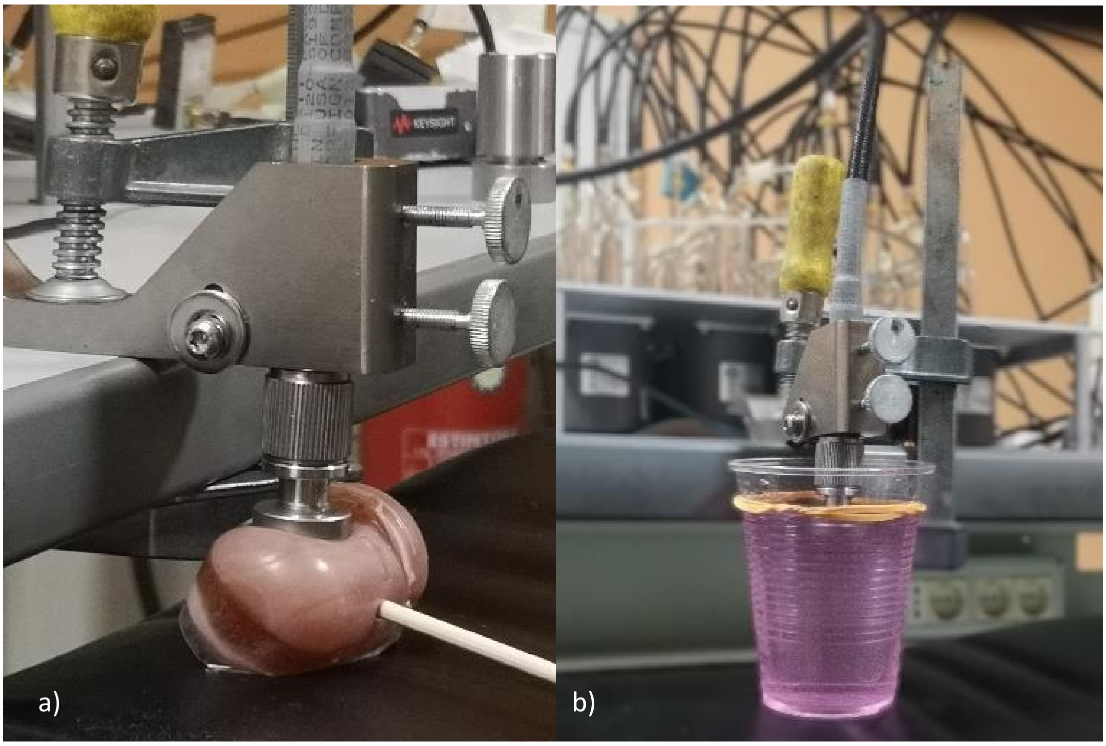


Figure 4.8. Coaxial probe method for measuring dielectric properties: a) balloon filled with liquid that simulates the hemorrhagic stroke, b) average brain.

Chapter 5

Machine Learning Theory

The use of Machine Learning (ML) techniques combined to the new Microwave Brain Imaging system (MWI) represents a very promising approach for stroke detection. ML can be an interesting alternative to deterministic imaging algorithms, reducing data processing time.
The aim is to train algorithms that can define the region affected by ischemia or hemorrhage, simply by "looking" at the S parameters measured by the MWI system.
This chapter aims to illustrate the theory which is behind the ML algorithms used in this thesis project, to solve the classification problem (Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and K-Nearest Neighbors (K-NN)) [47,52].

5.1 Basic Concepts

Machine learning (ML) is a subfield of the artificial intelligence (AI) that has become quite relevant since the 1980s.
ML can be understood as the science capable of extracting knowledge from data. It essentially uses statistical methods to improve its performance on identifying patterns. The data that describe an element belonging to a given class are called *features* (or *instances*), and they are the input of the ML algorithms. The class is the *target (*or *label*). Each model has a very precise structure and operating logic: these characteristics can be set by the user simply by defining some hyper-parameters.
Generally, in classification problems there are two types of data: those whose class is known a priori and those for which the class is not known, but which must be predicted. The first is the *training set* and the latter is called *test set*. Some elements of the training set are used to evaluate the performance of the trained algorithm and to fine-tune the hyperparameters. These elements are part of the so-called *validation set*.
In the literature, the terms "*test*" and "*validation*" are often interchangeable.
There are three types of learning on which ML techniques are based:

- **Supervised learning**: the features given as input to the algorithm are labeled. In this way the algorithm builds a model based on the knowledge of the class.
- **Unsupervised learning**: the inputs provided have neither a defined structure nor associated outputs. The algorithms analyzes and cluster unlabeled datasets, discovering hidden patterns or data groupings independently.

- **Reinforcement learning**: it is based on feedback or rewards that the algorithm receives from the external environment. [53]

The algorithms used in this work are all part of the first category.

5.3 Support Vector Machine (SVM)

Support Vector Machines (SVM) are binary classifiers that can identify elements belonging to two classes. These models were introduced by Cortes and Vapnik in 1995 [54]. A SVM is based on the concept of decisional boundary. The first step taken by the algorithm is to map the training data in a higher dimensional space which is called *feature space*. In this new space, the elements belonging to the two classes are linearly separable by a hyperplane (Fig.5.1) [55].

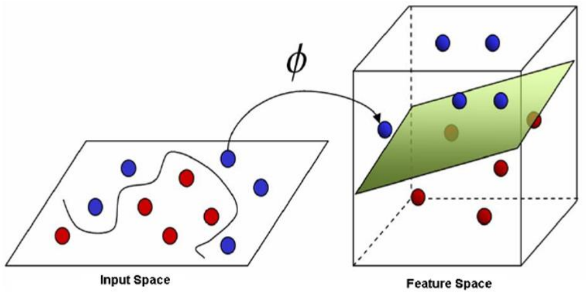


Figure 5.1 The elements of the two classes are linearly separable in the feature space by a hyperplane [56].

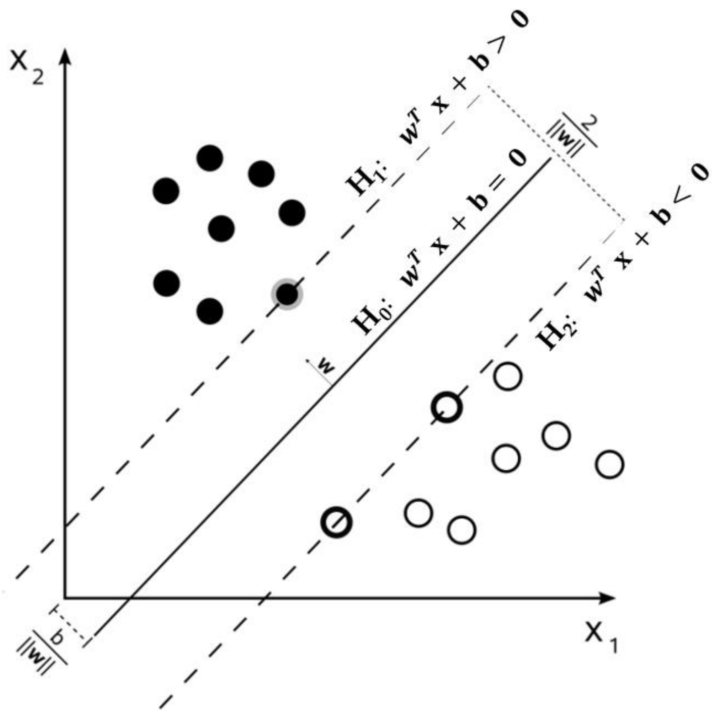


Figure 5.2 An example of optimal separating hyperplane with maximum margin in the Euclidian space [57]

The support vectors (SV) are the elements belonging to the two classes, closest to the decision boundary (the optimal separation hyperplane in the feature space). Referring to Figure 5.2, the solid black line is the decision boundary, and the elements present on the two parallel dashed lines are the SVs. Of course, the latter are the most critical to be classified as they are easily confused. Training the SVM algorithm translates into identifying, by means of an optimization method, the hyperplane that has the greatest margin from the SVMs. The decision margin is defined as the smallest distance between the decision boundary and any of the samples. The hyperplane H_0 can be defined with the following equation (5.1):

$$w^T \mathrm{x} + \mathrm{b} = 0 \tag{5.1}$$

where w is the weight vector, normal to the optimal hyperplane, x is the input vector and b is the bias.

If the space is two dimensions (like the one in Fig.5.2.), the hyperplane equation is reduced to the equation of a line. H_1 and H_2 are the two hyperplanes parallel to H_0 and passing through the SVMs (one per class).

All points above the H_0 hyperplane belong to H_1 and satisfy the following equation (5.2):

$$w^T \mathrm{x} + \mathrm{b} > 0 \tag{5.2}$$

while the points below the H_0 hyperplane belong to H_2 and satisfy the following equation (5.3):

$$w^T \mathrm{x} + \mathrm{b} < 0 \tag{5.3}$$

It is possible to calculate the margins (the distance between H_0 and H_1, and the distance between H_0 and H) with this formula (5.4):

$$\frac{w^T \mathrm{x} + \mathrm{b}}{||w||} = \frac{1}{||w||} \tag{5.4}$$

The total distance between H_1 and H_2 is (5.5):

$$\frac{2}{||w||} \tag{5.5}$$

This function must be maximized by the algorithm to maximize the distance of the SVs from the decision boundary. Very often, maximization problems can be reformulated in terms of minimization, as a quadratic optimization problem subject to linear constraints. The maximization of the equation (5.5) is rewritten as follows (5.6):

$$\frac{1}{2}w^Tw \quad \rightarrow \quad \frac{1}{2}||w||^2 \tag{5.6}$$

Minimizing the equation allows to maximize the margins, to find the optimal hyperplane, and SVs.

In real cases, where it is not possible to make a linear separation between elements, there are two options: to map the original problem in a new space or to admit a certain number of errors in the classification (relaxation of the margins).

In the second case, the so-called *slack variables* ($\xi > 0$) are added to the model.

$M(w)$ is the new function to be minimized (5.7):

$$(w) = \frac{1}{2}\, w^Tw + c\sum \xi_i \tag{5.7}$$

The SVM algorithm must identify the hyperplane that maximizes the margins and minimizes the errors on the training set, at the same time.

ξ_n depends on the position and distance of each sample with respect to the decision boundary:

• $\xi_n = 0$: the sample is classified correctly, that means it is on the correct side of the decision boundary and it is outside the margin;

• $0 < \xi_n \leq 1$: the sample is inside the margin, even though it is on the correct side of the decision boundary;

• $\xi_n > 1$: the sample is misclassified because it is on the wrong side of the decision boundary [47].

Hyperparameter c is a penalty term. It balances the compromise between maximizing margins and minimizing error: small *c* means large margins and high number of mistakes, while large *c* means narrow margins and low number of mistakes.

The original input space can always be mapped to some higher-dimensional feature space where the training set is separable. The dimensionality of the mapped space can be arbitrarily chosen.

The *kernel function* transforms the data from the original space to the feature space. There are several types of kernel functions (5.8 -5.11):

• Linear: $\Phi(x_i, x_j) = {x_i}^T x_j$ (5.8)

• Polinomial: $\Phi(x_i, x_j) = (1 + {x_i}^T x_j)^p$ (5.9)

• Sigmoidal: $\Phi(x_i, x_j) = \tanh(\beta_0 {x_i}^T x_j + \beta_1)$ (5.10)

• Gaussian Radial Basis Function (RBF): $\Phi(x_i, x_j) = e^{-\frac{||x_i - x_j||^2}{2\sigma^2}}$ (5.11)

Once the kernel function (ϕ) to be used has been chosen, the algorithm determines the optimal decision boundary through an optimization process, guided by the variables c and ξ_n. At the end of the training phase, the weights of the hyperplane (w_s), the bias and the SVs (x^s) are known.
When a new element x^{new} is provided as input to the algorithm, the classifier applies the following formula to obtain the output function y (5.12):

$$y = \sum_{s \in SV} \phi(x^{new}, x^s) + bias \tag{5.12}$$

The unknown element is mapped into the feature space. At this point, if the value of y is greater than zero, the element will be above the hyperplane, otherwise it will be below it. The class will be associated according to the chosen convention [58].
As mentioned at the beginning, SVM is a binary classifier, however most classification problems require distinction for more than two categories. In this work, as will be seen better later, the classes to be identified are nine.
The multi-class problem is broken down into multiple binary classification problems. For the choice of the label, the winner takes all or the majority voting principles are used [59].
One-vs-rest rest (one against all) approach consists of building k separate binary classifiers for k classes. For each binary classifier, one class is positive, and the rest are negative (Fig.5.3.). The classifier with the highest output function y assigns the class (winner takes all strategy) [60].

	SVM1	SVM2	SVM3
Class 1	+1	-1	-1
Class 2	-1	+1	-1
Class 3	-1	-1	+1

Figure 5.3. One vs rest approach: the number of classifiers is equal to the number of classes (k=3, in this example). For each binary classifier one class is positive and the rest are negative [60].

If all classes are made up of the same number of training examples, the ratio of positive to negative examples would be 1: (k 1). In this case, the symmetry of the original problem is lost. To overcome this limit the Pairwise (one-vs-one) approach is adopted.
It evaluates all possible pairwise classifiers and thus it uses $\frac{k(k-1)}{2}$ individual binary classifiers. For each binary classifier, one class is positive, another is negative, and the rest are ignored (Fig.5.4.). The class with the maximum number of votes is the winner (majority voting strategy).

	SVM1	SVM2	SVM3	SVM4	SVM5	SVM6
Class 1	+1	+1	+1	0	0	0
Class 2	-1	0	0	+1	+1	0
Class 3	0	-1	0	-1	0	+1
Class 4	0	0	-1	0	-1	-1

Figure 5.4. Pairwise approach: number of classifiers equal to $\frac{k(k-1)}{2}$, where k is the number of classes (6 in this example). For each binary classifier, one class is positive, one is negative, and the rest are ignored.

The size of classifiers created by the one versus one approach is much larger than that of the one versus rest approach [61].

5.4 K-Nearest Neighbors (K-NN)

The K-Nearest Neighbors (K-NN) is a very simple method that classifies unlabeled cases based on their similarity to known ones in the training set [62]. K-NN does not require any learning phase, it simply calculates the distance between the test element and those of the training set. The training set samples are sorted according to the distance they have with the test element, from the closest to the furthest. The most represented class among the first k nearest examples of the training set is assigned to the test element.

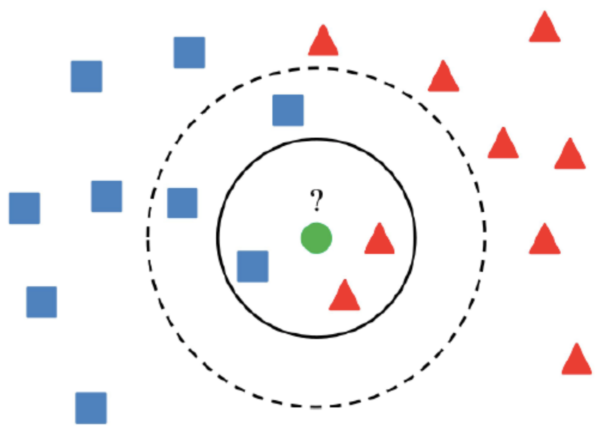


Figure 5.5. K-NN algorithm for binary classification, considering k = 3 (solid line) and k = 5 (dashed line). [63]

Figure 5.5. shows an example of classification problem solved with k-NN algorithm. The training set instances belong to two classes: that of blue squares and that of red triangles. The circular green element (*query point*) is the test element, which must be classified.

If k = 3, the neighbors are two red and one blue: by majority, the element is classified as red triangle. But if k = 5, the neighbors are three blue and two red and, in this case, the green circle is classified as blue square.
The only two hyperparameters that must be set are k, that is the number of neighboring training elements to be considered, and the metric to be used for the distance calculation. To operate correctly, the elements must have the same scale of values and must therefore be normalized in the preprocessing phase.
K is usually an integer, greater than one, odd to eliminate the probability of an exactly even vote.
Low values can make the classifier susceptible to overfitting and sensitive to noise. High values, on the other hand, can lead the classifier to erroneously predict a label because it considers points that are very distant as neighbors. This risk can be reduced by weighing the contribution of neighbors by distance.
The distance ($d(x,y)$) that defines the similarity between elements of the training (vector x) and elements of the test (vector y) can be calculated with the following metrics (5.13-5.14) (Fig.5.6.):

- *Euclidean distance*:

$$d(x,y) = ||x-y|| = \sqrt{(x-y)^T(x-y)} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{5.13}$$

- *Manhattan distance (L1 norm):*

$$d(x,y) = ||x-y|| = |(x-y)| = \sum_{i=1}^{k}|x_i - y_i| \tag{5.13}$$

- *Chebyshev distance (L∞ norm):*

$$d(x,y) = max_k|x_i - y_i| \tag{5.14}$$

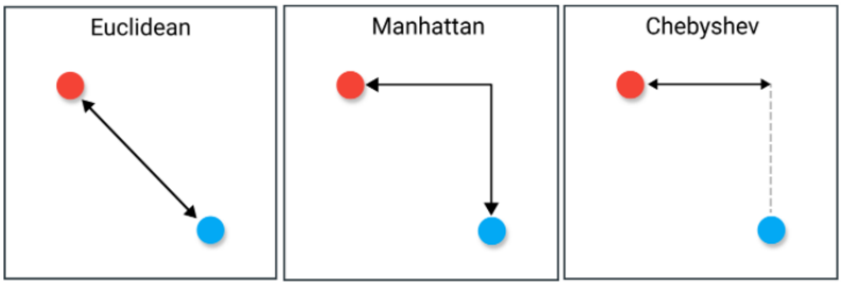
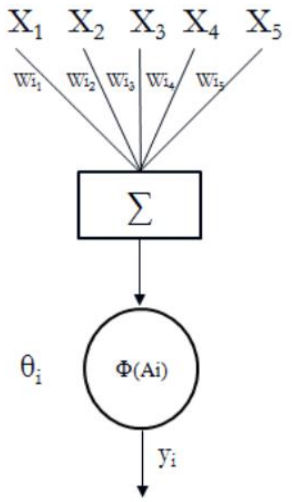


Figure 5.6. Graphical representation of the metrics used for distance calculation [63].

The k-NN algorithm uses local information and lends itself very easily to parallel implementations. However, it has very high memory requirements (it must memorize the entire training set and all the distances between elements).

5.5 Multi-Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP) algorithm is a feed forward artificial neural network (NN), inspired by the functioning of the human nervous system [64].
Biological neurons are connected to each other via so-called dendrites and depending on the incoming signals, they can be activated. The branching that transmits the information is called axon and each axon can be connected to one or more dendrites at intersections called synapses.
The basic element of the MLP is the "*neuron*" also called "*node*". The perceptron is the simplest model of artificial NN (Fig.5.7).



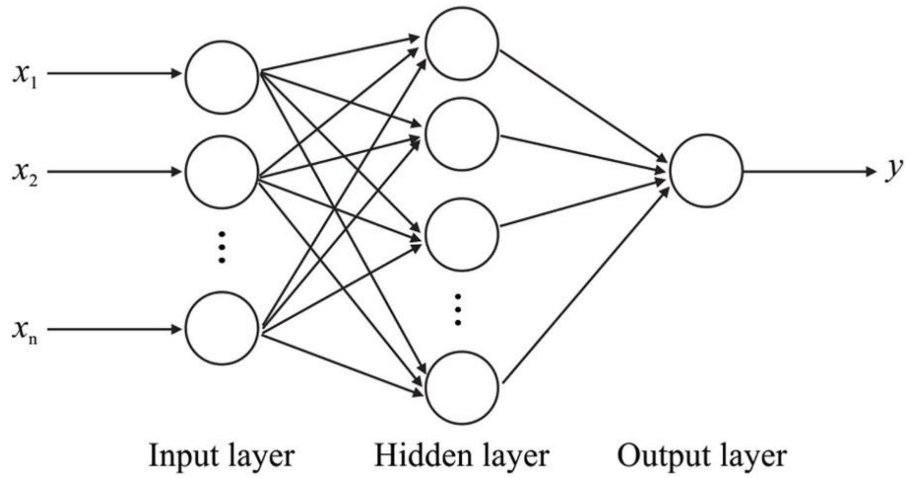
Figures 5.7. Diagram of a perceptron [64]

A weighted sum of the inputs is performed. The neuron determines its output value y by analyzing the result of the linear combination of the inputs, through a special activation function θ.
The input is represented by the N features (x_1, x_2, x_3, …. , x_n).
The behavior of a perceptron can be briefly described with this formula (5.15):

$$y = \theta\left(\sum_{i=1}^{N}\left(w_i x_i + b_j\right)\right) \tag{5.15}$$

Where w_i are the weights and b_j are the bias.
During the training phase, the weights, based on the data, are repeatedly adjusted, until the system returns the expected output. Many neural networks can hold hundreds of billions of weights, which requires high computing power.
Multi-Layer Perceptron (MLP) is a modified and complex version of the perceptron: it presents between the input and output layers, different layers of neurons, called *hidden layers* (Fig.5.8).

Each neuron of a layer is connected to each neuron of the next layer (feed forward NN).



Figures 5.8. Diagram of a multi-Layer Perceptron (MLP) [65].

The output of an intermediate layer is unknown, and it does not coincide with the desired output. Therefore, it is difficult to calculate the error rate. To solve the problem, a technique called *back-propagation* is applied. The error $\delta_k(p)$ at the output of neuron k, at iteration p_{th}, is defined by this formula (5.16):

$$\delta_k(p) = y_{d,k}(p) - y_k(p) \tag{5.16}$$

Based on this difference, the algorithm updates the network weights, in an attempt to gradually converge the output results with those expected. The variation in the value of the weights can be calculated with the following equation (5.17):

$$\Delta w_{jk}(p) = \eta\delta_k x_{jk} - \alpha\Delta w_{jk}(p-1) \tag{5.17}$$

where $\Delta w_{jk}(p)$ is the weight update performed during the p_{th} iteration, α is the regularization term, δ_k is the error and η is the learning rate.
In this way, the weights in the next iteration will be (5.18):

$$w_{jk}(p+1) = w_{jk}(p)+\Delta w_{jk}(p) \tag{5.18}$$

The activation functions are mathematical equations that determine the output of a neural network. Generally, artificial neural networks replicate the behavior of biological ones, using continuous, non-linear, and differentiable functions [67]. The most common activation functions are (Fig.5.9)(5.19-5.23):

$$\textit{Step function: } f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{5.19}$$

$$\textit{Linear function: } f(x) = k \cdot x \tag{5.20}$$

$$\textit{Sigmoid (Standard Logistic) function: } f(x) = \frac{1}{(1+e^{-x})} \tag{5.21}$$

$$\textit{Hyperbolic tangent function: } f(x) = \tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \tag{5.22}$$

$$\textit{Rectified Linear Unit (ReLU) function: } f(x) = \max(0, x) \tag{5.23}$$

The step function is based on the existence of an activation threshold value, if the weighted sum of the input values is greater than this threshold, then the neuron is activated and transmits the same signal to the next layer. The linear function generates output values proportional to the inputs. The ReLu function is usually used when the input data cause a particular slowness in reaching the optimal solution of the model.

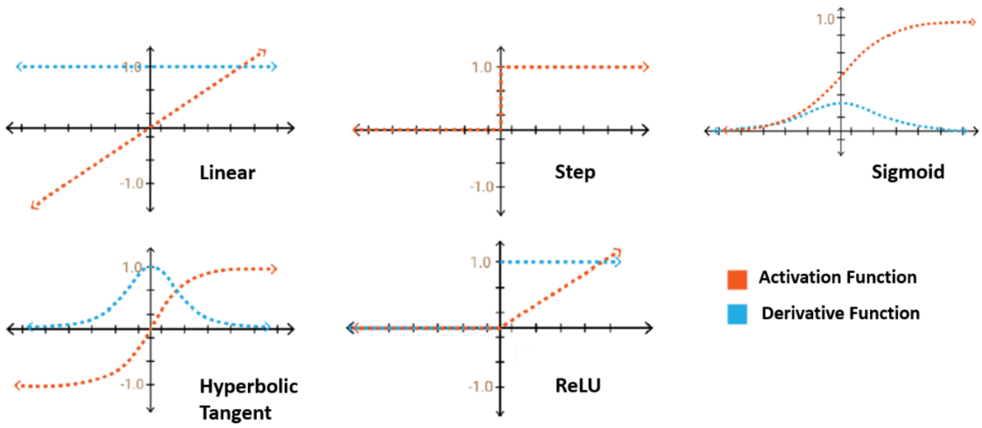


Figure 5.9. The most common activation functions [67].

5.5 Overfitting problem

The accuracy of a prediction learned through machine learning can be very different in the training data and in the separate test data. This phenomenon is called *overfitting* [69].

Overfitting a training set means that the classifier memorizes training patterns and consequently loses the ability to generalize. This means that the algorithm may perfectly fit past data and may not actually work with future test data.

Machine learning methods are particularly prone to overfitting as they can try out a huge number of different "rules" until they find one that fits all training data perfectly. This happens especially for those particularly flexible models.

Considering SVMs, overfitting occurs when c is too large, that is, when the penalty hyperparameter narrows the margins and minimizes the number of prediction errors for the training set elements.

For k-NN method, overfitting occurs for very small values of k.

In neural networks, overfitting occurs when architecture is too large, for instance the NN has too many weights.

A validation set, obtained from the training set, is usually used to get an idea of the level of generalization achieved during the training phase [68].

5.6 Performance evaluation Metrics

The metrics that will be illustrated in this section, allow a quantitative and standardized evaluation of the performance of the models. These metrics are useful for comparing the results obtained by different classifiers.

A binary classifier problem involves the distinction of elements into two categories: positive (1) or negative (0). The predictions can be correct (true) or wrong (false).

An element can be defined in 4 different ways, based on its true label, and based on the classifier's prediction (Fig.5.10) [52].

In particular:

- A true positive (TP) occurs when the model correctly classifies the positive element (the predicted class is positive and the true label is positive).
- True negative (TN) occurs when the model correctly classifies the negative element (the predicted class is negative and the true label is negative).
- A false positive (FP) occurs when the model incorrectly classifies the element as positive when the latter is negative (the predicted class is positive and the true label is negative).

- A false positive (FN) occurs when the model incorrectly classifies the element as negative when the latter is positive (the predicted class is negative and the true label is positive).

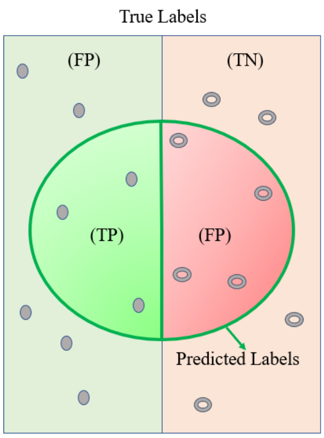


Figure 5.10. Schematic representation of TP, TN, FP, FN.

Starting from these definitions it is possible to formulate the so-called accuracy, that is the ratio between the exact predictions on the total number of predictions made by the classifier (5.24):

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{TOT\ right\ predictions}{TOTpreditions} \tag{5.24}$$

In case of an unbalanced class data set, where the number of positives is significantly higher or lower than the number of negatives or vice versa, the number of FN or FP would not be evaluated correctly. To take into account the proportion between the number of elements of one class and another, within the dataset, two other metrics have been introduced: precision and recall.

Precision (or Positive Predicted Value (PPV)) suggests how good the model is at classifying true positives: it indicates how many times the classifier predicted the correct class whenever the element to be classified belonged to the positive category (5.25). The higher the precision, the lower the FP rate.

$$Precision\ (PPV) = \frac{TP}{TP+FP} \tag{5.25}$$

The recall (or sensitivity or True Positive Rate (TPR)) expresses how many times the classifier has predicted the correct class every time the element to be classified belonged to the negative category (5.26). The higher the recall, the lower the FN rate.

$$Recall\ (TPR) = \frac{TP}{TP+FN} \tag{5.26}$$

Other metrics (5.27-5.32), combining the relative ratios between TN, TP, FN, FP, can be calculated [52]:

$$F_1 = 2\frac{Precision \cdot Recall}{Precision+Recall} = \frac{2TP}{2TP+FP+FN} \quad (5.27)$$

$$TNR\ (true\ negative\ rate\ or\ specificity) = \frac{TN}{TN+FP} \quad (5.28)$$

$$NPV\ (\text{Negative Predicted Value}) = \frac{TN}{TN+FN} \quad (5.29)$$

$$FPR\ (\text{False Positive Rate Fall out}) = \frac{FP}{FP+TN} \quad (5.30)$$

$$FNR\ (\text{False Negative Rate}) = \frac{FN}{TP+FN} \quad (5.31)$$

$$FDR\ (\text{False Discovery Rate}) = \frac{FP}{TP+FP} \quad (5.32)$$

An intuitive and effective method to represent the performance of a model is through a confusion matrix, in which the four possible outcomes are reported (Fig.5.11) [70].

In the main diagonal there are the correct predictions TP and TN, while off the diagonal there are the wrong predictions FP and FN.

Predicted / True Label	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Figure 5.11. Confusion matrix for a binary classifier [56].

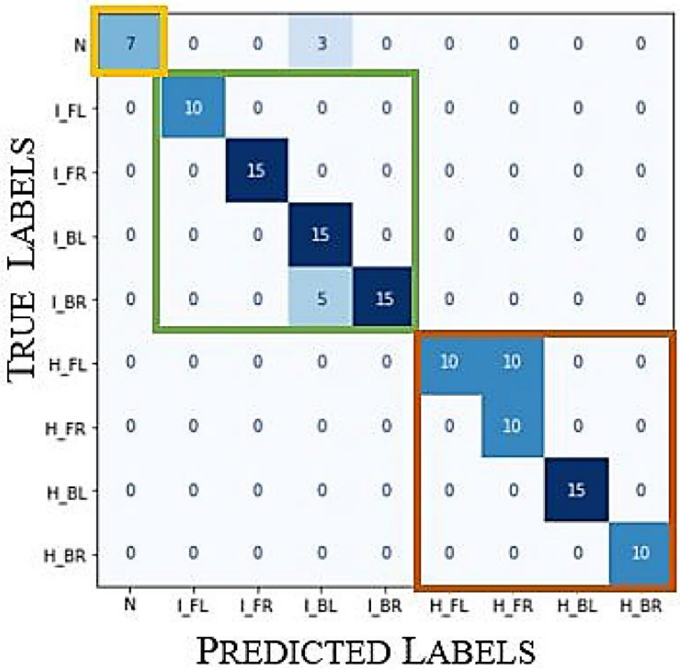


Figure 5.12. Example of a non-normalized multi-class (nine) confusion matrix.

Considering a multi-class problem, the confusion matrix appears as in Figure 5.12. The true labels are present along the rows, while the predicted labels are read on the columns. In a multiclass problem only exact predictions (true (T)) and wrong predictions (false (F)) are considered: FP and FN lose their meaning, as there are no positive and negative categories.

Referring to the confusion matrix in the Figure 5.12, the ML algorithm incorrectly classifies five I_BR elements and three N elements as I_BL.

The user who supervises the model can manipulate the input data or modify the hyper-parameters based on the number and type of errors.

For example, a solution could be to provide more training examples for those categories that the algorithm fails to classify.

The matrix shown in Figure 5.12 is not normalized: to get a more accurate view of the model's performance, it is useful to report the number of correct or incorrect predictions relative to the total number of instances per class.

Chapter 6

Training Set – creation of synthetic data

In this thesis project, Machine Learning classifiers are used to recognize the position and type of stroke. However, to solve the classification problem, the ML algorithms need to be trained on a very large dataset.
A training set of the scattering parameters can be obtained through clinical trials, laboratory measurements with the anthropomorphic phantom or with electromagnetic (EM) simulations. Unfortunately, all these options are not feasible as they require a great effort in terms of time.
To overcome this limit, an algorithm based on the Born approximation and linearization of the scattering operator, has been implemented on MatlabR2021b.
The method considers two scenarios (healthy and pathological) and allows to derive the S parameters starting from the dielectric contrast.
In about 16 hours it creates a large dataset made up of 10000 different examples.
The Matlab code was executed on the *micenea* server of the Politecnico di Torino, Intel (R) Xeon (R) CPU E5-2670 @ 2.60 GHz with a memory of 256 GB and 24 cores.
An accurate description of the steps followed is given below [71].

6.1 Reference model: background scenario

To create the training set, it is necessary to have a reference model, a background scenario, which can be modified to generate different stroke conditions.
A 3D model that reproduces the head, the brain (considered as a homogeneous medium) without target and the MWI system, has been recreated using GiD CAD software [48].
A discretization of this volume with a mesh, created 19221164 tetrahedra. At this point the discretized model has been subjected to a full-wave FEM simulation. This operation will be detailed in paragraph 8.1.
The FEM returns in output two type of files associated to the background scenario:
- ".VTK" : it contains the EM field inside the entire simulated system.
- ".txt": it is a list of the real and imaginary part of the S parameters for each antenna pair.

Both files are read using Matlab scripts. First, the domain of interest (DOI) is defined: it consists of 794502 tetrahedra of which 459277 are related to the brain and the remaining 335225 to the head.
For each tetrahedron, the spatial coordinates *x*, *y*, *z* of the barycenter and of the vertices that constitute it, are always known.
Among the EM fields in the .VTK file, only those relating to the defined DOI are considered.
The scattering parameters are arranged inside an *m x n* matrix called *S.inc*, where n and m are respectively the transmitting and receiving antenna ports (*n = m = 24*).
The 3D model of the head and brain without target, the electric fields and the S parameters associated with each tetrahedron of the DOI, are used for the creation of the scattering operator that will be described in paragraph 6.3.

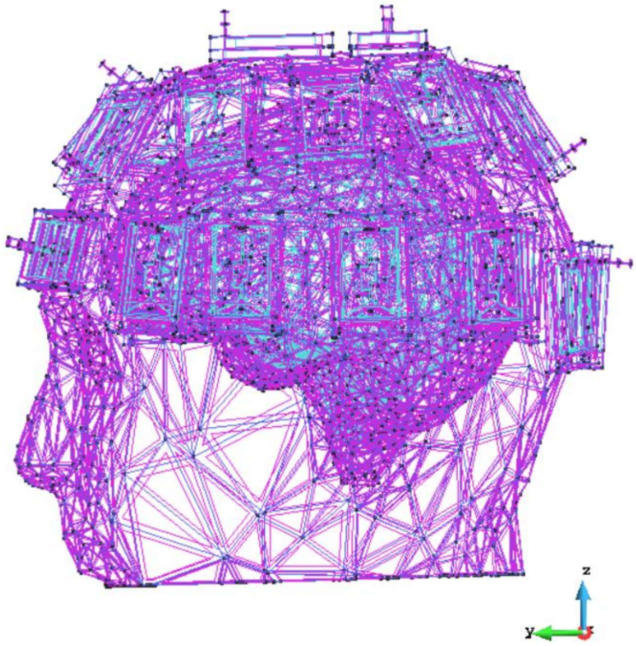


Figure 6.1. GiD model of the head, brain, and MWI system: background scenario.

6.2 Dielectric contrast

Dielectric contrast is what allows MWI system to distinguish healthy tissue from pathological tissue. In particular, at the working frequency of 1 GHz, the dielectric permittivity of ischemic stroke and hemorrhagic stroke is respectively lower and greater, than the permittivity of the surrounding brain tissue [33]. In the Table 6.1. the dielectric characteristics of healthy, ischemic, and hemorrhagic cases, are reported. The permittivity and conductivity values refer to measurements obtained from the phantom mixtures, that have been recreated in the laboratory (section 4.2). The healthy case, called also *average brain*, represents the homogeneous compound whose dielectric properties have been obtained as the average of those of the gray and white matter.

Case	ε_r	**σ (S/m)**
Healthy (average brain)	42.58088	0.787368
Ischemic	36	0.72
Hemorrhagic (blood)	63.41461	1.576326

Table 6.1. Dielectric properties of healthy, ischemic, and hemorrhagic cases, at 1 GHz. The healthy case which represents the background scenario, refers to a homogeneous medium called "average brain" (Paragraph 4.2).

To generate new cases with strokes, a certain number of tetrahedra inside the healthy head is selected: the properties of those tetrahedra are modified, with respect to the homogeneous background.
It is possible to define the dielectric contrast as the relative difference between the permittivity value related to a background scenario without a target (healthy case), and the permittivity value associated with a pathological scenario (ischemic or hemorrhagic case). It can be calculated with the following formula (6.1):

$$\Delta\chi(\bar{r}) \triangleq \frac{\varepsilon_s(\bar{r}) - \varepsilon_b(\bar{r})}{\varepsilon_b(\bar{r})} = \frac{\varepsilon_s(\bar{r})}{\varepsilon_b(\bar{r})} - 1 \tag{6.1}$$

Where $\bar{r}$ is a generic point in the DoI, $\varepsilon_s(\bar{r})$ and $\varepsilon_b(\bar{r})$ are the complex permittivity, located in r, of the corresponding average brain and of the (ischemic or hemorrhagic) stroke, respectively. The dielectric contrast for the three possible cases, at 1 GHz, is shown in the Table 6.2.

Case	**Dielectric contrast**
Healthy (average brain)	0
Ischemic	- 0.14769 - j 0.020649
Hemorrhagic (blood)	+ 0.54028 - j 0.15347

Table 6.2. Dielectric contrast associated to healthy, ischemic, and hemorrhagic cases, at 1 GHz.

Note that the contrast can also be defined considering the same scenario with stroke, but in two different instants of time [71].

6.3 Linearized Integral Operator

Now consider a scenario with target. The differential scattering matrix can be defined with the following expression (6.2):

$$\Delta S_{p,q} = S^t_{p,q} - S^b_{p,q} \tag{6.2}$$

where $\Delta S_{p,q}$ is the differential scattering matrix for each p, q antenna pair ($p = q = 1,\ldots, 24$), obtained as a value-by-value subtraction, between the elements that make up the S matrix related to the case with target ($S^t_{p,q}$), and those of the S matrix related to the background scenario ($S^b_{p,q}$). $S^t_{p,q}$ is the unknown factor of the problem. $\Delta S_{p,q}$ is due to the contrast variation $\Delta\chi$, introduced by the presence of the target in the reference scenario. Differential scattering matrix and dielectric contrast variation are related to each other through the following non-linear equation (6.3) [72,73]:

$$\Delta S_{p,q} = \frac{-j\omega\varepsilon_b}{2\, a_p a_q}\int_{DOI} \overline{E}_p^{\,b}(\overline{r})\cdot \overline{E}_q^{\,t}(\overline{r})\Delta\chi(\overline{r})\, d^3\overline{r} \tag{6.3}$$

where:

- j is the imaginary unit;
- ω is the angular frequency of the antennas ($2\pi f$, with $f = 1GHz$);
- ε_b is the complex "background" dielectric permittivity (*average brain*);
- $a_p a_q$ are the power waves at the *p,q* antennas ports, respectively;
- DOI is the domain of interest: the brain and the head;
- $\overline{r}$ is the position vector: it indicates the distance between a point r in the DOI and the positions of the *p, q* transmitting and receiving antennas;
- $\cdot$ is a dot product;
- $\overline{E}_p^{\,b}$ is the electric field radiated in each point r of the DOI by the antenna p in the *background scenario*;
- $\overline{E}_q^{\,t}$ is the electric field radiated in each point r of the DOI by the antenna *q*, when in the DOI there is a target (*test scenario*) [52].

If it is assumed that the difference between the test scenario and the background one is negligible, and this assumption is valid only if the stroke has a limited extent, the Born approximation can be applied.
The latter is valid for "weak" scattering phenomena, for which $\chi \ll 1$ and $d \ll \lambda$, where χ is the contrast, d is the size of the target, and λ is the wavelength of the field EM within the area occupied by the stroke.

The electric field radiated by the antenna p in a background scenario is approximated to the electric field radiated by the antenna q in the case of a test scenario (6.4):

$$\overline{E}_q^{\,t}(\overline{r}) \cong \overline{E}_p^{\,b}(\overline{r}) \qquad (6.4)$$

Then the equation (6.3) that links the differential scattering matrix to the variation of the dielectric contrast, depends only on the electric field radiated by the antennas p and q in a background scenario, and can be written as (6.5):

$$\Delta S_{p,q} \cong \mathcal{S}\{\Delta\chi(\overline{r})\} \qquad (6.5)$$

Where $\mathcal{S}$ is the linearized integral operator, which maps the relationship between the contrast and the scattered field.
If this procedure is applied to the discretized 3D model, the equation (6.5) can be written as (6.6):

$$[\Delta S] \cong [\mathcal{S}]\,[\Delta\chi] \qquad (6.6)$$

Where $[\mathcal{S}]$ is now the discretized linear scattering operator, which can be calculated with the following equation (6.7) [18]:

$$[\mathcal{S}]_{m,n} = \frac{-j2\pi f \varepsilon_b}{2\,a_p a_q}\,\overline{E}_p^{\,b}(\overline{r_n})\cdot\overline{E}_q^{\,b}(\overline{r_n})\,\Delta V_n = -j\pi f \varepsilon_b Z_r\,\overline{E}_p^{\,b}(\overline{r_n})\cdot\overline{E}_q^{\,b}(\overline{r_n})\Delta V_n \qquad (6.7)$$

where Z_r is the impedance of the monopolar antenna used for the MWI system, $\overline{r_n}$and ΔV_n are the coordinates (x,y,z) of the barycenters and the volume of each n tetrahedra (794502) that make up the DOI. The discretized linear scattering operator will therefore have dimensions $m\ x\ n$ (576 x 794502): m is the number of each p and q antenna pair, that make up the MWI system.
The truncated singular value decomposition (TSVD), a particular factorization of a matrix based on the use of eigenvalues and eigenvectors, is applied to the scattering operator (6.8):

$$[\mathcal{S}]_{m,n} = U_{m,n}\,S_{n,n}\,V_{m,m}^{*} = \sum_{\mathrm{i}=1}^{n} u_{\mathrm{i}}\,\sigma_{\mathrm{i}}\,v_i^{*} \qquad (6.8)$$

Where:

- U is the unitary matrix with dimensions $m\ x\ n$ (794502 x 24): the u_{i} elements of this matrix consist of orthonormal vector columns, called left singular vectors of $\mathcal{S}$.
- S is the diagonal matrix with dimensions $n\ x\ n$ (794502 x 794502): the diagonal elements of S are called single values of $\mathcal{S}$.

- V* is the conjugate transpose of V, unitary matrix with dimensions $m\ x\ m$ (24 x 24): the v_{i} elements of this matrix consist of orthonormal vector columns, called right singular vectors of $\mathcal{S}$.

The discretized and linearized scattering operator $[\mathcal{S}]_{m,n}$ depends only on the incident electric field related to the background scenario, and therefore must be calculated only once.
At this point, equation (6.6) can be rewritten as follows (6.9) [76]:

$$[\,\Delta S\,] = \sum_{\mathrm{n}\in\mathrm{DOI}} \sigma_{\mathrm{n}} \langle \Delta\chi, [v_{\mathrm{i}}{}^{T}]\rangle [u_n] \tag{6.9}$$

Multiplying the S, V, D vectors of the discretized linear scattering operator, by Δχ (the vector that contains the contrast variation for each tetrahedron of the brain), the differential scattering matrix due to the presence of the target is calculated. Finally, from equation (6.2) the scattering matrix related to the scenarios with the target is obtained (6.10):

$$S^{t}_{p,q} = \Delta S_{p,q} + S^{b}_{p,q} \tag{6.10}$$

The S parameters of the case without target ($S^{b}_{p,q}$) are computed, like the TSVD of the discretized linear scattering operator, only once.
Summing up: from the FEM simulations performed on the reference scenario, the indices related to the 3D model of the head and brain without target, the electric fields associated with each point in the DOI, and the S parameters are imported.
The S matrix related to the test scenario is obtained as the sum of $S^{b}_{p,q}$ and $\Delta S_{p,q}$. This latter is calculated by multiplying the contrast $\Delta\chi$ by the linear discretized scattering operator. Generating different scenarios basically consists of generating different $\Delta S_{p,q}$ with the procedure just explained [71].

6.4 Division into classes

ML algorithms must be able to recognize the type and position of stroke. In this regard, a subdivision of the head on the sagittal and frontal planes is considered . In this way there are four regions of the head that identify 4 different positions: front-left (FL), front-right (FR), back-left (BL), and back-right (BR) (Fig.6.2) [71].

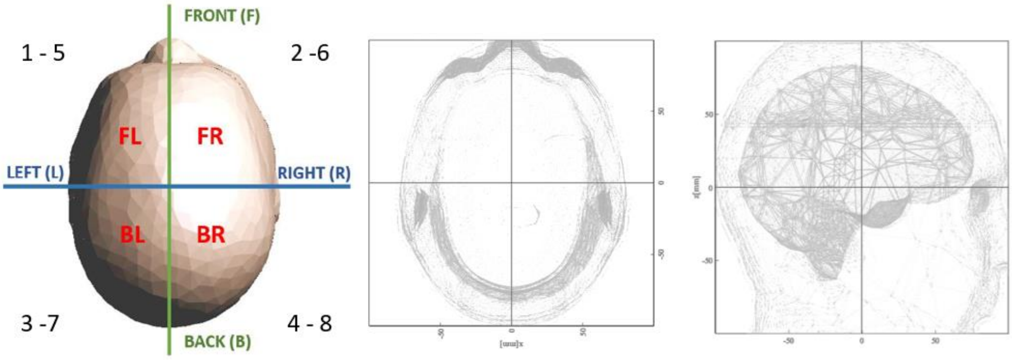


Figure 6.2. Division of the head into 4 regions: front-left (FL), front-right (FR), back-left (BL), back-right (BR).

The classification problem therefore considers nine classes (Tab.6.3). Class 0 corresponds to the healthy case, the one without target (N). The classes 1,2,3,4 correspond to the case of ischemic stroke (I) extended to the areas FL, FR, BL, and BR, respectively. The classes 5,6,7,8 correspond to the case of hemorrhagic stroke(H) extended to the areas FL, FR, BL, and BR respectively.
The position of the stroke that determines the class, is chosen randomly by the algorithm. The class is assigned based on the position that the center of the stroke occupies with respect to the head. The center of the head is the origin of the axes of the Cartesian x,y,z plane.

Classes	
0	N
1	I_FL
2	I_FR
3	I_BL
4	I_BR
5	H_FL
6	H_FR
7	H_BL
8	H_BR

Table 6.3. The nine classes that the ML algorithm should be able to classify.

6.5 Stroke size

The stroke is obtained by selecting a set of tetrahedra that make up the model of the healthy brain and modifying their dielectric properties, making them different from the surrounding homogeneous tissue.

The stroke generated has a spherical shape and the center is randomly chosen among the barycenters of the tetrahedra within the brain.

A clinical study has shown that 54 cm^3 can be a reasonable estimate of the mean final volume of an ischemic stroke, with a variability ranging from 19 cm^3 to 100 cm^3. The high variability is due to the influence of several factors such as the type of scanner and the imaging sequences used. For example, computed tomography (CT) scans tend to underestimate the size of the infarcted area, while T2-weighted MRI images tend to overestimate [74,75].

Furthermore, the instant in time in which the image is acquired has a significant impact. In the first days, ischemic stroke has a greater volume due to tissue swelling, while starting from the second week the volume tends to decrease [6].

In this perspective, five possible rays were considered in the Matlab code, from 1 to 3 cm, with a step of 0.5 cm, which correspond to a volume between 4.2 cm^3 and 113.1 cm^3 (Fig.6.3).

The lower limit of the stroke radius corresponds to the maximum resolution of the MWI system, while the upper limit makes the approximation of Born still valid [71].

If some of the selected tetrahedra are outside the brain they are discarded, while if more than half of the selected tetrahedra fall outside the brain, that stroke case is discarded.

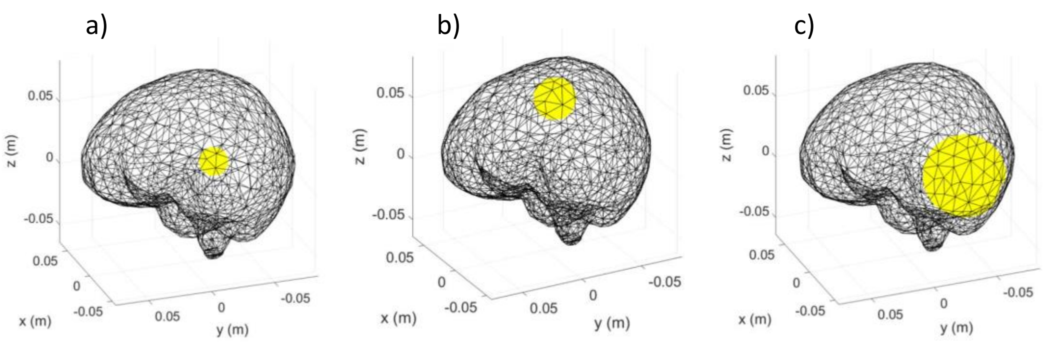


Figure 6.3. Examples of spherical strokes generated with volume equal to: a) $4.2\ cm^3(radius = 1\ cm), b)$ $10.7\ cm^3(radius = 2\ cm)$ *and* $c)$ $113.1\ cm^3(radius = 113.1\ cm)$ *[71].*

6.6 Noise thresholds

Four noise thresholds have been added to the dielectric contrast in order to obtain four corresponding noise levels on scattering parameters equal to: -110 dB, -105 dB, -95 dB and -90 dB .
The four noise thresholds applied to the dielectric contrast are shown in the table in the first two columns (Tab.6.4).

Noise levels in the dielectric contrast space		Noise levels in the scattering parameters space
Re($\delta\chi$)	Im($\delta\chi$)	
0.001583333	0.00033	-90 dB
0.000863636	0.00018	-95 dB
0.000513514	0.000107027	-105 dB
0.000283582	5.91045E-05	-110 dB

Table 6.4. Noise levels in the dielectric contrast space and in the scattering parameters space.

The first and second columns contain the values to be added or subtracted from the real and imaginary part of the dielectric contrast for each of the N tetrahedra, to obtain the four corresponding noise thresholds reported in the third column.
The passage from one space (dielectric contrast) to another (S parameters) is made possible by the discretized linear scattering operator $[\mathcal{S}]$ (see equation 6.8), with the following formula (6.11):

$$[\delta S] \cong [\mathcal{S}][\delta\chi] \tag{6.11}$$

where $[\delta\chi]$ is an N x 1 array collecting the random noise associated to each tetrahedron in the dielectric contrast space, while $[\delta S]$ is an M x 1 array that collects the corresponding noise in the scattering parameters space.
The minimum chosen noise level (-110 dB) is comparable with the noise floor of a medium-quality vector network analyzer [45] like that used in the laboratory, and the maximum noise level (-90dB) is up to 20 dB higher than the noise floor.
The addition of the noise is performed randomly, each time choosing a variation of the dielectric contrast.
The addition of the noise allows to have, starting from an example that reproduces a type of stroke in a specific position, five cases of the same class but slightly different, to be given as input to the ML algorithm during training [71].

6.7 Dataset size

The generated dataset counts 10000 examples. Of these, 1000 are without target, 4500 are ischemic, and 4500 are hemorrhagic. Except for class 0 which contains 1000 elements, the remaining classes (1,2,3,4,5,6,7,8) consist of 1125 examples of ictus with different sizes and noise thresholds applied.

6.8 Feature extraction: Amplitude and Complex dataset

Usually before the implementation a machine learning model, the dataset is analyzed and manipulated to extract meaningful properties from those already available. This process is called *feature extraction*.

The implemented method returns a dataset consisting of 10000 differential scattering matrices 24x24, containing the S parameters in the form of complex numbers. Since the S matrix is symmetric, the elements above the diagonal and those below it, are equal. The use of data that does not add new information, would only slow down the processing time by the ML algorithms. For this reason, only the elements belonging to the upper triangular matrix are considered for each of the 10000 examples.

The features of an element are arranged in a single line, following the order shown in Figure 6.4. The resulting array contains 300 input features for the ML algorithms. However, the machine learning does not work with complex number.

For this reason, the dataset is split into its real and imaginary parts, thus obtaining a row of 600 elements. The 601st column is occupied by the labels. The training dataset is called *complex dataset*. The *amplitude dataset* instead contains the module of the complex number associated with the scattering parameter. In this way, for each of the 10000 cases produced, a line of 300 features is obtained, to which the label column is added. The 10000x601 and 10000x301 matrices are finally converted into two ".csv" file.

1	25	49	73	97	121	145	169	193	217	241	265	289	313	337	361	385	409	433	457	481	505	529	553
2	26	50	74	98	122	146	170	194	218	242	266	290	314	338	362	386	410	434	458	482	506	530	554
3	27	51	75	99	123	147	171	195	219	243	267	291	315	339	363	387	411	435	459	483	507	531	555
4	28	52	76	100	124	148	172	196	220	244	268	292	316	340	364	388	412	436	460	484	508	532	556
5	29	53	77	101	125	149	173	197	221	245	269	293	317	341	365	389	413	437	461	485	509	533	557
6	30	54	78	102	126	150	174	198	222	246	270	294	318	342	366	390	414	438	462	486	510	534	558
7	31	55	79	103	127	151	175	199	223	247	271	295	319	343	367	391	415	439	463	487	511	535	559
8	32	56	80	104	128	152	176	200	224	248	272	296	320	344	368	392	416	440	464	488	512	536	560
9	33	57	81	105	129	153	177	201	225	249	273	297	321	345	369	393	417	441	465	489	513	537	561
10	34	58	82	106	130	154	178	202	226	250	274	298	322	346	370	394	418	442	466	490	514	538	562
11	35	59	83	107	131	155	179	203	227	251	275	299	323	347	371	395	419	443	467	491	515	539	563
12	36	60	84	108	132	156	180	204	228	252	276	300	324	348	372	396	420	444	468	492	516	540	564
13	37	61	85	109	133	157	181	205	229	253	277	301	325	349	373	397	421	445	469	493	517	541	565
14	38	62	86	110	134	158	182	206	230	254	278	302	326	350	374	398	422	446	470	494	518	542	566
15	39	63	87	111	135	159	183	207	231	255	279	303	327	351	375	399	423	447	471	495	519	543	567
16	40	64	88	112	136	160	184	208	232	256	280	304	328	352	376	400	424	448	472	496	520	544	568
17	41	65	89	113	137	161	185	209	233	257	281	305	329	353	377	401	425	449	473	497	521	545	569
18	42	66	90	114	138	162	186	210	234	258	282	306	330	354	378	402	426	450	474	498	522	546	570
19	43	67	91	115	139	163	187	211	235	259	283	307	331	355	379	403	427	451	475	499	523	547	571
20	44	68	92	116	140	164	188	212	236	260	284	308	332	356	380	404	428	452	476	500	524	548	572
21	45	69	93	117	141	165	189	213	237	261	285	309	333	357	381	405	429	453	477	501	525	549	573
22	46	70	94	118	142	166	190	214	238	262	286	310	334	358	382	406	430	454	478	502	526	550	574
23	47	71	95	119	143	167	191	215	239	263	287	311	335	359	383	407	431	455	479	503	527	551	575
24	48	72	96	120	144	168	192	216	240	264	288	312	336	360	384	408	432	456	480	504	528	552	576

a)

1	25	26	49	50	51	. . . .	576	Target
x1	x2	x3	x4	x5	x6	. . . .	x300	

b)

Re(1)	Im(1)	Re(25)	Im(25)	Re(26)	Im(26)	. . . .	Im(576)	Target
x1	x2	x3	x4	x5	x6		x600	

Figure 6.4 S parameters belonging to the upper triangular scattering matrix are arranged in a single row. Organization of the features (x_1, x_2, x_3, …. , x_n) that make up the Amplitude dataset (a) and the Complex dataset (b).

Chapter 7

Flowchart of the implemented algorithm

The next paragraphs show schematically the steps that have been carried out for the creation of the synthetic training dataset. For a more complete explanation, the reader is referred to the previous Chapter 6.

The Matlab code sequentially performs the following steps:

- creation of the linear discretized scattering operator

- creation of the dataset made up of healthy cases

- creation of the dataset composed of pathological cases. The procedure used to create the ischemic stroke is identical to that used to create the hemorrhagic case, so they are reported only once. What changes is the associated dielectric contrast value with respect to the background scenario which is the *average brain*. Radius, position, and noise floor are randomly chosen within a fixed range.

- features extraction: creation of the *amplitude* dataset and the *complex* dataset.

7.1 Linearized Integral Operator

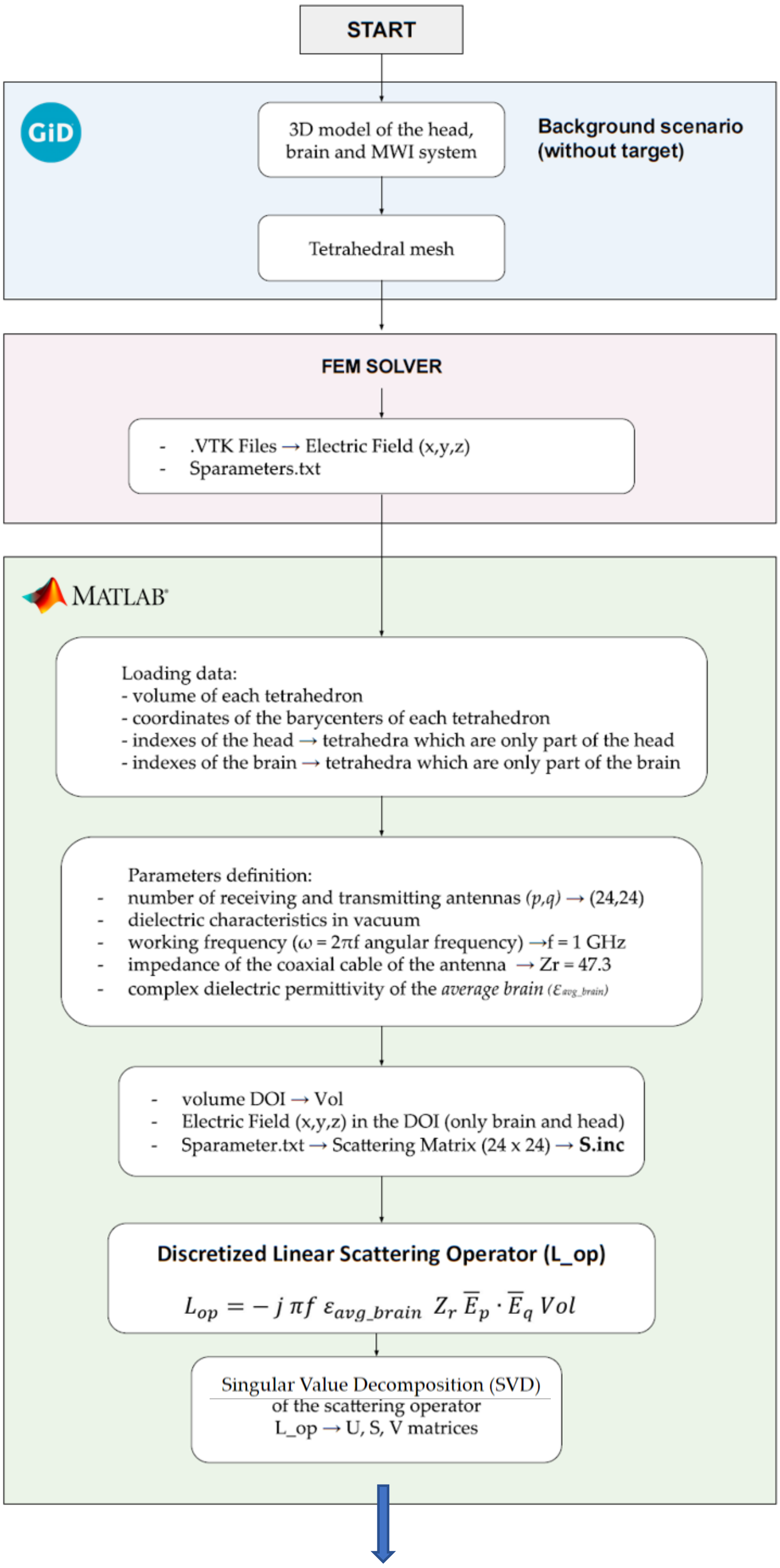


Figure 7.1. Flowchart: creation of discretized linear Scattering Operator.

7.2 Healthy cases

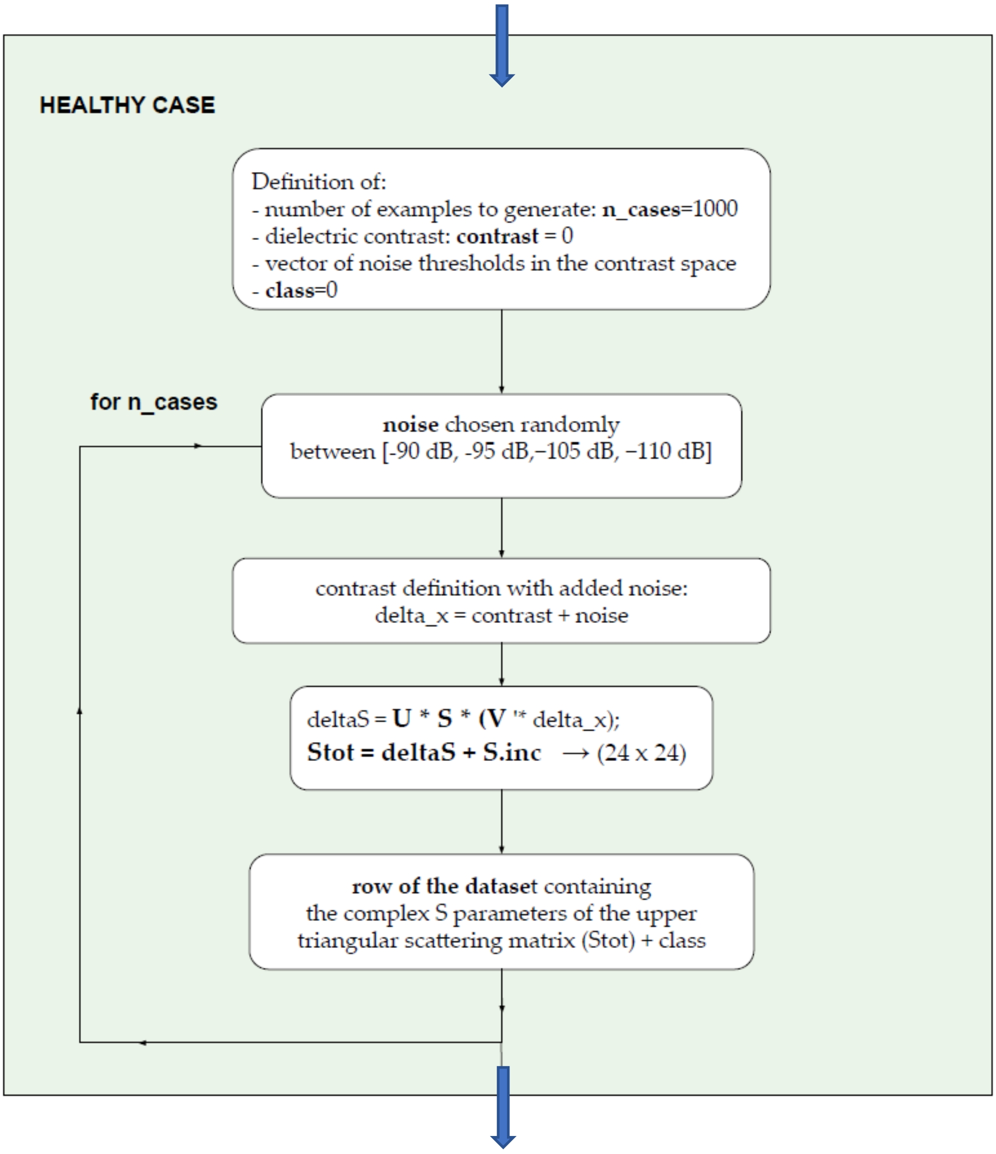


Figure 7.2. Flowchart generation of healthy cases.

7.3 Cases with stroke

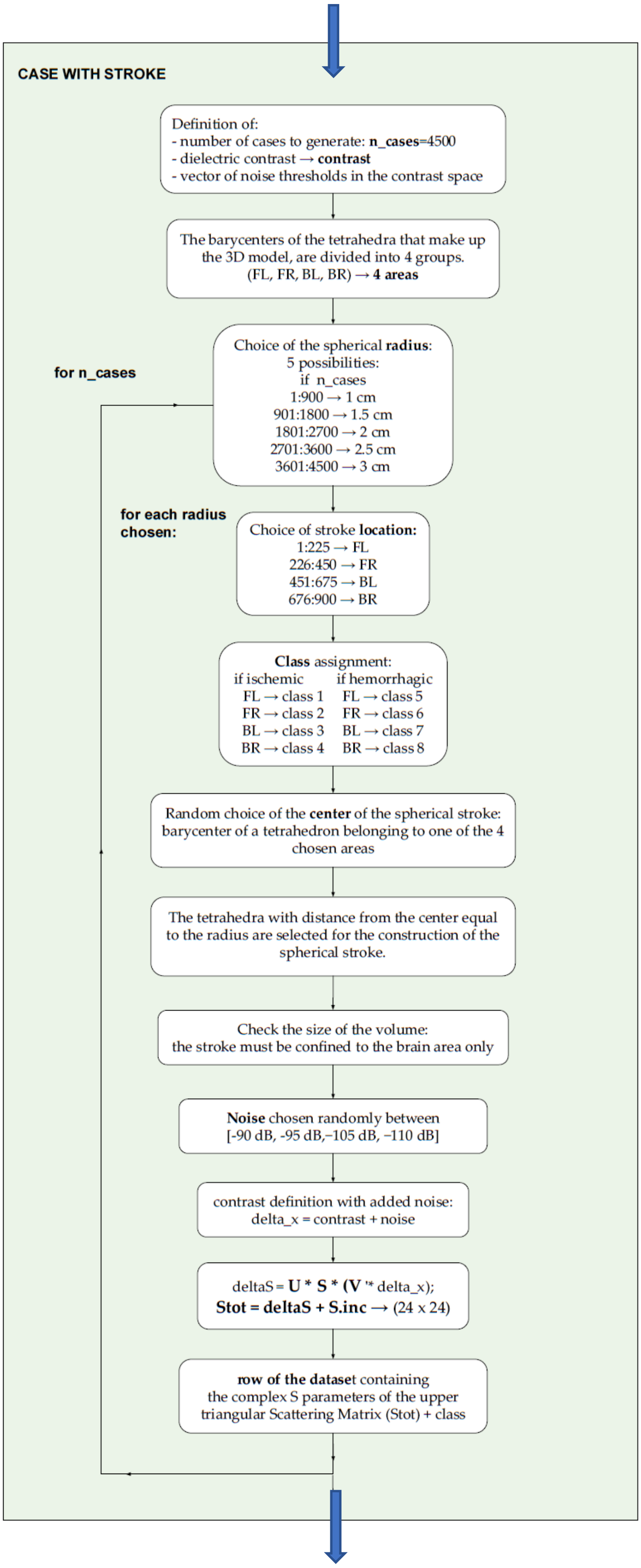


Figure 7.3. Flowchart: generation of cases with stroke.

7.4 Features extraction

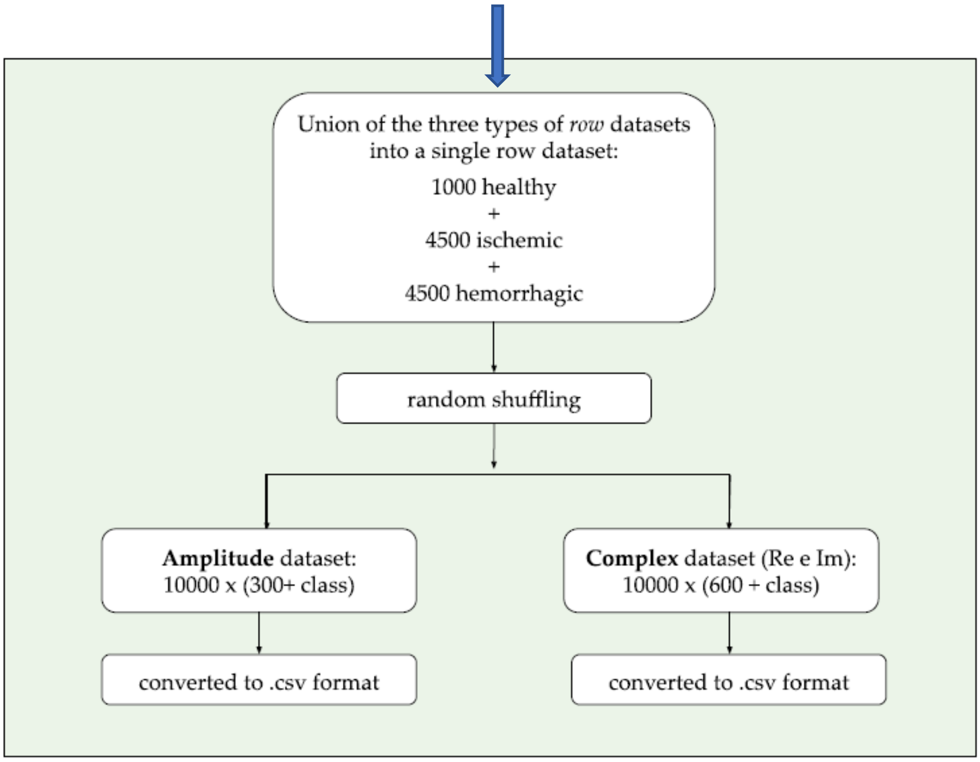


Figure 7.4. Flowchart: features extraction.

Chapter 8

Testing Set – creation of synthetic data via Full-Wave FEM simulations

The full-wave Finite Element Method (FEM) simulations are adopted to produce a testing set that will be used to evaluate the performance of the trained ML algorithms. Based on the performance achieved, it will also be possible to establish whether the method implemented for generating the training set data, is effective or not.

The scattering parameters of the test set were obtained by combining the use of a processing software for numerical simulations (GiD), a programming and numerical calculation platform (Matlab2021b) and an internal Full-Wave software, based on the finite elements method (FEM) [18].

The figure below briefly shows the steps followed (Fig.8.1).

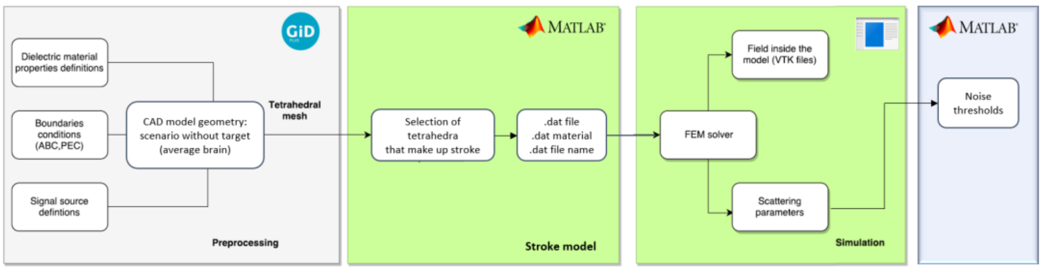


Figure 8.1. Flowchart related to the steps followed for the creation of test set.

8.1 Preprocessing (CAD model without target)

The CAD model developed on GiD software includes the volume of the head and the antennas of the MWI system. The brain is intended as a homogeneous material having the dielectric characteristics of the *average brain* (paragraph 4.2).

The model refers to a healthy-background scenario, so it is without target.

After building the geometry, the dielectric properties of the materials, the boundary conditions and the signal sources are defined.

At this point it is possible to perform a discretization of the model. This step consists in dividing the entire volume into many tetrahedra. The greater the number of tetrahedra, the faster the problem will converge towards a solution. In case the mesh is not precise enough, it is likely that some elements of the model will be ignored. A fine mesh size will provide more accurate solutions, but it will take longer to compute [48, 77].

8.2 Stroke model

The head and brain model, and the mesh processed on GiD, are imported in ".stl" format on Matlab2021b.

The implemented function selects a spherical or ellipsoidal volume, consisting of a certain number of tetrahedra, close to each other, that will have different dielectric properties with respect to the background. The volume is within the range used for the training set (from 4.2 cm^3 to 113.1 cm^3).

First, the x, y and z coordinates of the point that will act as the center of the hemorrhage or the infarcted area, must be set. Then, the radius of the sphere or the length of the semiaxis of the ellipsoid are defined.

Figure 8.2. shows four examples of generated strokes, relating to the four different areas (BR, BL, FR, FL), with different shapes and sizes. The central point of the head that separates the four zones, has coordinates (0,0,0).

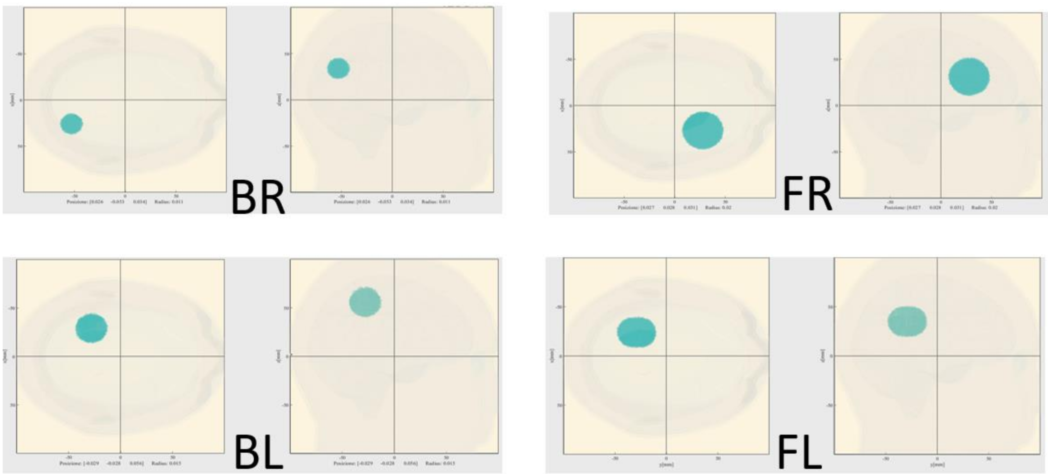


Figure 8.2. Examples of generated stroke models, related to the four areas of the brain (BR, BL, FR, FL), with different shape and size.

In total, 29 different cases were created: for some of them, the stroke falls in a specific portion of the head, while for others the stroke center is located near the border that divides the BL, BR, FL, FR areas. Their position is indistinguishable with the naked eye (Fig.8.3).

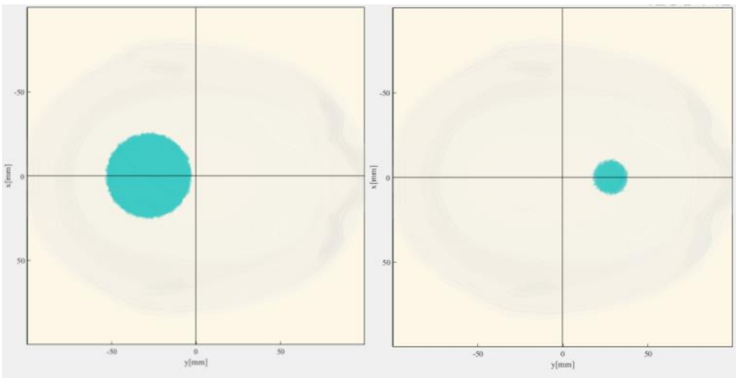


Figure 8.3. Examples of generated stroke models, with the center of the sphere near the border that divides the brain into the four areas. The position of the stroke is indistinguishable to the naked eye.

After verifying, through a plot, that the stroke is confined to the brain area only, a .DAT file is drawn up.

This text file essentially describes the generated example, and it will be one of the FEM solver inputs. It is composed by several sections:

- *Coordinates section* includes the number and Cartesian coordinates of the vertices of each point in the DOI
- *Element section* returns the number of faces of tetrahedra, vertexes coordinates and barycenter. Each face is associated with a number that identifies the related material.
- *PEC section* includes perfect electric conductor faces and information about the input/output ports.
- *Output definition section* indicates what the simulation output parameters should be.

Thanks to this file, FEM simulator can assign materials and conditions at each element of the mesh [78].

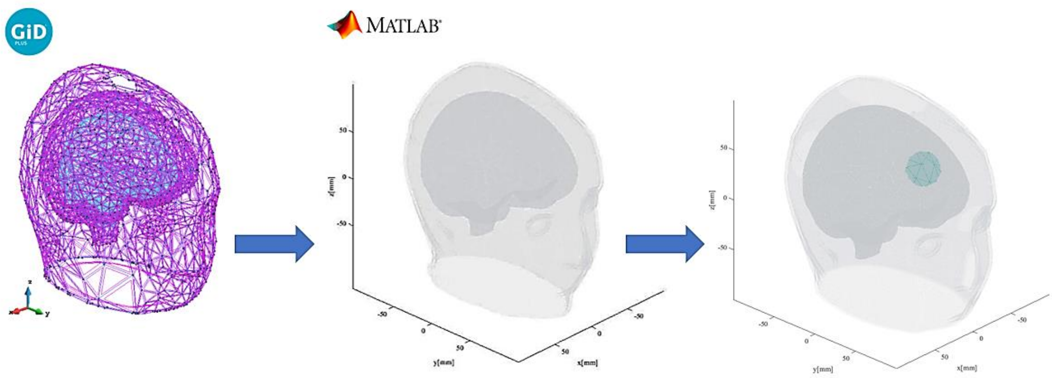
For the drafting of the new .DAT file, the .DAT file related to the no-stroke case exported by GiD, is taken as a reference, and modified according to the radius and position choices made previously.

Since only the scattering parameters are required for the ML algorithm, the output definition within the .dat file is set to 0.

In this way, the output will return the parameter file s directly, leaving out the fields and saving a lot of memory space, making the simulation faster.

Until now, the type of stroke has not yet been defined. The only thing known is that some tetrahedra of the brain model have different dielectric characteristics than the *average brain*, as they have been assigned another material. To characterize the latter, it is necessary to act through the "*material.dat*" file: add a unit to the number of materials and write ε_r and σ values at the end of the file itself. Again, the dielectric properties refer to the values obtained experimentally in the laboratory (Paragraph 4.2, Paragraph 6.2-Tab.6.1). The.dat file is renamed and at this point the model is subjected to a simulation with the finite element method

(FEM). Figure 8.4. summarizes the steps needed to add the stroke to the background scenario.



Figures 8.4. Steps performed to create a test example with stroke.

8.3 Finite Element Method (FEM) simulations

The simulation represents the intermediate level between an ideal and a realistic situation.
Finite Element Method (FEM) is a numerical algorithm which aim is to convert partial differential equations into a set of linear algebraic equations, used to obtain approximate solutions to mathematical problems [79].
It can manage complex geometries and accurately calculate the radiated and reflected fields. Each tetrahedron is a sub-domain, that composes the model, and represents a finite element, where the method finds the approximated solution.
A full-wave analysis consists in solving the complete set of Maxwells equations without any simplifying assumptions. The "full wave" connotation associated with the type of simulation, indicates that all the components of the fields are considered: E_x, E_y, E_z, H_x, H_y, H_z.
Before starting the FEM simulation with the in-house software, the inputs must be properly arranged.
Five elements coexist within the same folder: the executive file, two libraries, the file name, the .DAT file and the material file.
The file name tells the program which is the .DAT file containing all the information useful for the simulation. The material file reports the number of points in frequency, number of materials, the values of the frequencies to be considered (in this case one, equal to 1 GHz) and finally the relative permittivity and conductivity of each material.

The executable is opened directly from the shell and the simulation takes about an hour to perform the calculations.
In output from the simulation a text file is obtained containing for each pair of antenna ports (m x n), the real part and imaginary part of the S parameters, at the considered working frequency.
The file is read by a Matlab script, whose output return a matrix n x m (24x24).
Since the FEM solver does not implement the equation for the case $m = n$, the -1 of the formula (see equation (8.1)) is corrected later for the elements of the diagonal of the S matrix [48].

$$S_{mn} = \begin{cases} \dfrac{\iint_{S_p} \underline{E}_m \cdot \underline{E}_n^{inc}\, dS}{\iint_{S_p} |\underline{E}_n^{inc}|^2\, dS} & if\ m \neq n \\ \dfrac{\iint_{S_p} \underline{E}_m \cdot \underline{E}_n^{inc}\, dS}{\iint_{S_p} |\underline{E}_n^{inc}|^2\, dS} - 1 & if\ m = n \end{cases} \tag{8.1}$$

Where S_p is the port section, $\underline{E}_n^{inc}$ is the electric field transmitted by port n and $\underline{E}_m$ is the electric field evaluated at the port m.
The same procedure was applied to all twenty-nine examples of generated strokes.

8.4 Addition of noise

The generated S matrices ($S_{full-wave}$) are contaminated with white noise. This allows to have data variety also in the case without target and to have a certain specularity with the training set.
Four levels of dielectric contrast noise are chosen, in order to have four corresponding noise levels on S parameters equal to: -110 dB, -105 dB, -95 dB and -90 dB.
The procedure applied is identical to that adopted for the training examples (Paragraph 6.6), with the difference that, having no dielectric contrast $\Delta\chi$ in output from the FEM simulations, the variation of the dielectric contrast $\delta\chi$ is calculated based on an initial dielectric contrast equal to zero.
As a result, the dielectric variation $\delta\chi$ is equal to the set noise only.
By applying the linear scattering operator ($\mathcal{S}$), calculated for the construction of the training set (see equation 6.8), to the values of the matrix $\delta\chi$, for each noise threshold considered, the corresponding δS matrix (8.2) is obtained:

$$\delta S\left(\mathrm{r_p}, \mathrm{r_q}\right) = \mathcal{S}(\delta\chi) \tag{8.2}$$

This noise has been added to the S matrix of each test example generated (8.3):

$$[S_{with\ noise}] = \left[S_{full-wave}\right] + [\delta S] \tag{8.3}$$

The scattering matrices and the values obtained were always compared, to verify the correctness of the steps (Fig.8.5).

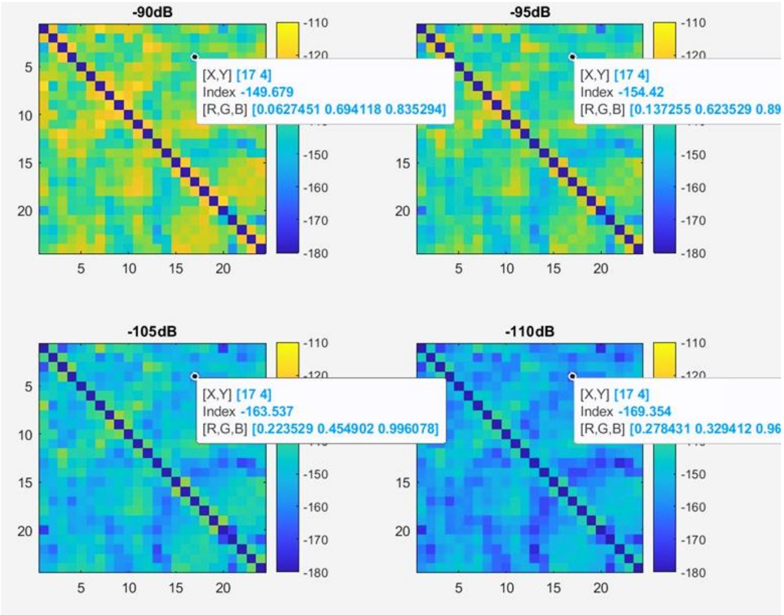


Figure 8.5. The four δS matrices (24x24) added to the noise-free S matrices (24x24) generated via full-wave FEM simulations.

With the addition of the four noise thresholds, the number of cases goes from 29 to 145.

As regards the training set, the dataset was further elaborated to obtain a testing set whose features are constituted by the amplitude (*Amplitude Dataset*) and by the real and imaginary part (*Complex Dataset*), of the scattering parameters. Finally, these two datasets were converted to .csv file.

The Table 8.1. summarizes the characteristics of the generated test set.

CLASS			*n° of cases without noise*	*TOT n° of cases*	SHAPES AND DIMENSIONS					
					sphere					*ellipsoid*
0	**Healthy**	N	2	10						
1	**Ischemic**	FL	3	15	1 cm	2 cm	2 cm			
2		FR	3	15	1 cm	1.5 cm	2.5 cm			
3		BL	4	20	1 cm	1 cm	2.5 cm			0.75 cm x 1 cm x 1 cm
4		BR	4	20	1 cm	2.5 cm	2.5 cm	3 cm		
5	**Hemorrhagic**	FL	5	25	1 cm	1 cm	1.5 cm	2.5 cm	3 cm	
6		FR	2	10	1 cm	2 cm				
7		BL	4	20	1.5 cm	2.5 cm	2.5 cm			0.75 cm x 1 cm x 1.5 cm
8		BR	2	10	2.5 cm	2.5 cm				
			29	**145**						

Table 8.1. Characteristics of the generated test set.

Chapter 9

Application of Machine Learning

This section will briefly describe the steps that were followed to apply the ML algorithms to the previously prepared datasets. The training set is the one generated by the method based on the Born approximation, considering a linear scattering operator (Chapter 6), while the test set is generated through full-wave FEM simulations (Chapter 8). Each row of the datasets contains the S parameters of the scattering matrices, organized as explained in Paragraph 6.8. The algorithms, written in Python 3.7 language, have been implemented, trained, and tested on a platform called Google Colaboratory.

9.1 Libraries for Machine Learning

Python is the most widely used programming language in the ML community due to its versatility: it is a high-level, object-oriented, powerful, easy-to-use, dynamic and flexible multi-paradigm language. It has an essential syntax that allows to execute many commands while writing very few lines of code [80].
For the purpose of this project, it was decided to use Python on *Google Colaboratory* (Colab), a free platform that allows to write, run and share Python code on an interactive environment, through the simple use of a browser.
With Google Colab no special downloads or configurations are required: the Python code runs on Google's cloud servers, thus exploiting the power of Google hardware, including GPU and TPU.
Programs written on Colab are called Notebooks and are made up of cells that can contain one or more lines of Python code that will run together. Each cell can be renamed, and this facilitates the reading of the program. Colab allows to run the cells that make up the entire program in sequence or in random order: this feature allows users to modify small pieces of code and see the effects directly, without having to relaunch the whole program. The only precaution that must be taken into account, is that the variables are overwritten each time. Notebooks and input/output data are automatically saved in a Google Drive folder [81].
Colab takes full advantage of the power of Python libraries to analyze, process and visualize data. A brief description of the libraries used (Fig.9.1) will be provided below.

Figure 9.1. Common open-source library for ML application.

- *NumPy* (`np`): is a numerical computing tool. It allows users to work with large matrices and multidimensional arrays. It offers a large collection of mathematical functions [82].
- *pandas* (`pd`): is a fast, powerful, versatile, and easy to use open-source data analysis tool. This library is used to read excel files in .csv format and create complex data structures, organized in easy to manipulate numerical tables [83].
- *matplotlib* (`plt`): is a library that processes data in many formats to create static, animated, and interactive visualizations. In the script used, it is adopted to present the performance of the models through the confusion matrices [84].
- *TensorFlow* (`tf`): is an open platform specific for ML which collects a series of models already developed and optimized. It represents the starting point for many applications [85].
- *Scikit-learn:* is an open-source ML library that supports supervised and unsupervised learning. It also provides different tools for model fitting, data preprocessing, model selection and model evaluation [86].

9.2 Loading and splitting the Dataset

The training set counts 10000 samples, while the test set has 145 records: both consist of a .csv file. The input features to the ML algorithms are provided in the form of amplitude (Amplitude Dataset) and real and imaginary part (Complex Dataset). In the first case they are 300, in the second 600.

First of all, the dataset is manipulated to separate the features array from the array which contains the labels.

At this point, the `train_test_split` function, which is part of the `model_selection` of the scikit-learn library, is imported.

The dataset is split randomly: 80% of the elements are actually used to train the models, while the remaining 20% is used to test the performance of the trained algorithm. The proportion of target is preserved as in original dataset, in the *training* and *validation* set as well.

This technique tries to avoid the phenomenon of overfitting and allows to have an idea of the generalization level of the algorithm. Before making the division, the elements of the dataset are further mixed [86].

9.3 Features scaling and standardization

Feature scaling consists in transforming the input features in such a way that they have a similar scale of values. This operation is used to ensure that the ML algorithms put all the features on the same level without "having preferences". This could happen when there are some features that have greater numeric range than others. If a model is trained with this kind of data, parts of its calculations could have numerical problems.
To apply effectively feature scaling, the same scaling method needs to be adopted for both training and test data [86]. There are several ways to scale features. The one used, is called *standardization method*.
The model `StandardScaler()` imported from scikit-learn calculate mean (μ) and standard deviation (σ) from the training set, and perform a standardization by centering and scaling all the features with the following formula (9.1):

$$x_{i,new} = \frac{x_i - \mu}{\sigma} \tag{9.1}$$

Where $x_{i,new}$ is the new value of the feature x_i.

9.4 Hyperparameters and Grid Search Method

Hyperparameters are parameters that define a ML algorithm and characterize its behavior and performance. They always depend on the dataset to be classified, for this reason there are no default settings suitable for all kinds of classification problems. The construction of the model is performed by setting and modifying the parameters from time to time, and evaluating the results obtained after several attempts. The procedure is not easy because the parameters are many, and the performance of a classifier often depends on their combination. The hyperparameter modification phase is called *tuning*. Manually, it can be a long and cumbersome process. There are automatic estimator methods that allow to maximize performance and save time and resources. The optimization method used in this project, is called Grid Search method [86].
A grid of values, to be tested, for each hyperparameters, is defined by the user.
The `GridSearchCV` function, imported from scikit-learn, performs every possible hyperparameter combination and records, and compares performances, based on accuracy score [87]:

```
grid = GridSearchCV(model (), hyperparameters, scoring='accuracy')

grid.fit(X_train. y_train).
```

Processing time is quite long: a couple of hours for the SVM and k-NN, while the MLP requires about ten and a half hours.

9.5 Construction of models

The hyperparameters used for SVM, K-NN, MLP classifiers, obtained following the application of the Grid Search method on the training set and a subsequent manual fine-tuning phase on the validation set, are listed below. For a more detailed explanation of the meaning of the hyperparameters, the reader is referred to Chapter 5.

SVM

- `kernel='rbf'`: the function that performs the transformation from the input data space to the feature space, is the Gaussian radial basis function.
- `C=300`: the penalty term that adjusts the slack variables (ξ) and serves to balance the trade-off between maximizing margins and minimizing errors, is set to 300.
- `decision_function_shape='ovo'`: the one-versus-one (ovo) approach is adopted to solve the multiclass problem, creating $\frac{k(k-1)}{2}$ different binary classifiers (k = 9), for each class pair. The class with the maximum number of votes is the winner (majority voting strategy).
- `class_weight='balanced'` : in unbalanced problems, it allows to give an appropriate importance even to the less represented class [88].

K-NN

- `metric='manhattan'`: it defines the type of metric for the distance calculation.
- `n_neighbors=3`: it indicates the size of the neighborhood. The 3 closest elements are evaluated, and the class most represented, corresponds to the prediction of the model.
- `weights='distance'`: it means that closer neighbors of a query point will have a greater influence than neighbors which are further away [89].

MLP

- `activation='tanh'`: the hyperbolic tangent function is used as the activation function of each neuron in the hidden layer.
- `alpha= 0.0001`: the term of penalty is set to 0.0001.
- `hidden_layer_sizes=[1000, 500, 250, 100, 50]` means that there are 5 hidden layers in the model: the first consists of 1000 neurons, the second of 500, the third of 250, the fourth of 100, and the last of 50.
- `solver='sgd'`: indicates that for weight adjustment, Stochastic Gradient Descent is used as optimization method.
- `learning_rate='adaptive'`: the learning rate η controls the step-size in updating the weights. If η is small, the weight adjustments are accordingly small. The initial value is set by default equal to 0.001. If the performance of the model tends to improve by two consecutive epochs, the current learning rate is divided by 5. It means that the algorithm is starting to converge towards a local optimum. [90].

9.6 Training, prediction, and accuracy

In Python the training phase is as delicate as it is simple to perform [86]. Once the hyperparameters of the model have been set, just a few commands are enough: the set of features (`X_train`) and the correct labels (`y_train`) are given as input to the ML model.

```
model.fit(X_train, y_train)
```

The algorithm will look for patterns within the features that will lead it to understand the existing link with the assigned class.
To classify a dataset whose class is unknown (`X_test`), the `predict` function is used:

```
model.predict(X_test)
```

Finally, to evaluate the performance of the ML algorithm, the `accuracy_score` metric is imported from scikit-learn. This function receives the correct labels (`y_test`) and the predicted ones (`p_test`) as input, and automatically calculates the accuracy:

```
accuracy=accuracy_score(y_test_p_test)
```

9.7 Elapsed time for training

The time required to train the algorithms depends on the number of features provided as input. The training set consists of 8000 examples, but the features associated with each record in the *complex* dataset are twice that of the *amplitude* dataset. However, what significantly affects the elapsed time for training is the type and architecture of the algorithm used to classify the data.
The SVM and the K-NN take approximately 13 s and 10 s to train the amplitude dataset, and 15 s and 48 s for the elements of the complex dataset, respectively.
The MLP is the slowest: 11 minutes for the amplitude dataset and 13 minutes for the complex dataset. The reason is related to the fact that the structure of the neural network is quite articulated and there are many neurons within the hidden layers. The adjustment of the weights via back propagation, through this dense network, requires a considerable computational time.

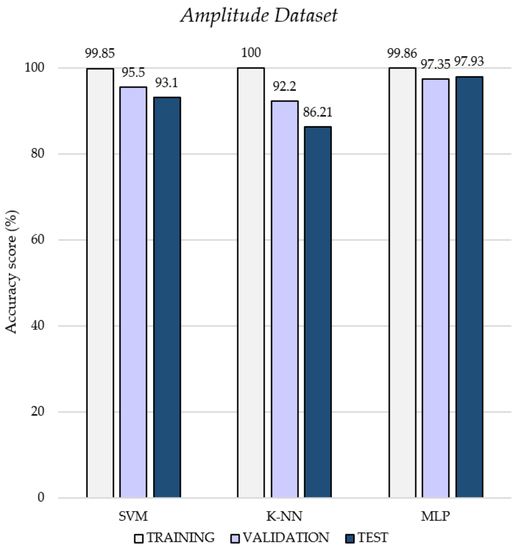
Chapter 10

Numerical Results

This section reports the results obtained during the training, validation and testing phase of the ML algorithms used: SVM, K-NN and MLP.
For each classifier, the features have been given in input as amplitude (*amplitude* dataset) or in the form of real and imaginary part (*complex* dataset).
The training dataset was split in two sub-set (Paragraph 9.3) thus obtaining the *training* set and *validation* set.
The first contains 80% of the data (8000 cases) and was actually used to train the ML algorithms, while the remaining 20% (2000 cases) was used for the tuning of the hyperparameters and to evaluate the degree of generalization achieved.
The *test* set is instead made up of 145 examples generated through full-wave FEM simulations.
The accuracy values obtained are summarized in Table.10.1, and in the graphs below (Fig.10.1-Fig.10.2). For the class labels, see Table 6.3.

	Amplitude Dataset			*Complex* Dataset		
	SVM	**K-NN**	**MLP**	**SVM**	**K-NN**	**MLP**
Training set	99.85	100	99.86	99.89	100	99.95
Validation set	95.5	92.2	97.35	96.05	92.35	97.7
Test set	93.1	86.21	97.93	84.83	85.52	93.1

Table 10.1. Accuracy values expressed as a percentage, reached by the three ML algorithms (SVM, K-NN, MLP), for training, validation and test set. On the left the results for the amplitude dataset and on the right those for the complex dataset.



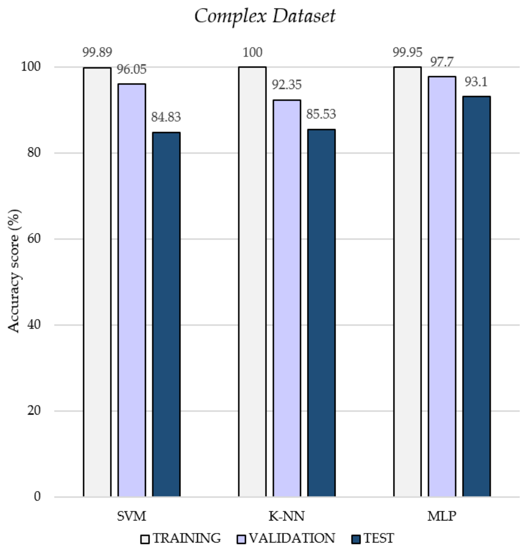


Figure 10.1. Bar diagrams showing the accuracy value in percentage, reached by the three algorithms (SVM, K-NN, MLP), for training, validation, and test set. On the left the results for the amplitude dataset and on the right those for the complex dataset.

Note how the performance trend of the classifiers decreases on the three types of datasets: it is maximum for the training set, and gradually decreases for the validation, and the test set. The MLP is among the algorithms the one that has obtained the best results, followed by the SVM, and the K-NN which instead shows some limitations. All three classifiers achieved higher accuracy values when tested on the amplitude dataset.
In general, all algorithms have always succeeded in correctly classifying healthy cases in training, validation, and testing phases, whether it was the complex dataset or the amplitude dataset. Most misclassification are attributable to positional errors. Clearly, for the purposes of therapeutic treatment, it is much more serious to confuse the type of stroke, rather than its location inside the head. Furthermore, many times the stroke does not occupy a precise position among the four areas: in this case the example in question becomes difficult to classify.
The following paragraphs contain the normalized and non-normalized confusion matrices for each type of dataset (see Fig.10.2. – Fig.10.19).

10.1 SVM classifier (Amplitude Dataset)

TRAINING SET: accuracy score 99.85%

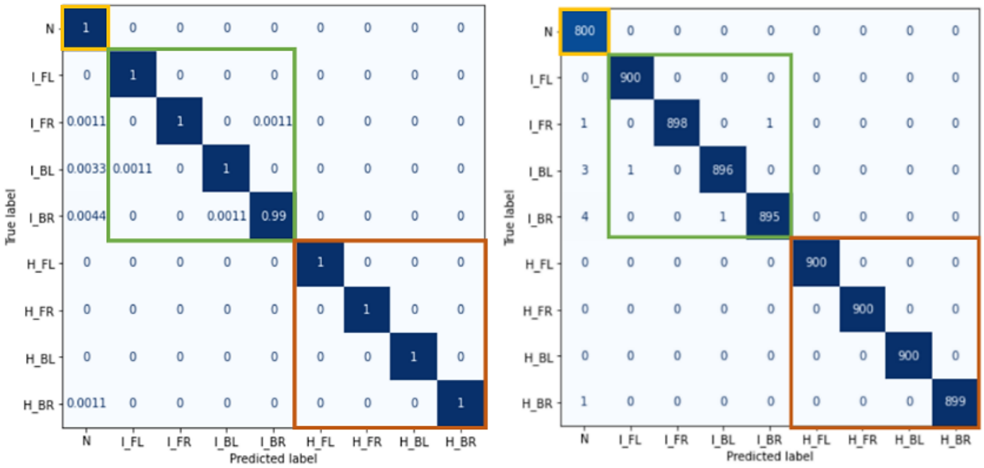


Figure 10.2. Confusion matrices showing the results obtained by the SVM classifier on the training set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 95.50%

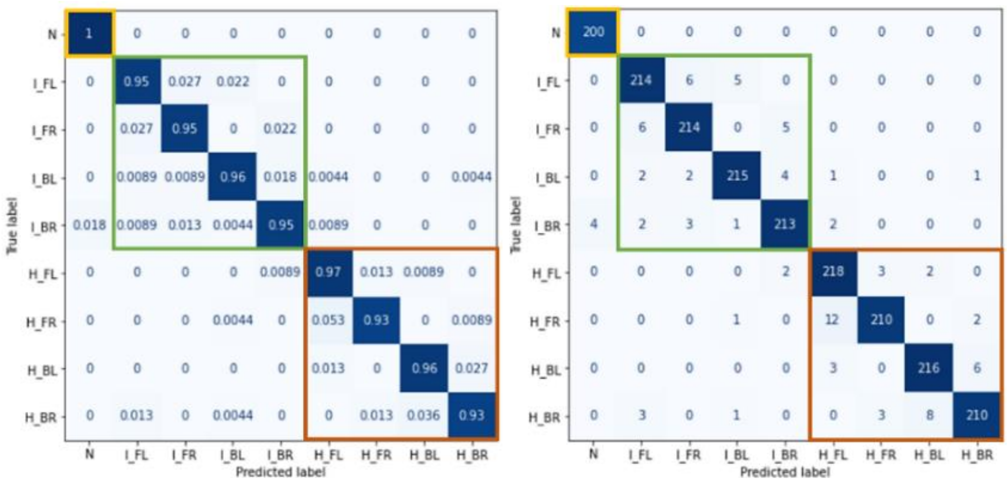


Figure 10.3. Confusion matrices reporting the results obtained by the SVM classifier on the validation set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

TEST SET: accuracy score 93.10%

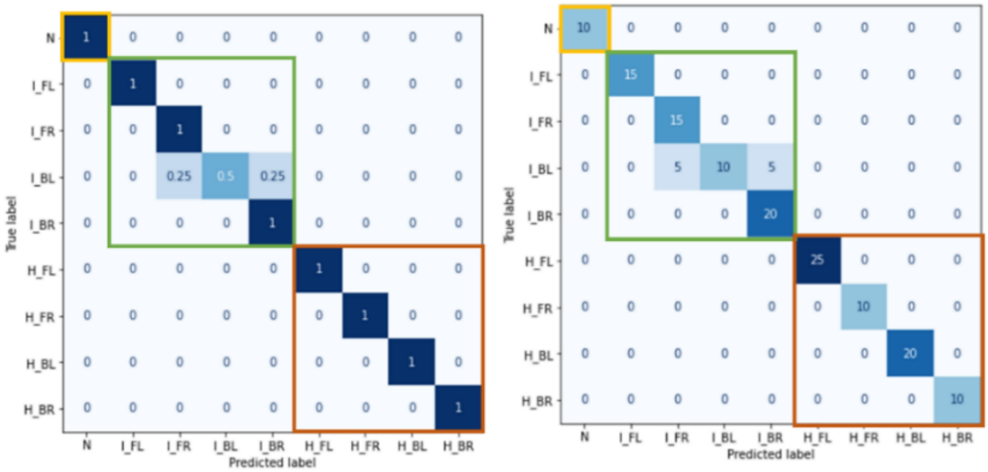


Figure 10.4. Confusion matrices reporting the results obtained by the SVM classifier on the test set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The SVM classifier manages to correctly identify all healthy cases in the training set. It does not recognize only 1 hemorrhagic element belongs to H_B class, which is identified as a healthy case. Among the 3600 ischemic examples in the training set, 11 are misclassified: for the most part they are labelled as cases without target. The accuracy achieved on the training set is very high and it is equal to 99.85%.
As regard the validation set, as expected, the performance drops slightly. Despite this, all healthy cases are recognized in the right way. The SVM classifier correctly identifies between 93% and 97% of cases belonging to other classes. 12 elements of the H_FR class are labelled as H_FL, and 4 I_BR cases are classified as healthy records (N).
For what concerns the test set, the SVM classifier correctly recognizes all healthy cases and hemorrhagic cases. The only mistakes made, concern errors on the position: 10 I_BL elements are classified as I_FR or as I_BR. The accuracy score achieved is 93.10%.

10.2 SVM classifier (Complex Dataset)

TRAINING SET: accuracy score 99.89%

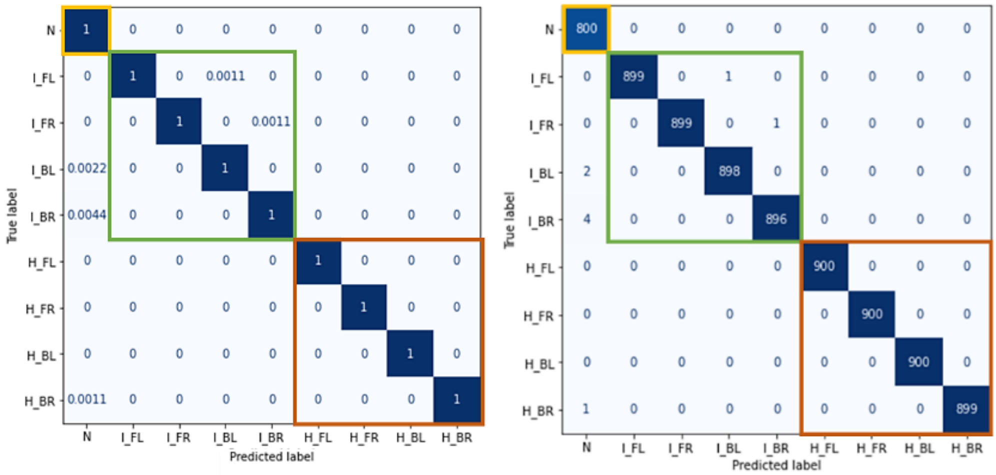


Figure 10.5. Confusion matrices reporting the results obtained by the SVM classifier on the training set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 96.05%

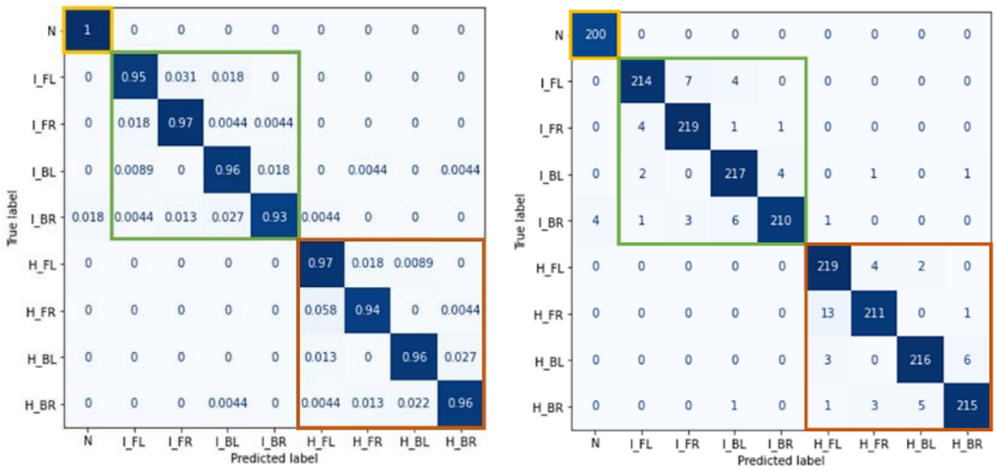


Figure 10.6. Confusion matrices reporting the results obtained by the SVM classifier on the validation set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

TEST SET: accuracy score 84.83%

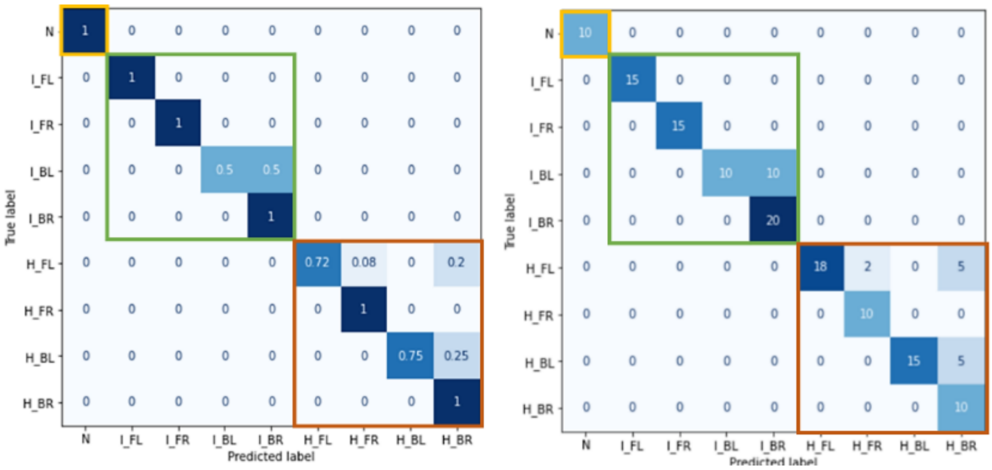


Figure 10.7. Confusion matrices reporting the results obtained by the SVM classifier on the test set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The SVM classifier manages to correctly identify all healthy cases in the training set. As for the amplitude dataset, it does not recognize only 1 hemorrhagic element belongs to H_BR class, which is identified as a healthy case. Of the 3600 ischemic examples in the training set, 10 are misclassified: most are labelled as cases without target and 2 are location errors. The accuracy achieved on the training set is very high and it is equal to 99.89%.

As regard the validation set, all healthy cases are recognized in the right way. The SVM classifier correctly recognizes between 93% and 97% of cases belonging to other classes. 13 elements of the H_FR class are labelled as H_FL and 4 I_BR cases are classified as healthy (N).

There are 4 errors affecting the distinction between macro classes: 1 H_BR record is classified as I_BL, 1 I_BR element is identified as H_FL, 2 I_BL cases are incorrectly classified, 1 as H_FR and, 1 as H_BR.

For what concerns the test set, the SVM classifier correctly recognizes all healthy cases. The errors for the ischemic macro class concern the position: 10 elements of the I_BL category are classified as I_BR. As for hemorrhagic cases, 7 H_FL examples are misrecognized: 2 as H_FR and 5 times as H_BR. The accuracy score achieved is 84.83%.

Ultimately, the best results for the SVM classifier, on the test set, were obtained when the features were given as input in the form of amplitude.

10.3 K-NN classifier (Amplitude Dataset)

TRAINING SET: accuracy score 100%

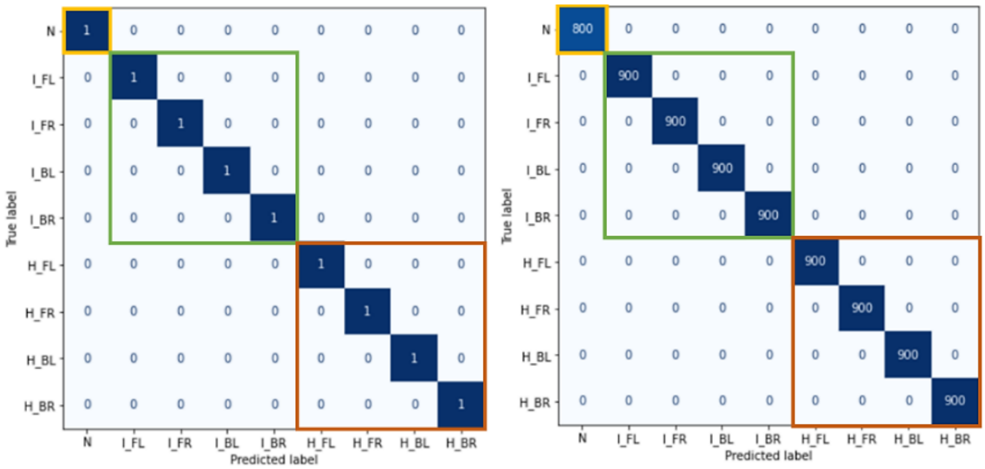


Figure 10.8. Confusion matrices reporting the results obtained by the K-NN classifier on the training set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 92.20%

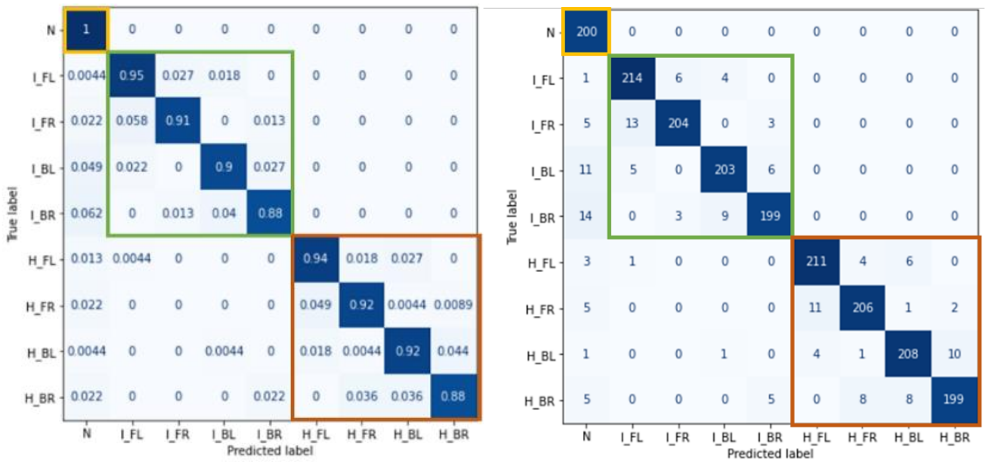


Figure 10.9. Confusion matrices reporting the results obtained by the K-NN classifier on the validation set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

TEST SET: accuracy score 86.21%

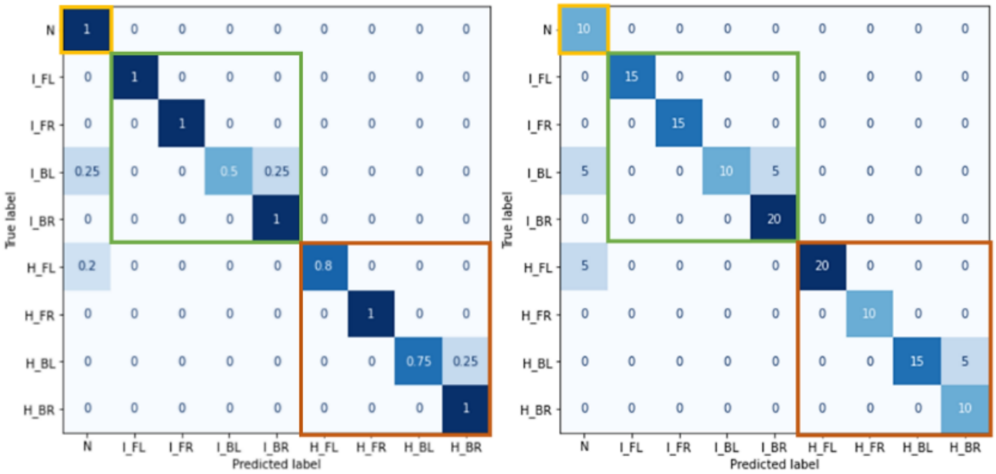


Figure 10.10. Confusion matrices reporting the results obtained by the K-NN classifier on the test set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The K-NN classifier is able to correctly identify all the examples of the training set. The performance drops considerably for the validation set: an accuracy value of 92.20% is reached.

Nevertheless, all 200 healthy cases are recognized in the right way. The classifier has difficulty recognizing the I_BR and the H_BR examples, for which the maximum accuracy achieved is 88%.

The I_BR class is labelled as N wrongly 14 times out 225. Sixteen out of 225, the class H_BR is misclassified as H_FR (eight times), and as H_BL. As regard the test set, the K-NN classifier correctly recognizes all healthy cases. The errors for the ischemic macro class concern the position: five elements of the I_BL are misclassified as I_BR, while others five elements of I_BL are labelled as healthy cases. Finally, five H_BL examples are misrecognized as H_BR. The accuracy score achieved is 86.21%.

10.4 K-NN classifier (Complex Dataset)

TRAINING SET: accuracy score 100%

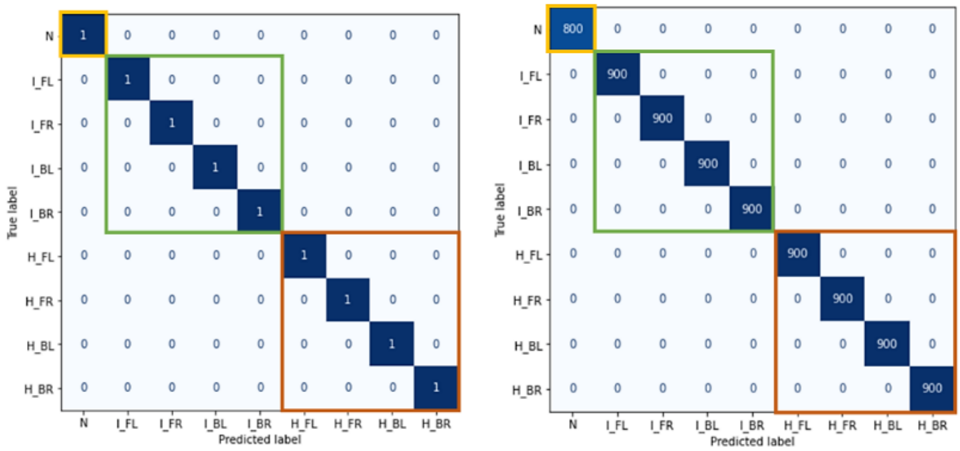


Figure 10.11. Confusion matrices reporting the results obtained by the K-NN classifier on the training set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 92.35%

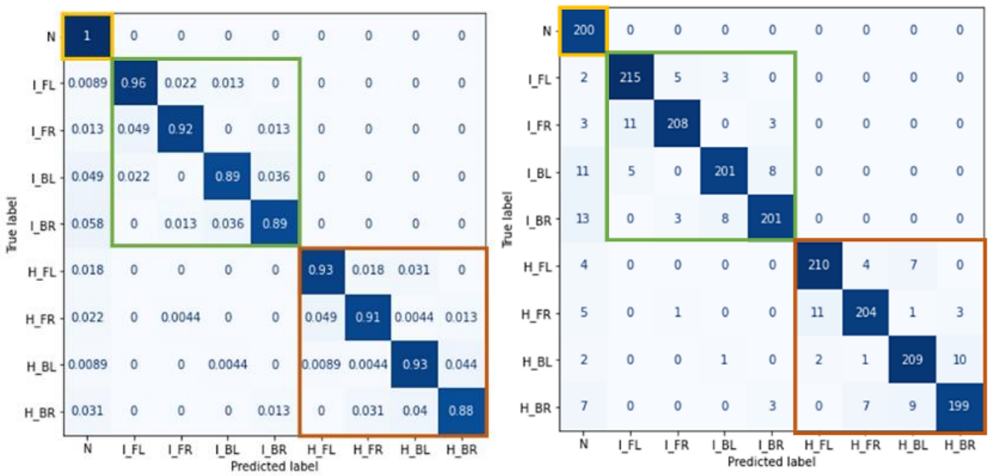


Figure 10.12. Confusion matrices reporting the results obtained by the K-NN classifier on the validation set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

TEST SET: accuracy score 85.52%

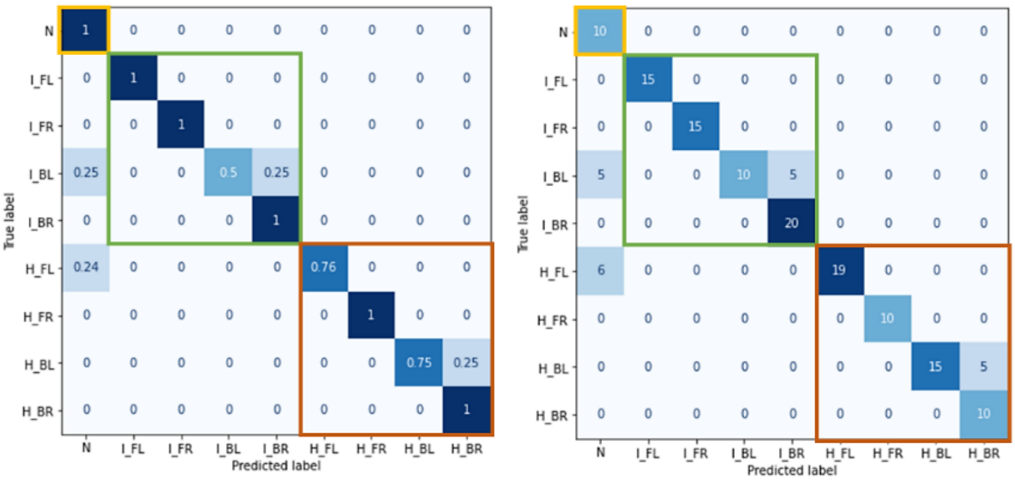


Figure 10.13. Confusion matrices reporting the results obtained by the K-NN classifier on the test set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The K-NN classifier is able to correctly identify all the examples in the training set. The performance drops considerably for the validation set: an accuracy value of 92.35% is reached.

Nevertheless, all 200 healthy cases are recognized in the right way. The classifier has difficulty in recognizing H_BR class: out of 225, 7 times the K-NN assigns the N class, 3 times the I_BR class, 7 times the H_FR class, and 9 times the H_BL class. The I_BR class is labelled as N wrongly 13 times out 225. Ischemic cases are never labeled as hemorrhagic, while 5 hemorrhagic examples are confused as ischemic cases. As regard the test set, the K-NN classifier correctly recognizes all healthy cases. The errors for the ischemic macro class concern the position: 10 elements of the I_BL are misclassified as I_BR (5 times) and as N (5 times). 6 H_FL examples are wrongly labelled as healthy cases. Finally, 5 H_BL records are misrecognized as H_BR. The accuracy score achieved is 85.52%.

Ultimately, the best results for the K-NN classifier, on the test set, were obtained when the features were given as input in the form of amplitude.

10.5 MLP classifier (Amplitude Dataset)

TRAINING SET: accuracy score 99.86%

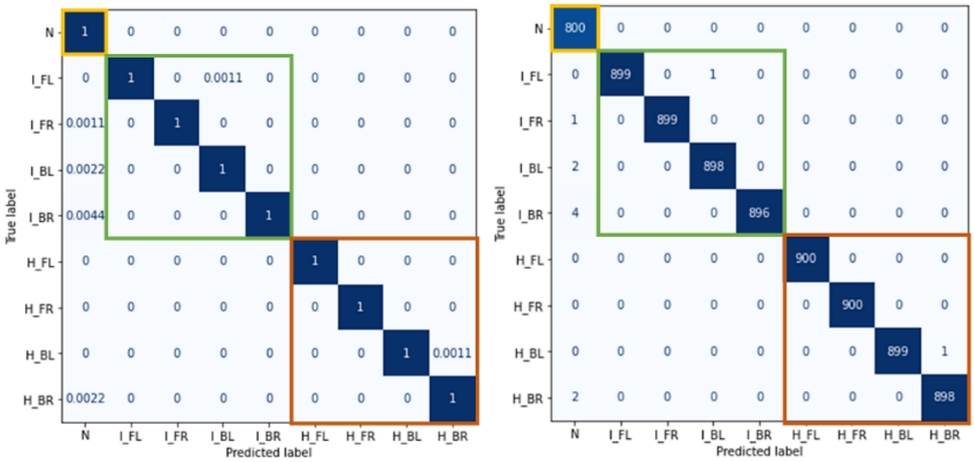


Figure 10.14. Confusion matrices reporting the results obtained by the MLP classifier on the training set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 97.35%

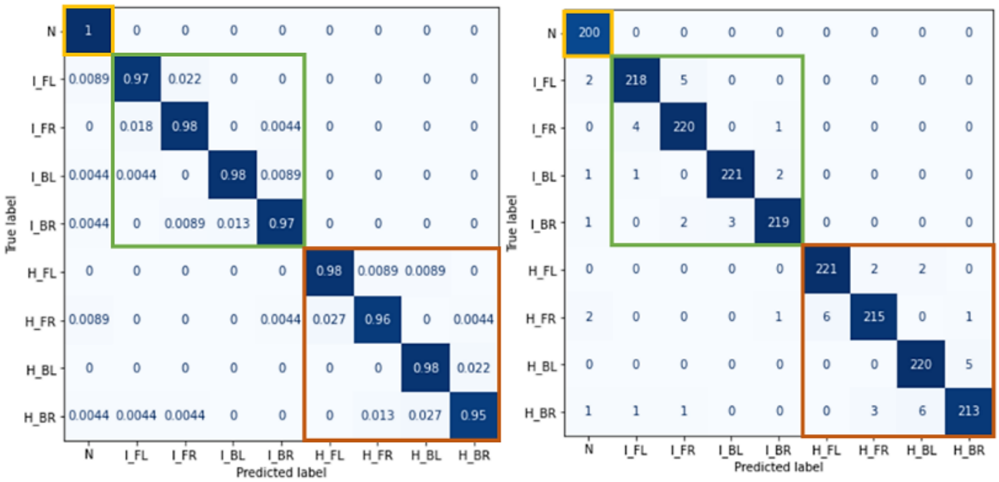


Figure 10.15. Confusion matrices reporting the results obtained by the MLP classifier on the validation set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

TEST SET: accuracy score 97.93%

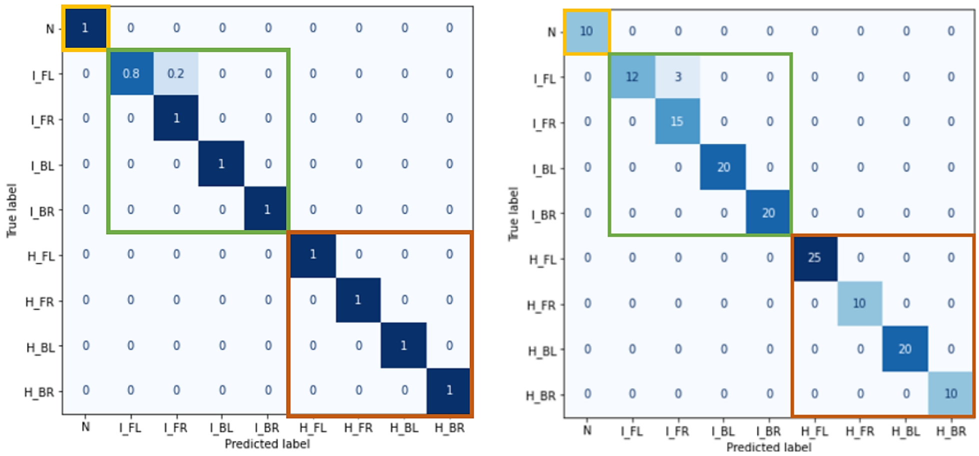


Figure 10.6. Confusion matrices reporting the results obtained by the MLP classifier on the test set (amplitude dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The MLP classifier is able to correctly identify all the healthy cases of the training set, with 11 total errors out of 7200 examples for the other remaining classes. There are 2 position errors: an I_FL is confused with an I_BL, while an H_BL is labelled as H_BR. The classifier misidentifies 2 elements of the N class, 2 as hemorrhagic cases and 7 times as ischemic cases.
As regard the validation set, MLP correctly identifies all 200 healthy records.
The lowest performances, with accuracy values equal to 95%, are recorded for the H_BR class, where 6 elements are recognized as H_BL. Ischemic cases are never confused as hemorrhagic, while 3 hemorrhagic examples are misclassified as ischemic records. As regard the test set, the MLP classifier correctly recognizes all healthy and hemorrhagic cases. The errors for the ischemic macro class concern the position: 3 elements of the I_FL are classified as I_FR. The accuracy score achieved is high and it is equal to 97.93%.

10.6 MLP classifier (Complex Dataset)

TRAINING SET: accuracy score 99.95%

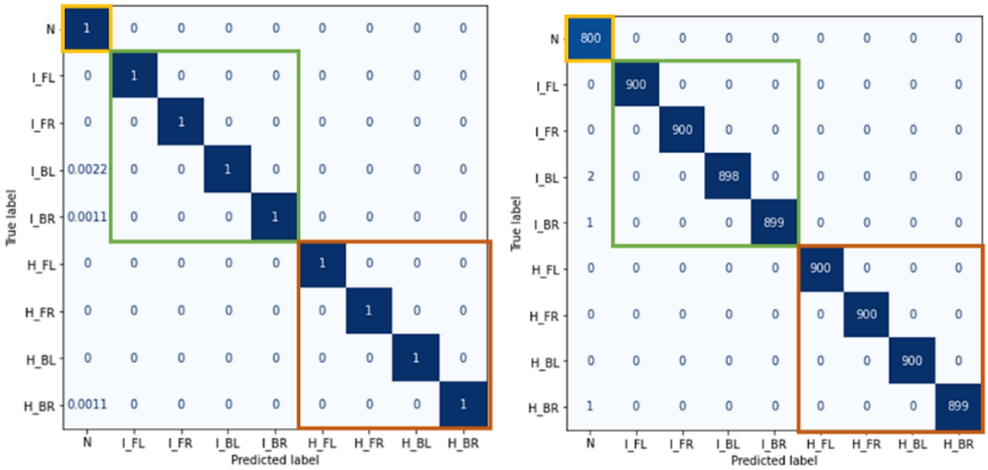


Figure 10.17. Confusion matrices reporting the results obtained by the MLP classifier on the training set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

VALIDATION SET: accuracy score 97.70%

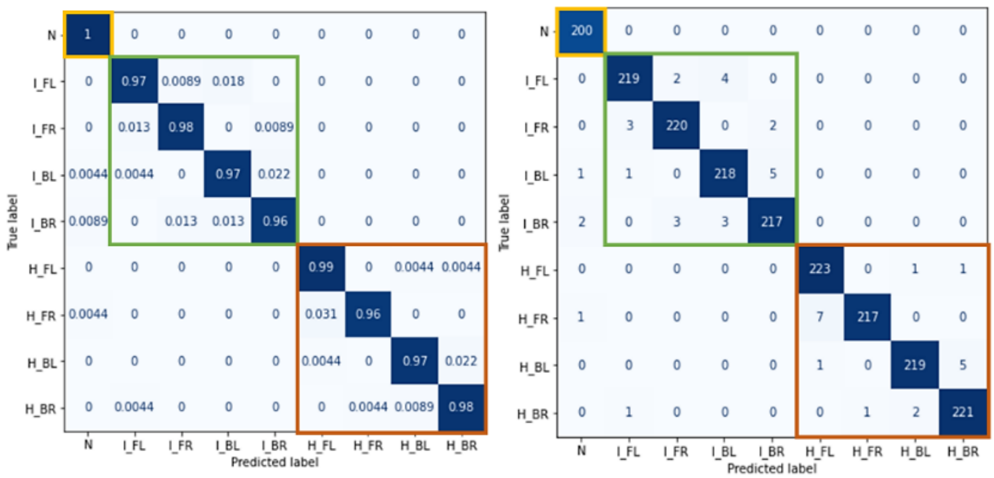


Figure 10.18. Confusion matrices reporting the results obtained by the MLP classifier on the validation set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix

TEST SET: accuracy score 93.10%

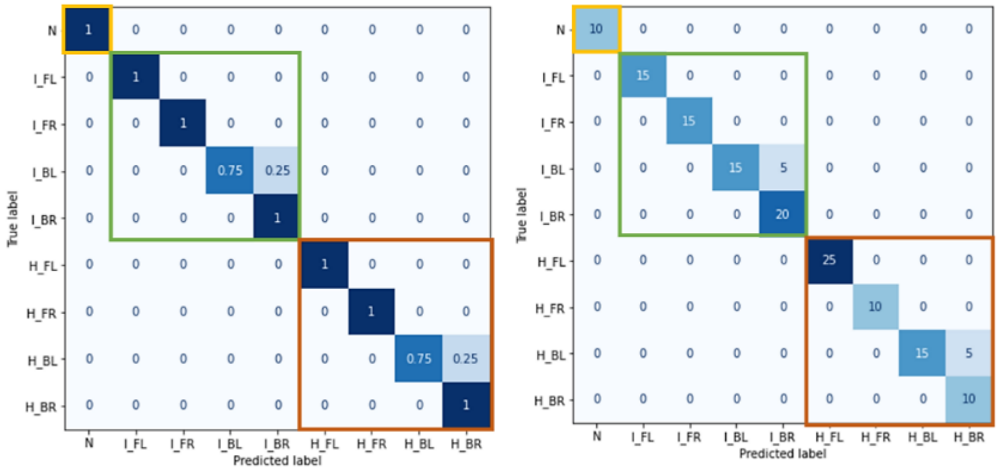


Figure 10.19. Confusion matrices reporting the results obtained by the MLP classifier on the test set (complex dataset). On the left the normalized matrix, on the right the non-normalized matrix.

The MLP classifier always correctly identifies the macro classes, making only four errors out of 8000 total predictions. In particular, it incorrectly assigns the N class, 2 to I_BL class and 1 to I_BR and H_BR classes.
As regard the validation set, the MLP correctly identifies all 200 healthy cases.
The highest performance, with accuracy values of 98%, is recorded for the H_BR class, with only 4 total errors out of 225. Ischemic cases are never confused as hemorrhagic, while an ischemic H_BR case is mistakenly recognized as I_FL. 4 times the classifier MLP incorrectly assigns the N class: 3 are ischemic cases (two I_BL and one I_BL) and 1 is hemorrhagic (H_FR). 7 H_FR records are misclassified as H_FL. As regard the test set, the MLP classifier correctly recognizes all healthy cases. The errors concern only the position and not the type of stroke: 5 elements of the I_BL are classified as I_BR, while 5 H_BL examples are labelled as H_BR. The accuracy score achieved is equal to 93.10%.

Ultimately, the best results for the MLP classifier, on the test set, were obtained when the features were given as input in the form of amplitude.

Chapter 11

Conclusion and future development

Brain stroke is one of the most common cardiovascular diseases worldwide. If it is not recognized and treated in time, hemorrhage or ischemia cause irreparable damage that can lead to severe disability or death. The new Microwave Imaging (MWI) system, portable, non-invasive and low-cost, is proposed as a complementary tool to the already available CT and MRI imaging techniques, allowing early diagnosis and continuous monitoring. The MWI system measures the scattering parameters at the ports of 24 receiving and transmitting antennas, placed on a helmet that covers the upper part of the patient's head. The image reconstruction algorithm exploits the dielectric contrast between healthy and pathological tissues at the microwave frequencies.

In this thesis project, the potential of the MWI system has been combined with those of artificial intelligence (AI).

Supervised Machine Learning (ML) algorithms such as Support Vector Machine (SVM), K-Nearest Neighbour (K-NN) and Multi-Layer Perceptron (MLP) has been used to solve the stroke classification problem. The implemented models have been trained to identify the region, among the four areas of the head, affected by ischemia or bleeding having as input the scattering parameters, in the form of amplitude (amplitude dataset) or of real and imaginary part (complex dataset).

Clinical data, laboratory measurements or full-wave simulations require a great effort in terms of time and therefore they are not a feasible option for building large dataset, indispensable for the training phase of the ML algorithms.

A method based on the Born approximation and on the Linearized Integral Operator has been used to create the training set. The implemented algorithm proved to be fast and effective: it was able to generate 10000 cases with stroke in different conditions, in about 16 hours.

The hyperparameters of each ML model have been defined following the application of an optimization method called Grid Search. After implementing and training the algorithms, to evaluate their performances, a test set consisting of 145 examples, created via full-wave Finite Element Method simulations, has been used.

The relative permittivity and conductivity values of the healthy brain and of ischemic, and hemorrhagic stroke, used both for the creation of the training and test set, are those obtained from the measurements on the mixtures created in the laboratory.

All the algorithms, in the training and testing phase have always been able to correctly recognize records without target, both for the amplitude dataset and for the complex dataset, and most importantly they always manage to distinguish the macro classes (healthy, ischemic, and hemorrhagic). Most of the classification errors concern the position.

Among the algorithms used, the MLP, with an accuracy score of 97.93% for the amplitude dataset and 93.10% for the complex dataset, proved to be the most performing, followed by SVM and k-NN. The latter is particularly sensitive to noise, and it is prone to the problem of overfitting: the accuracy score for the test set does not exceed 87%. The performance for the amplitude dataset, during the test phase, are slightly higher than those obtained for the complex dataset. It suggests that even if the number of input features is halved, the information necessary to identify the correct class is preserved. This can be an interesting aspect to consider because, with comparable or even better performance, both tuning and training times can be reduced. It would also allow the development of even simpler and cheaper MWI systems because they could be made with less sophisticated instruments capable of measuring only the amplitude of the signals.

From the results obtained, it is evident that the method based on the born approximation and, on the linear scattering operator, can be a valid solution to generate large amounts of data for the ML training phase. The levels of accuracy achieved on the test set are proof that the synthetic S parameters of the training set are comparable with those obtained through simulations.

In order to have further confirmation of the validity of the implemented method and to consider the application of AI as a valuable resource for the diagnosis of stroke, it is necessary to test the ML trained algorithms, on measurements performed on the 3D human head phantom.

As regards the method implemented for the training set creation, the improvements to be made to obtain a variable set of examples more similar to reality, could consist in adding complex and irregular shapes of stroke, and in combining more noise thresholds together. The classification problem can be extended to a multi-tissue model and solved by training other types of ML algorithms.

In view of a possible clinical application, a heterogeneous training set built considering different head models as reference, can certainly increase the performance of the ML algorithms. The use of artificial intelligence combined with the MWI system, can allow the diagnostic examination to be carried out in a very short time, directly in the ambulance. An early diagnosis makes it possible to limit the damage caused by stroke injuries and save lives.

Bibliography

[1] T. Truelsen, S. Begg, and Colin Mathers, "The Global Burden of Cerebrovascular Disease", (2006). [Online]. Available: https://www.researchgate.net/publication/228551377_The_Global_Burden_of_Cerebrovascular_Disease.

[2] A. L. Abbott, M. Silvestrini, R. Topakian, J. Golledge, A.M. Brunser et all, " Optimizing the definition of stroke, transient ischemic attack, and the infarction for research and application in clinical practice". *Frontiers in Cellular Neuroscience*, vol. 8, article 537, (2017). [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29104559/

[3] A. P. Coupland, A. Thapar, M. I Qureshi, H. Jenkins, and A. H. Davies, "The definition of stroke". *Journal of the Royal Society of Medicine* vol.110, n. 1 (2017). [Online]. Available: https://journals.sagepub.com/doi/full/10.1177/0141076816680121

[4] D. Belov Kirdajova, J. Kriska, J. Tureckova and M. Anderova1, "Ischemia-Triggered Glutamate Excitotoxicity from the Perspective of Glial Cells". *Frontiers in Cellular Neuroscience*, vol. 13, (2020).

[5] Julio H. Garcia, Niels A. Lassen, Cornelius Weiller, Bjørn Sperling and Jyoji Nakagawara, "Ischemic Stroke and Incomplete Infarction". *Stroke,* vol. 27, pp. 761–765, (1996).
[Online]. Available: https://www.ahajournals.org/doi/epub/10.1161/01.STR.27.4.761

[6] Jeffrey L. Saver, MD , "Time Is Brain—Quantified". *American Heart association: Stroke,* vol. 37, pp. 263:266 (2006). [Online]. Available: https://www.ahajournals.org/doi/epub/10.1161/01.STR.0000196957.55928.ab

[7] Ministero della Salute, "Ictus". *La nostra salute,* (2022). [Online]. Available: https://www.salute.gov.it/portale/salute/p1_5.jsp?area=Malattie_cardiovascolari&id=28&lingua=italiano

[8] A. K. Boehme, C. Esenwa and Mitchell S. V. Elkind, "Stroke risks factors, genetics, and prevention". *Circulation Research*, vol. 120, pp. 471-495, (2018). [Online]. Available: https://www.ahajournals.org/doi/epub/10.1161/CIRCRESAHA.116.308398

[9] A. Arboix, "Cardiovascular risk factors for acute stroke: risk profiles in the different subtypes of ischemic stroke". *World Journal of Clinical Cases*, 418-429, (2015).
[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419105/pdf/WJCC-3-418.pdf

[10] M. K. Hayes et all, "Influence of age and health behaviors on stroke risk: lessons from longitudinal studies". *J Am Geriatr Soc*, (2018).
[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006180/

[11] T. Sabbouh, M. T. Torbey, " Malnutrition in stroke patients: risk factors, assessment, and management". *Neurocritical Care*, vol. 29, pp. 374-384, (2018).
[Online]: Available: https://pubmed.ncbi.nlm.nih.gov/28799021/

[12] L. K. Kwah and J. Diong, "National Institutes of Health Stroke Scale (NIHSS)". *Journal of Physiotherapy*, vol. 60, p.61, (2014).
[Online]: Available: https://pubmed.ncbi.nlm.nih.gov/24856948/

[13] Vivek P. Patel, Jeremy J. Heit, "Ischemic Stroke Treatment Trials: Neuroimaging Advancements and Implications". *Top Magn Reson Imaging,* pp. 133-139, (2017).
[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28277455/

[14] D. J. Miller, J. R. Simpson and B. Silver, "Safety of thrombolysis in acute ischemic stroke: a review of complications, risk factors, and newer technologies". *The Neurohospitalist*, pp. 138-147, (2011).
[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23983849/

[15] M. P. Lin and D. S. Liebeskind, "Imaging of Ischemic Stroke". *American Academy of Neurology, Continuum,* vol. 22(5), pp. 1399-1423, (2016). [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/27740982/

[16] M. A. Naji, S. El Filali, K. Aarika, EL H. Benlahmar, R.A. Abdelouuhahid, O. Debauche, " Machine Learning algorithms for breast cancer prediction and diagnosis". *Procedia Computer Science*, vol. 191, pp. 487-492, (2021). [Online]. Available:
https://www.sciencedirect.com/science/article/pii/S1877050921014629

[17] M. Ricci, B. Stitic, L. Urbinati, G. Di Guglielmo, J. A. T. Vasquez, L. P. Carloni, F. Vipiana, and M. R. Casu, "Machine-learning based microwave sensing: A case study for the food industry," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, (2021).
[Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9489295

[18] J. A. Tobon Vasquez, R. Scapaticci, G. Turvani, G. Bellizzi, D. O. Rodriguez Duarte, N. Joachimowicz, R. Duchene, E. Tedeschi, M. R. Casu, L. Crocco and F. Vipiana "A Prototype Microwave System for 3D Brain Stroke Imaging". *Sensors,* vol. 20(9), (2020):
[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32375220/

[19] J. Vymazal, A. M. Rulseh, J. Keller and L. Janouskova, " Comparison of CT and MR imaging in ischemic stroke". *Insight into Imaging,* vol. 3(6), pp. 619-627, (2012)
[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23055115/

[20] [Online]. Available: https://www.domaltech.eu/it/mostra-prodotti/g-e-brightspeed-elite-16-slices-19

[21] E. Seeram, " Computed Tomography: physical principles and recent technical advance". *Journal of Medical Imaging and radiation sciences,* vol. 41(2), pp. 87-109, (2010).
[Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31051822/

[22][Online]. Available: https://en.wikipedia.org/wiki/Hounsfield_scale

[23] R. V. Kummer, L. L. Allen, R. Holle, L. Bozzao, S. Bastianello, C. Manelfe, E. Bluhmki, P. Ringleb, D. H. Meier and W. Hackle, " Acute stroke: usefulness of early CT findings before thrombolytic therapy". *Radiology*, vol. 205(2), (1997).
[Online]. Available: https://pubs.rsna.org/doi/10.1148/radiology.205.2.9356611

[24] J. H. W Pexman, P.A. Barber, M. D. Hill, R. J. Sevick, A. M, Demchuck, M. E. Hudon, W. Y. Hu and A. M. Buchan, “Use of the Alberta Stroke Program Early CT Score (ASPECTS) for Assessing CT scans in Patients with Acute Stroke”, *American Journal of Neuroradiology*, vol. 22(8), pp. 1534-154, (2001).
[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7974585/

[25] T. Popolizio, T. Scarabino, “Ictus ischemico”, cap. 31, pp. 279-293, (2006). [Online]. Available: http://eknygos.lsmuni.lt/springer/395/279-293.pdf

[26] M. Resta, C. Andreula, M. Resta and M. Donatelli, “Ictus Emorragico”, Radiologia geriatrica cap. 32, pp. 295-303, (2006). ISBN: 9788847004863
[Online]. Available: http://eknygos.lsmuni.lt/springer/395/295-303.pdf

[27] [Online]. Available: https://www.jackson-hospital.com/medical-services/radiology-x-ray/mri-magnetic-resonance-imaging/

[28] S. Kathiravan and J. Kanakaraj, “ A Review on Potential Issues and Challenges in MR Imaging”. *The Scientific World Journal.* (2013).
[Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3863452/

[29] P. Rodrigues, “The role of CT and MR in stroke patients”. *Qwenta*. (2018)
[Online]. Available: https://www.qmenta.com/blog/the-role-of-ct-and-mr-in-stroke-patients

[30] S. Choi, D.H. Ha, M. J. Kang, J. H. Lee, S. and K. Yook, “Does the ADC Map have Additional Clinical Significance Compared to the DWI in the Brain Infarction?”. *Journal of the Korean Society of Magnetic Resonance in Medicine*, vol. 17(4), pp. 267-274, (2013).
[Online]. Available:
https://synapse.koreamed.org/upload/synapsedata/pdfdata/0040jksmrm/jksmrm-17-267.pdf

[31] G. Thomalla, B. Cheng, M. Ebinger, Q. Hao, T. Tourdias, O. Wu, J. S. Kin, L. Breur, O. C Singer, S. W et all, “DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4·5 h of symptom onset (PRE-FLAIR): a multicentre observational study”. *The Lancet Neurology*, vol. 10, pp. 978–86 (2011).
[Online]. Available: https://www.thelancet.com/journals/laneur/article/PIIS1474-4422(11)70192-2/fulltext

[32] F. Bonci, “Stroke e visione”. *Rio*, vol. 33(4), pp. 154-162, (2010). [Online]. Available https://www.researchgate.net/publication/264129615_Stroke_e_Visione/link/53cf99610cf2f7e53cf81679/download

[33] I. Sarwar, G. Turvani, M. R. Casu, J. A. Tobon Vasquez, F. Vipiana, R. Scapaticci, and L. Crocco. “Low-Cost Low-Power Acceleration of a Microwave Imaging Algorithm for Brain Stroke Monitoring.” *Journal of Low Power Electronics and Applications*, vol. 8, n. 4 (2018). [Online]. Available https://doi.org/10.3390/jlpea8040043.

[34] D. O. Rodriguez Duarte, J. A. Tobon Vasquez, R. Scapaticci, G. Turvani, M. Cavagnaro, M. R. Casu, L. Crocco and F. Vipiana, “Experimental Validation of a Microwave System for Brain Stroke 3-D Imaging”. *Diagnostics*, vol. 11(7), (2021). [Online]. Available: https://doi.org/10.3390/diagnostics11071232

[35] R. Orta, "Lecture Notes on Electromagnetic Field Theory", Preliminary version, Department of Electronics and Telecommunications, Politecnico di Torino, (2017).

[36] R. Orta, "Lecture Notes on Transmission Line Theory", Department of Electronics and Telecommunications, Politecnico di Torino, (2017).

[37] https://paminerva.blogspot.com/2020/08/11-phong-shading.html

[38] L. Crocco, I. Keranasiou, M. L. James and R. C. Conseicao, "Emerging Electromagnetic Technologies for Brain Diseases Diagnostics, Monitoring and Therapy". Springer Link (2018).

[39] G. L. Morales, M. E. Luna and C. F. Reyes, "Optical characterization of amber from Chiapas, Mexico". *Proceedings of SPIE - The International Society for Optical Engineering,* vol. 60(3), pp. 217-221, (2014). [Online]. Available:
https://www.researchgate.net/publication/260868047_Optical_characterization_of_amber_from_Chiapas_Mexico

[40] D.O. Rodriguez Duarte, J. A. Tobon Vasquez, R. Scapaticci, L. Crocco and F. Vipiana, "Brick Shaped Antenna Module for Microwave Brain Imaging Systems". *Antennas and Wireless Propagation Letters*, vol. 19(12), pp. 2057 – 2061, (2020). DOI 10.1109/LAWP.2020.3022161.

[41] R. Scapaticci, L. Di Donato, I. Catapano and L. Crocco, " A feasibility study on microwave imaging for brain stroke monitoring". *Progress In Electromagnetics Research B,* vol. 40, pp. 305-324, (2012).
[Online]. Available: http://www.jpier.org/PIERB/pier.php?paper=12022006

[42] Microwaves1O1.com, (2022). [Online]. Available:
https://www.microwaves101.com/encyclopedias/sparameters#:~:text=Definition%20of%20S-parameters,a%20signal%20at%20port%201

[43] F. Passos, E. Roca, R. C. Lopez and F. V: Fernandez, "Automated Hierarchical Synthesis of Radio-Frequency Integrated Circuits and Systems: A Systematic and Multilevel Approach", (2020).

[44] F. Caspers, "RF engineering basic concepts: S-parameters", CERN, (2012)

[45] P9375A Keysight Streamline USB Vector Network Analyzer, 26.5 GHz. Data Sheet. [Online]. Available: https://www.keysight.com/it/en/support/P9375A/streamline-usb-vector-network-analyzer-26-5-ghz.html

[46] J. A. Tobon Vasquez, R. Scapaticci, G. Turvani, G. Bellizzi, N. Joachimowicz, B. Duchene, E. Tedeschi, M. R. Casu, L. Crocco and F. Vipiana, "Design and Experimental Assessment of a 2-D Microwave Imaging system for Brain Stroke Monitoring". *International Journal of Antennas and Propagation,* (2019)
[Online]. Available: https://www.hindawi.com/journals/ijap/2019/8065036/

[47] L. Urbinati, "Detection of food contaminants with Microwave Sensing and Machine Learning", Master's Degree Thesis in Electronic Engineering, Politecnico di Torino (2019).

[48] 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". *Journal of Electromagnetism, RF and microwaves in medicine and biology.* DOI 10.1109/JERM.2020.3049071.

[49] A. La Gioia, E. Portes, I. Merunka, A. Shahzad, S. Salahuddin, M. Jones and M. O'Halloran, "Open-Ended Coaxial Probe Technique for Dielectric Measurement of Biological Tissues: Challenges and Common Practices". *Diagnostics*, vol. 8(40), doi:10.3390/diagnostics8020040, (2018).

[50] M. T. Jilani, P. W. Wong, M. Z. Rehaman and L. Y. Cheong, "Microwave sensor for non-destructive dielectric characterization of biological systems", *DEStech Transaction on Materials Science and engineering*, (2016).

[51] Keysight 85070E Dielectric Probe Kit 200 MHz to 50 GHz- Data Sheet
[Online]. Available:
https://www.keysight.com/it/en/assets/7018-01196/technical-overviews/5989-0222.pdf

[52] G. Marchioni, "Machine learning techniques for microwave food contamination detection". Master's Degree Thesis in Electronic Engineering (2021). [Online]. Available: https://www.researchgate.net/publication/295833637_Microwave_sensor_for_nondestructive_dielectric_characterization_of_biological_systems

[53] "Security machine learning algorithms", ENISA, (2021). ISBN: 978-92-9204-543-2 DOI: 10.2824/874249.

[54] C. Cortes and V. Vapnik "Support-Vector Networks". *Machine learning,* vol. 20(3), pp. 273-297

[56] C. Nantasenamat and T. Tantimongcolwat, "Data mining of magnetocardiograms for prediction of ischemic heart disease". *EXCLI Journal,* vol. 9, pp. 82-95, (2010).

[57] [Online]. Available: https://stackoverflow.com/questions/21694855/what-does-the-support-mean-in-support-vector-machine

[58] C. Bishop, "Pattern recognition and machine learning" *Information Science and Statistics*. (2006).

[59] C. Rosè, Y. C. Wang, Y. Cui, J. Arguello, K. Stegmann and F. Fischer, "Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning". *Computer-Supported Collaborative Learning*, vol. 3, pp. 237–271, (2008).

[60] Y. Ma and G. Guo, "Support Vector Machine Applications", Springer Link, (2014)

[61] Z. Wang and X. Xue, "Multi-Class Support Vector Machine", Springer Link, pp. 23-48, (2014).

[62] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," *International Conference on Intelligent Computing and Control Systems* (ICCS), pp. 1255–1260, (2019).

[63] https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa

[64] M. A. Nielsen, Neural Networks and Deep Learning, (2019). [Online]. Available: Http://neuralnetworksanddeeplearning.com/index.html

[65] G. Susi, "Asynchronous Spiking Neural Networks: paradigma generale e applicazioni", (2014). [Online]. Available:
https://www.researchgate.net/publication/260793937_Asynchronous_Spiking_Neural_Networks_paradigma_generale_e_applicazioni

[66] M. M Real and R. Salvador, "Physical Side-Channel Attacks on Embedded Neural Networks: A Survey". *Applied Sciences*, vol. 11(15), (2021).
[Online]. Available: https://www.mdpi.com/2076-3417/11/15/6790/htm

[67] F. Careglio, "Design di un trading system azionario di tipo trend reversal basato su modelli di Machine Learning", Master's degree Thesis in Computer Engineering, Politecnico di Torino, (2020).

[68] A. P Engelbrecht, "Computational Intelligence. An Introduction", Second Edition, ISBN: 978-0-470-03561-0, (2017).
https://staff.fmi.uvt.ro/~daniela.zaharie/ma2017/books/Computational%20Intelligence%20An%20Introduction%20-%20Andries%20P.%20Engelbrecht.pdf

[69] J. A. Cook and J. Ranstam, "Overfitting". *British Journal of Surgery*, vol. 103(13), (2016).

[70] I. Markoulidakis, I. Rallis, I. Gerogoulas, G. Kopsiaftis, A. Doulamis and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem". *Technologies, (2021).*
[Online]. Available: https://doi.org/10.3390/technologies9040081

[71] V. Mariano, J. A. Tobon Vasquez, M. R. Casu F. Vipiana, "Brain Stroke Classification via Machine Learning Algorithms Trained with a Linearized Scattering Operator", submitted to IEEE Transactions on Antennas Propagation. Special issue: "*Artificial Intelligence: New Frontiers in Real-Time Inverse Scattering and Electromagnetic Imaging*", (2022).

[72] A. S. Beaverstone, D. S. Shumakov, and N. K. Nikolova, "Frequencydomain integral equations of scattering for complex scalar responses," *IEEE Transactions on Microwave Theory and Techniques*, vol. 65(4), pp. 1120–1132, (2017).

[73] D. Tajik, F. Foroutan, D. S. Shumakov, A. D. Pitcher, and N. K.Nikolova, "Real-time microwave imaging of a compressed breast phantom with planar scanning," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 2(3), pp. 154–162, (2018).

[74] C. Laredo, Y. Zhao, S. Rudilosso, A. Renu, J. Pariente, A. Chamorro, and X. Urra, "Prognostic significance of infarct size and location: The case of insular stroke," *Scientific Reports*, vol. 8, (2018)

[75] J. L. Saver, K. C. Johnston, D. Homer, R. Wityk, W. Koroshetz, L. L. Truskowski, and E. C. Haley, "Infarct volume as a surrogate or auxiliary outcome measure in ischemic stroke clinical trials," *Stroke,* vol. 30, p. 293–298, (1999).

[76] D. O. Rodriguez-Duarte, J. A. Tobon Vasquez and F. Vipiana, "Multi-shot calibration technique for microwave imaging system". *IEEE Conference on Antenna Measurements and Application*, 978-1-7281-9697-8/21/, (2021).

[77] E. A. Attardo, A. Borsic, G. Vecchi, and P. M. Meaney, "Whole-system electromagnetic modeling for microwave tomography," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 1618–1621, (2012).

[78] A. Giordano, "Microwave Imaging Technology for food contamination monitoring", Master Thesis in Electronic Engineering (2018).

[79] J.M. Jin, "Theory and Computation of Electromagnetic Fields", (2010).
Print ISBN:9780470533598 |Online ISBN:9780470874257 |DOI:10.1002/9780470874257

[80] [Online]. Available: https://www.python.org/

[81] [Online]. Available:
https://colab.research.google.com/notebooks/welcome.ipynb?hl=it

[82] [Online]. Available: https://numpy.org/about

[83] [Online]. Available: https://pandas.pydata.org/

[84] [Online]. Available: matplotlib.org/

[85] [Online]. Available: https://www.tensorflow.org/

[86] [Online]. Available: https://scikit-learn.org/stable/index.html

[87] [Online]. Available:
https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

[88] [Online]. Available: https://scikit-learn.org/stable/modules/svm.html

[89] [Online]. Available:
https://scikitlearn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

[90] [Online]. Available:
https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html