
PREDICTOR OF NEOPLASMS AND BODY COMPOSITION WITH MACHINE LEARNING MODELS

TECHNICAL REPORT

M^a Jesús Fuentes Sebio
Dra. Fuentes Medical Nutrition Center
Pontevedra, Spain
fsebio@cmpont.es

May 6, 2023

ABSTRACT

Background: The tissue microenvironment of neoplastic diseases differs from that of normal cells. Their extracellular matrix, innervation, metabolism, as well as interstitial fluid and intercellular interconnections mark clear physical differences between normal and cancerous cellular ecosystems. Detecting these physical changes early without using diagnostic methods that are harmful and uncomfortable for the patient is a major challenge for the medical-scientific community. Electrical bioimpedance supported by machine learning techniques can provide clues to incipient preneoplastic tissue changes. **Methods:** In this study, 7 predictive machine learning models were developed using a database with bioimpedanciometric data from a group of healthy individuals and another group of patients who had or were suffering from cancer at the time of measurement. **Results:** The *Random Forest* was the model that reported the best Accuracy, reaching over 90% of hits.

Conclusions: These results open the door to future research linking changes in body composition and pretumoral tissue environments using machine learning tools.

Keywords Electric Impedance · Neoplasms · Prevention · Neuronal Networks

Contents

1	Introduction	2
2	Objectives	4
3	Methodology	4
4	Exploratory data analysis	4
5	Training of machine learning models	5
5.1	k-Nearest Neighbors (kNN)	5
5.2	Logistic Regression	5
5.3	Linear Discriminant Analysis	5
5.4	<i>Single C5.0 Tree</i>	5
5.5	Random Forest	6
5.6	Stochastic Gradient Boosting	6

5.7	Neural Network	6
6	Evaluation of the results of all the models	7
6.1	Accuracy and average Kappa of each model	7
6.2	Differences between the training and test set according to the predicted results	8
7	Discussion	8
8	Conclusion	9
	Bibliography	9
	Supplementary material	14
	Variables used to train the neural network	14
	Frequency of different types of tumors	14
	Phases of the training of the different models	14
	Analysis of variance of the models	15
	Graphical representation of the exploratory analysis of the data	16

1 Introduction

Cancer encompasses a group of diseases characterized by the abnormal and uncontrolled division of anarchic cells that infiltrate and destroy normal tissues [56]. Its development requires a three-dimensional microenvironment that favors it, in which mechanical forces as well as biomolecular gradients and non-organic components come into play [52]. Neoplastic cells have an impressive capacity to evade therapy directed against them, and it is necessary to find medical clues indicating the development of favorable conditions for their development.

The study of the electrical properties of the organism to estimate body composition in both healthy and diseased individuals has been used for years with **bioimpedancemetry** (BIA) [53, 54], one of the most widely used methods to analyze body composition due to its ease of use, safety, reproducibility, low cost and good accuracy [29], although without reaching the precision of other more sophisticated techniques such as Computerized Axial Tomography (CT), Magnetic Resonance Imaging, Dual X Absorptiometry [9], Hydrometry, Isotope Dilution, etc., whose use is limited by their cost and side effects [4][18]. Bioimpedance can be used at any age and, except in some situations that contraindicate it, such as the use of pacemakers, it is a safe technique.

The physical basis of BIA is based on the resistance of the different body structures to the passage of a low-voltage electric current [11, 10]. This means that if we modify the electrical frequency or the physical-chemical conditions of the organism, the resistance to the passage of the current also changes [8]. Based on these properties, we define the **conductance** as the facility of a material to allow the flow of electric current and, on the contrary, the **resistance** is the difficulty to the passage of the same one.

There is no doubt that our organism is not a simple [48] circuit and as such, other physical components come into play, as is the case of the **capacitors**, whose function is to store energy to release it later. Our body capacitors are the cell membranes and interfaces, thus, we define the **electrical capacitance** as their capacity to store energy and the **reactance**, the resistance offered by these cell capacitors to the passage of electric current.

$$Z = \sqrt{R^2 + Xc^2}$$

Z impedance
 R Resistance
 Xc Reactance

Figure 1: Corporal impedance.

The **resistance** and the **reactance**, that together we call **Impedance (Z)**, are the factors that oppose resistance to the passage of electric current (Figure 1). The values of the Impedance depend on the voltage and the intensity of the electric current, as we can see in Figure 2.

$$Z = \frac{V}{I}$$

Z impedance
 V voltage
 I intensity

Figure 2: Impedance formula.

Bioimpedance is a simple and reliable method to identify groups of individuals with alterations in the quantity and distribution of body fluids as well as to assess the state of cell membranes [13]. The most commonly used equipment today is in the frequency range of 1 kHz - 1 MHz. The electric current flows through the extracellular water at low frequencies (1 - 5 kHz), which determines that the reactance is minimal [70]. If we increase the frequency (50 - 100 kHz), electricity penetrates into the intercellular space, but the membranes hinder it by acting as capacitors. If the frequency continues to rise, electricity flows through the extra and intracellular water as well as through the intervening compartments [39].

The angle formed by the above three vectors is called the “phase angle” (Figure 3) [12, 22, 21] and is one of the most widely used parameters in research. The **phase angle** shows a positive relationship with the integrity and optimal functionality of the body cells (reflected in the reactance) and is negatively associated with the degree of hydration of the various tissues (resistance) [47, 48].

To obtain the value of the phase angle we have to apply the Formula 1 [26]:

$$\text{Phase angle} = \arctan\left(\frac{\text{Reactance}}{\text{Resistance}}\right) \times \frac{180}{\pi} \quad (1)$$

It is usually calculated with frequencies of 50 kHz on the right side of the body, however multi-frequency equipment calculates it on both sides [60].

The phase angle value reflects the integrity of cell membranes as well as intracellular and extracellular water concentrations. Low values indicate damage, destructuring of the cell wall and even cell death [14, 15, 20], being considered very useful to evaluate inflammatory abnormalities [69, 27]. Low phase angle values are a good prognostic marker in cardiac, renal, hepatic, infectious pathologies and cancer [51, 55], among others [47]. They give us clues to poor clinical evolution in individuals with cancer [59, 33], speculating that the alteration of the corporal homeostasis as well as an increase of inflammatory cytokines constitute the underlying basis [36, 37, 38]. High values, on the contrary, reflect optimal cellular integrity.

In the scientific literature, several research papers used impedance-based techniques to distinguish tumors from healthy tissues [30, 31, 32]. Anarchic and uncontrolled cell proliferation as well as a misconfiguration of the cell stroma at the tumor level results in alterations of the electrical charges; in fact, the most aggressive tumors entail cellular alterations at the systemic level that reduce survival [35].

Detecting minimal precancerous tissue changes is a major challenge for the scientific community, and new artificial intelligence tools are a breakthrough in discovering patterns and correlations in a data set in order to make predictive

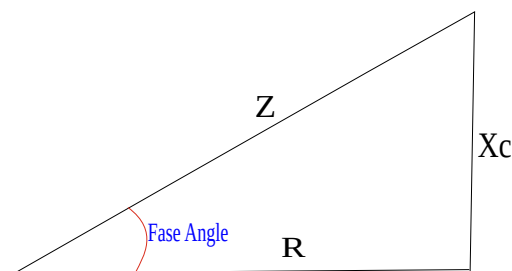


Figure 3: Relations between the components of bioimpedance.

analyses of the data, much in the same way as human reasoning. **Machine learning** learns from the data, from a training dataset and generates a predictive model [23, 58], thus improving statistical analysis [57].

2 Objectives

The aim of this research work is to predict with machine learning algorithms whether a person, object of the study, is in a clinical situation with a high probability of suffering a neoplastic degeneration based on bioimpedanciometric studies.

3 Methodology

Observational, retrospective, cross-sectional study including bioimpedanciometric records of 912 healthy individuals and 1139 records of 88 individuals who suffers or suffered some type of neoplasms at the time of analysis, attended at Dra. Fuentes Medical Nutrition Center's office in the period between 2015 and 2022. Pregnant women and individuals under 18 years of age were excluded. The multi-frequency BIA equipment TANITA MC-980MA with 8 electrodes [28] was the bioimpedance meter at 1, 5, 50, 250, 500 and 1000 kHz frequency.

The design and training of the neural network was carried out using the open source programming language **R** using the package **caret** (classification and regression training), developed by *Max Kuhn* [46]. This interface has tools for the development of predictive models based on classification and regression processes and it is available at <http://CRAN.R-project.org/package=caret>.

The evaluation of the quality of the algorithm when making predictions is carried out with two metrics: *Accuracy* and *Kappa*. The former reports the percentage of correctly predicted observations and the *Kappa*, or *Cohen's Kappa metric*, is *Accuracy's* standardized result for accurate random results.

The machine learning models used were:

- **k-Nearest Neighbors (knn)**: Supervised learning method that classifies new data based on the distances of nearest neighbor values [42].
- **Logistic regression**: Used to model data that does not fit the linear regression model. Calculates the probability of a binary qualitative variable as a function of different quantitative variables [72].
- **Linear Discriminant Analysis**: Supervised classification method in which groups are organized according to their characteristics [72].
- **Single C5.0 Tree**: Model made up of simple classification trees [74].
- **Random Forest**: This algorithm created in 2001 by Leo Breiman uses multiple decision trees, training different parts of them, thus reducing the variance [45]. It has great flexibility and optimal predictive results, which is why it is widely used [50].
- **Stochastic Gradient Boosting**: It is based on the construction of several decision trees with random data from subsets of the data [74].
- **Neural Networks**: Inspired by the structure and functioning of our brain neural networks [45].

4 Exploratory data analysis

The participants were divided into healthy and pathological: 912 records from healthy individuals and 1139 bioimpedance measurements from 89 pathological individuals. The smaller number of pathological individuals and the need for machine learning to be able to recognize as many "unhealthy" individuals as possible for learning, motivated the selection of the records in this way.

The types of tumors are specified in the Table 1, with breast, uterus and thyroid neoplasms accounting for the largest number of individuals. Exploratory analysis of the data reflects a lower phase angle value in individuals in the pathological group compared to individuals in the normal group 10. A reduced phase angle value reflects an alteration in tissue electrical properties [59], which suggests an unfavorable evolution of their health status

Individuals in the pathological group have a higher Body Mass Index (BMI), as can be seen in Figures 9, 10 and 11. This confirms the relationship between excess body mass and the development of different types of cancer reported in analyses conducted by the World Cancer Research Fund (WCRF) [77, 78]. What is most striking from the exploratory analysis of the data is that normally phase angle declines with age in healthy individuals. However, this does not occur

in individuals in the pathological group in which the relationship between phase angle and age is not linear (Figures 12 and 13).

5 Training of machine learning models

The various machine learning algorithms were applied to the training data in order to achieve the best result.

5.1 k-Nearest Neighbors (kNN)

This is a very simple algorithm that attempts to identify similar observations. It uses k as the only parameter that tells us the number of neighboring observations used. In this case the values of k that were included were: 1, 2, 5, 10, 15, 20, 30 and 50. The highest Accuracy was achieved with k = 1, as we can see in the result:

k	Accuracy	Kappa
1	0.8738130	0.7410694

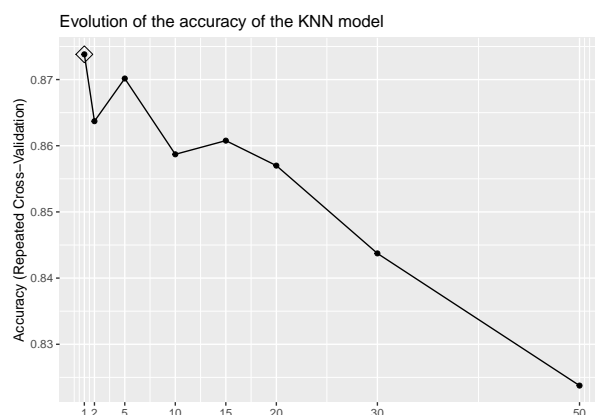


Figure 4: Graphical representation of the training KNN with the values de k used.

5.2 Logistic Regression

Logistic regression is a type of regression in which the dependent variable is categorical. It does not use hyperparameters and is a good technique for estimating probabilities[75].

The best training result was:

Accuracy	Kappa
0.8221619	0.6388024

5.3 Linear Discriminant Analysis

Machine learning classification method based on Bayes' theorem. It calculates the probability that an observation belongs to one of the groups to be studied. It does not use any hyperparameter.

The result obtained was:

Accuracy	Kappa
0.8186268	0.6307069

5.4 Single C5.0 Tree

Developed by J.Ross Quinlan [75], this algorithm is simple but very efficient with the ability to exclude unhelpful features [75] and is very productive even on large data sets. It does not need hyperparameters.

Final result taking into account the best Accuracy:

Accuracy	Kappa
0.8774582	0.7499857

5.5 Random Forest

This method, although also simple, is capable of processing tables with unbalanced data [23]. Its algorithm uses the parameters: **mtry** and **min.node.size**, the former representing the number of randomly selected predictor variables in each tree and the parameter “min.node.size” setting the minimum size of the node to be split.

The hyperparameters used were $mtry = c(3, 4, 5, 7)$ and $min.node.size = c(2, 3, 4, 5, 10, 15, 20, 30)$. Final result taking into account the best Accuracy:

mtry	min.node.size	Accuracy	Kappa
3	2	0.9025617	0.8019504

5.6 Stochastic Gradient Boosting

This model works with a set of decision trees, in which sequentially one tree learns from the preceding one, thus improving the prediction. It uses random sampling of observations without replacement in the training data. This method consists of 4 parameters:

- *shrinkage*: It controls the influence of each model on the final prediction. The values of $shrinkage = c(0.001, 0.01, 0.1)$ were used in training.
- *interaction.depth*: It monitors the number of divisions the tree has. The values of $interaction.depth = c(1, 2)$ were selected.
- *n.minobsinnode*: It sets a minimum number of observations for the node to be split. In this training a $n.minobsinnode = c(2, 5, 15)$ was used.
- *n.trees*: Number of models used in the process. The selected values were $n.trees = c(500, 1000, 2000)$.

Final result taking into account the best Accuracy:

shrinkage	interaction.depth	n.minobsinnode	n.trees	Accuracy	Kappa
0.001	1	2	500	0.8069453	0.6059745

5.7 Neural Network

Neural networks are designed with groups of interconnected nodes. Each neural network has 3 types of neurons: input, hidden layer and output. The neurons are interconnected with each other and the strength of this connection is called “weight”. It has the advantage of working well with non-linear relationships between variables and can handle large data diligently. This model has 2 hyperparameters, on the one hand “**size**” which tells us the number of neurons in the hidden layer and “**decay**”, the parameter in charge of regularization during training. In this case we used: $size = c(1,3,5)$ and $decay = c(0e+00, 1e-01, 1e-04)$. Final result and representative graph in Figure 5.

size	decay	Accuracy	Kappa
1	0e+00	0.8415314	0.6814546

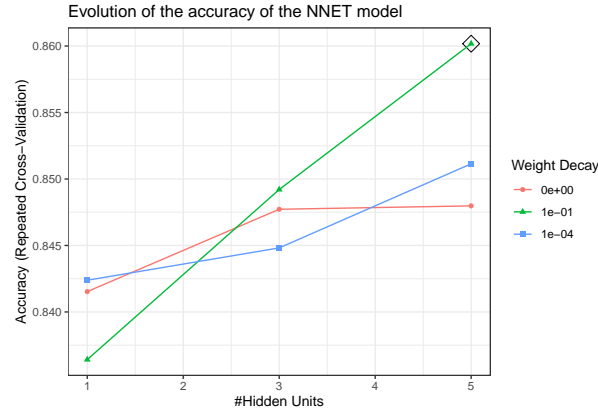


Figure 5: NNET model training.

6 Evaluation of the results of all the models

6.1 Accuracy and average Kappa of each model

Number	Model	Accuracy	Kappa
1	Rf	0.910	0.817
2	Boosting	0.885	0.766
3	Tree	0.877	0.750
4	KNN	0.874	0.741
5	NNET	0.860	0.717
6	Logistic	0.822	0.639
7	LDA	0.819	0.631

All models exceed the baseline level (0.62) marked on the Figure 6 as a dashed line.

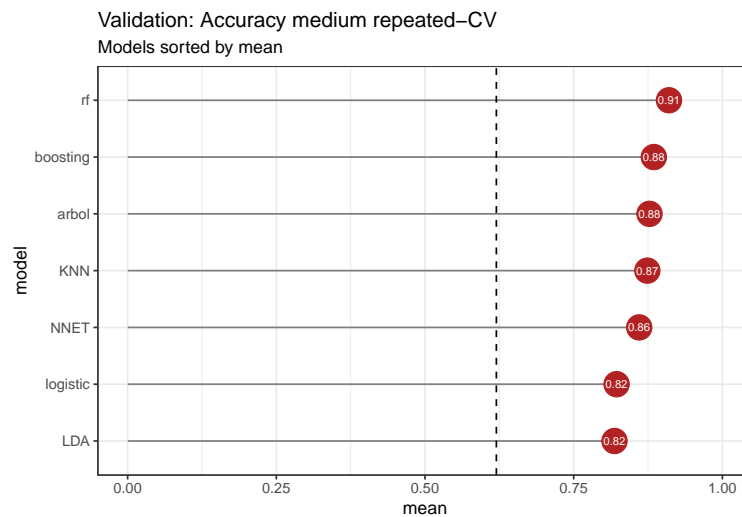


Figure 6: Final results of the training of the different models.

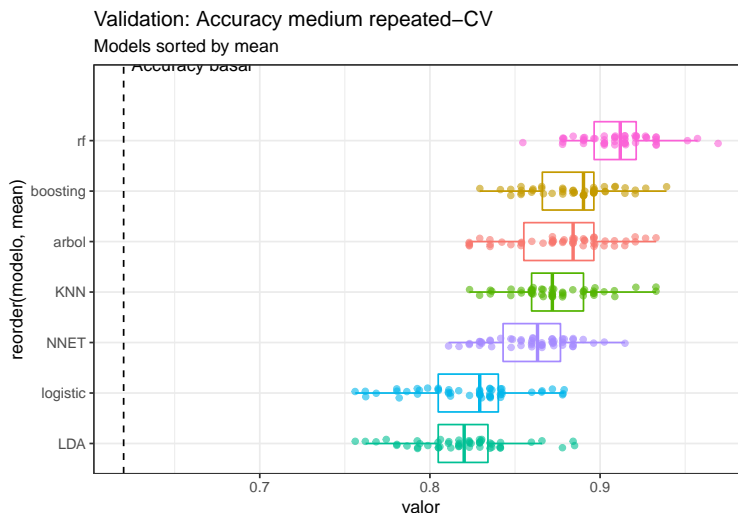


Figure 7: Boxplots of final model results.

The **Random Forest** model is the one that achieved the best result (Figures 6 and 7). However, we need to compare the variances to determine if the differences are significant (Figure 8).

6.2 Differences between the training and test set according to the predicted results

When training the data test with the different models, accuracy values higher 82 % were reported:

Number	Model	Test	Training
1	Rf	0.905	0.999
2	KNN	0.868	1
3	Boosting	0.866	0.979
4	Tree	0.858	0.926
5	NNET	0.851	0.876
6	Logistic	0.824	0.826
7	LDA	0.822	0.818

7 Discussion

Cancer is not one but a variety of complex diseases whose cellular components are derived from normal cells. Abnormalities of their membranes as well as cell junctions, increased lactate production and the appearance of novel antigens clearly mark the differences between cancer cells and the rest of the cellular pool. These modifications cause changes in the electrical properties of the tissue [68], anticipating even the state of cachexia [63, 62]. This means that alterations in the physiology of the organism's tissue components are reflected in their electrical properties [67].

The main objective of this study was to use multifrequency bioimpedance analysis with machine learning algorithms in order to predict a high predisposition to neoplastic degeneration based on tissue electrical changes. Machine learning is a discipline in which algorithms and models are applied to a set of data and in this case, the **Random Forest** algorithm reported the best results (according to the accuracy metric) with very good performance. The result achieved in the test was 90 % correct (Figure 8). This model works very well with large databases and is able to handle and catalog numerous predictors [49].

In this study, it is reported that the mathematical algorithms used in machine learning and bioimpedance data are very useful tools to predict an alteration in body homeostasis, a risk situation that may favor the development of neoplastic

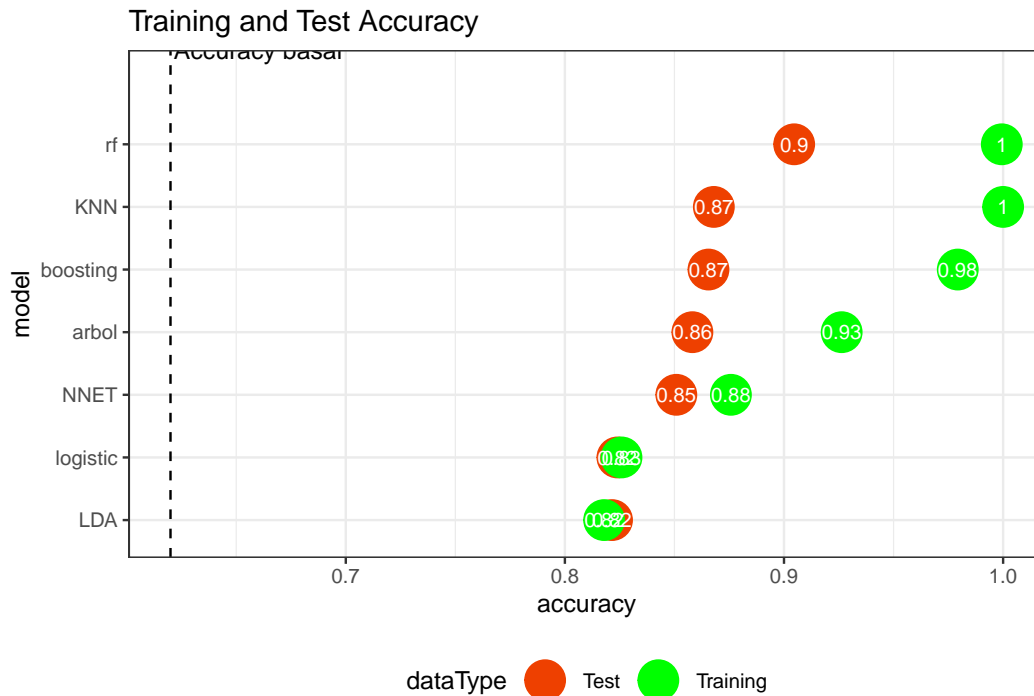


Figure 8: Results of applying the different models to the training and test data.

pathologies or even to predict relapses in patients after neoplastic treatment. There is no doubt that this work has limitations, since a larger sample of people with neoplasias would be needed, as well as a differentiation of them in their different types, stages and treatments received.

8 Conclusion

The machine learning-bioimpedance duo is a useful screening tool to detect patients at high risk of neoplastic degeneration, allowing early preventive and therapeutic measures to be used if appropriate.

Bibliography

- [1] Dittmar M. Reliability and variability of bioimpedance measures in normal adults: effects of age, gender and body mass index. *Am J Physiol Anthropol.* 2003;122:361-70.
- [2] Ofer Reizes and Nathan A. Adipocytokines, Energy Balance and Cancer.12,DOI 10.1007/978-3-319-41677-9_1 Berger Editors. Editorial Springer International Publishing Switzerland.
- [3] Bellido Castañeda, Virginia. Relación del ángulo de fase determinado por bioimpedanciometría con factores de riesgo cardiovascular, adipocitoquinas, antropometría e ingesta dietética de pacientes obesos. Tesis doctoral, 2014, Universidad de la Coruña.
- [4] Yamada, Y.; Masuo, Y.; Yokoyama, K.; Hashii, Y; Ando, S.; Okayama, Y.; Morimoto, T.; Kimura, M.; Oda, S. Proximal electrode placement improves the estimation of body composition in obese and lean elderly during segmental bioelectrical impedance analysis. *Eur. J. Appl. Physiol.* 2009, 107, 135–144.
- [5] Kyle, U.G.; Bosaeus, I.; De Lorenzo, A.D.; Deurenberg, P.; Elia, M.; Gómez, J.M.; Heitmann, B.L.; Kent-Smith, L.; Melchior, J.C.; Pirlich, M.; et al. Bioelectrical impedance analysis—Part I: Review of principles and methods. *Clin. Nutr.* 2004, 23, 1226–1243.
- [6] Bosy-Westphal, A.; Jensen, B.; Braun, W.; Pourhassan, M.; Gallagher, D.; Müller, M.J. Quantification of whole-body and segmental skeletal muscle mass using phase-sensitive 8-electrode medical bioelectrical impedance devices. *Eur. J. Clin. Nutr.* 2017.

- [7] MSc. Lidyce Quesada Leyva; Dra. Cira Cecilia León Ramentol; MSc. José Betancourt Bethencourt; MSc. Elizabeth Nicolau Pestana. Theoretical and practical facts about health electric bioimpedance. *Rev. Arch Med Camagüey* Vol20(5)2016
- [8] Bedogni G, Malavolti M, Severi S, Poli M, Mussi C, Fantuzzi AL, Battistini N. Accuracy of an eight-pandersonoint tactile-electrode impedance method in the assessment of total body water. *Eur J Clin Nutr.* 2002;56(11):1143–8.
- [9] Anderson LJ, Erceg DN, Schroeder ET. Utility of multifrequency bioelectrical impedance compared with dual-energy x-ray absorptiometry for assessment of total and regional body composition varies between men and women. *Nutr Res.* 2012;32(7):479–85.
- [10] Coppini LZ, Waitzberg DL, Campos AC. Limitations and validation of bioelectrical impedance analysis in morbidly obese patients. *Curr Opin Clin Nutr Metab Care.* 2005;8:329–332.
- [11] Scharfetter H, Schlager T, Stollberger R, Felsberger R, Hutten H, Hinghofer-Szalkay H. Assessing abdominal fatness with local bioimpedance analysis: basics and experimental findings. *Int J Obes Relat Metab Disord.* 2001;25:502–511. doi: 10.1038/sj.ijo.0801556.
- [12] Barbosa-Silva MC, Barros AJ, Wang J, Heymsfield SB, Pierson RN Jr. Bioelectrical impedance analysis: population reference values for phase angle by age and sex. *Am J Clin Nutr.* 2005;82(1):49-52. doi: <https://doi.org/10.1093/ajcn/82.1.49>.
- [13] Luigi Barrea, Giovanna Muscogiuri, Gabriella Pugliese, Daniela Laudisio, Giulia de Alteriis, Chiara Graziadio, Annamaria Colao and Silvia Savastano. Phase Angle as an Easy Diagnostic Tool of Meta-Inflammation for the Nutritionist. *Nutrients* 2021, 13, 1446. <https://doi.org/10.3390/nu13051446> <https://www.mdpi.com/journal/nutrients>.
- [14] Kumar, S.; Dutt, A.; Hemraj, S.; Bhat, S.; Manipadybhima, B. Phase angle measurement in healthy human subjects through bio-impedance analysis. *Iran. J. Basic Med. Sci.* 2012, 15, 1180–1184. *Nutrients* 2021, 13, 1446 14 of 16
- [15] Buffa, R.; Floris, G.; Marini, E. Migration of the bioelectrical impedance vector in healthy elderly subjects. *Nutrition* 2003, 19, 917–921.
- [16] Norman K, Stobaus N, Zocher D, Bosy-Westphal A, Szramek A, et al.: Cutoff percentiles of bioelectrical phase angle predict functionality, quality of life, and mortality in patients with cancer. *Am J Clin Nutr* 3, 612–619, 2010.
- [17] Dana A. Pérez Camargo, Silvia R. Allende Pérez, Monica M. Rivera Franco, Nelson E. Alvarez Licon, Víctor I. Urbalejo Cenicerros and Lilian E. Figueroa Baldenegro. Phase Angle of Bioelectrical Impedance Analysis as Prognostic Factor in Palliative Care Patients at the National Cancer Institute in Mexico. *Nutrition y Cáncer.* 2017, VOL. 0, NO. 0, 1–6 <http://dx.doi.org/10.1080/01635581.2017.1299880>.
- [18] Robinson Cruz. Composición Corporal a través del análisis de Biomedancia Bioeléctrica (BIA). *Renut(2010)4(13)667-680*.
- [19] Quetelet A. *Physique sociale, ou essai sur le développement des facultés de l’homme.* Volume 1. 1st ed. Bruxelles: Muquardt; 1869.
- [20] Digant Gupta, Carolyn A Lammersfeld, Pankaj G Vashi, Jessica King, Sadie L Dahlk, James F Grutsch and Christopher G Lis. Bioelectrical impedance phase angle as a prognostic indicator in breast cancer. *BMC Cancer* 2008, 8:249 doi:10.1186/1471-2407-8-249.
- [21] Maria Mantzorou, Maria Tolia, Antigoni Poultsidi, Eleni Pavlidou, Sousana K. Papadopoulou, Dimitrios Papan-dreou and Constantinos Giaginis. Can Bioelectrical Impedance Analysis and BMI Be a Prognostic Tool in Head and Neck Cancer Patients? A Review of the Evidence. *Cancers* 2020, 12, 557; doi:10.3390/cancers12030557.
- [22] Morlino, D.; Cioffi, I.; Marra, M.; Di Vincenzo, O.; Scalfi, L.; Pasanisi, F. Bioelectrical Phase Angle in Patients with Breast Cancer: A Systematic Review. *Cancers* 2022, 14, 2002. <https://doi.org/10.3390/cancers14082002>.
- [23] AshishSingh Bhatia Yu-Wei, Chiu (David Chiu). *Machine Learning with R Cookbook.* 2^aEdition. Published by Packt Publishing Ltd.ISBN 978-1-78728-439-5.
- [24] Mayane Marinho Esteves Pereira et al. The Prognostic Role of Phase Angle in Advanced Cancer Patients: A Systematic Review. *Nutrition in Clinical Practice* Volume 00 Number 0 April 2018 1–12 C 2018 American Society for Parenteral and Enteral Nutrition. DOI: 10.1002/ncp.10100 wileyonlinelibrary.com.
- [25] Da Silva BR, Gonzalez MC, Cereda E, Prado CM. Exploring the potential role of phase angle as a marker of oxidative stress: a narrative review. *Nutrition.* 2021;93:111493.
- [26] Institute of Medicine (US) Committee on Military Nutrition Research. Emerging technologies for nutrition research: Potential for assessing military performance capability. Carlson-Newberry S, Costello R, editors. *Emerg. Technol. Nutr. Res.* Washington (DC): National Academies Press (US); 1997.

- [27] Bruna Ramos da Silva et al. Phase angle and cellular health: inflammation and oxidative damage. *Reviews in Endocrine and Metabolic Disorders* <https://doi.org/10.1007/s11154-022-09775-0>.
- [28] Jiang, F.; Tang, S.; Eom, J.-J.; Song, K.-H.; Kim, H.; Chung, S.; Kim, C.-H. Accuracy of Estimated Bioimpedance Parameters with Octapolar Segmental Bioimpedance Analysis. *Sensors* 2022, 22, 2681. <https://doi.org/10.3390/s22072681>
- [29] A. Piccoli, L. D. Nescolarde y J. Rosell. Análisis convencional y vectorial de bioimpedancia en la práctica clínica. *Nefrología*. Vol. XXII. Número 3. 2002.
- [30] H.-G. Jahnke, A. Heimann, R. Azendorf, and et al., Impedance spectroscopy, an outstanding method for label-free and real-time discrimination between brain and tumor tissue in vivo. *Biosens. Bioelectron.*, vol. 46, pp. 8–14, Aug. 2013.
- [31] G. Qiao, W. Duan, C. Chatwin, and et al. Electrical properties of breast cancer cells from impedance measurement of cell suspensions. *Journal of Physics: Conference Series*, 2010, vol. 224, p. 012081.
- [32] A. J. Barrow and S. M. Wu. Impedance measurements for cervical cancer diagnosis. *Gynecologic oncology*, 2007, vol. 107, pp. S40–S43.
- [33] L. Llamas, V. Baldomero, M. L. Iglesias y L. P. Rodota. Valores del ángulo de fase por bioimpedancia eléctrica; estado nutricional y valor pronóstico. *Revisión. Nutr Hosp*. 2013;28(2):286-295
- [34] Rosa Noguera, Rebeca Burgos-Panadero, Esther Gamero-Sandemetro, Luis de la Cruz-Merino y Tomás Álvaro Naranjo. Una visión integral del cáncer (I). Microambiente tumoral: estudio, clasificación y reprogramación. *Rev Esp Patol*. 2019. <https://doi.org/10.1016/j.patol.2018.11.003>.
- [35] Axelsson L, Silander E, Bosaeus I and Hammerlid E. Bioelectrical phase angle at diagnosis as a prognostic factor for survival in advanced head and neck cancer. *Eur Arch Oto-Rhino-Laryngology [Internet]* 2018;275(9):2379-86. DOI: 10.1007/s00405-018-5069-2.
- [36] Pena NF, Mauricio SF, Rodrigues AMS, Carmo AS, Coury NC, Correia MITD, et al. Association Between Standardized Phase Angle, Nutrition Status, and Clinical Outcomes in Surgical Cancer Patients. *Nutr Clin Pract* 2019;34(3):381-6. DOI: 10.1002/ncp.10110.
- [37] Greten FR, Grivennikov SI. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity*. 2019;51(1):27-41. DOI: 10.1016/j.immuni.2019.06.025.
- [38] Rute Mattos Dourado Esteves Justa et al. Tumor aggressiveness is associated with cell integrity changes in breast cancer-surviving women: a follow-up study. *Nutr Hosp* 2022;39(1):138-146.
- [39] María Molina Vega, José Manuel García Almeida, Isabel Vegas Aguilar, Araceli Muñoz Garach, Ana María Gómez Pérez, Isabel Cornejo Pareja, Cristina Díaz Perdignes y Diego Bellido Guerrero. Revisión sobre los fundamentos teórico-prácticos del ángulo de fase y su valor pronóstico en la práctica clínica. *Nutr Clin Med* 2017; XI (3): 129-148 DOI: 10.7400/NCM.2017.11.3.5055.
- [40] A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144-152. ACM, 1992.
- [41] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273-297, 1995.
- [42] Cristina García and Irene Gómez Moreno. Algoritmos de aprendizaje: knn y kmeans. Technical report.
- [43] Hefin I. Rhys. *Machine Learning with R, the tidyverse, and mlr*. Manning Publications Co. ISBN: 9781617296574.
- [44] Jared P. Lander. *R for Everyone Advanced Analytics and Graphics*. Second Edition. Addison-Wesley ISBN-13: 978-0-13-454692-6.
- [45] Philippe J.S. De Brouwer. *The Big R-Book from data science to learning machines and big data*. 2021 John Wiley and Sons, Inc. ISBN: 9781119632771.
- [46] Max Kuhn. *The caret Package: A Unified Interface for Predictive Models*. Pfizer Global RyD Nonclinical Statistics Groton, CT max.kuhn@pfizer.com February 26, 2014
- [47] Bellido D, Carreira J, Bellido V. Evaluación del estado nutricional: antropometría y composición corporal. *Tratado de Nutrición: Nutrición Humana en el estado de salud*. 3a. Madrid: Panamericana; 2017. p. 99-132.
- [48] Kyle UG, Bosaeus I, De Lorenzo AD, Deurenberg P, Elia M, Gómez JM, et al. Bioelectrical impedance analysis—part I: review of principles and methods. *Clin Nutr Edinb Scotl* 2004;23(5):1226-43. DOI: 10.1016/j.clnu.2004.06.004.
- [49] Cánovas, F., Alonso, F., Gomariz, F. y Oñate, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Computers y Geosciences*, 103, 1-11. <https://doi.org/10.1016/j.cageo.2017.02.012>.

- [50] Robin Genuer and Jean-Michel Poggi. *Random Forests with R*. Springer. ISBN 978-3-030-56484-1.
- [51] Pereira MME, Wiegert EVM, Oliveira LC and Lima LC. Ângulo de Fase e Estado Nutricional em Indivíduos com Câncer Avançado em Cuidados Paliativos. *Revista Brasileira de Cancerologia* 2019; 65(1): e-02272. <http://dx.doi.org/10.32635/2176-9745.RBC.2019v65n1.272>.
- [52] White et al. *The Physics of Cancer*. *Cancer Res*; 79(9) May 1, 2019. www.aacrjournals.org
- [53] Brantlov S, Ward LC, Jødal L, Rittig S, Lange A. Critical factors and their impact on bioelectrical impedance analysis in children: a review. *J Med Eng Technol*. 2017;41(1):22–35. <https://doi.org/10.1080/03091902.2016.1209590>.
- [54] Piccoli A, Nescolarde LD, Rosell J. Análisis convencional y vectorial de bioimpedancia en la práctica clínica. *Nefrología*. 2002;22(3):228–38. <http://www.revistanefrologia.com/es-pdf-X0211699502014897>.
- [55] Luna et al. *Translational Medicine Communications* (2020) 5:10. <https://doi.org/10.1186/s41231-020-00062-1>.
- [56] Rita Fior and Rita Zilhão. *Molecular and cell biology of Câncer*. Springer. ISBN 978-3-030-11811-2.
- [57] Jason Brownlee. *Machine Learning Mastery with R Get Started, Build Accurate Models and Work Through Projects Step-by-Step*.
- [58] Hefin I. *Rhys. Machine Learning with R, the tidyverse, and mlr*. Manning Publications Co. ISBN: 9781617296574.
- [59] Norman K, Smoliner C, Kilbert A, Valentini L, Lochs H, Pirlich M. Disease related malnutrition but not underweight by BMI is reflected by disturbed electric tissue properties in the bioelectrical impedance vector analysis. *Br J Nutr* 2008;100:590e5.
- [60] David Hui, MD, MSc, Rony Dev, MD, Lindsay Pimental, BSN, Minjeong Park, PhD, Maria A. Cerana, MD, Diane Liu, MS, and Eduardo Bruera, MD. Association Between Multi-frequency Phase Angle and Survival in Patients With Advanced Cancer. *Journal of Pain and Symptom Management*. Vol. 53 No. 3 March 2017.
- [61] Tyagi R, Mishra S, Gaur N, Panwar A, Saini D, Singh K, Kumar D, Jeelani G. Bioelectric impedance phase angle in carcinoma prostate, a hospital-based study. *Int J Med Sci Public Health* 2016;5:1826-1830.
- [62] Trump BF, Heatfield BM, Phelps. PC, Sanefuji H, Shamsuddin AK. *Europe Pub Med Central* 1980;3:43-60.
- [63] Toso S, Piccoli A. Bioimpedance vector pattern in cancer patients without disease versus locally advanced or disseminated disease. *Nutrition* 2003;19:510-14.
- [64] Małecka-Massalska T, Smoleń A and Morshed K. Tissue electrical properties in head and neck tumors before and after surgery: Preliminary observations. *Indian J Cancer* 51: 209-213, 2014. PMID: 25494106. DOI: 10.4103/0019-509X.146717.
- [65] Władysiuk MS, Mlak R, Morshed K, Surtel W, Brzozowska A and Małecka-Massalska. Phase angle is considered to be a nutritional indicator in patients with head and neck cancer in detecting malnutrition. *Curr Oncol* 23(5): e481-e487, 2016. PMID: 25859939. DOI: 10.1007/s00405-015-3626-5.
- [66] Ozorio GA, Barão K and Forones NM: Cachexia stage, patient generated subjective global assessment, phase angle, and handgrip strength in patients with gastrointestinal cancer. *Nutr Cancer* 69(5): 772-779, 2017. PMID: 28524706. DOI: 10.1080/01635581.2017.1321130.
- [67] Czerniec SA, Ward LC, Refshauge KM, Beith J, Lee MJ, et al. (2010) Assessment of breast cancer-related arm lymphedema. Comparison of physical measurement methods and self-report. *Cancer Invest* 28: 54-62.
- [68] Tuorkey MJ (2012). Bioelectrical Impedance as a Diagnostic Factor in the Clinical Practice and Prognostic Factor for Survival in Cancer Patients: Prediction, Accuracy and Reliability. *J Biosens Bioelectron* 3:121. doi:10.4172/2155-6210.1000121.
- [69] Barbosa-Silva MC, Barros AJ. Bioelectrical impedance analysis in clinical practice: A new perspective on its use beyond body composition equations. *Curr Opin Clin Nutr Metab Care* 2000;8:311-7.
- [70] Bosity-Westphal A, Danielzik S, Dörhöfer RP, Later W, Wiese S, Müller MJ. Phase angle from bioelectrical impedance analysis population reference values by age, sex, and body mass index. *JPEN J Parenter Enteral Nutr*. 2006;30:309–16.
- [71] Bennett JM, Reeves G, Billman GE, Sturmberg JP. Inflammation-nature’s way to efficiently respond to all types of challenges: implications for understanding and managing “the epidemic” of chronic diseases. *Front Med*. 2018;5:1–30.
- [72] Julian J. Faraway. *Extending the Linear Model with R*. 2^a edition. 2016. CRC Press. Taylor and Francis Group.
- [73] Análisis discriminante lineal (LDA) y análisis discriminante cuadrático (QDA) by Joaquín Amat Rodrigo, available under a Attribution 4.0 International (CC BY 4.0) at <https://www.cienciadedatos.net/documentos/28> linear discriminant analysis lda y quadratic discriminant analysis qda.

- [74] Max Kuhn and Kjell Johnson. Applied Predictive Modelling. Springer.
- [75] Raghav Bali et al. R: Unleash Machine Learning Techniques. Packt Publishing Ltd ISBN 978-1-78712-734-0.
- [76] Thomas Mailund. Beginning Data Science in R. Data Analysis, visualización and modelling for data scientist. Ed Apress. DOI 10.1007/978-1-4842-2671-1.
- [77] World Cancer Research Fund/American Institute of Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington, DC: AICR; 2007.
- [78] World Cancer Research Fund/American Society for Cancer Research. Policy and action for cancer prevention. Food, nutrition and physical activity: a global perspective. Washington, DC: AICR; 2009.

Supplementary material

Variables used to train the neural network

- **Cod**: Classification variable defining the state of health of the individual (pathological or not).
- **Sex**
- **Age**
- **Weight (kg)**
- **ACT**: Total Body Water
- **Fat**: Total fat
- **IMC (BMI)**: Body Mass Index.
The calculation was based on the Quetelet body mass index formula [19]:
$$\text{BMI} = \text{Weight (kg)} / \text{Height}^2(\text{m})$$
- **Visceral**: Visceral Fat
- **ImpedanceArmLeft**: Impedance Left Arm
- **ImpedanceArmRight**: Impedance Right Arm
- **ImpedanceLegLeft**: Impedance Left Leg
- **ImpedanceLegRight**: Impedance Right Leg
- **PhaseangleLeftLeg50**: Phase angle Left Leg 50 kHz
- **PhaseangleRightArm50**: Phase angle Right Arm 50 kHz
- **PhaseangleLeftArm50**: Phase angle Left Arm 50 kHz
- **PhaseangleLL50**: Phase angle LL 50 kHz
- **PhaseangleRLLL50**: Phase angle RLLL 50 kHz

Frequency of different types of tumors

<i>Neoplasm</i>	<i>No. Cases</i>
Breast	33
Thyroid	12
Uterus	11
Melanoma	8
Colon	5
Lymphoma	5
Lung	3
Pancreas	3
Ovary	2
Prostate	2
Bladder	1
Bone	1
Kidney	1
Testicle	1

Table 1

As can be seen in Table 1, the malignant tumors with the highest number of cases in this study were breast, thyroid and uterus.

Training phases of the different models

- Study and preparation of the data.
- Normalization of numerical variables.
- Creation of training and test data.
- Application of machine learning algorithms to the training data.
- Evaluation of the results of the different models.
- Prediction

Analysis of variance of the models

To determine the most efficient machine learning method, we should not only rely on the metrics, but it is necessary to analyze the variance to establish if there is a clear difference between the different models. In Table 2 we see that the difference in the results of the different models is significant at a significance level of $\alpha=0.05$.

Model A	Model A	p value
Boosting	Tree	1.158947e-01
KNN	Tree	6.302123e-01
KNN	Boosting	2.196969e-02
LDA	Tree	1.998722e-08
LDA	Boosting	1.998722e-08
LDA	KNN	4.122123e-08
Logistic	Tree	1.998722e-08
Logistic	Boosting	1.998722e-08
Logistic	KNN	4.890528e-08
Logistic	LDA	1.158947e-01
NNET	Tree	6.910109e-04
NNET	Boosting	1.225374e-06
NNET	KNN	1.019842e-02
NNET	LDA	1.998722e-08
NNET	Logistic	2.451687e-08
Rf	Tree	8.395390e-08
Rf	Boosting	8.243194e-08
Rf	KNN	4.078512e-08
Rf	LDA	1.581434e-08
Rf	Logistic	1.581434e-08
Rf	NNET	1.581434e-08

Table 2

To complete the analysis, we compared only the Random Forest (RF) model with the rest of the models, finding that the difference between them is still significant (Table 3).

Model A	Model B	p value
Rf	Tree	8.395390e-08
Rf	Boosting	8.243194e-08
Rf	KNN	4.078512e-08
Rf	LDA	1.581434e-08
Rf	Logistic	1.581434e-08
Rf	NNET	1.581434e-08

Table 3

Graphical representation of the exploratory analysis of the data

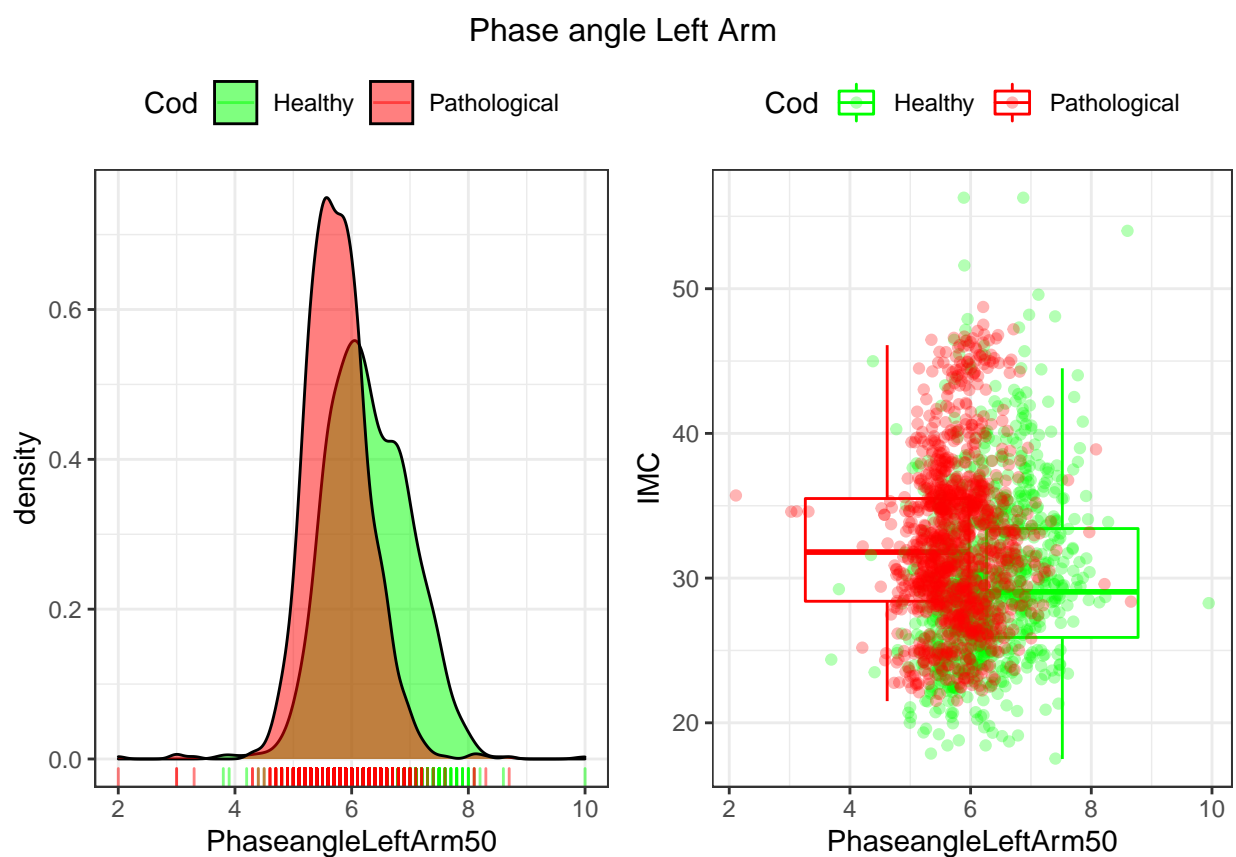


Figure 9: Density and IMC of the two groups based on the phase angle of the left arm with frequencies of 50 kHz.

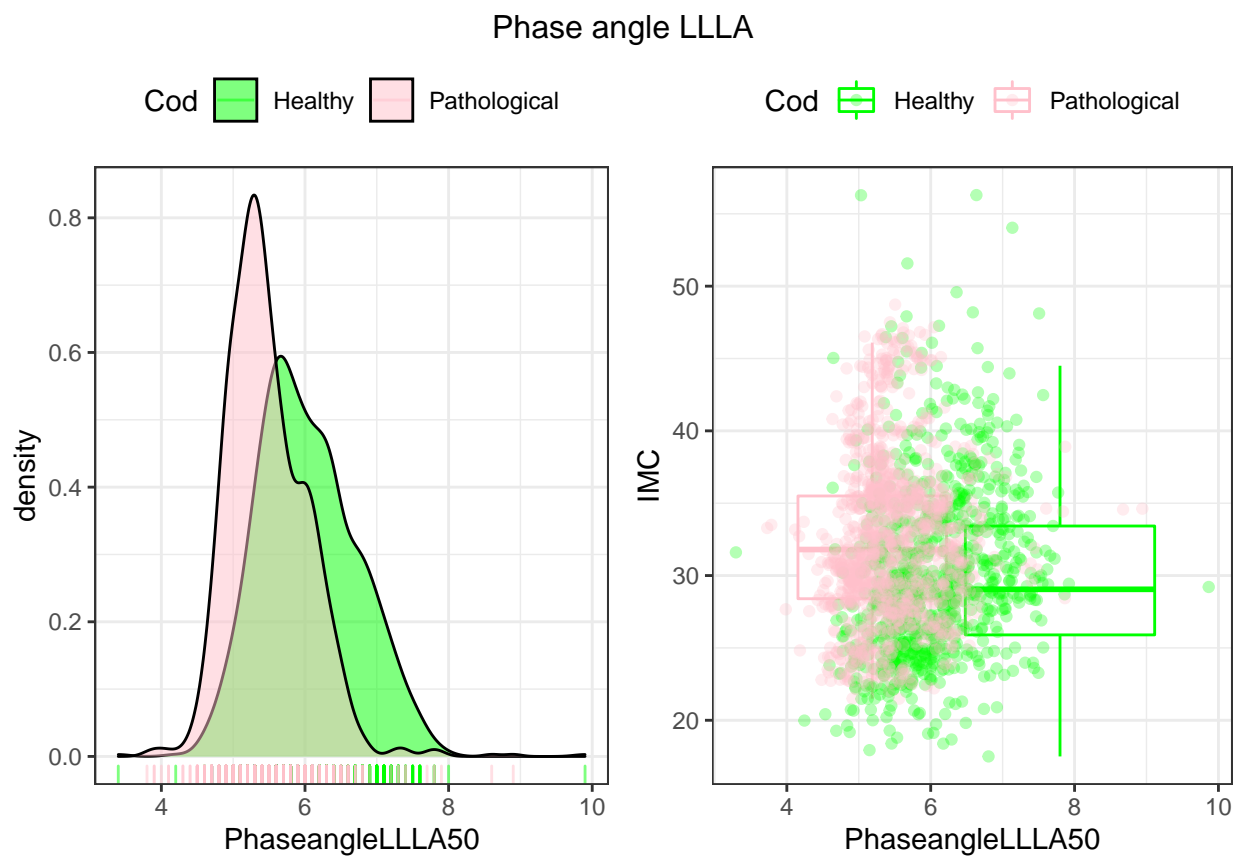


Figure 10: Density and IMC of the two groups based on the phase angle between the left leg and left arm with frequencies of 50 kHz.

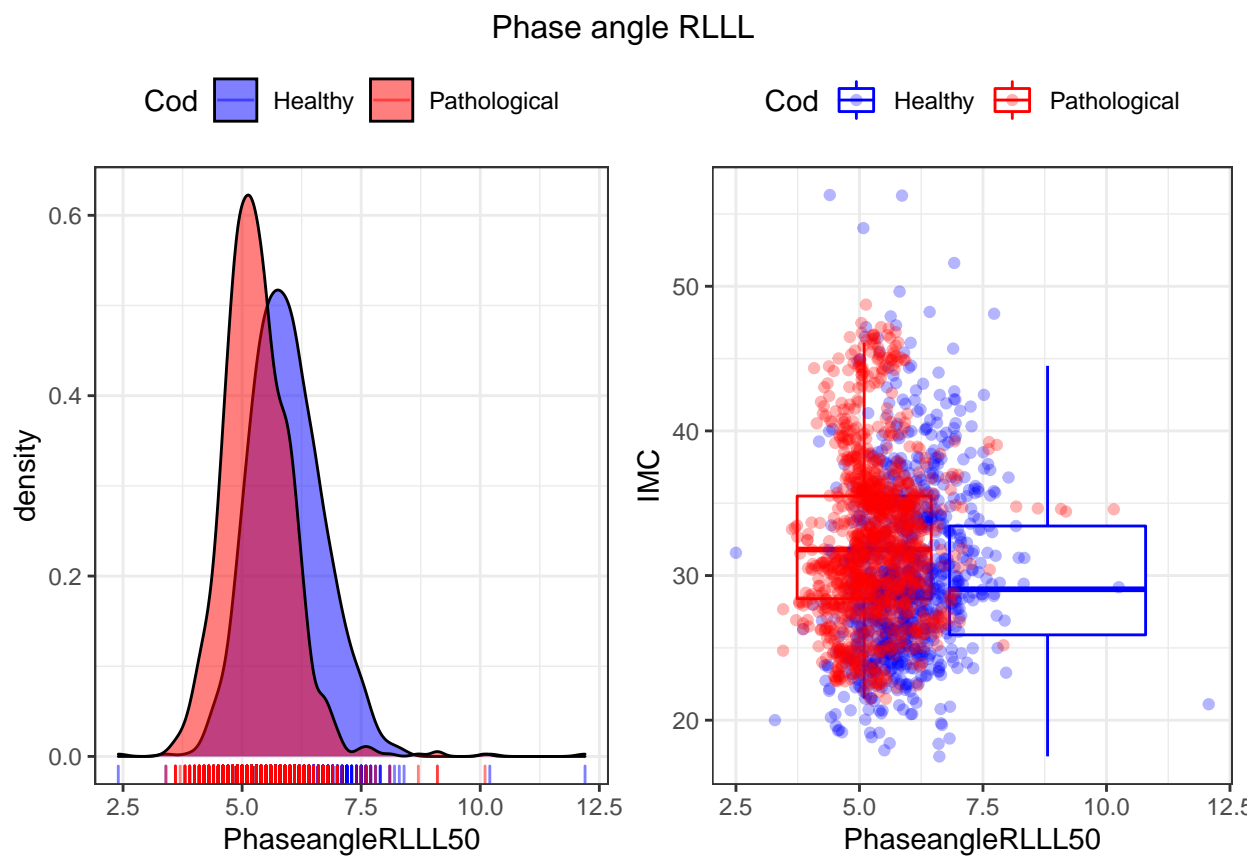


Figure 11: Density and IMC of the two groups based on the phase angle between the right leg and left leg with frequencies of 50 kHz.

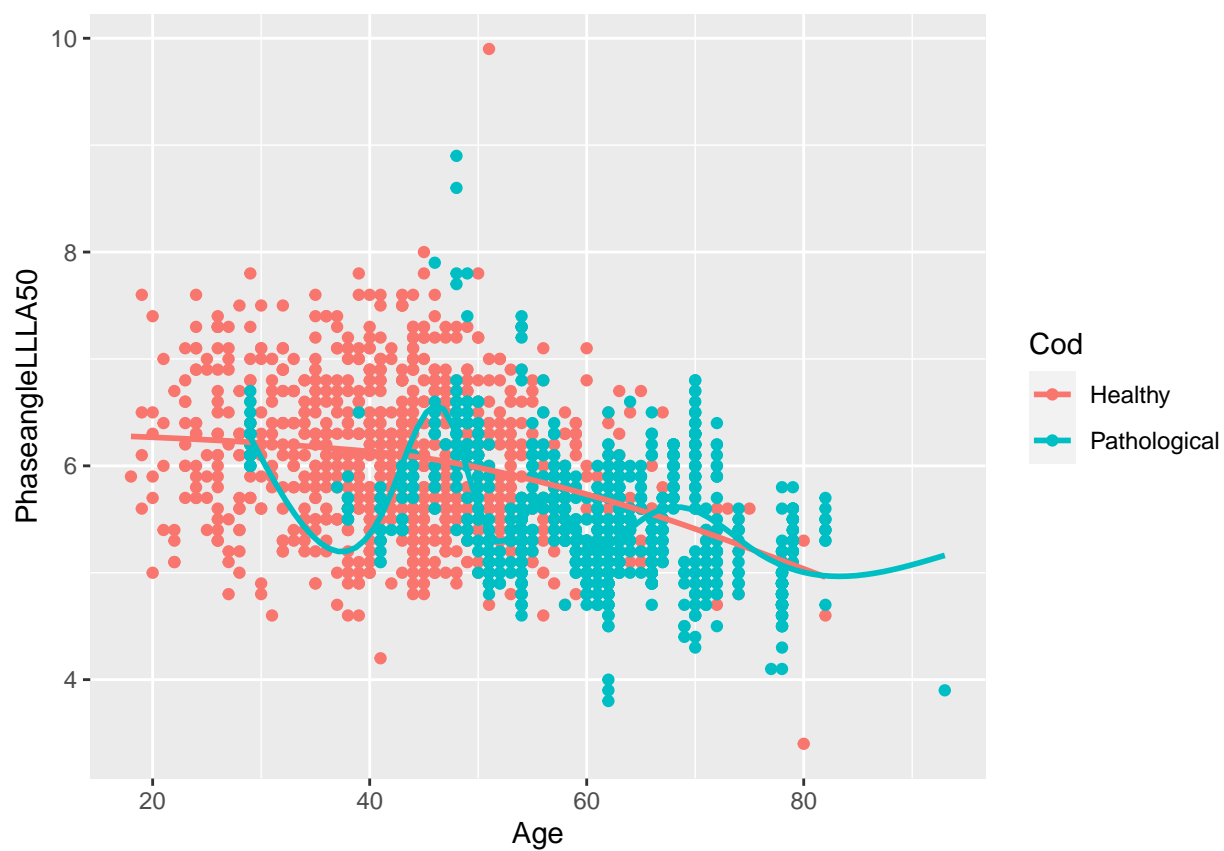


Figure 12: Graphical representation of the phase angle between left leg and left arm according to age in the different groups with frequencies of 50 kHz.



Figure 13: Graphical representation of the phase angle between right leg and left leg according to age in the different groups with frequencies of 50 kHz.