

Predictive classification and regression models for bioimpedance vector analysis: Insights from a southern Cuban cohort

Jose Luis García Bello^{1,9}, Taira Batista Luna², My Phuong Pham-Ho^{3,4}, Minh Tho Nguyen^{5,6}, Alcibiades Lara Lafargue⁷, Héctor Manuel Camué Ciria⁷ and Yohandys A. Zulueta^{8,9}

1. *Autonomous University of Santo Domingo (UASD), San Francisco de Macorís Campus, Dominican Republic.*
2. *Autonomous University of Santo Domingo (UASD), UASD Nagua Center, Dominican Republic.*
3. *Faculty of Chemical Engineering, Ho Chi Minh City University of Technology (HCMUT), District 10, Ho Chi Minh City, 70000 Vietnam*
4. *Vietnam National University Ho Chi Minh City, Linh Trung, Thu Duc, Ho Chi Minh City, 70000 Vietnam*
5. *Lab. for Chem. Comput. and Modelling, Institute for Comput. Science and Art. Intell., Van Lang University, Ho Chi Minh City, 70000 Vietnam*
6. *Faculty of Applied Technology, School of Technology, Van Lang University, Ho Chi Minh City, 70000 Vietnam*
7. *National Center for Applied Electromagnetism (CNEA), Universidad de Oriente CP 90500, Santiago de Cuba, Cuba.*
8. *Departamento de Física, Facultad de Ciencias Naturales y Exactas, Universidad de Oriente, CP 90500, Santiago de Cuba, Cuba*
9. *E-mail any correspondence to: yzulueta@uo.edu.cu, jgarcia44@uasd.edu.do*

Abstract

This study used predictive models to explore the link between bioparameters at characteristic frequency and their positions within tolerance ellipses in a southern Cuban cohort. The database includes 367 individuals (235 females, 132 males) aged 18–86. Among them, 61 had cancer, while 306 were healthy. After balancing the data, the analysis used 16 bioimpedance-based characteristics along with other anthropometric and location factors. The results showed that characteristic frequency bioparameters (Z_c , θ_c , X_{cc} , and R_c) are key for assessing health and location. There was a strong agreement between experimental and predicted values for Z_c , θ_c , X_{cc} , and R_c across various categories. Cancer patients showed higher Z_c and slightly lower θ_c and X_{cc} values, attributed to unbalanced body composition and cell membrane deterioration. Females exhibited higher Z_c and X_{cc} values, indicating better cell membrane integrity. Predictions are consistent across quartiles and percentiles, with lower θ_c observed in higher quartiles and centiles where more cancer patients are located. Variations in R_c values across different BIVA statuses demonstrated the model's robustness in estimating impedance parameters in diverse physiological conditions. These predictive models are significant for assigning locations without developing BIVA methods, enhancing clinical assessments and health monitoring.

Keywords: BIA; BIVA; characteristic frequency; characteristic reactance; characteristic resistance.

Introduction

Bioimpedance technology presents a versatile and non-invasive method for cancer detection and monitoring. Its use in detecting and tracking mammary, neck, skin, and breast cancers showcases its potential to enhance early diagnosis, monitor treatment efficacy, and improve patient outcomes. As research advances, bioimpedance could become a crucial component of personalized cancer care, equipping clinicians with essential tools for more accurate diagnosis and effective management [1-12].

Bioimpedance technologies, which measure the resistance and reactance of biological tissues to an applied electrical current, have demonstrated significant potential in oncology [1-13]. This non-invasive method offers valuable insights into tissue composition and physiological changes, making it a promising tool for cancer detection and monitoring [1-3]. For instance, mammary cancer, commonly known as breast cancer, is one of the most prevalent cancers among women [13-16]. Early detection and accurate monitoring are essential for improving patient outcomes [17-19]. Recent studies have explored the application of bioimpedance in various types of cancer, including mammary, neck, skin, and breast cancers [13-16].

Bioimpedance techniques, including single-frequency (SF-BIA) and multi-frequency (MF-BIA) bioimpedance, are used to measure physiological parameters. SF-BIA, which utilizes a single frequency (50 kHz) of electrical current, is a simple and cost-effective method often employed in healthcare and fitness settings [20-24]. However, it may not accurately measure extracellular and intracellular water content due to varying electrical properties of tissues and fluids at different frequencies [20-24]. MF-BIA and impedance spectroscopy (SBIA) uses multiple frequencies disclosing more detailed information on the cellular and molecular properties of tissues and fluids [20-24]. Another bioimpedance method is the BIA vector analysis (BIVA). BIVA is a non-invasive method that evaluates body composition and hydration status by measuring electrical impedance at a low frequency. The BIVA method standardizes Bioelectrical Impedance Analysis (BIA) measurements by height and represents them as bivariate vectors with confidence intervals, depicted as ellipses on the R-Xc plane. This technique's key advantage is its ability to provide simultaneous information about changes in tissue hydration and soft-tissue mass, independent of regression equations or body weight. BIVA has been extensively utilized to study hydration across various diseases [20-28] and to conduct general body composition assessments in patients with lung cancer [27, 29] and head and neck cancers [25-28].

Studies have demonstrated that BIVA can effectively differentiate between cancer patients and healthy individuals by analysing parameters such as impedance, phase angle, and reactance [25-28]. For instance, cancer patients often exhibit lower phase angles and higher impedance values, indicating altered cellular integrity and body composition due to the disease. In addition, BIVA has been utilized to assess hydration status in advanced cancer patients, revealing significant associations between hydration levels and clinical outcomes. Research has shown that lower hydration status is linked to increased symptom intensity and shorter survival times in cancer patients [25,29,30]. The integration of BIVA into routine assessments highlights its potential as a valuable tool in oncology and general health monitoring [25,29,30].

In a previous work, the integration of various bioimpedance modalities was used to evaluate the health status of southern Cuban populations. We demonstrated that this integrated methodology serves as a sensitive complementary tool, adept at distinguishing between healthy individuals and cancer patients. In addition, it confirms that the phase angle value at the characteristic frequency may be a robust indicator of overall health status of individuals, similar to the reported at 50 kHz [31].

Determining the location of individuals within the tolerance ellipses derived by BIVA is crucial for accurately assessing their health and nutritional status [25-31]. Location variables contemplate the BIVA status, quartile and centile. These ellipses provide a graphical representation of an individual's impedance vector relative to a reference

population [26-29]. By pinpointing where an impedance vector falls within these ellipses, clinicians can gain valuable insights into their body composition, hydration levels, and cellular health [26-29]. For example, individuals positioned within the upper regions of the ellipses typically exhibit better cell membrane integrity and hydration status, while those in the lower regions may indicate dehydration or cellular function [25-30]. Moreover, the location within the tolerance ellipses can help identify early signs of malnutrition, overhydration, or other health issues that may not be immediately apparent through conventional assessments [26-30]. This proactive approach is particularly beneficial for managing chronic conditions, such as heart failure, renal disease, and cancer. The ability to determine an individual's location within the BIVA tolerance ellipses enhances the precision and effectiveness of health assessments, leading to better-informed clinical decisions and improved patient outcomes [25-31].

Interest in artificial intelligence (AI) has grown significantly in recent years, driven by continuous advancements in computer science. AI has proven particularly useful in data management and various scientific fields, including medicine [32-36]. One area where AI holds great potential is diagnostic imaging, as it can aid in identifying various conditions, including body composition [32,33]. This capability may further contribute to predicting health outcomes across different diseases [32,36]. For instance, a combination of bioimpedance measurements and machine learning analysis was conducted in an infant-juvenile cohort from the eastern Cuban region [34,35]. The classification model demonstrated that, aside from body mass index, alternative indicators can serve as predictors of weight status [34,35]. Additionally, the regression learner model accurately predicted the weight status of the subjects with high precision [34,35].

Recently, a cancer predictive model was developed using bioimpedance measurements and machine learning methods [36]. The results revealed that the classification model identified two robust parameters for predicting health status: impedance, total body water, and phase angle, which showed significant relevance [36]. The phase angle predictions align with previous reports on other pathologies, indicating that higher phase angle values are associated with better health status. Furthermore, males tend to have higher phase angle values compared to females [36].

To our knowledge, there is no reports combining machine learning with BIVA studies for prediction of location classes. In the present work, we employed predictive classification and regression learner models to investigate the association between bioparameters of subjects, derived at the characteristic frequency, and their locations within tolerance ellipses for a cohort from the southern Cuban region. The model developed in this study plays a crucial role in location assignment without relying on BIVA methods, improving clinical evaluations and health monitoring.

Materials and Methods

Population recruitment and ethical approval

A descriptive, retrospective, randomized study was conducted on patients who visited the "Conrado Benítez" Teaching Oncology Hospital in Santiago de Cuba for suspected cancer during the periods of March–April and July–August 2002. The hospital was chosen for its suitable facilities and accommodations, despite participants hailing from various locations across eastern Cuba. The study comprised 367 subjects (235 females and 132 males) aged 18 to 86 years. Among them, 61 were diagnosed with different types of cancer pathologically, while 306 were healthy. Participants were recruited following strict ethical guidelines and medical practices as outlined by the Health General Law of the Ministry of Public Health of the Republic of Cuba (Number 41, July 13, 1983, updated in 2010).

Following the Helsinki Declaration, the research was approved by the ethics committees and scientific councils of the Oncological Hospital "Conrado Benítez." Additionally, data on bioelectrical parameters of cancer patients and healthy individuals were retrieved from a database archived at the National Center for Applied Electromagnetism (CNEA); (ISBN: 978-959-207-679-2). The diagnosis of the patients was made through pathological anatomy. The samples were taken using the fine needle aspiration cytology technique. Anatomical pathology laboratory diagnoses revealed that female patients had various types of cancer with different stages, including breast and cervical cancer, whereas male patients were diagnosed with more aggressive cancers, such as skin melanoma, colon cancer, lung neoplasm. Further details can be found in Ref [11].

Informed consent

Informed consent has been obtained from all individuals included in this study.

Ethical approval

The research related to human use has been complied with all relevant national regulations, institutional policies and in accordance with the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

Machine learning models

Achieving balanced data is essential for successful machine learning studies [32-41]. In our study, we ensured data balance by implementing random oversampling for the minority class (cancer) and random undersampling for the majority class (healthy). The final dataset included 621 individuals, aged between 18 and 86 years (382 females and 239 males), with 316 diagnosed with cancer and 306 healthy participants. To prevent overfitting, a cross-validation technique is used, in which 95% of the data is designated for training and the remaining 5% is reserved for validation. This approach ensures

that the model generalizes well to unseen data rather than memorizing patterns specific to the training set [37-41].

For the machine learning study, the features include: health status (cancer, healthy), sex, high frequency resistance (R_{inf}), characteristic frequency (f_c), Cole parameter (α), corrected resistance (r_c), capacitive reactance X_c , height, weight, age, resistance (R_c), capacitive reactance (X_{cc}), impedance absolute value $Z_c = \sqrt{R_c^2 + X_{cc}^2}$ and phase angle (θ_c) at the characteristic frequency and location variables (BIVA status, quartile and centile). For the classification models, the response are the location features and the health status, while for the regression, the bioelectrical parameters at the characteristic frequency (R_c , X_{cc} , θ_c and Z_c). In this sense, one can explore the role of the anthropometric and bioelectric parameters for location assignments at the tolerance ellipses, usually made by bioimpedance vector analysis.

Several metrics can be used to assess the accuracy of machine learning models [39-41]. Metrics for classification models are derived from the confusion matrix [39-41]. For classification models, the accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

where (TP) is true positives, (TN) is true negatives, (FP) is false positives, and (FN) is false negatives.

The precision, which indicates the amount of the predicted positive instances that are actually positive is defined as follows [39,40]:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

The recall describes how well the model captures all the positive instances [39,40]:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score is defined as [39,40]:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

which estimate the balance between Precision and Recall.

In addition, for regression models the common metrics are: the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (MSE) and its root mean square error (RMSE) defined by equations (5) [39,40]:

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ MAE &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ MSE &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ RMSE &= \sqrt{MSE} \end{aligned} \quad (5)$$

where n represents the number of elements in the database, y_i the observed value of the dependent variable, \hat{y}_i the predicted value of the dependent variable, and \bar{y}_i the mean value of the observations. When $R^2 \cong 1$ the model tends to be perfect, analogously for MAE, MSE and RMSE values [39,40].

Bioimpedance Measurements

Bioimpedance parameters were measured using a BioScan 98® model bioimpedance analyzer (Biológica Tecnología Médica S.L., Barcelona, Spain) with a tetrapolar whole-body configuration. Participants, who fasted for at least 3 hours, emptied their bladders, and refrained from exercise and alcohol for 12 hours prior, were included in the study. MF-BIA measurements were taken at frequencies ranging from 10 to 250 kHz using Ag/AgCl electrodes model 3M Red Dot 2560 (3M, Ontario, Canada). The study was conducted in a room maintained at 23°C with 60-65% relative humidity. Volunteers lay on a non-conductive surface without clothing or pillows, with their arms positioned 30° away from the chest and legs spread 45° apart. The injector electrodes were placed (after cleaning the skin with 70% alcohol) on the inner side of the dorsal surfaces of the hands and feet, near the metatarsophalangeal and third metacarpal joints. The detector electrodes were positioned between the distal ends of the ulna and radius at the pisiform prominence and at the midpoint between both malleoli. A 5 cm gap was maintained between detector and injector electrodes during measurements.

Results and Discussion

Estimation of subject location on the tolerance ellipse with the aid of classification models

With the balanced database, we perform classification models to predict the location features. Table 1 collects the best model and the metrics describing the model performance and generality of each location feature. The Fine Tree model is the most effective in describing health and BIVA statuses, with accuracies of 92.80% and 99.50%, respectively. Meanwhile, the Linear Support Vector Machine and Random under-sampling Boosted Tree approach (RUS Boosted Tree) are the top models for describing quartile and centile responses, achieving accuracies of 100% and 97%, respectively.

The confusion matrices presented in Figure 1 illustrate the model's performance across various classification tasks. For health status (Figure 1a), the matrix reveals a high true positive rate (TPR) for healthy individuals, with 96.9% correctly identified and only 3.1% misclassified. The model shows excellent accuracy in distinguishing between healthy individuals and those with health issues, demonstrated by the 100% correct classification of the cancer (C) class. For BIVA status (Figure 1b), which encompasses 44 classes, most achieved 100% classification accuracy. Minor misclassifications were observed between classes 33 and 34, with a small percentage (4.5%) incorrectly classified.

Table 1: Response models and their respective metrics for the classification of health status and location variables. The column Class includes Health Status (Cancer, Healthy), Quartile (1, 2, 3, 4), Centile (50, 75, 95, and 100%), and BIVA status (11, 12, 13, 14, 21, 22, 23, 31, 32, 33, 41, 42, 43, 44).

Response	Model	Accuracy	Class	Precision	Recall	F1-score
Health status	Fine Tree	92.80%	Cancer	0.912	0.951	0.931
			Healthy	0.947	0.905	0.926
BIVA status	Fine Tree	99.50%	11	1.000	1.000	1.000
			12	1.000	1.000	1.000
			13	1.000	1.000	1.000
			14	1.000	1.000	1.000
			21	1.000	1.000	1.000
			22	1.000	1.000	1.000
			23	0.857	0.750	0.800
			31	1.000	1.000	1.000
			32	1.000	1.000	1.000
			33	0.889	0.941	0.914
			41	1.000	1.000	1.000
Quartile	Linear SVM	100%	2	1.000	0.983	0.991
			3	0.990	1.000	0.995
			4	1.000	1.000	1.000
			50%	0.988	0.984	0.986
Centile	RUS Boosted Tree	97%	75%	0.959	0.953	0.956
			95%	0.948	0.938	0.943
			100%	0.957	1.000	0.978

The overall TPR of 99.5% highlights the model's high precision in identifying BIVA statuses correctly. The Quadrant Responses matrix (Figure 1c) displays perfect classification, achieving 100% accuracy across all four classes, indicating an exceptional ability to differentiate quadrant responses without any errors. For Centile Responses (Figure 1d), the matrix shows slightly lower but still high performance, with true classes 50%, 75%, and 95% having some minor misclassifications. Notably, class 100% is perfectly classified. The overall TPR of 95.3% demonstrates robust model performance, though there is some room for improvement in distinguishing closely related centiles. Collectively, these matrices underscore the models' effectiveness in accurately classifying various health statuses, BIVA parameters, quadrant responses, and centile categories, with high true positive rates and minimal misclassifications.

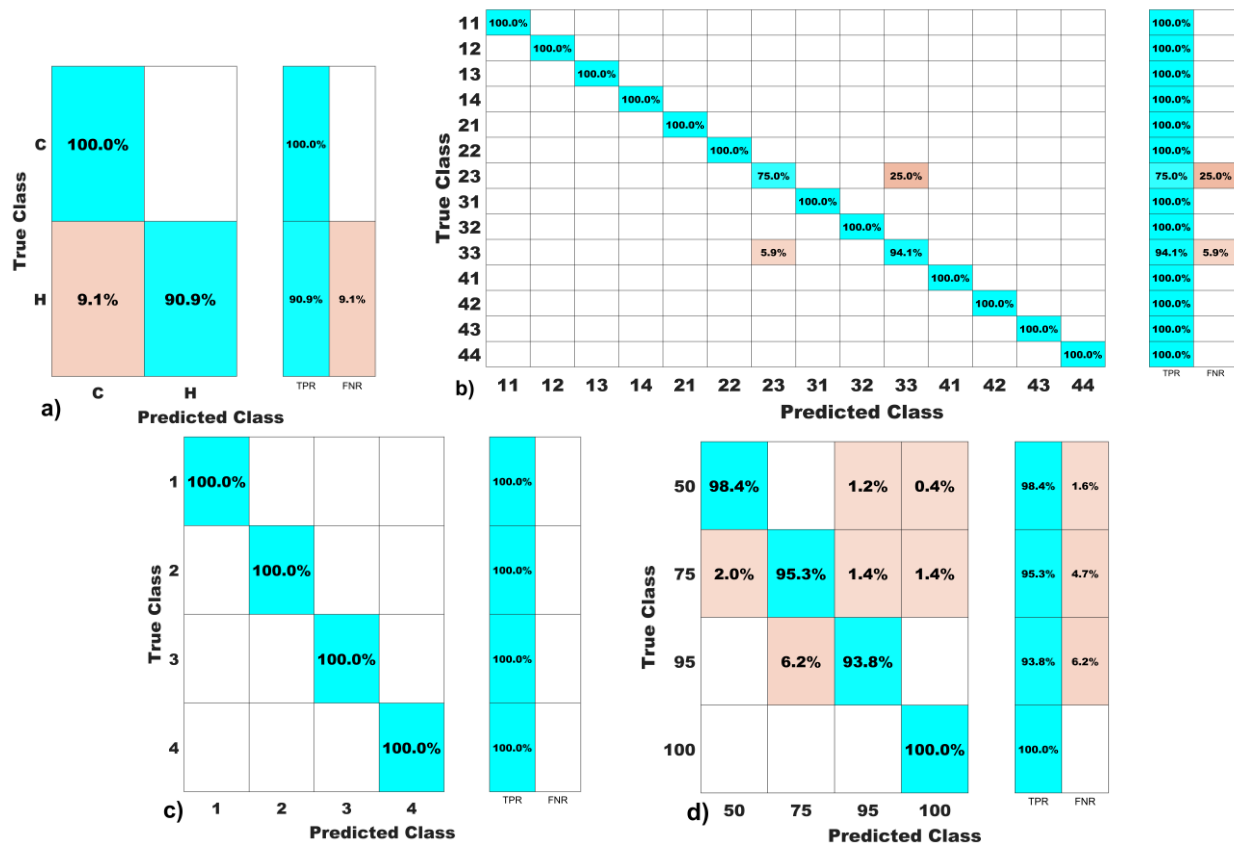


Fig.1: Confusion matrix of trained models: a) Health status, b) BIVA status, c) quartile and d) centile responses.

Feature importance analysis using a Pareto chart is a powerful method for identifying and visualizing the most significant predictors in a classification model [37-39]. By ranking features according to their importance scores and displaying them in a Pareto chart, one can easily discern the few key features that contribute most to the model's predictive power [37-39]. Applied to feature importance, this means that a small number of features typically account for the majority of the model's performance. In the Pareto chart, features are ordered from highest to lowest importance, with a cumulative line illustrating their collective contribution. This visual representation aids in prioritizing features for model improvement, simplifying complex datasets, and focusing efforts on the most impactful predictors [37-39].

Figure 2 displays the feature importance of each response. For health status model (Figure 2a), it can be noted that the most influencing feature is the age with a 95% of importance score, followed by α , f_c , BIVA status (BIVA), R_{inf} , θ_c , X_c , weight and R_c , in descending order. These findings align with the natural progression of life, influenced by biological changes, environmental factors, and lifestyle habits that accumulate over time. Cole parameter (α) and f_c with a $\sim 18\%$ and $\sim 8\%$ of importance, respectively, also align with life evolution. Both parameters undergo logical variations in consequence of biological changes. In the case of BIVA status response (Figure 2b), there are two top ranking features: quartile ($\sim 97\%$) and centile ($\sim 31\%$). The quartile (Figure 2c) is highly influenced by the BIVA status ($\sim 96\%$) and centile response (Figure 2d) by the

BIVA ($\sim 96\%$), X_{cc} ($\sim 5\%$) and R_{inf} ($\sim 2\%$) bioparameters. These findings align well with the BIA vector analysis location in the tolerance ellipse, where the quartile determine the body composition and hydration status and the centile offers insights into cell integrity, muscle mass and fluid balance [25-31].

The phase angle is a well-established marker for assessing cell membrane health and integrity, bearing significant implications for cancer prognosis [41-44]. A comprehensive study with 2625 participants demonstrated a positive and substantial correlation between phase angle and cancer survival rates. Specifically, patients exhibiting low phase angle values were found to be 23% less likely to survive compared to those with higher values [41]. Additionally, phase angle at the time of diagnosis is recognized as a crucial prognostic factor for survival among patients with advanced head and neck cancer [42]. Previous research has indicated that phase angle shows a moderate to strong correlation with body composition and physical function, while its correlation with nutritional status, complications, survival, quality of life, and symptoms tends to be weaker. These findings underscore the utility of phase angle as a robust indicator in clinical assessments, particularly in the context of cancer prognosis [43,44].

In addition, phase angle, X_c and R determine the location in the tolerance ellipses in BIVA studies [25-27]. By analysing the position of an individual's impedance vector within these ellipses, researchers can assess hydration status, cell membrane integrity, and overall cellular health.

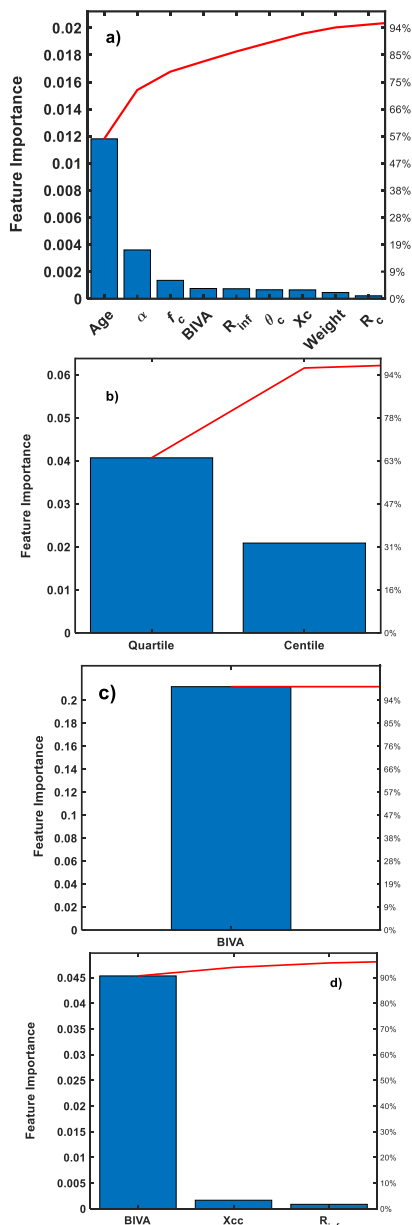


Fig.2: Feature importance of trained models: a) Health status, b) BIVA Status, c) quartile and d) centile responses.

Previous reports used single-frequency bioimpedance for BIVA studies. In our previous work, we demonstrated the use of impedance spectroscopy combined with BIVA for accessing the health status by constructing the tolerance ellipses at the characteristic frequency instead of the common 50 kHz [31]. In the present study, we found that the bioparameters derived at the characteristic frequency play a crucial role determining the health status and the location parameters.

The role of bioelectrical parameters at the characteristic frequency estimating the location on the tolerance ellipse

This section focuses on developing regression models to predict Z_c , R_c , X_{cc} , and θ_c in order to assess their robustness in determining the location on the tolerance ellipses. Table 2 collects the results of the model for each response and their respective accuracy values. The response vs. predicted plots in Figure 3 provides a visual assessment of the models' performance in predicting Z_c , R_c , X_{cc} , and θ_c . For Z_c , the top-

left plot indicates that the model predictions are closely aligned with the actual values, as most data points fall near the line of perfect prediction. This high level of accuracy suggests that the model is highly reliable in predicting impedance values. The characteristic phase angle (θ_c), shows a similarly strong agreement between predicted and actual values. The close alignment of data points with the line of perfect prediction underscores the model's effectiveness in estimating phase angle, a critical indicator of cell membrane health and integrity [25-30]. In the bottom-left plot, which compares true and predicted values for reactance X_{cc} , the data points again closely follow the line of perfect prediction. This suggests that the model is adept at estimating reactance, an important measure for assessing cellular health and hydration status [25-31]. In the case of the characteristic resistance (R_c), shows that the model's predictions are also highly accurate. The data points close alignment with the perfect prediction line indicates that the model can reliably predict resistance values, essential for evaluating total body water and extracellular water content [25-30].

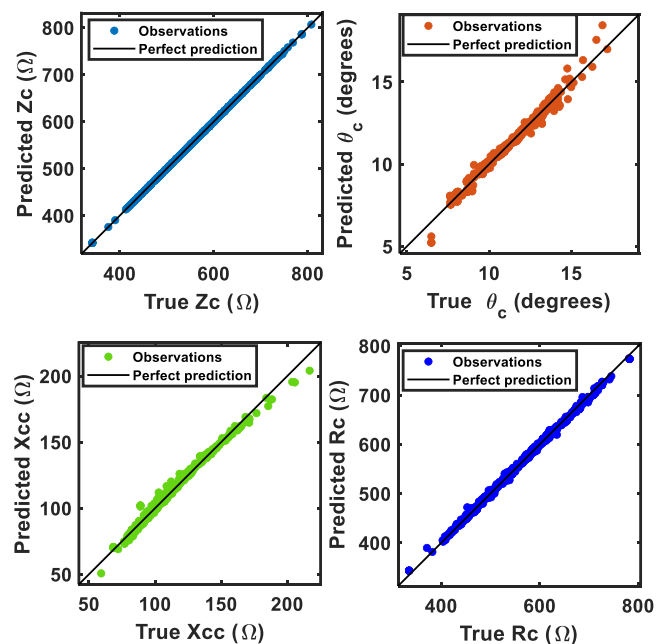


Fig.3: Response vs predicted plot of the selected responses.

The accuracy parameters listed in Table 2 further corroborate the visual findings from Figure 3. The Fine Tree model for health and BIVA statuses demonstrates an R^2 of 1.00 for characteristic impedance and resistance, indicating perfect prediction. The RMSE, MSE, and MAE values are very low for these models, further emphasizing their high accuracy. The phase angle has an $R^2 = 0.98$, with similarly low RMSE, MSE, and MAE values, highlighting its robust prediction accuracy.

For the characteristic reactance, the Linear SVM model shows an $R^2 = 0.99$, indicating slightly lower but still high accuracy. The RMSE, MSE, and MAE values reflect this slight decrease in precision but still demonstrate the strong performance.

Table 2: Accuracy parameters of each model.

Response	Model	R ²	RMSE	MSE	MAE
Z _c (Ω)	Linear	1.000	0.351(Ω ²)	0.123 (Ω)	0.233 (Ω)
θ _c (°)	Linear	0.980	0.237 (° ²)	0.056 (°)	0.166 (°)
X _{cc} (Ω)	Linear SVM	0.990	2.378 (Ω ²)	5.652 (Ω)	1.674 (Ω)
R _c (Ω)	Linear SVM	1.000	4.216 (Ω ²)	17.773 (Ω)	3.223 (Ω)

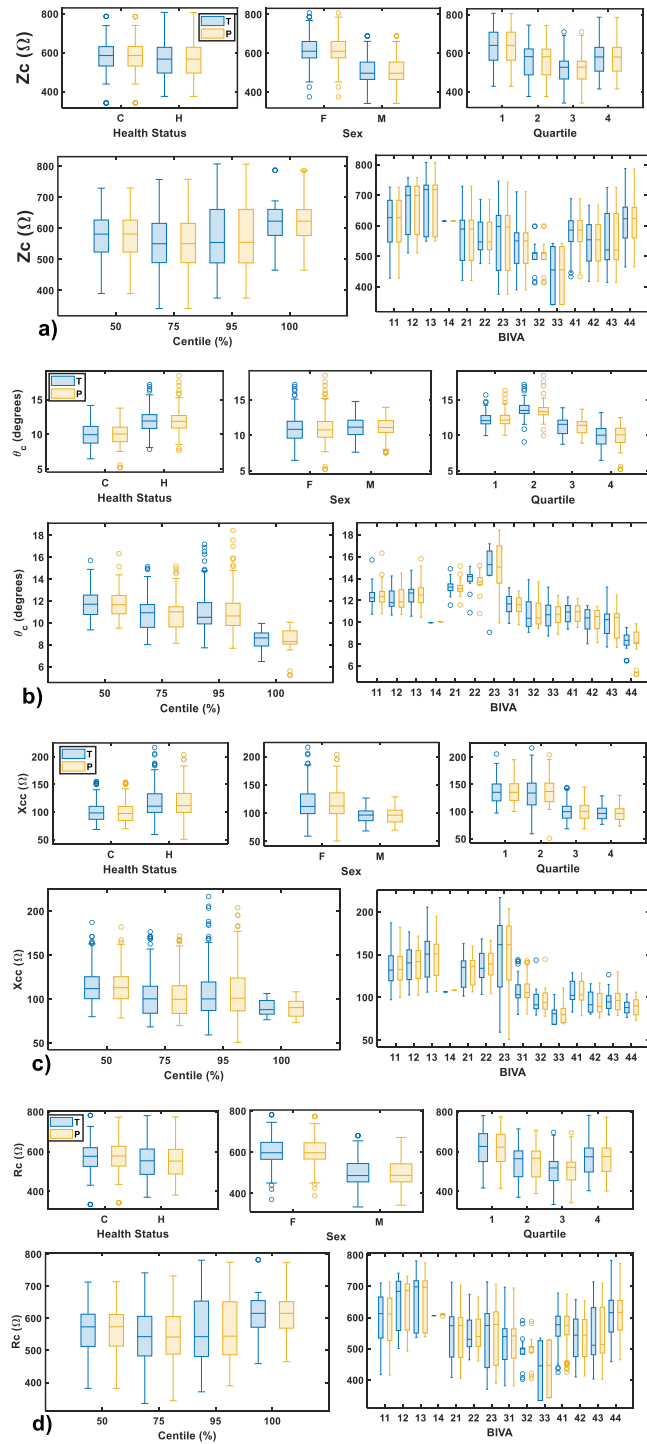


Fig.4: observable and predictions of Z_c, θ_c, X_{cc} and R_c across various categories: Health Status (Cancer, Healthy), Sex, Quartile (1, 2, 3, 4), Centile (50, 75,95 and 100%), and BIVA status (11, 12, 13, 14, 21, 22, 23, 31, 32, 33, 41, 42, 43, 44).

Figure 4 displays the observable and predictions of Z_c, θ_c, X_{cc} and R_c across various categories: Health Status (Cancer, Healthy), Sex, Quartile (1, 2, 3, 4), Centile (50, 75,95 and 100%), and BIVA status (11, 12, 13, 14, 21, 22, 23, 31, 32, 33, 41, 42, 43, 44).

The box plots in Figure 4 illustrate the close alignment between the experimental and predicted values of Z_c, θ_c, X_{cc} and R_c. The blue box plots represent experimental data, while the yellow box plots depict the predicted values. From Figure 4, a strong agreement is observed between the experimental and predicted values across all categories, suggesting that the model effectively captures the characteristic electrical parameters, with high accuracy. For instance, the median value of Z_c of observable cancer patients is 585.9 Ω and 567.3 Ω for healthy individuals, and the predicted values are 585.8 Ω and 567.6 Ω for cancer and healthy subjects (Figure 4a). It can be attributed that cancer patients have more unbalanced body composition due to the proper evolution of the pathology [25,29,30]. By sex, female have higher median values of 610.5 Ω as compared with male individuals 497.2 Ω with similar predicted values. These findings can be attributed to the fact that woman accumulates more adipose tissues than man in general. Concerning the characteristic phase angle (Figure 4b), healthy subjects (11.94 degrees) has higher θ_c than cancer patients (9.96 degrees), in agreements with previous reports [25,29,30]. In terms of sex, the predicted θ_c values for both female and male participants are also highly consistent with the experimental data, indicating robust performance regardless of gender. Larger characteristic phase angle is observed in male subjects, in line with previous works [25,29,30,31]. The quartile categories show a similar level of accuracy. Lower θ_c values are found in subject located at quartile 3 and 4, where the majority of the cancer patients are located. For centile categories (50, 75, 95 and 100%), the effectiveness in predicting phase angles across different percentiles is demonstrated, subjects with lowest θ_c are located at 100% and the higher at 50% centiles. The BIVA status categories highlights the model's ability to predict θ_c across different BIVA statuses. In addition, from 31 to 44 BIVA statuses the characteristic phase angle decrease gradually, reaching its lower value at 44.

From Figure 4c, again, lower characteristic reactance is reasonably found for cancer patients, in accordance with cell membrane deterioration reported [25-31]. By sex, female subjects having higher X_{cc} indicates better cell membrane integrity and overall well-being [25-30]. Quartiles 3 and 4 have lower X_{cc}, where the majority of the cancer patients are located. Higher X_{cc} is encountered for subjects located at 50% centile, while 75% and 95% have similar values and at 100% centile has the lowest X_{cc}, at 50-100% centiles the cancer patients are located. Similar behaviour is found for the BIVA statuses. From Figure 4d, the characteristic resistance in cancer patients is slightly higher compared to that of the healthy individuals, which can be attributed to unbalanced

body composition in cancer patients. Notably, lower R_c values are observed in male individuals, those in the third quartile, and within the 75-95% centile range. Additionally, R_c exhibits variations across different BIVA statuses.

In order to evaluate the model's performance, a schematic representation of BIVA is presented in Figure 5. The maximum and minimum values of Z_c , θ_c , X_{cc} and R_c (relative to their respective medians) are located on the tolerance ellipses. As shown in the figure, the model identifies lean individuals with maximum Z_c (Z_c^{max}) and R_c (R_c^{max}) at quartile 1, centile 100%, indicating less cellular mass and structure. Additionally, athletic individuals with maximum θ_c (θ_c^{max}) and X_{cc} (X_{cc}^{max}) are located at quartile 2, centile 50%, reflecting balanced hydration and cellular mass. The model predicts that obese subjects have minimum Z_c (Z_c^{min}), R_c (R_c^{min}) and X_{cc} (X_{cc}^{min}) located at quartile 3 between the 75th and 95th percentiles. Meanwhile, cachectic subjects are located at quartile 4, centile 100%, with lower θ_c (θ_c^{min}) and X_{cc} (X_{cc}^{min}).

Previous studies have reported that individuals with pathologies associated with hydration status are located in quartile 1, while athletes or people who exercise regularly are in quartile 2. The model can also assign the location of healthy individuals between quartiles 1 and 2, centile 50%. Additionally, obese subjects, hemodialysis patients, and individuals with chronic renal failure undergoing hemodialysis are found in quartile 3. Patients diagnosed with cancer, anorexia nervosa, human immunodeficiency virus (in various stages), COVID-19, etc., are located in quartile 4 [2,29]. Our predictive model aligns with these previous studies. Thus, the model developed in this work is significant for assigning locations without the need for BIVA methods, enhancing clinical assessments and health monitoring.

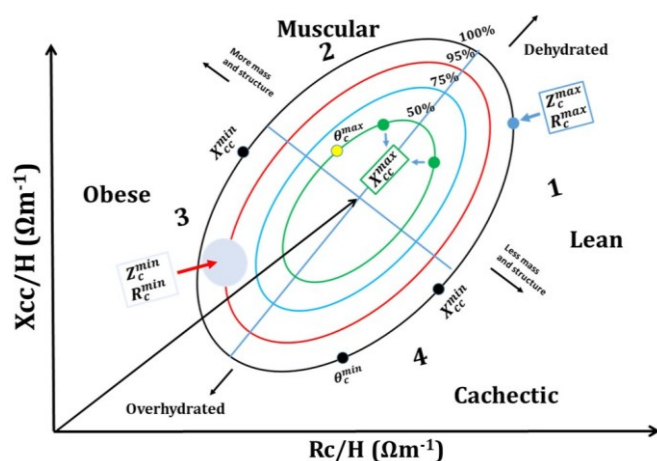


Fig.5: Schematic representation of BIA vector analysis (BIVA), presenting the maximum and minimum values of Z_c , θ_c , X_{cc} and R_c (relative to their respective medians).

Conclusion

In this work, a predictive classification and regression learner models is used to study the association between bioparameters at the characteristic frequency with the location at the tolerance ellipses of a cohort from the southern Cuban region. We used 16 characteristics (features) derived from bioimpedance measurements, including other physical parameters. The classification model shows that the bioparameters derived at the characteristic frequency play a crucial role determining the health status and the location parameters. From the regression models, we found a strong agreement between experimental and predicted values for Z_c , θ_c , X_{cc} and R_c (resistance) across various categories, including health status, sex, quartile, centile, and BIVA status. The model accuracy is proved, with minimal deviation between predicted and experimental values, indicating its effectiveness in capturing these characteristic electrical parameters. Cancer patients exhibit higher Z_c and slightly lower θ_c and X_{cc} values due to unbalanced body composition and cell membrane deterioration. Females generally show higher Z_c and X_{cc} values, indicating better cell membrane integrity compared to males. The predictions are consistent across quartiles and percentiles, with lower θ_c observed in higher quartiles and centiles where more cancer patients are located. Lastly, variations in R_c values across different BIVA statuses highlight the robustness in estimating impedance parameters, derived at the characteristic frequency, in diverse physiological conditions. The predictive models developed in this study are highly valuable for location assignment without the need to implement bioimpedance vector analysis methods.

While the predictive models presented in this study demonstrate strong agreement between experimental and predicted values, certain limitations must be acknowledged. The dataset is specific to a cohort from the southern Cuban region, which may limit the generalizability of the findings to other populations with different physiological or environmental conditions. For future directions, expanding the dataset to include diverse ethnic, geographic, and physiological groups could enhance the robustness of the models across populations.

Acknowledgement

The authors would like to thank the Ministry of Higher Education, Science and Technology (MESCYT for its initials in Spanish) of the Dominican Republic and the 2020-2A2-102 project for their invaluable assistance in the completion of this work. The authors also express our gratitude for the volunteer's participation and contribution to the research.

Conflict of interest

Authors state no conflict of interest.

References

- Alves EAS, Salazar TC do N, Silvino VO, Cardoso GA, dos Santos MAP. Association between phase angle and adverse clinical outcomes in hospitalized patients with COVID-19: A systematic review. *Nutr Clin Pract*. 2022;37:1105-1116. <https://doi.org/10.1002/ncp.10901>.
- Bellido D, García-García C, Talluri A, Lukaski HC, García-Almeida JM. Future lines of research on phase angle: Strengths and limitations. *Rev Endocr Metab Disord*. 2023;24:563–83. <https://doi.org/10.1007/s11154-023-09803-7>.
- Peng Z, Xu D, Li Y, Peng Y, Liu X. Phase Angle as a Comprehensive Tool for Nutritional Monitoring and Management in Patients with Crohn's Disease. *Nutrients*. 2022;14:2260. <https://doi.org/10.3390/nu14112260>.
- Stupin DD, Kuzina EA, Abelit AA, Emelyanov AK, Nikolaev DM, Ryazantsev MN, et al. Bioimpedance Spectroscopy: Basics and Applications. *ACS Biomater Sci Eng*. 2021;7:1962-1986. <https://doi.org/10.1021/acsbiomaterials.0c01570>.
- Pérez-Morales R, Donate-Correa J, Martín-Núñez E, Pérez-Delgado N, Ferri C, López-Montes A, et al. Extracellular water/total body water ratio as predictor of mortality in hemodialysis patients. *Ren Fail*. 2021;43:821-829. <https://doi.org/10.1080/0886022X.2021.1922442>.
- Campa F, Gobbo LA, Stagi S, Cyrino LT, Toselli S, Marini E, et al. Bioelectrical impedance analysis versus reference methods in the assessment of body composition in athletes. *Eur J Appl Physiol*. 2022;122(3):561-589. <https://doi.org/10.1007/s00421-021-04879-y>.
- Carobbio ALC, Cheng Z, Gianiorio T, Missale F, Africano S, Ascoli A, et al. Electric Bioimpedance Sensing for the Detection of Head and Neck Squamous Cell Carcinoma. *Diagnostics*. 2023;13:2453. <https://doi.org/10.3390/diagnostics13142453>.
- Schotman JM, Van Borren MM, Kooistra MP, Doorenbos CJ, de Boer H. Towards personalized hydration assessment in patients, based on measurement of total body electrical resistance: Back to basics. *Clinical nutrition ESPEN*. 2020;35:116-122. <https://doi.org/10.1016/j.clnesp.2019.10.018>.
- AlDisi R, Bader Q, Bermak A. Hydration Assessment Using the Bio-Impedance Analysis Method. *Sensors*. 2022;22(17):6350. <https://doi.org/10.3390/s22176350>.
- Karavetian M, Salhab N, Rizk R, Poulia KA. Malnutrition-inflammation score VS phase angle in the era of GLIM criteria: A cross-sectional study among hemodialysis patients in UAE. *Nutrients*. 2019;11. <https://doi.org/10.3390/nu11112771>.
- Lukaski HC, García-Almeida JM. Phase angle in applications of bioimpedance in health and disease. *Reviews in Endocrine and Metabolic Disorders*. 2023;24(3):367-370. <https://doi.org/10.1007/s11154-023-09799-0>.
- Van der Sande FM, Van de Wal-Visscher ER, Stuard S, Moissl U, Kooman JP. Using bioimpedance spectroscopy to assess volume status in dialysis patients. *Blood Purif*. 2020; 49(1):178-184. <https://doi.org/10.1159/000504079>.
- Bello JLG, Luna TB, Lara Lafargue A, Ciria HMC, Zulueta YA, Bioimpedance formalism: A new approach for accessing the health status of cell and tissues, *Bioelectrochemistry*. 2024;160:108799. <https://doi.org/10.1016/j.bioelechem.2024.108799>.
- Chaunzwa TL, Qian JM, Li Q, Ricciuti B, Nuernberg L, Johnson JW, Weiss J, Zhang Z, MacKay J, Kagiampakis I, Bikiel D, Di Federico A, Alessi JV, Mak RH, Jacob E, Awad MM, Aerts HJWL. Body Composition in Advanced Non-Small Cell Lung Cancer Treated with Immunotherapy. *JAMA Oncol*. 2024;10:773–783. <https://doi.org/10.1001/JAMAONCOL.2024.1120>.
- Braun RP, Mangana J, Goldinger S, French L, Dummer R, Marghoob AA. Electrical Impedance Spectroscopy in Skin Cancer Diagnosis. *Dermatol. Clin*. 2017;35:489–493. <https://doi.org/10.1016/j.det.2017.06.009>.
- Baidillah MR, Riyanto R, Busono P, Karim S, Febryarto R, Astasari A, Sangaji D, Taruno WP. Electrical impedance spectroscopy for skin layer assessment: A scoping review of electrode design, measurement methods, and post-processing techniques. *Measurement*. 2024;226:114111. <https://doi.org/10.1016/J.MEASUREMENT.2023.114111>.
- da Silva BR, Rufato S, Mialich MS, Cruz LP, Gozzo T, Jordão AA. Phase angle is related to oxidative stress and antioxidant biomarkers in breast cancer patients undergoing chemotherapy. *PLoS One*. 2023;18:e0283235. <https://doi.org/10.1371/journal.pone.0283235>.
- Jung M, Jeon JY, Yun GJ, Yang S, Kwon S, Seo YJ. Reference values of bioelectrical impedance analysis for detecting breast cancer-related lymphedema. *Med. (United States)*. 2018;97:1–6. <https://doi.org/10.1097/MD.00000000000012945>.
- Mansouri S, Alhadidi T, Ben Azouz M. Breast cancer detection using low-frequency bioimpedance device. *Breast Cancer Targets Ther*. 2020;12:109–116. <https://doi.org/10.2147/BCTT.S274421>.
- Aljarrah M, Salman F. A Simple Analysis of Impedance Spectroscopy: Review. *J Inst Eng*. 2021;102(1):237-242. <https://doi.org/10.1007/s40033-021-00252-7>.
- Pingel J, Harrison A, Von Walden F, Hjalmarsson E, Bartels EM. Multi-frequency bioimpedance: a non-invasive tool for muscle-health assessment of adults with cerebral palsy. *J Muscle Res Cell Motil*. 2020;41:211-219. <https://doi.org/10.1007/s10974-020-09579-2>.
- Kanoun O, Kallel AY, Nouri H, Atitallah BB, Haddad D, Hu Z, et al. Impedance spectroscopy: applications, advances and future trends. *IEEE Instrumentation & Measurement Magazine*. 2022;25(3):11-21. <https://doi.org/10.1109/MIM.2022.9759355>.
- Wang LC, Raimann JG, Tao X, Preciado P, Thwin O, Rosales L, et al. Estimation of fluid status using three multifrequency bioimpedance methods in hemodialysis patients. *Hemodial Int*. 2022;26(4):575-587. <https://doi.org/10.1111/hdi.13034>.
- Abasi S, Aggas JR, Garayar-Leyva GG, Walther BK, Guiseppi-Elie A. Bioelectrical Impedance Spectroscopy for Monitoring Mammalian Cells and Tissues under Different Frequency Domains: A Review. *ACS Measurement Science*. 2022;2(6):495-516. <https://doi.org/10.1021/acsmesureciau.2c00033>.
- Nwosu AC, Mayland CR, Mason S, Cox TF, Varro A, Eilershaw J. The association of hydration status with physical signs, symptoms and survival in advanced cancer - The use of Bioelectrical Impedance Vector Analysis (BIVA) Technology to evaluate fluid volume in palliative care: An observational study. *PLoS One*. 2016;11:e0163114. <https://doi.org/10.1371/journal.pone.0163114>.

26. Piccoli A, Rossi B, Pillon L, Bucciante G. A new method for monitoring body fluid variation by bioimpedance analysis: The RXc graph. *Kidney Int.* 1994;46:534–539. <https://doi.org/10.1038/ki.1994.305>.
27. Campa F, Toselli S. Bioimpedance Vector Analysis of Elite, Subelite, and Low-Level Male Volleyball Players. *Int. J. Sports Physiol. Perform.* 2018;13:1250–1253. <https://doi.org/10.1123/IJSP.2018-0039>.
28. Martins PC, Gobbo LA, Silva DAS. Bioelectrical impedance vector analysis (BIVA) in university athletes. *J. Int. Soc. Sports Nutr.* 2021;18:1–8. <https://doi.org/10.1186/S12970-020-00403-3/FIGURES/3>.
29. Nwosu AC, Mayland CR, Mason S, Cox TF, Varro A, Stanley S, Ellershaw J. Bioelectrical impedance vector analysis (BIVA) as a method to compare body composition differences according to cancer stage and type. *Clin. Nutr. ESPEN.* 2019;30:59–66. <https://doi.org/10.1016/j.clnesp.2019.02.006>.
30. Limon-Miro AT, Valencia ME, Lopez-Teros V, Guzman-Leon AE, Mendivil-Alvarado H, Astiazaran-Garcia H. Bioelectric impedance vector analysis (Biva) in breast cancer patients: A tool for research and clinical practice. *Med.* 2019;55:663. <https://doi.org/10.3390/medicina55100663>.
31. Bello JLG, Lafargue AL, Ciria HC, Luna TB, Leyva YZ. Methodology for integrated analysis of vector- and spectroscopic bioimpedance methods. *J. Electr. Bioimpedance.* 2024;15:154–161. <https://doi.org/10.2478/JOEB-2024-0018>.
32. Koh DM, Papanikolaou N, Bick U, Illing R, Kahn CE, Kalpathi-Cramer J, et al. Artificial Intelligence and Machine Learning in Cancer Imaging. *Commun Med.* 2022;2(1):1-14. <https://doi.org/10.1038/s43856-022-00199-0>
33. Trivizakis E, Papadakis GZ, Souglakos I, Papanikolaou N, Koumakis L, Spandidos DA, et al. Artificial Intelligence Radiogenomics for Advancing Precision and Effectiveness in Oncologic Care (Review). *Int J Oncol.* 2020;57(1):43-53. <https://doi.org/10.3892/ijo.2020.5063>.
34. Luna TB, Bello JLG, Carbonell AG, Montoya A de la CR, Lafargue AL, Ciria HMC, Zulueta YA. The role of various physiological and bioelectrical parameters for estimating the weight status in infants and juveniles cohort from the Southern Cuba region: a machine learning study. *BMC Pediatr.* 2024;24:313. <https://doi.org/10.1186/s12887-024-04789-w>.
35. Luna TB, Bello JLG, Carbonell AG, Montoya A de la CR, Lafargue AL, Ciria HMC, Zulueta YA. Integrating classification and regression learners with bioimpedance methods for estimating weight status in infants and juveniles from the southern Cuba region. *BMC Pediatr.* 2024;24:1–17. <https://doi.org/10.1186/s12887-024-04841-9>.
36. Bello JLG, Luna TB, Carbonell AG, Román Montoya A de la C, Lara Lafargue A, Ciria HMC, Zulueta YA. Cancer predictive model derived from bioimpedance measurements using machine learning methods. *Clin. Nutr. Open Sci.* 2024;58:131–145. <https://doi.org/10.1016/j.nutos.2024.10.006>.
37. Charilaou P, Battat R. Machine learning models and over-fitting considerations. *World J Gastroenterol.* 2022;28(5):605-607. <https://doi.org/10.3748/wjg.v28.i5.605>
38. Bagui S, Li K. Resampling imbalanced data for network intrusion detection datasets. *J. Big Data.* 2021;8:6. <https://doi.org/10.1186/s40537-020-00390-x>.
39. Chakraborty P, Rafiammal SS, Tharini C, Jamal DN. Influence of Bias and Variance in Selection of Machine Learning Classifiers for Biomedical Applications. In *Smart Data Intelligence: Proceedings of ICSDMI 2022* (pp. 459-472). Singapore: Springer Nature. https://doi.org/10.1007/978-981-19-3311-0_39.
40. He H, Ma Y. Imbalanced learning: Foundations, algorithms, and applications. *Imbalanced Learn. Found. Algorithms, Appl.* (2013) 1–210. <https://doi.org/10.1002/9781118646106>.
41. Refaeilzadeh P, Tang L, Liu H. Cross-Validation. *Encycl Database Syst.* 2009:532–8. https://doi.org/10.1007/978-0-387-39940-9_565.
42. Arab A, Karimi E, Vingrys K, Shirani F. Is phase angle a valuable prognostic tool in cancer patients' survival? A systematic review and meta-analysis of available literature. *Clin. Nutr.* 2021;40:3182–3190. <https://doi.org/10.1016/j.clnu.2021.01.027>.
43. Axelsson L, Silander E, Bosaeus I, Hammerlid E. Bioelectrical phase angle at diagnosis as a prognostic factor for survival in advanced head and neck cancer. *Eur. Arch. Oto-Rhino-Laryngology.* 2018;275:2379–2386. <https://doi.org/10.1007/s00405-018-5069-2>.
44. Amano K, Bruera E, Hui D. Diagnostic and prognostic utility of phase angle in patients with cancer. *Rev. Endocr. Metab. Disord.* 2023;24:479–489. <https://doi.org/10.1007/s11154-022-09776-z>.