Machine Learning-Driven Optimization of Bioelectrical Impedance Analysis for Intracellular Fluid Prediction

Jannatul Nayem Novera[1]*, Md. Eyakub Ali[2], Mohammad Mahmudul Alam Mia[3], Aynal Haque[4], Tahasin Ahmed Fahim[5], Israt Jahan[6]

- [1,2,4,6]Noakhali Science and Technology University, Noakhali-3814, jannatulnovera987@gmail.com, meyakub.nstu@gmail.com, aynalhaque.eee2017@gmail.com, israt.eee@nstu.edu.bd.
- [3]RTM Al-Kabir Technical University, Sylhet-3100, mahmud_ece_ku@yahoo.com.
- [5]University of Chittagong, Chittagong-4331, tahasin@cu.ac.bd.

*Corresponding author [Jannatul Nayem Novera]

Keywords:

- Intracellular fluid;
- Bioelectrical impedance analysis;
- Machine learning algorithms;
- Linear regression;
- Ridge regression;

Abstract: Bioelectrical Impedance Analysis (BIA) is a widely used non-invasive method for body composition assessment, including intracellular fluid (ICF) estimation. While traditional methods rely on empirical equations, machine learning (ML) offers a more robust and data-driven approach to enhance predictive accuracy. This study aims to develop and validate ML-based models for ICF prediction. A dataset of 2,520 participants from the National Health and Nutrition Examination Survey (NHANES) was utilized to train and validate multiple ML algorithms, including Linear Regression (LR), Ridge Regression (RR), Random Forest (RF), Gradient Boosting (GB), AdaBoost (AB), Support Vector Regression (SVR), Decision Tree (DT), K-Neighbors Regressor (KNR), Lasso Regression (LaR), and Neural Network (NN). Model performance was assessed using Mean Squared Error (MSE), R^2 score, precision, recall, and F1 score. Results indicate that Ridge Regression and Linear Regression outperformed other models, achieving the lowest MSE (~3,916) and the highest R^2 score (~0.964) on the validation dataset. This study demonstrates the potential of ML techniques in improving ICF estimation from BIA, offering a scalable and accurate alternative for body composition analysis. Future research will explore deep learning approaches and optimized feature selection to enhance predictive accuracy further.

1. Introduction

Body fluids are aqueous liquids that carry a wide range of solutes and metabolic products and contain the cellular and ionic components required for a healthy body. It controls body temperature, preserves electrolyte balance, and modifies normal osmotic pressure (Zhang et al. 2019). A two-to-one ratio differentiates the total body fluid into two fluid compartments: the intracellular fluid (ICF) compartment and the extracellular fluid (ECF) compartment. An important component of the cytoplasm and cytosol, intracellular fluid is a material found inside live cells that is mostly composed of water and other compounds like dissolved ions. Approximately 40% of the body weight is made up of intracellular fluid. Unbalanced changes in body fluids are not only good indicators for identifying dehydration and water intoxication, but they are also risk factors for a number of illnesses, including kidney ailments, diabetic ketoacidosis, cardiovascular disease in hemodialysis patients, and others.

Body composition can be evaluated by bioelectrical impedance analysis (BIA), which enables the assessment of important body components such as total body fluid (ICF and ECF), fat mass, and fat-free mass (Mialich, Sicchieri, and Junior, n.d.). Despite being the most effective method for body composition assessment, the dual-energy X-ray absorptiometry (DEXA) approach has many drawbacks, including the size of the equipment and the costs of the measurement. This is the reason bioelectrical impedance analysis (BIA) has become a popular non-invasive alternative for body composition assessment in a variety of healthcare environments (Przytula and Popiolek-Kalisz 2025).

Advanced healthcare systems now heavily rely on machine learning and sensor technology (Liaqat et al. 2020). As intelligent, self-governing, and pervasive decision-making systems, they are anticipated to diagnose and treat illnesses. Since the model for intracellular fluid is relevant to the parameters obtained from bioelectrical impedance analysis, it becomes a significant problem to develop an accurate model based on the appropriate parameters. There has been some research on this topic, but no such model development using machine learning algorithms has been done so far. Therefore, the application of machine learning algorithms could provide the appropriate model based on the exact required parameters. Therefore, this study focuses on using a number of machine learning algorithms for the development and validation of the model, including Random Forest (RF), Gradient Boosting (GB), AdaBoost (AB), Support Vector Regression (SVR), Linear Regression (LR), Ridge Regression (RR), Lasso Regression (LaR), Decision Tree (DT), K-Neighbors Regressor (KNR), and Neural Network (NN). The aim of this research is to develop machine learning-based models for intracellular fluid estimation and evaluate the best-performing machine learning algorithms.

In this study, new mathematical models for intercellular fluid with machine learning algorithms have been developed and presented, and the newly established models have been validated to provide clarity. The rest of the paper is organized in the following sections: Section 2 represents the materials and methods, Section 3 presents the results and discussion of the newly established models, Section 4 shows the contributions of the work, and Section 5 concludes the paper.

2. Materials and Methods

2.1 Subjects

To develop the ICF model, a data set including 2520 participants (1,200 males and 1,320 females) was utilized. The database arrives from the National Health and Nutrition Examination Survey (NHANES), administered to US citizens from 2003 to 2004 (version 7, updated in July 2016) (Statistics 2016). The models were developed using 1275 participants (males and females). The established model was verified using the remaining 1245 participants (males and females). Several machine learning algorithms were used for predicting and validating the model of ICF using this data. Feature selection was based on domain knowledge and previous studies. Key features included age, height, BMI, impedance at 1 MHz, height squared resistance, and height squared impedance. These variables were chosen due to their established relevance in BIA-based body composition analysis.

2.2 BIA

The Bioelectrical Impedance Analyzer (BIA) uses an 800-microamp current to assess the electrical resistance in the deeper layers of tissue of the human body (Waller and Lindinger 2006). BIA involves placing electrodes across a region of interest on multiple body regions of the person (Brantlov et al. 2017). With the first set of electrodes introduced to a safe, undetectable alternating electrical current (AC), the impedance meter measures the resistance of the body to the current passing through the second set of electrodes. Body fluids serve as resistors, while cell membranes act as capacitors. By dividing the injected alternating current into resistive (fluid and electrolytes) and capacitive (cell membranes and tissue interfaces) pathways, the body can be viewed as a parallel resistor-capacitor (RC) equivalent circuit in Figure 1. (Chinen et al. 2015).

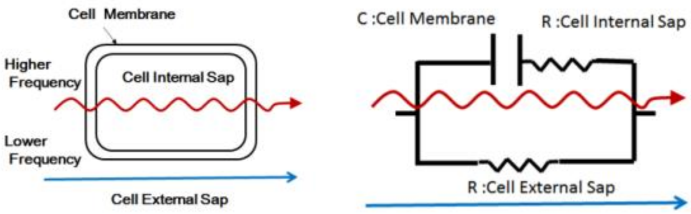

Figure 1. (a) electrical conductance through the cell, (b) an equivalent circuit.

2.3 Machine learning based algorithms

To develop the model, a number of machine learning algorithms were used in this study. AdaBoost, Random Forest, Gradient Boosting, Support Vector Regression, Linear Regression, Ridge Regression, Lasso Regression, Decision Tree, K-Neighbors Regressor, and Neural Network are among the algorithms. ML models were selected based on their predictive power, interpretability, and computational efficiency. Linear Regression and ridge regression were chosen due to their strong performance in regression tasks. Tree-based models such as Random Forest and Gradient Boosting were included for their ability to capture non-linear relationships. Support Vector Regression and Neural Networks were tested to assess their applicability in ICF prediction.

From these algorithms, model performance has been evaluated using the parameters mean squared error (MSE), mean absolute error (MAE), and R^2 score. The classification metrics, such as precision, recall, and F1 score has also been computed for further validation of the model.

Classification metrics and model performance parameters can be defined as below:

$$\text{Mean Square Error (MSE)} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x_i})^2 \quad (1)$$

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}|(x_i - \hat{x_i})| \quad (2)$$

$$R^2 \text{ score} = 1- \frac{\sum_{i=1}^{n}(x_i-\hat{x_i})^2}{\sum_{i=1}^{n}(x_i-\bar{x_i})} \quad (3)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (4)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive}+\text{False Negative}} \tag{5}$$

$$\text{F1 score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \tag{6}$$

3. Result and Discussions

3.1 Proposed Mathematical Models for ICF Measurement

Two mathematical models have been developed to measure intracellular fluid. To develop the model, age, height, BMI, impedance at 1 MHz, height to impedance ratio, and height to resistance ratio have been considered as parameters.

Prediction Equation (Linear Model):

$$\textbf{ICF} = 1238.5245 + (-6.3048) * \text{Age} + (3.5925) * \text{BMI} + (-34.6848) * \text{HT2} + (352.7262) * \text{Z1M} + (-2134.3120) * \text{HT2R1M} + (2200.5031) * \text{HT2Z1M} \tag{7}$$

Prediction Equation (Ridge Model):

$$\textbf{ICF} = 1238.7675 + (-6.0871) * \text{Age} + (5.0681) * \text{BMI} + (-35.4050) * \text{HT2} + (351.1390) * \text{Z1M} + (-143.0933) * \text{HT2R1M} + (207.5425) * \text{HT2Z1M} \tag{8}$$

3.2 Evaluation of Performance

Table 1. Performance Evaluation of Different Machine Learning Models for Prediction

Algorithms	Machine Learning Algorithms									
	Model performance				Classification metrics					
	Train		Validation		Train			Validation		
	R2	MSE	R2	MSE	Precision	Recall	F1 score	Precision	Recall	F1 score
LR	0.94	9000.16	0.96	3946.23	0.85	0.90	0.88	0.91	0.97	0.94
RR	0.94	9204.43	0.96	3916.88	0.85	0.91	0.87	0.91	0.97	0.94
RF	0.88	17685.37	0.85	16547.20	0.87	0.88	0.88	0.90	0.88	0.89
GB	0.89	15665.89	0.89	15305.41	0.89	0.90	0.89	0.88	0.92	0.90
AB	0.87	18951.54	0.91	17525.35	0.81	0.93	0.87	0.85	0.89	0.88
SVR	0.23	119093.3	0.65	99873.26	0.81	0.88	0.84	0.79	0.86	0.87
DT	0.79	31099.37	0.80	30479.89	0.84	0.87	0.85	0.87	0.89	0.75
LaR	0.93	10674.94	0.90	11509.81	0.85	0.90	0.87	0.88	0.90	0.85
KNR	0.86	20949.19	0.88	19309.11	0.87	0.88	0.88	0.89	0.92	0.89
NN	0.63	55967.81	0.71	54875.01	0.85	0.63	0.72	0.86	0.75	0.78

Shown in Table 1 that, the evaluation of machine learning algorithms shows Linear Regression (LR) and Ridge Regression (RR) as the best-performing models for both regression and classification. They achieved high R^2 values (0.94–0.96) with low MSE (~9000 in training, ~4000 in validation) and demonstrated strong classification metrics with F1-scores of 0.94 in validation. Ensemble models like Gradient Boosting (GB) and K-Nearest Neighbors (KNR) provided competitive classification performance (F1-score ~0.89–0.90) but had higher MSE in regression, limiting their effectiveness in continuous predictions. Decision Trees (DT) exhibited overfitting, while Neural Networks (NN) and Support Vector Regression (SVR) performed poorly, struggling with both regression and classification tasks. Overall, LR and RR are the most reliable models for both tasks.

It is seen from Figure 2. that the predicted models of linear regression and ridge regression almost resemble the actual values of ICF. So, the proposed models can be used for ICF estimation with good accuracy.

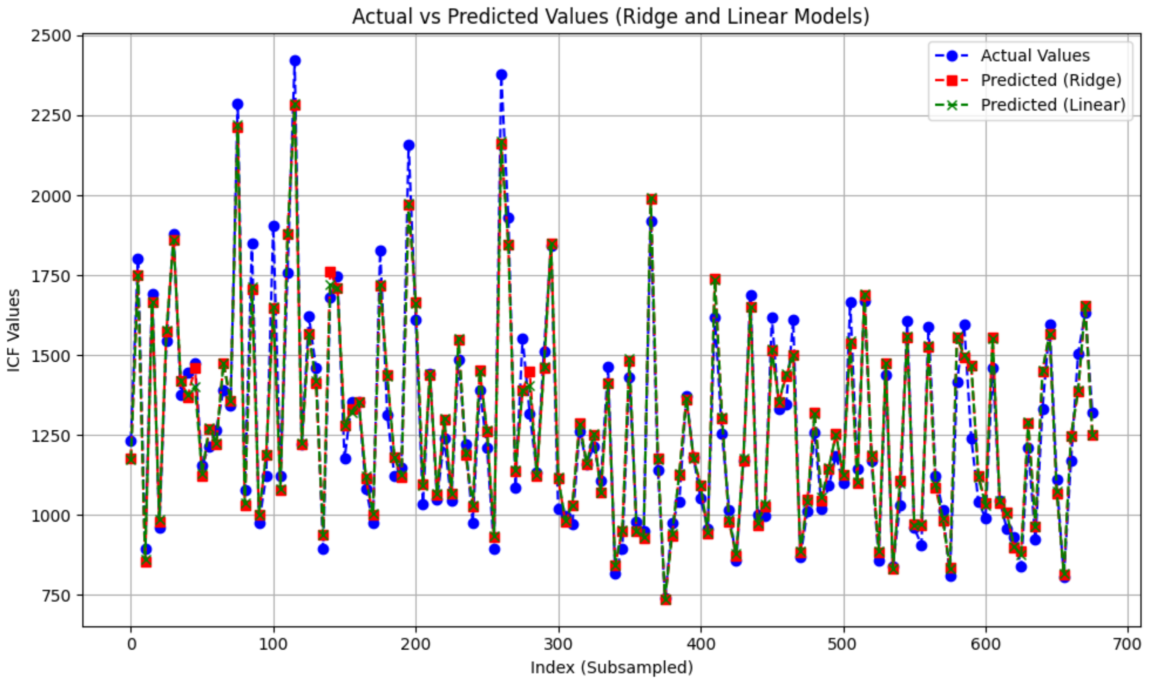

Figure 2. Actual vs Predicted graphs for Linear and Ridge (best two) Models.

4. Contributions

This study made significant contributions to the field of bioelectrical impedance analysis (BIA) and machine learning applications in body composition prediction. Multiple machine learning models have been implemented and evaluated, identifying Ridge Regression and Linear Regression as the most effective for ICF prediction. The findings highlight the potential of machine learning in improving non-invasive body composition analysis, paving the way for more accurate, scalable, and accessible health monitoring solutions.

5. Conclusion

This study underscores the effectiveness of machine learning techniques in predicting ICF from BIA-derived metrics, demonstrating that Ridge and Linear Regression provide the best predictive accuracy. Our models achieved high R^2 scores (>0.96) and low MSE values, making them highly reliable for body composition analysis. While other machine learning models, such as Random Forest and Gradient Boosting, performed well, they did not surpass the efficiency and interpretability of the regression-based approaches. The implications of this research are significant for health and nutrition assessment, offering a cost-effective and non-invasive method for monitoring body composition. Though ML models demonstrated improved predictive accuracy, certain limitations remain. The study relied on a dataset from NHANES, which may not fully represent diverse populations. Future work will expand the dataset to include diverse populations, explore deep learning approaches, enhance feature selection techniques, and incorporate of additional physiological markers to further improve prediction accuracy.

References

Brantlov, Steven, Leigh C. Ward, Lars Jødal, Søren Rittig, and Aksel Lange. 2017. "Critical Factors and Their Impact on Bioelectrical Impedance Analysis in Children: A Review." *Journal of Medical Engineering & Technology* 41 (1): 22–35. https://doi.org/10.1080/03091902.2016.1209590.

Chinen, Koyu, Ichiko Kinjo, Aki Zamami, Kotoyo Irei, and Kanako Nagayama. 2015. "New Equivalent-Electrical Circuit Model and a Practical Measurement Method for Human Body Impedance." *Bio-Medical Materials and Engineering* 26 (s1): S779–86. https://doi.org/10.3233/BME-151369.

Liaqat, Sidrah, Kia Dashtipour, Kamran Arshad, and Naeem Ramzan. 2020. "Non Invasive Skin Hydration Level Detection Using Machine Learning." *Electronics* 9 (7): 1086. https://doi.org/10.3390/electronics9071086.

Mialich, Mirele Savegnago, Juliana Maria Faccioli Sicchieri, and Alceu Afonso Jordao Junior. n.d. "Analysis of Body Composition: A Critical Review of the Use of Bioelectrical Impedance Analysis."

Przytula, Agata, and Joanna Popiolek-Kalisz. 2025. "Bioelectrical Impedance Analysis as a Helpful Tool in Pediatric Obesity Monitoring: A Case Report." *Reports* 8 (1): 15. https://doi.org/10.3390/reports8010015.

Statistics, United States Department of Health and Human Services Centers for Disease Control and Prevention National Center for Health. 2016. "National Health and Nutrition Examination Survey (NHANES), 2003-2004." Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR25503.v7.

Waller, A., and M. I. Lindinger. 2006. "Hydration of Exercised Standardbred Racehorses Assessed Noninvasively Using Multi-Frequency Bioelectrical Impedance Analysis." *Equine Veterinary Journal* 38 (S36): 285–90. https://doi.org/10.1111/j.2042-3306.2006.tb05554.x.

Zhang, Na, Jianfen Zhang, Songming Du, Hairong He, Xinyu Yan, and Guansheng Ma. 2019. "Association between the Content of Intracellular and Extracellular Fluid and the Amount of Water Intake among Chinese College Students." *Nutrition & Metabolism* 16 (1): 67. https://doi.org/10.1186/s12986-019-0397-9.