Postharvest Biology and Technology 222 (2025) 113358

Contents lists available at ScienceDirect

Postharvest Biology and Technology

journal homepage: www.elsevier.com/locate/postharvbio

Bioimpedance-based prediction of dry matter content and potato varieties through supervised machine learning methods

Ciro Allará [a,*], Roberto Moscetti [d], Giacomo Bedini [d], Manuela Ciocca [a], Alessandro Benelli [d], Paolo Lugli [a,b], Luisa Petti [a,c], Pietro Ibba [a,*]

[a] Faculty of Engineering, Free University of Bozen-Bolzano, Piazza Domenicani 3, Bozen-Bolzano, 39100, Italy
[b] Competence Centre for Mountain Innovation Ecosystems, Free University of Bozen-Bolzano, Piazza Universitá 1, Bozen-Bolzano, 39100, Italy
[c] Competence Centre for Plant Health, Free University of Bozen-Bolzano, Piazza Universitá 1, Bozen-Bolzano, 39100, Italy
[d] Department for Innovation in Biological, Agro-food and Forest System, University of Tuscia, Via S. Camillo de Lellis snc, Viterbo, 01100, Italy

ARTICLE INFO

Keywords:
Solanum tuberosum L
Bioimpedance spectroscopy
Dry matter
Chemometrics
Features extraction

ABSTRACT

This study explored bioimpedance measurements and their derived parameters as potential indicators of dry matter content and potato variety. We investigated how bioimpedance correlates with the dry matter composition of potatoes,through destructive tests on tissue slices. Using the machine learning approach, we identified patterns and associations that can aid in predicting dry matter content, using regression models, and discriminating among potato varieties, focusing on classification models. In particular, four feature selection methods (correlation matrix, minimum-redundancy–maximum-relevance, neighborhood component analysis, and t-test) were evaluated against a baseline with all features. It was found that the best regression model for dry matter was a neural network regression model with t-test features, which achieved R^2 of 0.62 and RMSE of 2.3% in testing, while the best classification model for potato variety was a neural network classification model with correlation matrix features, achieving an F_1 score of 0.92.

1. Introduction

The application of more sustainable production approaches in the agri-food sector is fundamental to ensure food safety and security, while preventing food waste, as highlighted by the UN Sustainable Development Goals (SDGs), especially SDG12 (Assembly, 2015). This is particularly relevant in the production and postharvest handling of staple foods such as the potato (*Solanum tuberosum* L.), which has a high cultural, economic, and environmental impact worldwide. In the EU, in 2022 potato production was around 374 mil. Mg, for a value of EUR 12.3 billion (3.1% of total EU agricultural output) (Food and Agriculture Organization of the United Nations (FAO), 2024), meeting a high internal demand for both direct consumption and food processing. To fulfill such a need, potato tubers are stored in controlled conditions to delay dormancy break, weight loss, and sweetening, thereby increasing their marketability and availability for the food industry (Kasampalis et al., 2021).

During its low-temperature storage (approx. 4-7 °C for fresh and processing market), the tuber is subject to the aforementioned physiochemical changes, varying with storage conditions, time, and cultivar type, and potentially altering its final quality and safety (Hertog et al., 1997; Broćić et al., 2016). In particular, the sweetening occurs due to cold-induced sucrose hydrolysis resulting in reducing sugar (RS) (i.e., glucose and fructose) accumulation. A high level of RS is responsible for (i) degraded quality of processed products (e.g., brown color, bitter taste) and (ii) acrylamide accumulation, a potential carcinogenic agent, during cooking (European food safety authority (EFSA), 2022). Thus, potato sweetening is a serious risk factor for consumer health, with its rate depending on the aforementioned factors (Escuredo et al., 2021). The chemical and physiological maturity of potatoes is reached when the sucrose content is the lowest and the dry matter (DM) content is the highest. In this context, DM and RS are considered the potatoes' most important quality features to be monitored during storage, determining their price, as well as their suitability for industrial applications, where large variations could lead to food losses (Wold et al., 2021).

Currently, industrial quality control consists of (i) oven or gravimetric methods to assess the DM content (Horwitz et al., 2000), and (ii) the United States Department of Agriculture (USDA) reference method for qualitative determination of RS, consisting of product frying, followed by a visual inspection of its chemical browning rate: i.e., the greater the browning, the higher the RS content and, consequently, the lower the quality and safety (Whitehead, 1954). While the DM assay is accurate, repeatable, and simple, the RS method is not. Nevertheless, both are

* Corresponding author.
E-mail addresses: callara@unibz.it (C. Allará), pietro.ibba@unibz.it (P. Ibba).

https://doi.org/10.1016/j.postharvbio.2024.113358
Received 28 June 2024; Received in revised form 2 December 2024; Accepted 7 December 2024
Available online 2 January 2025
0925-5214/

not ideal due to their time intensiveness, destructive nature, and, most importantly, the requirement for a laboratory (Bedini et al., 2023). For the above reasons and driven by the growing attention of companies and consumers towards food waste and healthy food, the potato industry has recently experienced an increased demand for rapid, and, if possible, non-destructive techniques for quality control. However, although portable, rapid, and simple technologies have already been proposed for the on-site determination of DM and RS in potatoes, they suffer from performance and robustness issues. For instance, Bedini et al. (2023) demonstrated that DM and RS gradients along tuber tissues and potato cultivars affect the quality of on-spot measurements carried out with portable instruments. Moreover, identifying the potato cultivar is crucial for industrial applications, because different varieties have unique DM and RS ranges at harvest and during storage that influence their suitability for various end-products, such as chips, fries, or mashed potatoes (Zaheer and Akhtar, 2016; Tolessa, 2018). Knowing the variety helps optimizing processing parameters, improving product consistency, and reducing waste. In this context, near-infrared spectroscopy (NIRS) has proven to be an effective technique for the prediction of RS and DM (Rady and Guyer, 2015; Wold et al., 2021; Bedini et al., 2024). NIRS allow to achieve optimal predictions of potato DM, with a root mean square error (RMSE) as low as 0.51% when combined with prediction model updating methods (Wang et al., 2023). Such technique demonstrated high validation accuracy also in the estimation of multiple tubers DM contents, such as yam, with a regression coefficient (R^2) in prediction of 0.958 and an RMSE of 0.898% (Adesokan et al., 2024) and cassava amylose content with an R^2 of 0.94 (Nuwamanya et al., 2024). Despite this, the study of the interaction of a potato tuber with different regions of the electromagnetic (EM) spectrum still offers various opportunities to tackle such issues in a fast and multifaceted way. In this direction, a promising technique is electrical impedance spectroscopy (EIS), working at low frequencies (10^1 – 10^6 Hz) in the EM spectrum. This method, already successfully used in many different contexts, from food product screening (Grossi and Riccò, 2017) to human body analysis (Zou and Guo, 2003), is based on the physico-chemical property of a target to oppose the flow of current induced by an external AC voltage applied at multiple frequencies (Martinsen and Heiskanen, 2023). The resulting bioimpedance output depends on various factors, including the geometry of the object, the electrical properties of the material, and the chemical processes occurring within it (Dean et al., 2008). The application of EIS to fruit quality has only recently gained popularity and, while providing promising results (Ibba et al., 2020), has yet to prove its effectiveness compared to conventional optical methods. In the context of potato quality estimation, bioimpedance has been employed to monitor the drying process (Ando et al., 2014), the presence of freeze damage (Zhang and Willison, 1992) and for the estimation of its moisture content (Feng et al., 2020). While these works provide a solid foundation for applying bioimpedance in the tuber quality sorting, they fall short in the volume of data used and, crucially, lack a comprehensive data analysis. The data extracted from EIS measurements, if acquired properly and in big numbers, can in fact be leveraged by machine learning algorithms to predict both the DM (using regression models) and the potatoes' varieties (using classification models). Regression models offer the advantage of providing continuous predictions, making them ideal for estimating the precise DM. Prediction of DM is important for quality control and product formulation, although they can be sensitive to outliers and require a large amount of high-quality data for accurate predictions. Classification models, on the other hand, are useful for categorizing potatoes into distinct varieties, offering simplicity and interpretability, even though they can struggle with imbalanced datasets and may require extensive feature engineering to improve accuracy. Thus, the hypothesis behind the present research activity was that the use of EIS has the potential to yield a good understanding, and thus modeling, of the potato quality. In detail, in the present study, the data extracted from EIS measurements were leveraged by machine learning algorithms to predict both the DM and the potatoes' varieties, using regression and classification models, respectively.

2. Materials and methods

2.1. Sample preparation

Potato (*Solanum tuberosum* L.) tubers were provided by the "Patata dell'Alto Viterbese" PGI consortia (C.C.OR.A.V. and Coop. Centro Agricolo Alto Viterbese, Viterbo, Italy). Nine cultivars were selected on the basis of a literature study focused on ensuring the widest DM data range (approx. 10-26 %) for regression and classification model development: *Actrice*, *Ambra*, *Constance*, *El Mundo*, *Fontane*, *Gaudi*, *Jelly*, *Monalisa* and *Universa*. Tubers were collected immediately after harvest and stored at 4 ± 1 °C. Twenty potato tubers per cultivar were randomly sampled. Moreover, tubers were selected to be of uniform size, without damage, defects, and/or blemishes. Samples were tempered to room temperature for 24 h before the analysis.

2.2. Bioimpedance data acquisition

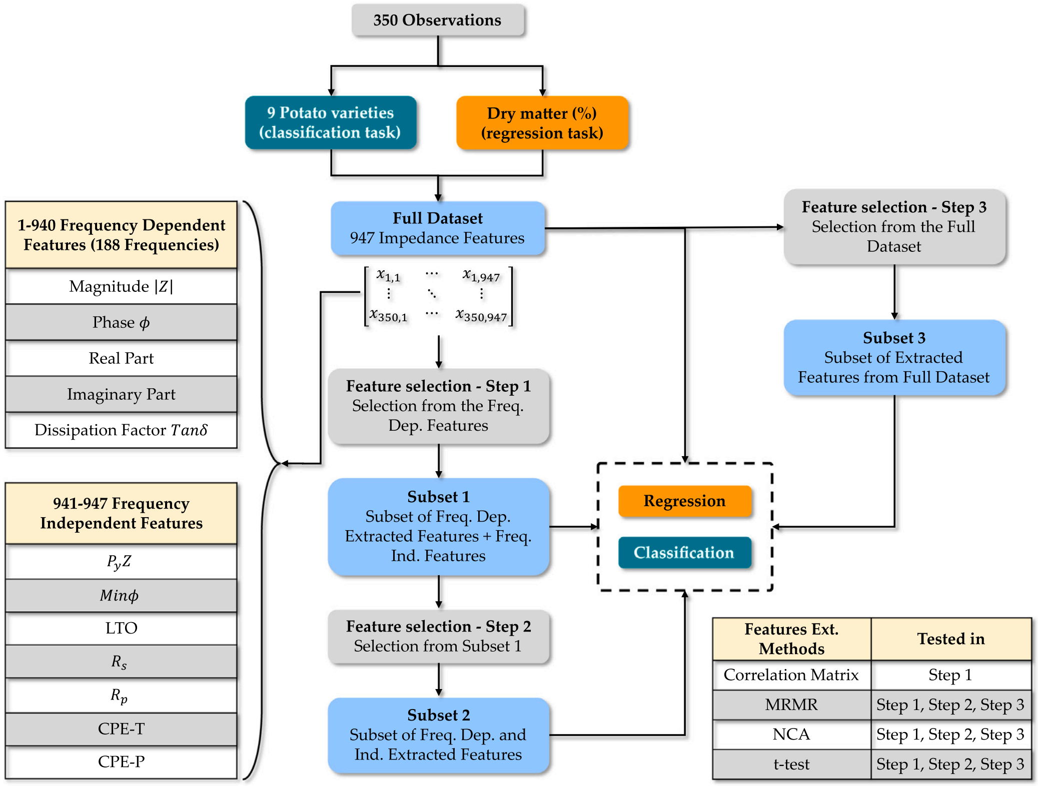
Bioimpedance spectral scans were performed on the middle slice of each sample tuber (2.5 cm thick), using a bench-top impedance analyzer (E4990 A, Keysight Technologies), in the two electrode configuration, directly inserting the stainless steel electrodes in the tuber pulp with a standardized 2 cm separation. Here, the impedance magnitude and phase data were acquired in the 20 Hz–20 MHz frequency range with 188 logarithmically spaced frequency points. The raw magnitude and phase angle spectra data were then used to extract additional bioimpedance parameters, such as (i) the minimum phase point of each spectrum, $Min\phi$, (ii) the ratio between the low- and high-frequency values of the impedance magnitude, P_yZ (Pliquett et al., 2003), (iii) the dissipation factor, $Tan\delta$ (Strand-Amundsen et al., 2016), (iv) the distance between the zero and the maximum value of the Nyquist plot, LTO (Watanabe et al., 2021) and (v) the Cole model equivalent circuit parameters (Cole, 1940) (i.e., R_s, R_p, CPE-P and CPE-T), which were organized in a dataset before the model development, described in Section 2.4. These parameters, commonly applied in the food industry, are valuable for correlating the moisture content of tubers with impedance measurements, enabling accurate predictive analyses. Their relevance to our study is highlighted by their ability to capture electrical properties related to tissue structure and composition, which are also pertinent to tubers. In particular, the P_y parameter, originally developed for meat, offers significant insights into moisture content and cell integrity, key factors in evaluating tuber quality as well (Ibba et al., 2021).

2.3. Dry matter determination

Immediately after the bioimpedance data acquisition, the samples were subjected to DM determination. Each potato slice was trimmed and dried in an electric oven at 105 °C for 72 h. DM was expressed as a percentage by mass (grams per 100 grams).

2.4. Feature extraction

The dataset, consisting of a total 947 features, was structured as follows: (i) frequency-dependent spectral data for magnitude ($|Z|$), phase (ϕ), real and imaginary part, and $Tan\delta$ calculated over a set of 188 frequencies, representing features from 1 to 940; (ii) frequency-independent features from 941 to 947 were represented by P_yZ, $Min\phi$, LTO and the parameters of the Cole equivalent circuit (R_s, R_p, CPE-T, CPE-P). The feature selection process began by applying separate feature selection techniques to each set of frequency-dependent variables, followed by combining the resulting dataset with the frequency-independent variables. Thus, Subset 1 consisted of the selected frequency-dependent variables along with all frequency-independent variables. Subset 2 was created by further refining the features in Subset 1. In contrast, Subset 3 was generated by selecting features from



Features Ext. Methods	Tested in
Correlation Matrix	Step 1
MRMR	Step 1, Step 2, Step 3
NCA	Step 1, Step 2, Step 3
t-test	Step 1, Step 2, Step 3

Fig. 1. Flowchart illustrating the dataset structure, where variables (the minimum phase point of each spectrum, $Min\phi$, the ratio between the low- and high-frequency values of the impedance magnitude, P_yZ, the dissipation factor, $Tan\delta$, the distance between the zero and the maximum value of the Nyquist plot, LTO and the Cole model equivalent circuit parameters (i.e., R_s, R_p, CPE-P and CPE-T)) were categorized based on frequency and other criteria. The four analyzed datasets were highlighted in blue.

the entire dataset without distinguishing between frequency-dependent and independent variables. This approach allowed us to evaluate the impact of frequency-dependent features separately. By employing this multi-step feature selection process, deeper insights into the relevance and importance of different variables were gained, possibly leading to the identification of a more robust and interpretative model.

In training and prediction phase, the MATLAB® 2023a `fitrauto()` regression and `fitcauto()` classification functions (The MathWorks Inc., Natick, MA, United States) were used to estimate the predictant values based on the selected features. Before prediction, the data were normalized (regression) (Bonamutial and Prasetyo, 2023) and autoscaled (classification) (Dzierżak, 2019) to ensure uniform distribution and reduce the risk of over-fitting. The normalization method was based on the minimum and maximum, transforming the values of each feature so that they fall within the range of 0 to 1. This was achieved by subtracting the minimum value from the feature and dividing it by the difference between the maximum and minimum. The autoscaling process was performed to ensure that the features have a mean of 0 and a variance of 1. Subsequently, four different feature extraction methods have been utilized and compared in this study. Three of these are conventional methods, namely: minimum-redundancy–maximum-relevance (MRMR) (Ding and Peng, 2005; Darbellay and Vajda, 1999), neighborhood component analysis (NCA) (Nash et al., 1994), and the univariate features selection (t-test) (Waugh, 1995). For MRMR and t-test methods, by analyzing the resulting ranking plots, the top 100 features were empirically selected for prediction for Subset 1, while only the top 40 features for Subset 2 and 3 were selected. The selection was driven by observing the ranking trends and aimed at ensuring an optimal balance between model performance and computational efficiency. The fourth method, instead, was based on computing the correlation matrix among the frequency-dependent features, to take into account and exclude the redundant information between neighboring frequency points, commonly having a high degree of autocorrelation. For the correlation analysis, ten separate correlation matrices were calculated, each corresponding to a group of frequencies (ranging from 1 to 10). After deriving these matrices, we examined feature relationships within each matrix. Features showing significantly different correlations with other variables (correlation values differing by at least 10 %) were prioritized for selection, as they are less likely to be redundant and more informative. In the end, the prediction of DM and potatoes' variety done after the features extraction process was compared with the prediction done on the entire dataset, without excluding any feature (Nicosia et al., 2024). In the end, (i) the full dataset (FD), (ii) the subset that underwent feature extraction considering only frequency-dependent variables (1-940) with the extracted features supplemented by the frequency-independent variables (Subset 1), (iii) the subset created by performing feature extraction on Subset 1 (Subset2), and (iv) the subset constituted by the extracted features by the entire dataset, without any differentiation between frequency-dependent and -independent variables (Subset 3), were considered.

2.5. Model calibration and validation

After feature extraction, the resulting four sets of data (i.e., one with all features and the other three with extracted features, Fig. 1) were randomly separated into the train and test datasets (75% and 25% respectively) used for the machine learning analysis and the following validation. This analysis was carried out using the default MATLAB® functions, `fitrauto()` for the regression model and `fitcauto()` for the classification model, which simplified and automated the performance optimization process. These two functions allowed the testing and comparison of different model architectures and numerous hyper-parameters, obtaining the best-performing predictive models with minimal computational effort. The algorithms selected for regression were the following: *(i) ensemble regression model (ERM)*, which combines multiple learning algorithms to improve predictive performance; *(ii)* *gaussian process regression model (GPRM)*, which are nonparametric kernel-based probabilistic models (Williams and Rasmussen, 2006); *(iii) kernel regression model (KRM)*, which uses kernel functions to capture nonlinear relationships in the data; *(iv) neural network regression model (NNRM)*, capable of learning intricate patterns through its layered architecture; *(v) support vector machine regression model (SVMRM)*, effective in high-dimensional spaces and for regression tasks with clear margin separation; and *(vi) binary decision tree (BDT)*, which splits the data into homogeneous subsets based on feature values, making it easy to interpret (Breiman, 2017).

The algorithms selected for classification were the following: *(i) discriminant analysis classifier (DAC)*, which models the decision boundary between classes based on statistical properties (Fisher, 1936); *(ii) ensemble classification model (ECM)*, leveraging the strength of multiple classifiers to enhance accuracy; *(iii) kernel classification model (KCM)*, employing kernel methods to handle non-linear classification tasks; *(iv) K-nearest neighbor model (KNM)*, which classifies based on the majority vote among the nearest neighbors in the feature space; *(v) naive bayes classifier (NBC)*, utilizing Bayes' theorem with strong (naive) independence assumptions between features (Manning, 2008); *(vi) neural network classifier (NNC)*, adept at capturing complex decision boundaries through its multi-layered network structure; *(vii) support vector machine classifier (SVMC)*, which finds the optimal hyperplane for class separation in high-dimensional space; and *(viii) binary decision tree classifier (BDTC)*, offering a straightforward and interpretable approach by recursively partitioning the data into classes based on feature splits (Breiman, 2017). Hyperparameters were set before the learning process and included elements such as the number of expansion dimensions *(NEM)*, the type of learner *(L)*, the basis function *(BF)*, the kernel function *(KF)*, the number of nodes *(NN)*, the layer size *(LS)*, and the activation function *(A)*. In this study, we employed the `OptimizeHyperparameters` option of `fitrauto()` and `fitcauto()` to automatically optimize these hyperparameters across different respectively regression and classification models. The performance of the regression models, Table 1, were evaluated in terms of coefficient of determination (R^2), root means squared error (RMSE), and systematic error (*BIAS*), while for classification, Table 2, the F_1 score is used (Brancato et al., 2023), in addition to accuracy, sensitivity and precision parameters. Each metric was computed for both the train and test phases.

3. Results and discussion

3.1. Descriptive data analysis

Prior to model development, a descriptive analysis on the DM content of potato samples was performed. Fig. 2 shows a detailed view of the DM (%) for each of the nine potato varieties.

The representation of the central tendencies and potential outliers of the distribution (Fig. S2) allows a more targeted identification of significant differences between the DM values among the varieties. In addition, the table associated with the plot in Fig. 2 shows the actual values of the statistical parameters of the DM for each potato variety. Here, it is possible to note that although some DM values appear to be similar to each other, there are varieties that tend to be very different from the others. Such distribution of DM values allows us to consider the acquired dataset as a good representation of the large potatoes' inter- and intra-varietal DM variability, thus validating its use for both the DM prediction and the variety discrimination tasks.

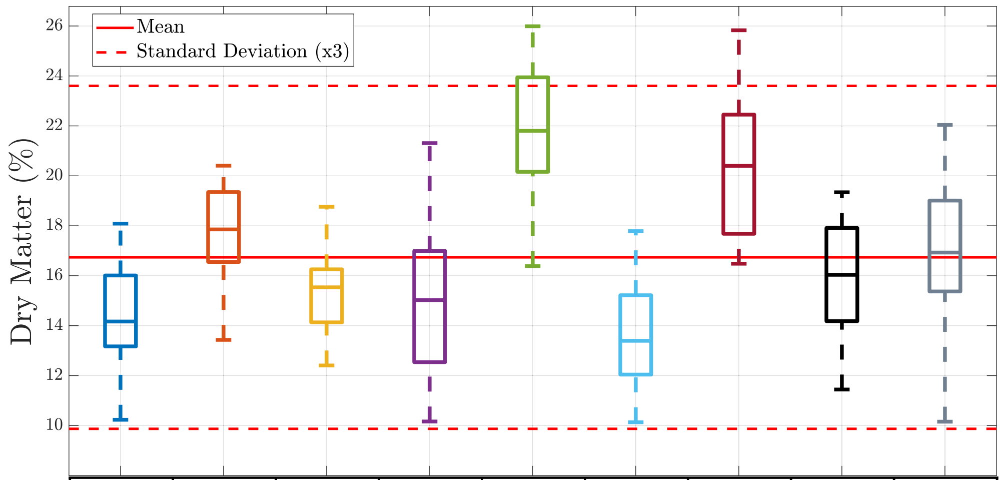
Prior to the development of the models and together with the response variable (i.e., DM), a preliminary evaluation of the predictor variable (i.e., the bioimpedance data) was necessary. This evaluation was essential to investigate possible problems arising from the data (e.g., noise, presence of outliers, etc.) and have an idea of the complexity of the subsequent modeling tasks. Fig. 3 shows the impedance amplitude and phase graphs for the potato varieties (in Fig. S1 the Nyquist plot is shown as well). In general, from the observation of both the amplitude Fig. 3a and phase angle Fig. 3b spectra, it can be seen that the nine potato varieties generate the so-called dispersion phenomenon (Ando et al., 2014). According tothis phenomenon the amplitude decreases as the frequency of the applied current increases, while the phase angle shifts towards capacitive behavior (i.e., towards −90°) at medium frequencies, to return to a more resistive behavior at higher frequencies (i.e., towards 0°). Such behavior is typical in the bioimpedance spectra of biological samples, i.e. animal tissues (Abie et al., 2021), plants (Hamed et al., 2023; Altana et al., 2024) and fruit (Riaz et al., 2023), and can be associated with the flow of current through the cell membranes and intra- and extra-cellular fluids in the tissues under test (Martinsen and Heiskanen, 2023). Here, the observed differences between the potato varieties can be related to various reasons, such as the relative chemical composition of the sample pulp, including its moisture content, and the integrity and distribution of its cellular structure. Fig. 3a and b did not show clear spectral differences among varieties that could be visually observed at higher frequencies; any possible speculation on possible differences between varieties in the shape and intensity of EIS spectra risks being detrimental compared to a machine learning approach.

As commonly recognized, the utilization of complex impedance analysis serves as a valuable tool for comprehending the intricacies associated with dielectric relaxation phenomena within a material. Numerous studies have focused on impedance measurements to investigate the dielectric properties of potatoes under various conditions, such as during drying processes (Ando et al., 2014), freezing processes (Feng et al., 2021) and both before and after a freezing-thaw cycles (Zhang and Willison, 1992). Compared to previous studies that focused on the dielectric properties of fresh potatoes under specific conditions such as drying or freezing, our approach offers a broader analysis by integrating bio-impedance data with machine learning techniques. This allows for more accurate classification of potato varieties and prediction of dry matter content, overcoming limitations of traditional methods based on visual inspection of impedance spectra.

3.2. Features reduction and selection

Fig. 4 depicts the features extracted through the four feature selection methods (i.e., CM, MRMR, NCA, and t-Test) for the three different subsets, before regression and classification analysis, respectively. The plot displays the extracted features from all four models and for all three datasets (Fig. 1). The red lines in the graphs represent the division of the x-axis, which denotes frequency values, for the five variables dependent on frequencies.

Using the CM method, the best results correspond to an analysis of the features extracted considering a frequency step (FS) of 3 for regression, while for classification an FS of 1. This feature selection process involved calculating correlation matrices among frequency-dependent features to eliminate redundant information typically found between neighboring frequency points due to high autocorrelation. Ten correlation matrices were generated, each incorporating an increasing number of grouped frequencies (from 1 to 10). Relationships among features within each matrix were then examined, identifying features with low correlation to others (differing by at least 10 %) as these were deemed more informative and less redundant (Section 2.4). In this case, the correlation matrix calculated by grouping the frequency by 3 is selected for the regression analysis, while the correlation matrix calculated by considering all the frequencies was selected for classification analysis. What can be observed from the figures is that features extracted through this method evenly cover almost the entire frequency range for all five frequency-dependent variables for both regression and classification analyses. In particular, for classification, it is possible to observe that the high-frequency features for the real part and dissipation factor variables have a higher impact concerning the others. With the application of the MRMR method, the extracted



	Actrice	Ambra	Constance	El Mundo	Fontane	Gaudi	Jelly	Monalisa	Universa
Mean	14.5	17.1	15.3	15.1	21.7	13.6	20.3	16.0	17.0
Std	1.9	1.7	1.5	3.1	2.6	2.2	2.7	2.2	2.6
Median	14.1	16.9	15.5	15.0	21.8	13.4	20.4	16.0	16.9
Max	18.1	20.4	18.8	21.3	26.0	20.4	25.8	19.3	22.0
Min	10.2	13.4	12.4	10.1	16.4	10.1	16.5	11.4	10.1

Fig. 2. Box plot of the DM distribution for the 9 potato varieties. The continuous red line represents the overall mean, while the dashed red lines represent 3 times the value of the overall standard deviation (99.7% of the dataset). The embedded table shows the descriptive statistics for each variety.

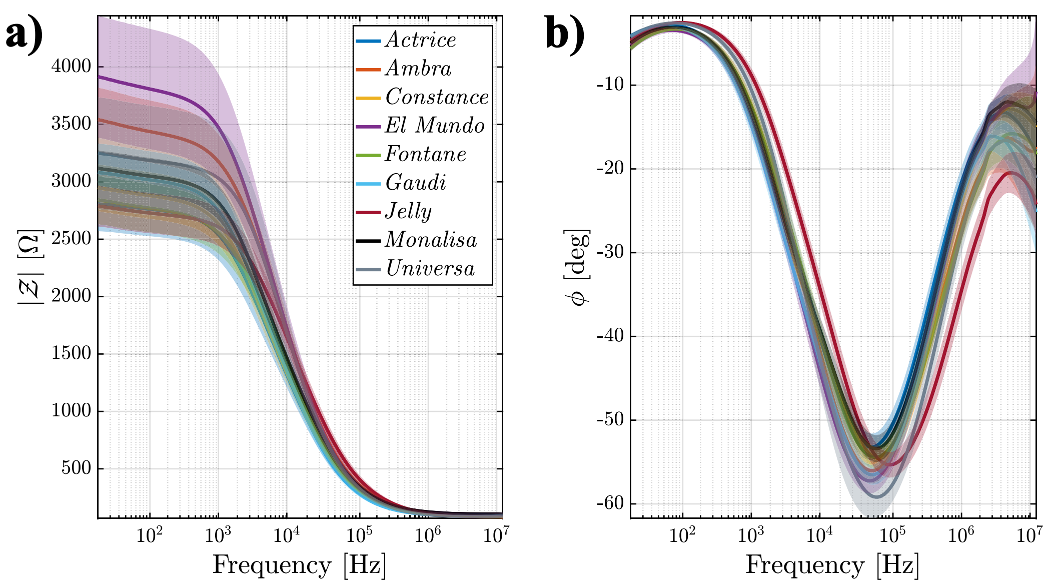


Fig. 3. Panel displaying impedance magnitude (A) and phase (B) for 9 different types of potatoes along with their corresponding standard deviations (shade). The *x*-axis is represented on a logarithmic scale to aid in data visualization. The impedance magnitude plot and phase plot each consist of nine curves, each representing a different potato type. Standard deviations are indicated in shaded form to illustrate variability within each potato type.

features also appear to be fairly evenly distributed along the frequency axis for each variable and subset, in both regression and classification. The only difference is that in regression, for the variable corresponding to impedance magnitude, in Subset 1, the features concentrate in a medium frequency range (10^3–10^4 Hz) while for Subset 2 and 3 they concentrate more at higher frequencies (10^7 Hz). In classification, however, a notable difference is observed where the variables corresponding to impedance magnitude and real part seem to carry more weight in Subset 1 compared to Subset 2 and 3. Finally, for classification purposes, this extraction method takes into account also the CPE-T variable, which is not frequency-dependent.

Regarding the NCA method, the extracted features also appear to be quite uniform along the frequency axis for the five variables, with a peculiar observation in both regression and classification: features corresponding to the magnitude and real part variables seem to be not extracted at all in Subset 2 and 3 concerning the Subset 1. Finally, when this method is employed for the features extraction for regression, it takes into account also the $Min\phi$ and CPE-P variables, which are not frequency-dependent.

For the t-test, on the other hand, what is evident from the plots is that in regression, the extracted features correspond to medium (10^2–10^4 Hz) to high frequency (10^7 Hz) intervals. In classification, the same trend is observed but with the frequency intervals corresponding to the extracted features being medium (10^2–10^4 Hz) to low (10^1–10^2 Hz). Finally, it can be noted that the features extracted for the magnitude

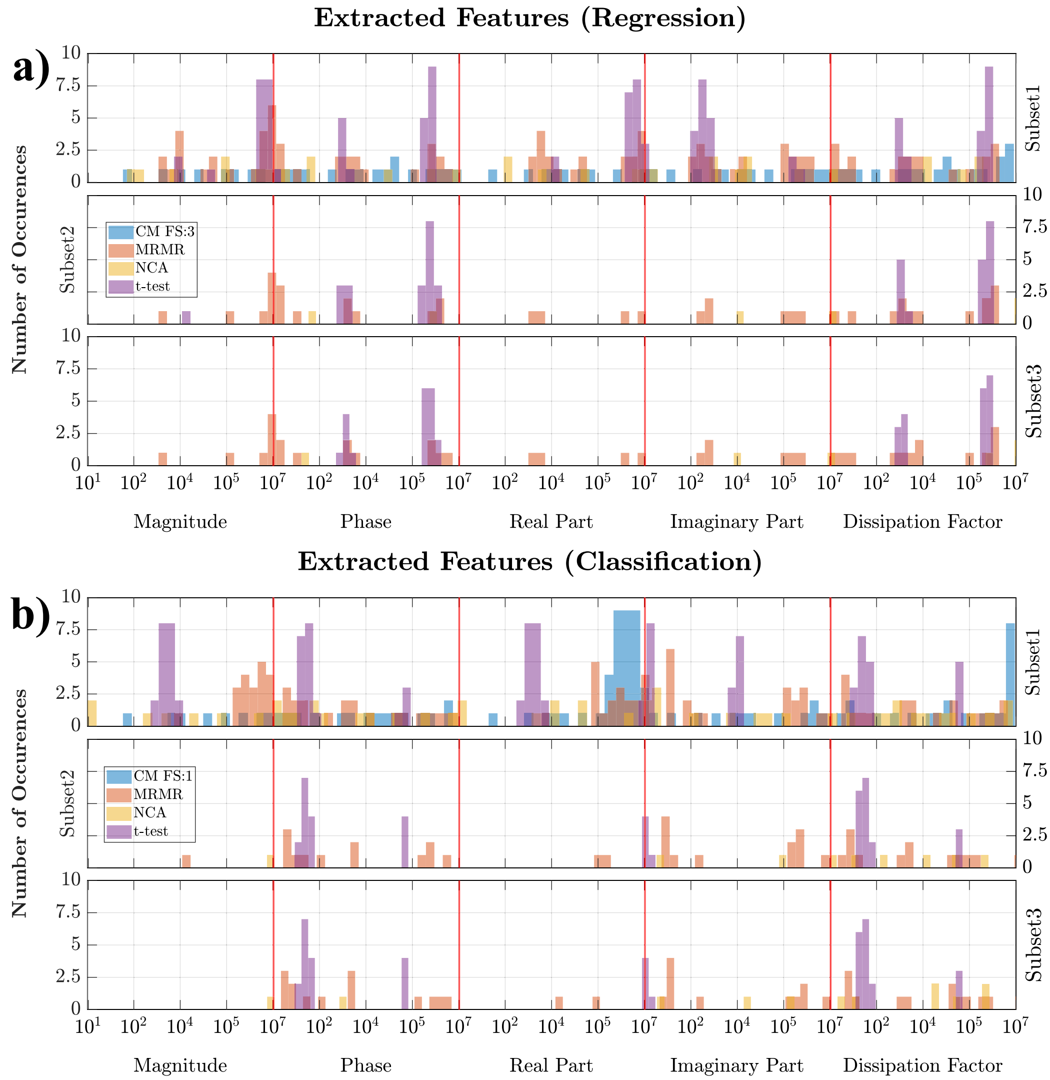


Fig. 4. Frequency (Hz) distribution of the extracted features by the four methods used (i.e., CM, MRMR, NCA, and t-test) across three different subsets (Subset 1, Subset 2, Subset 3), for regression (A) and classification (B) purposes. The plot is divided into five sections on the x-axis representing the frequency-dependent variables.

and real and imaginary part variables are significantly fewer compared to those extracted for the other variables in both regression and classification. It is important to note that most of the high-frequency features considered in this study are primarily in the lower MHz range, where the phase curves do not display any irregularities. In contrast, at frequencies above 10 MHz, anomalous behavior was observed in the phase curves. However, this limitation of the instrument (Fu and Freeborn, 2019) did not affect the development of most models, as the frequencies used for feature extraction were below this threshold. On a biological point of view, at low frequencies (kHz range), current mainly flows through the extracellular space due to the insulating effect of cell membranes. Intermediate frequencies (MHz range) allow current to penetrate the cells, providing insights into extracellular and intracellular environments. The current bypasses the membranes at high frequencies (MHz range) entirely, revealing intracellular properties. This broad frequency range captures different aspects of tissue impedance, offering a comprehensive analysis (Miklavčič et al., 2006). Moreover, an increase in the number of extracted features can elevate the risk of overfitting, as the model may capture noise rather than the underlying data patterns, reducing generalization performance (Yeom et al., 2018). This issue will be the subject of further investigation to enhance the model's performance by identifying optimal feature sets that minimize overfitting while preserving essential data patterns.

3.3. Regression

Fig. 5 shows the regression plots of the best model among the ones investigated, while Table 1, exhibits the best performance regression metrics (RMSE, R^2 and *BIAS*) for all the feature reduction methods (CM, MRMR, NCA, t-test). Here, the results were compared with those obtained from an analysis conducted on the full dataset (FD). The optimization of hyperparameters is a crucial step in the development of machine learning models, as these parameters significantly influence the model's performance (Feurer and Hutter, 2019). Thus, Table 1 also highlights the optimal hyperparameters for each model, which were carefully tuned to achieve the best predictive performance and minimize the risk of under-/over-fitting. The best prediction outcome was associated with the Subset 3 subjected to the t-test-based feature extraction method through a *NNRM* with a layers size of 5 and a *sigmoid* activation parameters (Supplementary Table S1 for all the results). The features extracted using this method for Subset 3 correspond to frequencies below the MHz range, which are well within the instrument's operational limits and do not suffer from the errors associated with higher frequencies, as detailed in Section 3.2. As depicted in Table 1, this method corresponds to the lowest RMSE value (2.3 %) in testing among those considered, leading to the fact that the difference between the predicted variables and the true ones is small (Alamoudi

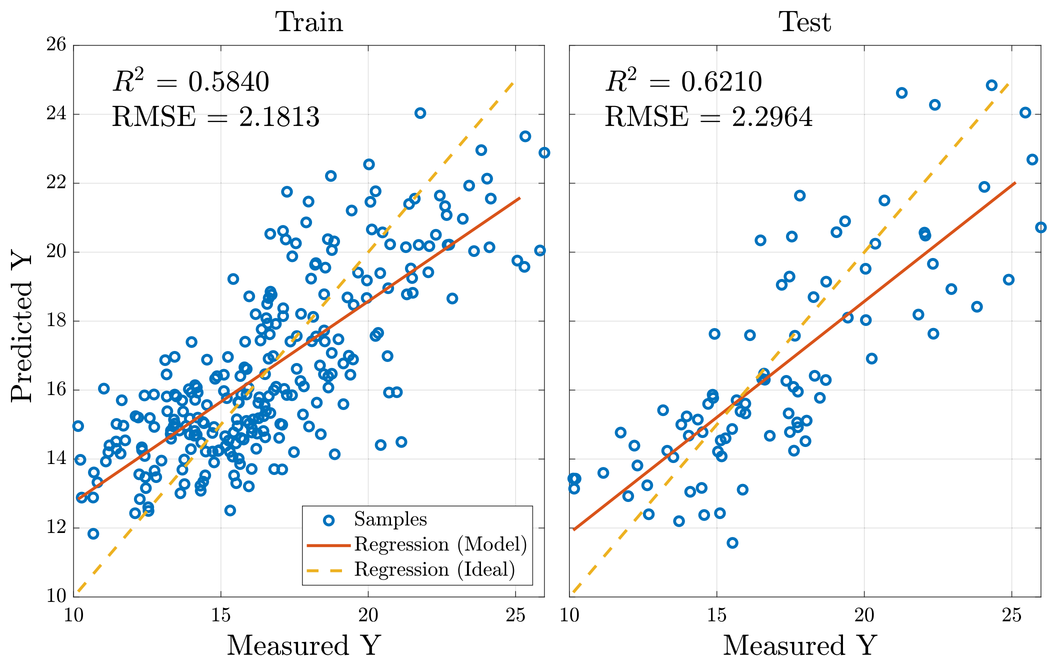


Fig. 5. Regression plot for the best regression model (i.e., Subset 3 - t-test - *NNRM* algorithm). The red line represents the regression model, while the dotted yellow line indicates the ideal regression function.

and Alamoudi, 2020). However, the difference between the training and test datasets (0.10) is not minimal, indicating that despite the low RMSE, it is not the most stable method among those analyzed. The R^2 value of this model in training is 0.58, while in testing 0.62, the higher between the R^2 testing value of the other models. The fact that our linear regression model attained the maximum R^2 in testing data suggests that the model effectively captures the relationships between the variables and generalizes well to new data, without suffering from overfitting (Alamoudi and Alamoudi, 2020). Moreover, this method exhibits a *BIAS* value in testing (0.51 %), suggesting accurate predictions with low systematic trends, although with a positive sign that indicates a tendency to overestimate values (Bilmes, 2020). Comparing our predictions of DM content in different potato varieties using machine learning techniques based on bioimpedance data with those based on near-infrared (NIR) spectra from the literature, the NIR spectra generally yield more accurate results (Wang et al., 2023; Helgerud et al., 2015, 2012; Camps and Camps, 2019). However, the NIR technique is less efficient in terms of time and cost. To enhance our prediction model, integrating datasets obtained from both impedance measurements and NIR spectra through data fusion could be a promising approach. This combined method may leverage the strengths of both techniques, potentially improving predictive accuracy while maintaining a balance between efficiency and cost-effectiveness. Different feature selection methods produced varying results in terms of regression performance. For instance, the model using the full dataset (FD) with 947 features achieved a lower RMSE during training (1.9%) and testing (2.4%) but showed a difference of 0.5% between training and testing RMSE, suggesting a potential risk of overfitting. Very similar is the MRMR (Subset 2) with 40 features, which exhibited slightly lower RMSE during training (1.7%) and testing (2.4%), but a more consistent difference of 0.7%. The NCA method on Subset 2, with only 6 features extracted, seems to be the worst one in terms of R^2, but it appears to be the most stable. Finally, it presents also the worst case in terms of RMSE.

3.4. Classification

In Table 2, classification analysis results for potato varieties are presented. These results were then compared with those obtained from an analysis conducted on the full dataset, without any feature selection. The table displays F1-score results, utilized in this study as a metric for identifying the optimal model (Vujović et al., 2021; Ibba et al., 2021), alongside accuracy, sensitivity, and precision metrics. Only the best results for each feature extraction method and subset are reported in the table (Supplementary Table S 2 for all the results). Upon comparison of the obtained results, it is evident that the optimal classification model pertains to the analysis conducted on Subset 1 following the CM features selection method considering all the frequencies (FS:1). This model exhibits an F1-score of 0.92 in testing, the highest among all, and 0.96 in training, close to the highest. However, considering the difference between training and testing of approximately 0.04 (−4.3%), it shows good stability. The algorithm that achieved the best classification model was *NNC*. This model was configured with three layer sizes of different weights and with *none* activation parameters. It also showed the best values, in testing, regarding the accuracy, sensitivity, and precision parameters. The confusion matrices presented in Fig. 6 offer a detailed visualization of the classification performance of the best model. In the training set, the model demonstrates robust classification accuracy (0.96), with high true positive rates in categories such as *Actrice* (34 correct classifications) and *Universa* (33 correct classifications). Notably, the model shows a consistent ability to distinguish between most categories with minimal confusion, as evidenced by the sparse off-diagonal elements, which represent confused classification. For instance, *El Mundo* was misclassified once as *Ambra* and once as *Fontane*, indicating slight confusion which could be attributed to overlapping features within these categories.

On the testing side, the matrix reveals some variances in performance. While *Jelly* maintained strong correct classifications (30 correct classifications in training and 10 in testing) with 0 % of misclassification in train and test, some categories like *Ambra* and *Monalisa* experienced a noticeable decrease in predictive accuracy compared to the training phase. For example, *Ambra* was correctly classified four times and misclassified two (33.33%) in the test set, down from twenty-two times correct classifications and two misclassifications (9.09%) in training, while *Monalisa* was correctly classified eleven times and misclassified two times (18.18%) in test, down from twenty-six times correct classifications and one misclassification (3.85%) in training. The model using FD with 947 features yielded a high training accuracy

Table 1
Summary results of the non-linear regression model evaluation using different feature selection methods for DM prediction in potato tubers using EIS spectra.

Regression Model											
FSM	*NoF*	*Algorithm*	*Optimized Hyperparameters*	*RMSE (%)*			R^2			*BIAS (%)*	
				Train	*Test*	*Diff.*	*Train*	*Test*	*Diff.*	*Train*	*Test*
FD	947	***KRM***	***NEM*** = 4341, ***L*** = *'Lastsquares'*, λ = 4.3e−06	1.9	2.4	0.50	0.69	0.59	0.10	6.2×10^{-5}	0.32
CM (FS:3)	64	***KRM***	***NEM*** = 445, ***L*** = *'Lastsquares'*, λ = 4.0e−06	2.1	2.5	0.40	0.60	0.56	0.040	1.2×10^{-4}	0.31
MRMR (Subset2)	40	***GPRM***	***BF*** = *'Constant'*, ***KF*** = *'RationalQuadratic'*, σ = 1.5	1.7	2.4	0.70	0.73	0.59	0.14	−0.0043	0.37
NCA (Subset2)	6	***ERM***	***NN*** = 3	2.6	2.7	0.10	0.40	0.48	0.080	2.1×10^{-9}	0.67
t-test (Subset3)	40	***NNRM***	***LS*** = 5, ***A*** = *'Sigmoid'*	2.2	2.3	0.10	0.58	0.62	0.040	−0.0040	0.51

FSM = Features Selection Method, ***NoF*** = Number of Features
NEM = NumExpansionDimensions, ***L*** = Learner, ***BF*** = BasisFunction, *KF* = KernelFunction, ***NN*** = NumNodes, ***LS*** = LayerSize, ***A*** = Activation

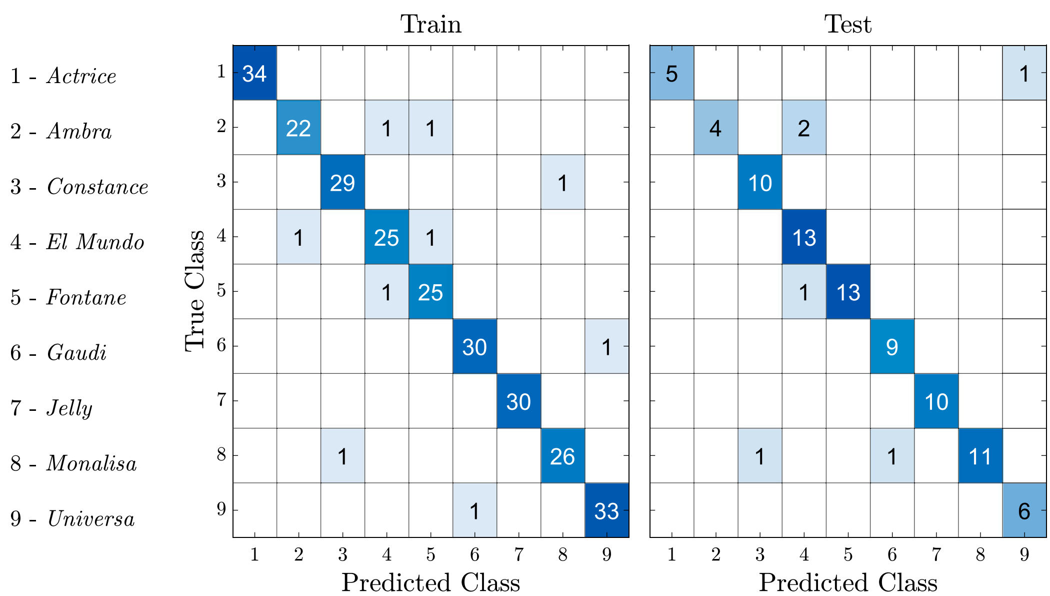


Fig. 6. Confusion matrices of the best (i.e. Subset 1 - CM, FS:1 - *NNC* algorithm) classification model for training (left) and test (right) data. On the left-hand side, there is the list of the potato varieties in which the corresponding class of the confusion matrix is indicated.

(94 %) and test accuracy (92 %) with a minor difference of 2 %, exhibiting reduced overfitting. In contrast, the model using MRMR (Subset 2) with 40 features showed slightly lower accuracy in testing (91 %) while in training it holds the same accuracy (94 %), but the difference was 3 %. Compared to existing studies, our approach to classifying potato varieties using bio-impedance measurements has shown comparable or even superior results in terms of accuracy and predictive capabilities. Previous works in the literature, such as those by Ibba et al. (2021), Lan et al. (2020), Bria et al. (2020), Zhu et al. (2019), have primarily focused on classification tasks using alternative techniques. However, these approaches often require more complex setups or pre-processing steps. In contrast, our method, leveraging bio-impedance data and machine learning, provides a simpler and more cost-effective solution, while maintaining high classification performance.

4. Conclusions

Prediction of DM and classification of nine potato varieties, based on impedance (and derived parameters) data were investigated in this study. Three different conventional feature extraction methods (MRMR, NCA, t-test) were involved in three different feature selection steps generating three different subsets (Subset 1, 2, and 3). Moreover, a features extraction method based on correlation matrices (CM) was applied to the frequency-dependent variables (Subset 1), taking into account different groups of frequencies (Figs. 1 and 2.4). The prediction analysis was done through two predefined MATLAB® functions (Section 2.4). The results obtained were then compared to the analysis done with the full dataset, without any feature extraction process. The results of the regression analysis reported that, among the various methods used, the model that achieved the best regression result was based on the analysis conducted on Subset 3, using a feature extraction method based on the t-test and the *NNRM* as regression algorithm. This model produced an RMSE of 2.3% and R^2 of 0.621 in testing. While not extremely low, the RMSE value was acceptable, indicating reasonable prediction accuracy. The R^2 coefficient of 0.621 suggested that the model can explain approximately 62.1% of the variation in potato dry matter, providing a good approximation of the observed data. Regarding the classification of potato varieties, the best result was obtained through a feature extraction method based on the *CM* with a frequency step of 1 (Section 2.4). This model produced an F_1 score of 0.964, indicating high precision and recall in classifying potato types.

Table 2
Summary results of the non-linear classification model evaluation using different feature selection methods for DM prediction in potato tubers using EIS spectra.

Classification Model													
FSM	*NoF*	*Algorithm*	*Optimized Hyperparameters*	F_1			*Accuracy*		*Sensitivity*		*Precision*	
				Train	*Test*	*Diff*	*Train*	*Test*	*Train*	*Test*	*Train*	*Test*
FD	947	***KCM***	***NED*** = 4460, ***L*** = *'logistic'*, λ = 5.5e−06	0.93	0.90	0.033	0.94	0.92	0.93	0.90	0.93	0.92
CM (FS:1)	98	***NNC***	***LS*** = [47,14,43], ***A*** = 'none'	0.96	0.92	0.041	0.97	0.93	0.96	0.92	0.96	0.94
MRMR (Subset 2)	40	***NNC***	***LS*** = [11,120,133], ***A*** = 'none'	0.94	0.89	0.043	0.94	0.91	0.94	0.89	0.94	0.92
NCA (Subset 1)	51	***KCM***	***NED*** = 628, ***L*** = *'logistic'*, λ = 6.7e−06	0.97	0.90	0.075	0.97	0.92	0.97	0.90	0.97	0.92
t-test (Subset 1)	40	***NNC***	***LS*** = [61,75], ***A*** = 'tanh'	0.96	0.90	0.054	0.96	0.91	0.96	0.90	0.96	0.92

FSM = Features Selection Method, ***NoF*** = Number of Features ***NEM*** = NumExpansionDimensions, ***L*** = Learner, ***LS*** = LayerSize, ***A*** = Activation

In conclusion, the application of machine learning methods for predicting potato DM and varieties, based on impedance data, holds promise, as they reach the same precision as conventional methods. Given that the results of the regression model are promising but not yet optimal, a more in-depth analysis could be undertaken. This deeper analysis should aim to reduce the root-mean-squared error and increase the coefficient of determination value, thereby enhancing the accuracy of dry matter data predictions. To achieve this, various techniques such as feature engineering, hyperparameter tuning, and advanced modeling approaches (e.g. convolutional neural networks) could be explored. Random splitting may introduce bias or inconsistency in model evaluation, and thus, employing methods like the Kennard-Stone algorithm, which selects representative samples based on distance criteria, could lead to more robust dataset partitioning. Additionally, incorporating data fusion with results obtained through infrared spectroscopy could further improve the model's performance. By integrating multiple data sources, we could take advantage of complementary information, which may lead to more robust and reliable predictive models. This integrated approach not only has the potential to refine the regression model but also to provide a more comprehensive understanding of the underlying patterns in the data.

CRediT authorship contribution statement

Ciro Allará: Writing – original draft, Methodology, Formal analysis, Data curation. **Roberto Moscetti:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Giacomo Bedini:** Writing – review & editing, Investigation, Conceptualization. **Manuela Ciocca:** Writing – review & editing. **Alessandro Benelli:** Writing – review & editing, Validation. **Paolo Lugli:** Supervision, Funding acquisition. **Luisa Petti:** Writing – review & editing, Supervision. **Pietro Ibba:** Writing – review & editing, Validation, Supervision, Data curation, Conceptualization.

Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used the free-to-use AI system ChatGPT (https://chat.openai.com/) in order to improve language and readability of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was carried out within the Free University of Bozen-Bolzano and received funding from the European Union Next GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) - Missione 4, Componente 1, Investimento 4.1 - D.M. 351/2022) and Agritech National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This work was supported by the Open Access Publishing Fund of the Free University of Bozen-Bolzano. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.postharvbio.2024.113358.

Data availability

Data will be made available on request.

References

Abie, S.M., Martinsen, Ø.G., Egelandsdal, B., Hou, J., Bjerke, F., Mason, A., Münch, D., 2021. Feasibility of using electrical impedance spectroscopy for assessing biological cell damage during freezing and thawing. Sensors 21 (12), 4129. http://dx.doi.org/10.3390/s21124129.

Adesokan, M., Otegbayo, B., Alamu, E.O., Olutoyin, M.A., Maziya-Dixon, B., 2024. Evaluating the dry matter content of raw yams using hyperspectral imaging spectroscopy and machine learning. J. Food Comp. Anal. 135, 106692. http://dx.doi.org/10.1016/j.jfca.2024.106692.

Alamoudi, R., Alamoudi, M., 2020. Development of linear regression models to estimate the margin of stability based on spatio-temporal gait parameters. IEEE access 8, 19853–19859. http://dx.doi.org/10.1109/ACCESS.2020.2969294.

Altana, A., Hamed, S., Lugli, P., Petti, L., Ibba, P., 2024. Monitoring iron stress in tomato plants through bioimpedance and machine-learning-enhanced classification based on circuit component analysis. IEEE Trans. AgriFood Electron. 1–8. http://dx.doi.org/10.1109/TAFE.2024.3411269.

Ando, Y., Mizutani, K., Wakatsuki, N., 2014. Electrical impedance analysis of potato tissues during drying. J. Food Eng. 121, 24–31. http://dx.doi.org/10.1016/j.fufo.2023.100241.

Assembly, U.G., 2015. UN General Assembly, Transforming Our World: the 2030 Agenda for Sustainable Development, 21 October 2015. Technical Report, A/RES/70/1.

Bedini, G., Chakravartula, S.N., Nardella, M., Bandiera, A., Massantini, R., Moscetti, R., 2023. Prediction of potato dry matter content by FT-NIR spectroscopy: impact of tuber tissue on model performance. Future Foods 8, 100241. http://dx.doi.org/10.1016/j.fufo.2023.100241.

Bedini, G., Haff, R.P., Benelli, A., Bandiera, A., Taormina, E., Massantini, R., Moscetti, R., 2024. Potatoes (*Solanum tuberosum* L.) grown at "patata dell'alto viterbese" PGI have different quality characteristics and storage responses. Postharvest Biol. Technol. 214, 112991. http://dx.doi.org/10.1016/j.postharvbio.2024.112991.

Bilmes, J., 2020. Underfitting and overfitting in machine learning. In: UW ECE Course Notes, vol. 5.

Bonamutial, M., Prasetyo, S.Y., 2023. Exploring the impact of feature data normalization and standardization on regression models for smartphone price prediction. In: 2023 International Conference on Information Management and Technology. ICIMTech, IEEE, pp. 294–298. http://dx.doi.org/10.1109/ICIMTech59029.2023.10277860.

Brancato, V., Brancati, N., Esposito, G., La Rosa, M., Cavaliere, C., Allarà, C., Romeo, V., De Pietro, G., Salvatore, M., Aiello, M., et al., 2023. A two-step feature selection radiomic approach to predict molecular outcomes in breast cancer. Sensors 23 (3), 1552. http://dx.doi.org/10.3390/s23031552.

Breiman, L., 2017. Classification and Regression Trees. Routledge, http://dx.doi.org/10.1038/nmeth.4370.

Bria, A., Cerro, G., Ferdinandi, M., Marrocco, C., Molinara, M., 2020. An IoT-ready solution for automated recognition of water contaminants. Pattern Recognit. Lett. 135, 188–195. http://dx.doi.org/10.1016/j.patrec.2020.04.019.

Broćić, Z., Dolijanović, Ž., Poštić, D., Milošević, D., Savić, J., 2016. Yield, tuber quality and weight losses during storage of ten potato cultivars grown at three sites in Serbia. Potato Res. 59, 21–34. http://dx.doi.org/10.1007/s11540-015-9311-7.

Camps, C., Camps, Z.-N., 2019. Optimized prediction of reducing sugars and dry matter of potato frying by FT-NIR spectroscopy on peeled tubers. Molecules 24 (5), 967. http://dx.doi.org/10.3390/molecules24050967.

Cole, K.S., 1940. Permeability and impermeability of cell membranes for ions. In: Cold Spring Harbor Symposia on Quantitative Biology, vol. 8, Cold Spring Harbor Laboratory Press, pp. 110–122. http://dx.doi.org/10.1101/SQB.1940.008.01.013.

Darbellay, G.A., Vajda, I., 1999. Estimation of the information by an adaptive partitioning of the observation space. IEEE Trans. Inform. Theory 45 (4), 1315–1321. http://dx.doi.org/10.1109/18.761290.

Dean, D., Ramanathan, T., Machado, D., Sundararajan, R., 2008. Electrical impedance spectroscopy study of biological tissues. J. Electrostat. 66 (3–4), 165–177. http://dx.doi.org/10.1016/j.elstat.2007.11.005.

Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 3 (02), 185–205. http://dx.doi.org/10.1109/CSB.2003.1227396.

Dzierżak, R., 2019. Comparison of the influence of standardization and normalization of data on the effectiveness of spongy tissue texture classification. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie środowiska 9, http://dx.doi.org/10.35784/iapgos.62.

Escuredo, O., Meno, L., Rodríguez-Flores, M.S., Seijo, M.C., 2021. Rapid estimation of potato quality parameters by a portable near-infrared spectroscopy device. Sensors 21 (24), 8222. http://dx.doi.org/10.3390/s21248222.

European food safety authority (EFSA), 2022. https://www.efsa.europa.eu/en/topics/topic/acrylamide. (Accessed: 16 November 2022).

Feng, L., Cheng, L., Zhang, B., Jiang, B., Dong, Z., 2020. Estimating moisture content and physical properties of sweet potato based on the multiparameter of electrical impedance spectroscopy. J. Food Process. Preserv. 44 (10), e14775. http://dx.doi.org/10.1111/jfpp.14775.

Feng, L., Hou, T., Zhang, B., 2021. A noninvasive method for detecting frozen injuries in potatoes based on electrical impedance spectroscopy. J. Food Process Eng. 44 (5), e13682. http://dx.doi.org/10.1111/jfpe.13682.

Feurer, M., Hutter, F., 2019. Hyperparameter optimization. In: Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing, pp. 3–33. http://dx.doi.org/10.1007/978-3-030-05318-5_1.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7 (2), 179–188. http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Food and Agriculture Organization of the United Nations (FAO), 2024. https://www.fao.org/faostat/en/#data/QCL. (Accessed: 01 June 2024).

Fu, B., Freeborn, T.J., 2019. Residual impedance effect on emulated bioimpedance measurements using keysight E4990a precision impedance analyzer. Measurement 134, 468–479. http://dx.doi.org/10.1016/j.measurement.2018.10.080.

Grossi, M., Riccò, B., 2017. Electrical Impedance Spectroscopy (EIS) for biological analysis and food characterization: a review. J. Sensors Sensor Syst. 6 (2), 303–325. http://dx.doi.org/10.5194/jsss-6-303-2017,2017.

Hamed, S., Ibba, P., Altana, A., Lugli, P., Petti, L., 2023. Towards tomato plant iron stress monitoring through bioimpedance and circuit analysis. In: 2023 IEEE Conference on AgriFood Electronics. CAFE, IEEE, pp. 20–24. http://dx.doi.org/10.1109/CAFE58535.2023.10292001.

Helgerud, T., Segtnan, V.H., Wold, J.P., Ballance, S., Knutsen, S.H., Rukke, E.O., Afseth, N.K., 2012. Near-infrared spectroscopy for rapid estimation of dry matter content in whole unpeeled potato tubers. J. Food Res. 1 (4), 55. http://dx.doi.org/10.5539/jfr.v1n4p55.

Helgerud, T., Wold, J.P., Pedersen, M.B., Liland, K.H., Ballance, S., Knutsen, S.H., Rukke, E.O., Afseth, N.K., 2015. Towards on-line prediction of dry matter content in whole unpeeled potatoes using near-infrared spectroscopy. Talanta 143, 138–144. http://dx.doi.org/10.1016/j.talanta.2015.05.037.

Hertog, M., Tijskens, L., Hak, P., 1997. The effects of temperature and senescence on the accumulation of reducing sugars during storage of potato (*Solanum tuberosum* L.) tubers: A mathematical model. Postharvest Biol. Technol. 10 (1), 67–79. http://dx.doi.org/10.1016/S0925-5214(97)87276-6.

Horwitz, W., Latimer, G.W., et al., 2000. Official Methods of Analysis of AOAC International, vol. 1, AOAC international Gaithersburg.

Ibba, P., Falco, A., Abera, B.D., Cantarella, G., Petti, L., Lugli, P., 2020. Bio-impedance and circuit parameters: an analysis for tracking fruit ripening. Postharvest Biol. Technol. 159, 110978. http://dx.doi.org/10.1016/j.postharvbio.2019.110978.

Ibba, P., Tronstad, C., Moscetti, R., Mimmo, T., Cantarella, G., Petti, L., Martinsen, Ø.G., Cesco, S., Lugli, P., 2021. Supervised binary classification methods for strawberry ripeness discrimination from bioimpedance data. Sci. Rep. 11 (1), 11202. http://dx.doi.org/10.1038/s41598-021-90471-5.

Kasampalis, D.S., Tsouvaltzis, P., Ntouros, K., Gertsis, A., Moshou, D., Siomos, A.S., 2021. Rapid nondestructive postharvest potato freshness and cultivar discrimination assessment. Appl. Sci. 11 (6), 2630. http://dx.doi.org/10.3390/app11062630.

Lan, H., Wang, Z., Niu, H., Zhang, H., Zhang, Y., Tang, Y., Liu, Y., 2020. A nondestructive testing method for soluble solid content in Korla fragrant pears based on electrical properties and artificial neural network. Food Sci. Nutr. 8 (9), 5172–5181. http://dx.doi.org/10.1002/fsn3.1822.

Manning, C.D., 2008. Introduction to Information Retrieval. Syngress Publishing.

Martinsen, O.G., Heiskanen, A., 2023. Bioimpedance and Bioelectricity Basics. Elsevier.

Miklavčič, D., Pavšelj, N., Hart, F.X., 2006. Electric properties of tissues. In: Wiley Encyclopedia of Biomedical Engineering. John Wiley & Sons, Inc. Hoboken, NJ, USA.

Nash, W.J., Sellers, T.L., Talbot, S.R., Cawthorn, A.J., Ford, W.B., 1994. The population biology of abalone (*Haliotis* species) in Tasmania. Blacklip abalone (*H. Rubra*) from the north coast and islands of bass strait. Sea Fish. Div., Tech. Rep. 48, p411.

Nicosia, G., Ojha, V., La Malfa, E., La Malfa, G., Pardalos, P.M., Umeton, R., 2024. Machine Learning, Optimization, and Data Science: 9th International Conference, lod 2023, Grasmere, uk, September 22-26, 2023, Revised Selected Papers. Springer, http://dx.doi.org/10.1007/978-3-031-53969-5.

Nuwamanya, E., Wembabazi, E., Kanaabi, M., Namakula, F.B., Katungisa, A., Lyatumi, I., Esuma, W., Alamu, E.O., Dufour, D., Kawuki, R., et al., 2024. Development and validation of near-infrared spectroscopy procedures for prediction of cassava root dry matter and amylose contents in Ugandan cassava germplasm. J. Sci. Food Agric. 104 (8), 4793–4800. http://dx.doi.org/10.1002/jsfa.12966.

Pliquett, U., Altmann, M., Pliquett, F., Schöberlein, L., 2003. Py—a parameter for meat quality. Meat Sci. 65 (4), 1429–1437. http://dx.doi.org/10.1016/S0309-1740(03)00066-4.

Rady, A.M., Guyer, D.E., 2015. Evaluation of sugar content in potatoes using NIR reflectance and wavelength selection techniques. Postharvest Biol. Technol. 103, 17–26. http://dx.doi.org/10.1016/j.postharvbio.2015.02.012.

Riaz, S., Ibba, P., Nadja, S., Rasheed, A., Lugli, P., Zanella, A., Petti, L., 2023. Exploring the potential of electrical impedance spectroscopy for predicting internal browning in apples. In: 2023 IEEE International Workshop on Metrology for Agriculture and Forestry. MetroAgriFor, IEEE, pp. 414–418. http://dx.doi.org/10.1109/MetroAgriFor58484.2023.10424025.

Strand-Amundsen, R., Tronstad, C., Kalvøy, H., Gundersen, Y., Krohn, C., Aasen, A., Holhjem, L., Reims, H., Martinsen, Ø.G., Høgetveit, J., et al., 2016. In vivo characterization of ischemic small intestine using bioimpedance measurements. Physiol. Meas. 37 (2), 257. http://dx.doi.org/10.1088/0967-3334/37/2/257.

Tolessa, E.S., 2018. Importance, nutrient content and factors affecting nutrient content of potato. Am. J. Food Nutr. Health 3 (3), 37–41.

Vujović, Ž., et al., 2021. Classification model evaluation metrics. Int. J. Adv. Comput. Sci. Appl. 12 (6), 599–606. http://dx.doi.org/10.14569/IJACSA.2021.0120670.

Wang, S., Yan, J., Tian, S., Tian, H., Xu, H., 2023. Vis/NIR model development and robustness in prediction of potato dry matter content with influence of cultivar and season. Postharvest Biol. Technol. 197, 112202. http://dx.doi.org/10.1016/j.postharvbio.2022.112202.

Watanabe, T., Nakamura, N., Sakamoto, K., Nagata, M., 2021. Bio-electrical impedance analysis of "silk sweet" sweet potato tissues under low-temperature storage using a novel indicator, LTO. Biosyst. Eng. 206, 1–5. http://dx.doi.org/10.1016/j.biosystemseng.2021.03.009.

Waugh, S.G., 1995. Extending and Benchmarking Cascade-Correlation: Extensions to the Cascade-Correlation Architecture and Benchmarking of Feed-Forward Supervised Artificial Neural Networks (Ph.D. thesis). University of Tasmania.

Whitehead, T., 1954. USDA issues grade standards for frozen french fried potatoes. Am. Potato J. 31, 384–385. http://dx.doi.org/10.1007/BF02866274.

Williams, C.K., Rasmussen, C.E., 2006. Gaussian Processes for Machine Learning, vol. 2, (no. 3), MIT press Cambridge, MA.

Wold, J.P., O'Farrell, M., Andersen, P.V., Tschudi, J., 2021. Optimization of instrument design for in-line monitoring of dry matter content in single potatoes by NIR interaction spectroscopy. Foods 10 (4), 828. http://dx.doi.org/10.3390/foods10040828.

Yeom, S., Giacomelli, I., Fredrikson, M., Jha, S., 2018. Privacy risk in machine learning: analyzing the connection to overfitting. In: 2018 IEEE 31st Computer Security Foundations Symposium. CSF, IEEE, pp. 268–282. http://dx.doi.org/10.1109/CSF.2018.00027.

Zaheer, K., Akhtar, M.H., 2016. Potato production, usage, and nutrition—a review. Crit. Rev. Food Sci. Nutr. 56 (5), 711–721. http://dx.doi.org/10.1080/10408398.2012.724479.

Zhang, M., Willison, J., 1992. Electrical impedance analysis in plant tissues: the effect of freeze-thaw injury on the electrical properties of potato tuber and carrot root tissues. Canadian J. Plant Sci. 72 (2), 545–553. http://dx.doi.org/10.4141/cjps92-068.

Zhu, H., Liu, F., Ye, Y., Chen, L., Liu, J., Gui, A., Zhang, J., Dong, C., 2019. Application of machine learning algorithms in quality assurance of fermentation process of black tea–based on electrical properties. J. Food Eng. 263, 165–172. http://dx.doi.org/10.1016/j.jfoodeng.2019.06.009.

Zou, Y., Guo, Z., 2003. A review of electrical impedance techniques for breast cancer detection. Med. Eng. Phys. 25 (2), 79–90. http://dx.doi.org/10.1016/S1350-4533(02)00194-7.