Non-Invasive Prediction of Sweet Cherry Soluble Solids Content Using Dielectric Spectroscopy and Down-Sampling Techniques

by Sacilik, K., Cetin, N., Ozbey, B and Cheein, F.A.

Harper Adams University

Journal Pre-proof

Non-Invasive Prediction of Sweet Cherry Soluble Solids Content Using Dielectric Spectroscopy and Down-Sampling Techniques

Kamil Sacilik , Necati Cetin , Burak Ozbey ,
Fernando Auat Cheein
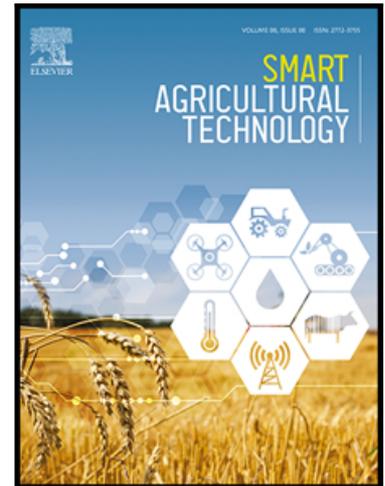
Please cite this article as: Kamil Sacilik , Necati Cetin , Burak Ozbey , Fernando Auat Cheein , Non-Invasive Prediction of Sweet Cherry Soluble Solids Content Using Dielectric Spectroscopy and Down-Sampling Techniques, *Smart Agricultural Technology* (2025), doi: https://doi.org/10.1016/j.atech.2025.100782

Non-Invasive Prediction of Sweet Cherry Soluble Solids Content Using Dielectric Spectroscopy and Down-Sampling Techniques

Kamil Sacilik[a], Necati Cetin[a*], Burak Ozbey[b], Fernando Auat Cheein[c*]

[a] *Department of Agricultural Machinery and Technologies Engineering, Faculty of Agriculture, Ankara University, Ankara, Türkiye*

[b] *Department of Electrical and Electronics Engineering, Faculty of Engineering, Ankara University, 06830 Ankara, Türkiye*

[c] *Department of Engineering, Harper-Adams University, Newport, United Kingdom*

*Correspondence: necati.cetin@ankara.edu.tr (Necati Cetin), FAuat@harper-adams.ac.uk (Fernando Auat Cheein)

Abstract

The soluble solid content (SSC) in fruits significantly influences consumers' taste, aroma, and flavor preferences. It also plays a crucial role for farmers and wholesalers in determining the optimal harvest period for marketing. Dielectric spectroscopy, an innovative and non-invasive technique, has shown promise for various applications in the food and agriculture sectors. This study introduces an open-ended coaxial line probe measurement system to non-invasively determine the SSC of sweet cherries at different radio and microwave frequencies. Key parameters such as the dielectric constant (ε'), loss factor (ε''), loss tangent (tan δ), and SSC of sweet cherries were measured across different harvest periods. The dielectric property frequency ranges were down-sampled from 300 MHz to 15 MHz. Using dielectric spectroscopy, we implemented predictive models: support vector regression (SVR) and multilayer perceptron (MLP), that demonstrated extremely low MAE and RMSE, with correlation coefficients (R) exceeding 0.97 for SVR and 0.96 for MLP. The down-sampled frequency ranges for dielectric

properties yielded consistently high performance across all subsets, demonstrating comparable results. These findings suggest that a dielectric measurement system designed for SSC estimation using fewer frequencies could effectively reduce costs while maintaining accuracy.

Key words: Sweet cherries; down-sampling; dielectric spectroscopy; soluble solid content; machine learning

Nomenclature

ANFIS	Adaptive neuro-fuzzy inference system
BAG	Bagging
BP-ANN	Backpropagation – artificial neural network
CARS	Competitive adaptive reweighted sampling
ELM	Extreme learning machine
GP	Gaussian processes
kNN	k-nearest neighbors
LSSVM	The least squares support vector machine
MAE	Mean absolute error
ML	Machine learning
MLP	Multilayer perceptron
MLR	Multiple linear regression
PCA	Principal component analysis
PI-RF-	Permutation importance based on the random forest
XGBoost	regression - extreme gradient boosting
PSO	Particle swarm optimization
PLSR	Partial least squares regression
PTFE	Polytetrafluoroethylene
PUK	Pearson VII kernel function
R	Correlation coefficient
REPTree	Reduced-error pruning tree
RF	Random forest
RMSE	Root mean square error
SMA	SubMiniature version A
SPA	Successive projection algorithm
SSC	Soluble solid content, /Brix
SVM	Support vector machine
SVR	Support vector regression
$tan\,\delta$	Loss tangent
Z_0	Impedance of free space, Ω
Z_r	Impedance of the material under test, Ω
ε^*	Complex relative permittivity
ε'	Dielectric constant
ε''	Loss factor

ε_0	The permittivity of free space (8.854×10^{-12} F/m)
ε	Absolute permittivity, F/m
μ	Magnetic permeability, H/m
μ_0	The magnetic permeability of free space, H/m
Γ	Fresnel reflection coefficient

1. Introduction

Cherries are rich in essential nutrients and bioactive compounds, including fiber, polyphenols, carotenoids, vitamin C, potassium, tryptophan, serotonin, and melatonin, while being low in calories (Hussain et al., 2021). They are classified into sweet and sour types. Sweet cherries are primarily consumed fresh, whereas sour cherries are used in various food products and as fruit juice (Li et al., 2018). In 2022, cherries were cultivated worldwide on 454,664 hectares, yielding an annual production of approximately 2,765,827.4 tons. The five leading cherry-producing countries were Türkiye, with the highest production of 656,041 tons, followed by Chile (443,067 tons), Uzbekistan (216,867 tons), the United States (210,190 tons), and Spain (116,070 tons) (FAO, 2024).

The soluble solid content (SSC) of cherries is a critical indicator of their nutritional value, influencing flavor, sweetness, and overall consumer acceptability. SSC primarily comprises sugars, acids, vitamins, and other soluble substances that enhance the fruit's nutritional profile and taste (Correia et al., 2017). The relationship between SSC and other physicochemical properties, such as firmness and acidity, further underscores its importance in determining the nutritional and sensory quality of cherries. Firmer cherries often exhibit higher SSC levels and appeal more to consumers (Hong et al., 2010). Harvesting fruits at optimal ripeness is crucial to ensure good eating quality. Consequently, SSC serves as a standard internal quality parameter, particularly for Prunus species (Escribano et al., 2017).

Hyperspectral imaging and visible near-infrared spectroscopy, often combined with advanced chemometric techniques, are the primary non-destructive methods for determining

fruit SSC (Li et al., 2018; Çetin et al., 2022). These techniques enable rapid and precise assessments without causing damage to the fruit. However, their limited penetration depth restricts their applicability despite their widespread use in assessing internal quality attributes (Cao et al., 2023). The irregular distribution of sugars within fruits, especially in stone fruits, and the effects of the peel necessitate greater penetration depth to accurately determine internal quality attributes (Cao et al., 2024a). Additionally, weather conditions—particularly temperature and humidity during the final month of fruit development—significantly affect SSC accumulation in cherries (Ivanova et al., 2021).

Due to their wide frequency range and substantial penetration depth, dielectric properties are frequently employed for non-destructive internal quality measurements in fruits (Nelson, 2015). Dielectric spectroscopy evaluates how a material stores and dissipates electrical energy by analyzing its dielectric properties, which include the dielectric constant ε′, and the dielectric loss factor ε″ (Ali et al., 2017). This technique is based on the interaction of an external electric field with the sample, with the complex permittivity providing a detailed description of that interaction (Sanchez et al., 2020). The complex relative permittivity ε* of a material is expressed as follows:

$$\varepsilon^* = \varepsilon' - j\varepsilon'' \quad\text{..}(1)$$

and

$$\varepsilon^* = \varepsilon / \varepsilon_0 \quad\text{...}(2)$$

Here, ε is the absolute permittivity, ε_0 is the free-space permittivity given as 8.854×10^{-12} F/m, and $j = \sqrt{-1}$. ε', the real part of ε^*, is called the dielectric constant and represents the stored energy when the material is exposed to an electric field. In contrast, the dielectric loss factor, ε'', which is the imaginary part, influences energy absorption and attenuation (Lleo et al., 2007). Loss tangent tanδ is also often used as a factor of power dissipation in a dielectric and can be expressed as follows:

$$\tan \delta = \frac{\varepsilon''}{\varepsilon'} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(3)$$

The most common methods used to determine the dielectric properties of fruits and vegetables are the open-ended coaxial probe (Zhu et al., 2016), resonant cavity (Sosa-Morales et al., 2010), and transmission line (Navarkhele et al., 2015) techniques. Among these, the open-ended coaxial probe is widely used for assessing fruit internal quality due to its broad spectral range, high precision, and simplicity of sample preparation and analysis (Zadeh et al., 2019; Cao et al., 2023). This method is particularly effective for measuring the dielectric properties of liquid and semi-solid biological and food materials, owing to its non-destructive nature, versatility, and capacity to operate across high-frequency ranges (La Gioia et al., 2018). The dielectric properties are determined based on the phase and amplitude of the signal reflected from the end of an open-ended coaxial line (Zadeh et al., 2019).

The open-ended coaxial probe technique has been successfully applied to predict the internal quality of various fruits (Shivamurthy et al., 2018; Zadeh et al., 2019; Cao et al., 2024b). Different mathematical models, such as those by Debye or Misra, convert the measured reflection coefficient into permittivity. Currently, this operation is automated using software embedded in measurement equipment like vector network analyzers or impedance analyzers. To improve prediction accuracy, studies are increasingly focusing on developing non-linear models for predicting internal fruit qualities from dielectric properties (Guo et al., 2015a). While linear regression models are commonly used for this purpose, their assumption of probabilistic data generation limits their ability to capture complex phenomena. As a result, machine learning (ML) algorithms have emerged as more suitable alternatives (Cavalcanti et al., 2024).

Dielectric spectroscopy, when combined with ML models, offers several advantages, including rapid, non-destructive, real-time monitoring and feasibility (Wang et al., 2024). Over recent years, researchers have applied ML techniques to predict the soluble solid content (SSC)

of fruits using dielectric properties. For instance, Guo et al. (2015b) measured dielectric constants and loss factors at 51 frequencies between 10 and 1800 MHz for apples and found that the extreme learning machine (ELM) combined with the successive projection algorithm (SPA) achieved the best SSC prediction performance, with R-values of 0.908 and 0.898 for prediction and calibration, respectively.

Similarly, Zhu et al. (2016) used a vector network analyzer and an open-ended coaxial probe to measure dielectric properties from 20 to 4,500 MHz, finding that ELM combined with principal component analysis (PCA) yielded the best SSC prediction performance, with an R-value of 0.6986 and RMSE of 0.7763. Liu and Guo (2017) demonstrated the feasibility of dielectric spectroscopy for SSC prediction by measuring dielectric loss factors and constants between 20 MHz and 4500 MHz. They found that least squares support vector machine (LSSVM) models outperformed partial least squares (PLS) models, with the competitive adaptive reweighted sampling-least squares support vector machine (CARS-LSSVM) achieving R and RMSE values of 0.970 and 0.494, respectively.

In further studies, Cao et al. (2024a) utilized backpropagation artificial neural networks (BP-ANN) and support vector machine (SVM) models to predict SSC in pears. Their findings revealed R^2p values of 0.83 and 0.79 for the BP-ANN and SVM models, respectively. Cao et al. (2023) also explored the dielectric properties of peaches during storage, showing positive correlations between ε', ε'', and SSC content. They achieved the highest Ri^2v value of 0.88 for SSC prediction using an LSSVM model. Liu et al. (2021) obtained similar performance outcomes for PLS and ELM models when predicting melon SSC from dielectric characteristics.

Lastly, Tang et al. (2024) investigated internal quality prediction in pears using a vector network analyzer and coaxial probe in the frequency range of 0.1 to 26.5 GHz. They calculated ε' and ε'' values and employed particle swarm optimization–least squares support vector regression (PSO-LSSVR), SVR, and PLSR algorithms. While SSC and dielectric properties

under a single frequency showed weak correlations, a PLSR model using ε'' as a variable successfully predicted SSC, achieving an R-value of 0.91 and RMSE of 0.09.

However, little effort has been made to predict the SSC of sweet cherries, considering the importance of the crop and its added value worldwide, as stated above. This study presents an open-ended coaxial line probe measurement system capable of non-invasively assessing the SSC of sweet cherries at different radio frequencies. The dielectric properties (ε', ε'', and tanδ) and SSC of sweet cherries are measured across different harvest periods. Additionally, down-sampling was applied to the frequency ranges of dielectric properties, reducing them from the control group of 15 MHz to a maximum of 300 MHz. Using these dielectric properties and the low dimensionality of the data, SSC is predicted with eight classical machine learning algorithms: Random Forest (RF), k-nearest neighbors (kNN), Support Vector Regression (SVR), Gaussian Processes (GP), Multilayer Perceptron (MLP), Multiple Linear Regression (MLR), Bagging (BAG), and Reduced-Error Pruning Tree (REPTree). The predictive performance of these models is then analyzed and compared.

The study innovatively combines down-sampling approaches with dielectric properties and machine learning algorithms to predict SSC in sweet cherries across various harvest periods, unlike the previous studies reported herein, aiming to reduce implementation costs. The remaining part of the article details each of the techniques implemented and their corresponding results.

2. Materials and Methods

An open-ended coaxial line probe measurement system was designed and fabricated based on dielectric spectroscopy principles. The process began with creating a dataset by measuring the dielectric properties and soluble solid content (SSC) of sweet cherries over 26 days. Using these dielectric properties, the SSC of cherries was predicted through machine learning models.

To ensure accuracy, min-max normalization was applied to the data, preserving the relative order and distances of data points while reducing variance and minimizing the influence of outliers. Additionally, a k-fold cross-validation technique was employed to enhance the model's generalizability by partitioning the dataset into multiple subsets for iterative training and testing. The extracted features were used to train various machine learning algorithms, and the models were evaluated based on key metrics: the correlation coefficient R, root mean square error (RMSE), mean absolute error (MAE), and model building time.

2.1. *Fresh product*

Freshly harvested sweet cherry fruits (*Prunus avium L. cv. Starks Gold*) were collected from the cherry orchard of the Faculty of Agriculture at Ankara University, Ankara Province, Turkey (39°57'49.0"N, 32°51'53.9"E). On the day of the experiment, fruits were hand-harvested from the same tree at 7:00 AM and immediately transported to a laboratory near the orchard.

Upon arrival, the cherries were cleaned to remove leaves and stems and sorted by size. Dielectric and soluble solid content (SSC) properties were then measured for the sorted cherries. The average measurements for key physical attributes of the fruits were as follows: major axis, 18.84 ± 0.13 mm; minor axis, 16.08 ± 0.15 mm; height, 17.59 ± 0.11 mm; mass, 3.36 ± 0.07 g; geometric mean diameter, 17.46 ± 0.11 mm; and sphericity, 0.93 ± 0.01.

A total of 180 cherries were tested during each daily trial. The trials were conducted at two-day intervals, three to four days per week.

2.2. *Determination of soluble solid content*

The soluble solid content (SSC) of sweet cherries used in the study was measured using a digital pocket refractometer (Atago PAL-3, Tokyo, Japan) with a 0–93% Brix measurement range. To prepare the samples, pitted cherries were crushed in a mortar, and the resulting liquid was filtered through filter paper. After calibrating the refractometer with pure water, approximately 1–2 mL of cherry juice was placed on the refractometer prism using a pipette. The SSC reading

was then recorded following the procedure described by Queb-González et al. (2020). Each experiment was performed with a minimum of five replicates.

2.3. *Determination of dielectric spectroscopy*

The schematic diagram of the dielectric measurement system is shown in Figure 1. The system comprises an Agilent E4991A impedance analyzer connected to an open-ended coaxial-line probe and a lifting platform. The impedance analyzer, capable of operating in the 1 to 3000 MHz frequency range, is interfaced with a desktop computer via an IEEE-488 (GPIB-625) connection. Custom-designed software installed on the computer controls the impedance analyzer and records the measurement data.

The coaxial probe, which measures the dielectric characteristics of the material, is mounted vertically on the stationary platform, as depicted in Figure 1. The probe comprises an SMA (SubMiniature version A) probe and a corresponding SMA socket. The SMA probe connector is securely screwed into the socket. Both the male and female SMA connectors feature gold-plated copper inner and outer conductor cylinders, separated by a Teflon (PTFE) insulating cylinder. The dielectric constant of the PTFE was considered in the design to ensure the required 50 Ω impedance between the inner and outer conductors. The dimensions of the inner conductor and the inner diameter of the outer conductor were calculated to meet this impedance specification.

A semi-rigid cable connects the probe to the test head of the impedance analyzer. The lifting platform is adjustable to accommodate the height of the sample and can hold additional weights if needed. The open-ended coaxial-line probe measures the complex relative permittivity (ε^*) by analyzing the reflection coefficient (Γ) when the probe is placed in contact with the sample under test (Hernandez-Gomez et al., 2021). The complex relative permittivity (ε^*) is determined using the following formula (Qin et al., 2021):

$$\varepsilon^* = \left(\frac{1-\Gamma}{1+\Gamma}\right)^2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(4)$$

Here, the material under test is assumed to be nonmagnetic, with its complex relative permittivity (ε^*) described by Eq. (1). The complex Fresnel reflection coefficient (Γ) at the air-dielectric material interface is used to determine ε^*. Γ is directly related to the material's impedance and can be expressed as follows (Pozar, 2011):

$$\Gamma = \frac{Z_r - Z_0}{Z_r + Z_0} \quad\dotfill(5)$$

Here, Z_r represents the impedance of the material under test, while Z_0 denotes the free-space impedance. For a nonmagnetic material, the impedance can be expressed as:

$$Z_r = \sqrt{\frac{\mu}{\varepsilon}} = \frac{Z_0}{\sqrt{\varepsilon^*}} \quad\dotfill(6)$$

Since $Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$, $\varepsilon = \varepsilon^*\varepsilon_0$, and $\mu = \mu_0$, where μ is the magnetic permeability. Thus, Equation (5) is transformed into:

$$\Gamma = \frac{Z_r(1 - \sqrt{\varepsilon^*})}{Z_r(1 + \sqrt{\varepsilon^*})} \quad\dotfill(7)$$

Equation (7) and Equation (4) can be easily obtained. Hence, by measuring the complex material impedance Z_r, it is possible to get first the reflection coefficient Γ from Equation (5) and then apply Equation (4) to extract the complex relative permittivity, from which the dielectric constant (ε') and the loss factor (ε'') are found.

2.4. Procedures

Before measurement, the impedance analyzer was powered on and allowed to warm up for at least 30 minutes. Calibration was performed sequentially using the standard open, short, and 50 Ω load procedures. The dielectric properties of the cherry samples were measured in the frequency range of 5 to 3005 MHz, with increments of 5 MHz. The coaxial probe was calibrated using the open, short, and deionized water at 22 °C. A preliminary test on deionized water and methanol was conducted to verify the accuracy of the measured dielectric constant and loss factor values.

Once the system calibration was complete, the dielectric properties of sweet cherries were measured non-destructively. Each sweet cherry was positioned on the probe's surface, resting on the stationary platform, and a liftable platform was used to ensure the probe maintained close contact with the cherry's surface. Measurements were taken at four equidistant points around the equatorial region of each cherry, spaced at 90° intervals.

To ensure accuracy and representative results, a total of 720 dielectric spectroscopy measurements were performed on 180 cherries during each one-day trial. After measuring the dielectric properties, the soluble solid content (SSC) of fruit juice extracted from six cherries was measured five times using a digital refractometer. The average of the five SSC readings was calculated as the reference SSC.

Dielectric and SSC measurements were carried out every two days over 14 trials following the initial measurements. Over this period, a total of 10,080 dielectric property measurements and 2,100 SSC measurements were conducted.

2.5. Down-sampling methodology

The high sensitivity of the dielectric measurement system enables the generation of dielectric features across a frequency range of 5 MHz increments, resulting in a total of 201 frequencies spanning 5 to 3005 MHz. This study focuses on identifying the frequency range that provides the highest performance in estimating the SSC based on dielectric properties. To achieve this, the frequency ranges of the dielectric properties were progressively reduced.

Predictions were initially conducted using the full control group range (15 MHz) and then by selecting frequencies at intervals of 15 MHz. Subsequently, the experiments were repeated with reduced frequency intervals of 30, 60, 90, 120, 150, 180, 210, 240, 270, and 300 MHz. Table 1 outlines the sampling frequency ranges, the total number of frequencies included in each range, and the percentage of information utilized in each sampling interval.

By applying this down-sampling approach, SSC predictions were carried out using various machine learning algorithms. A comparative analysis was then performed to evaluate the performance of the different models.

2.6. Machine learning modelling

SSC prediction was conducted using the dielectric properties ε', ε", tan δ, and their combinations across eight machine learning models: Random Forest (RF), k-Nearest Neighbors (kNN), Support Vector Regression (SVR), Gaussian Processes (GP), Multilayer Perceptron (MLP), Multiple Linear Regression (MLR), Bagging (BAG), and Reduced Error Pruning Tree (REPTree). The modeling process utilized WEKA software (Witten & Frank, 2005). Min-max normalization was applied to the raw data to ensure uniformity across features.

A 10-fold cross-validation technique was used to enhance the reliability of the models. This method divides the dataset into ten equal parts, maintaining a 9:1 ratio for the training and testing sets. During each iteration, nine segments were used for training while one segment was reserved for testing, ensuring that all data was used for validation (Ropelewska et al., 2023).

For the GP and SVR models, the Pearson VII kernel function (PUK) and the polynomial kernel were employed, respectively, to predict SSC. In the MLP model, the activation function was set to Sigmoid, with 500 epochs for training. The learning rate and momentum coefficient were configured at 0.1 and 0.2, respectively. The kNN model utilized a neighborhood size of 5, with the Euclidean distance rule applied for neighbor selection. The MLR model implemented the M5 method for feature selection, incorporating a ridge value of 1×10^{-8}. In the REPTree model, the minimum proportion of variance was set at 0.001, and pruning adjustments were carefully evaluated for optimal performance. The BAG and RF models ran with 10 and 100 iterations, respectively.

Model evaluation was performed both qualitatively and quantitatively using metrics such as the correlation coefficient (R), root mean square error (RMSE), mean absolute error (MAE), and model-building time. These metrics facilitated a comprehensive analysis of the performance of both the down-sampling strategies and the machine learning models.

3. Results

This section demonstrates the outcomes of techniques used to the data to predict soluble solid content from dielectric properties. Both machine learning algorithms and down-sampled information are analyzed.

3.1. *Soluble solid content analysis*

Figure 2 presents the SSC values of sweet cherry samples measured at two-day intervals over a 26-day period. The SSC increased from 14.53% to 21.09% Brix, demonstrating a significant rise as the fruit ripened. This progression reflects the natural biochemical changes, such as sugar accumulation, that occur during the maturation of cherries, particularly in the later stages of ripening (Ivanova et al., 2021).

Studied cherries were harvested at different ripening periods. Daily temperature and humidity changes may have slightly affected SSC accumulation during the trials. It has been reported that SSC is sensitive to microclimatic changes, especially during ripening (Ivanova et al., 2021). Although cherries were visually standardized in size and ripeness during sample selection, natural variation in the physiological ripeness of individual fruits may occur. This may explain the slight discrepancies in SSC measurements with a wider interquartile range for trial 4 (6th trial days) and trial 9 (16th trial days). In addition, although we used a calibrated refractometer in each test for SSC measurements, minimal deviations can also occur during the pre-measurement filtration process. However, the low standard error bars in Figure 2 show that these errors are minimal and do not significantly affect the overall trend. In addition, the cross-

validation processes applied in the study have eliminated the inconsistencies that can arise from these minor deviations.

3.2. *Dielectric spectroscopy analysis of sweet cherries*

Figure 3 illustrates the typical results for dielectric properties—dielectric constant, loss factor, and loss tangent—measured over 14 runs during a 26-day period for sweet cherries. The data highlights distinct trends in dielectric behavior across the frequency spectrum.

The dielectric constant and loss factor averages decreased sharply up to 20 MHz, while the loss tangent (tan δ) averages increased at a comparable rate. Between 20 and 80 MHz, the dielectric constant exhibited a more gradual decline, followed by a steady decrease as frequency increased beyond 80 MHz. Similarly, the loss factor and tan δ averages continued to decline with increasing frequency after 20 MHz. However, the tan δ averages showed a slight upward trend beyond 1280 MHz.

These results are consistent with findings from previous studies on fresh fruits and vegetables (Nelson et al., 2007; Guo et al., 2015a; Zhu et al., 2016; Liu & Guo, 2017). The observed changes in dielectric properties align with the increasing SSC of sweet cherries over time, reflecting the biochemical transformations associated with fruit ripening (Nelson et al., 2007; Cao et al., 2023).

3.3. *The results of the machine learning models for SSC prediction*

Figure 4 illustrates the correlation coefficients for the machine learning models, presented as rows in the heatmap. For SSC estimation using ε', the SVR algorithm achieved the highest R values of 0.93 and 0.92 at 5 and 30 MHz, respectively, followed by the MLP model at 60 MHz with an R of 0.91. The SVR model showed a slight decrease in R as the frequency range increased. Overall, SVR and MLP outperformed other models, while REPTree delivered the lowest R values, with a minimum of 0.83 at 5, 90, 180, and 240 MHz. Using ε", higher R values were observed compared to ε'. MLP performed best, with R values between 0.92 and 0.94,

followed by SVR at 0.89–0.94. MLR and REPTree produced the lowest results. SSC estimation with tanδ yielded the highest R values among all dielectric properties. SVR achieved a maximum of 0.97 at 5 MHz, while the lowest value of 0.88 was observed for kNN and REPTree at 30 MHz. SVR and MLP were the top-performing models. SVR and MLP again delivered the highest R values (0.95–0.97 and 0.95–0.96, respectively), whereas REPTree had the lowest (0.88–0.89) for combined dielectric properties. MLR also showed improved performance compared to its results with individual properties. Among dielectric properties, tanδ provided the best results.

For ε', the lowest RMSE values occurred at 60–270 MHz, with MLP and SVR performing best. The highest RMSE (0.09) was observed for REPTree at 15 MHz. For ε", RMSE values were lowest at 90–120 MHz, with MLP and SVR outperforming other models. SSC estimation with tanδ produced the lowest overall RMSE values, with SVR achieving the best result (0.04) at 5 MHz and REPTree the worst (0.08) at 30 MHz (Figure 5). For ε', MAE ranged from 0.05 (5 MHz SVR) to 0.07 (240 MHz REPTree). For ε", MLP and SVR had the lowest values, while REPTree had the highest. For tanδ, the lowest MAE was 0.03 (5 and 30 MHz SVR), with the highest at 0.06 for REPTree and MLR at 30 and 300 MHz (Figure 6). For ε', the longest time (50.37 s) was observed for MLP at 5 MHz, followed by RF and GP. kNN and REPTree were the fastest, completing within 0.01 s. Similar trends were observed for ε" and tanδ, with MLP consistently taking the longest time and kNN and REPTree being the most time-efficient. When all dielectric properties were used, the highest time (458.6 s) was observed for MLP at 5 MHz (Figure 7).

4. Discussion

During this research, a number of lessons were learned. The attributes ε', ε", tanδ, and SSC were measured from different harvests of sweet cherries over 26 days at two-day intervals using an open-ended coaxial probe technique. The down-sampling method systematically reduced the

frequency ranges of the dielectric properties. These properties, recorded across 11 frequency ranges, were initially used individually and then combined as inputs for machine learning (ML) models to estimate SSC.

The SSC of sweet cherries generally increased steadily throughout the harvest period (Figure 2). The mean SSC values rose from 14.53% at the start of the harvest to 21.09% at the final harvest. Minor decreases on the 6th, 16th, and 26th days did not significantly alter the overall trend. The late-harvest increase in SSC is attributed to a rise in fruit growth rates, while early harvests showed slower SSC accumulation due to the rapid growth of fruit mass and starch reserves. SSC accumulation peaks as growth slows, with the supply of soluble carbohydrates declining due to reduced photosynthetic activity in aging leaves. Two key moments influencing SSC accumulation are the decrease in starch concentration and the cessation of starch accumulation.

When examining dielectric properties, ε' decreased sharply from 5 MHz, levelling off between 20 and 40 MHz. Similarly, ε'' showed a linear decline, with a slower rate of decrease beyond 80 MHz. Tanδ initially increased linearly up to ~30 MHz, then decreased linearly up to 640 MHz, with smaller fluctuations observed between 1280 and 3005 MHz (Figure 3). These fluctuations are influenced by electric dipole movement (Skierucha et al., 2012; Zhu et al., 2019) and variations in water activity, ionic concentration, and ionic mobility (Cao et al., 2023).

Eight different ML models were fine-tuned for higher predictive accuracy and reduced processing time. The SVR and MLP models achieved the best performance, exhibiting high correlation coefficients and low error metrics. However, the MLP model required more time for training, primarily due to the increased number of epochs. In contrast, the SVR model demonstrated a balance between prediction efficiency and processing time, making it more practical. The success of SVR in SSC prediction aligns with studies on peaches and pears (Cao et al., 2023; Cao et al., 2024a), which highlighted the suitability of dielectric spectroscopy for

assessing SSC. However, Guo et al. (2015) found that SVR underperformed compared to ELM for apple SSC prediction across 51 frequency ranges (10–1800 MHz), potentially due to differences in produce type or applied frequencies.

Ionic conductivity and dipole relaxations have affected lower frequencies (15 MHz and 30 MHz). Changes in water content, ion concentration, and SSC due to ripening significantly affect the dielectric properties. Lower frequencies may penetrate the fruit more effectively. In this way, it can interact more intensely with soluble substances (sugars, acids) with free and bound water molecules. Due to ripeness, sugar accumulation, starch degradation, and changes in water distribution alter the dielectric properties of the fruit. At 15 MHz and 30 MHz frequencies, the dielectric properties better reflect changes in the structure and composition of the fruit, resulting in more accurate SSC predictions. High frequencies can be subject to increased signal attenuation and noise due to bound water dipoles and complex interactions at the molecular level. Therefore, low frequencies provide a trade-off between penetration depth and signal clarity, minimizing noise and ensuring reliable measurements.

This study demonstrated that fewer frequencies could sufficiently predict SSC, offering insights into fruit internal quality. Combining dielectric properties (ε', ε''', $\tan\delta$) yielded superior model performance compared to using them individually. Notably, models trained on $\tan\delta$ alone produced results comparable to combined properties. The down-sampled dielectric spectroscopy method presents a novel, efficient approach for rapid SSC determination across varying harvest times. An open-ended coaxial probe with ML algorithms offers a cost-effective and scalable solution for fruit SSC assessment.

Down-sampling reduced the information value to as little as 5.47% while maintaining reliable SSC estimations. Dielectric properties showed a general decline across 5 to 3005 MHz, with $\tan\delta$ initially increasing up to 30 MHz before decreasing. The most accurate predictions

were obtained in the 15, 30, and 60 MHz frequency ranges, with 300 MHz results closely matching these. Predictions based on tanδ alone were highly satisfactory, with SVR and MLP models achieving the highest correlation coefficients of 0.97 and 0.96, respectively.

For ε', ε", and ε'+ε"+tanδ, the best results were observed at 15 MHz, closely followed by the 30 MHz range. For tanδ, frequencies of 60, 90, and 120 MHz provided comparable outcomes. The proposed down-sampling method significantly influenced prediction performance, as results up to 300 MHz closely resembled those from the entire frequency range. This indicates that broader frequency ranges are not necessary for reliable SSC predictions, validating the efficacy of the down-sampling approach.

The highest R values and lowest RMSE and MAE values were obtained using models trained on all combined dielectric properties. However, the MLP model's time-intensive training posed practical challenges due to its large dataset requirements. This highlights the practical advantage of SVR in applications requiring efficient and accurate SSC prediction.

5. Conclusion

The results of predicting SSC by integrating machine learning models with an open-ended coaxial line probe measurement system were promising. For this study, selecting the proper machine learning algorithm and frequency values was critically important. Furthermore, the dielectric measurement system could perform the dielectric measurements in the 15 MHz frequency range. However, it was shown that devices with a less comprehensive measurement range may be sufficient for estimating SSC. The performance results of the SVR and MLP models were lower than those of the other models. However, the time taken to build the MLP model was very high, and the MLP showed remarkable performance in all indexes other than time.

In future studies, Optuna hyperparameter optimization for machine learning and Bayesian optimization and early stopping for deep learning can be explored to improve model performance. Furthermore, feature selection or elimination methods for SSC prediction can be modified to focus on optimizing different frequencies and how they affect different quality parameters. In addition, model accuracy can be improved by using spectral reflectance properties as input and dielectric properties.

CRediT authorship contribution statement

Kamil Sacilik: Conceptualization, Formal analysis, Investigation, Resources, Supervision. Writing-original draft preparation. **Necati Çetin:** Conceptualization, Investigation, Methodology, Software, Data curation, Visualization, Writing-original draft preparation. **Burak Ozbey:** Methodology, Visualization, Writing-original draft preparation. **Fernando Auat Cheein:** Conceptualization, Methodology, Data curation, Supervision, Visualization, Writing-original draft preparation.

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.

Data availability

The data presented in this study are available on request from the corresponding author.

Appendix A.: Data tables and figures

This appendix contains detailed data tables for the down-sampled frequency ranges and ML models. As stated in the main text, sampled data at frequency intervals of 15, 30, 60, 90, 120, 150, 180, 210, 240, 270, and 300 MHz were used to predict SSC using eight machine learning algorithms. This analysis resulted in an additional 17 tables and 3 figures, presented below.

ORCIDs

Kamil Sacilik https://orcid.org/0000-0001-5353-7328

Necati Çetin https://orcid.org/0000-0001-8524-8272

Burak Ozbey https://orcid.org/0000-0001-7485-2132

Fernando Auat Cheein https://orcid.org/0000-0002-6347-7696

References

Ali, M. M., Hashim, N., Bejo, S. K., & Shamsudin, R. (2017). Rapid and nondestructive techniques for internal and external quality evaluation of watermelons: A review. *Scientia Horticulturae*, *225*, 689-699. https://doi.org/10.1016/j.scienta.2017.08.012.

Cao, M., Zeng, S., Wang, J., & Guo, W. (2023). Dielectric properties of peaches with and without skin during storage and their relationship to internal quality. *Postharvest Biology and Technology*, *204*, 112433. https://doi.org/10.1016/j.postharvbio.2023.112433.

Cao, M., Zeng, S., Wang, J., & Guo, W. (2024a). Assessment of SSC and soluble sugar content of three pear cultivars during storage using dielectric method. *Postharvest Biology and Technology*, *212*, 112906. https://doi.org/10.1016/j.postharvbio.2024.112906.

Cao, M., Cao, C., Zhang, T., & Guo, W. (2024b). A dielectric method for predicting pear firmness combining deep data augmentation and ensemble learning. *Food Control*, *110*, 110790. https://doi.org/10.1016/j.foodcont.2024.110790.

Cavalcanti, R. N., Barbosa, V. P., Gut, J. A. W., & Tadini, C. C. (2024). Predicting dielectric properties of fruit juices at 915 and 2450 MHz using machine learning and physicochemical measurements. *Measurement*: *Food*, *14*, 100158. https://doi.org/10.1016/j.meafoo.2024.100158.

Correia, S., Schouten, R., Silva, A. P., & Gonçalves, B. (2017). Factors affecting quality and health promoting compounds during growth and postharvest life of sweet cherry (Prunus avium L.). *Frontiers in Plant Science*, *8*, 2166.

Çetin, N., Karaman, K., Kavuncuoğlu, E., Yıldırım, B., & Jahanbakhshi, A. (2022). Using hyperspectral imaging technology and machine learning algorithms for assessing internal

quality parameters of apple fruits. *Chemometrics and Intelligent Laboratory Systems*, *230*, 104650.

Escribano, S., Biasi, W. V., Lerud, R., Slaughter, D. C., & Mitcham, E. J. (2017). Non-destructive prediction of soluble solids and dry matter content using NIR spectroscopy and its relationship with sensory quality in sweet cherries. *Postharvest Biology and Technology*, *128*, 112-120. https://doi.org/10.1016/j.postharvbio.2017.01.016.

FAOSTAT, (2024). https://www.fao.org/faostat/en/#data/QCL.

Guo, W., Fang, L., Liu, D., & Wang, Z. (2015a). Determination of soluble solids content and firmness of pears during ripening by using dielectric spectroscopy. *Computers and Electronics in Agriculture, 117*, 226–233. https://doi.org/10.1016/j.compag.2015.08.012.

Guo, W., Shang, L., Zhu, X., Liu, Y., & Wang, S. (2015b). Nondestructive detection of soluble solids content of apples from dielectric spectra with ANN and chemometric methods. *Food and Bioprocess Technology*, *8*(6), 1126-1138. https://doi.org/10.1007/s11947-015-1477-0

Hernandez-Gomez, E. S., Olvera-Cervantes, J. L., Sosa-Morales, M. E., Corona-Vazquez, B., Corona-Chavez, A., Lujan-Hidalgo, M. C., & Kataria, T. K. (2021). Dielectric properties of Mexican sauces for microwave-assisted pasteurization process. *Comprehensive Reviews in Food Science and Food Safety*, *86*(1), 112-119. https://doi.org/10.1111/1750-3841.15555.

Hong, Y. P., Choi, S. Y., Cho, M. A., Choi, S. T., & Kim, S. J. (2010). Correlation between soluble solid content and physicochemical properties of 'Bing' cherry at different stages of ripening after harvest. *The Korean Society of Food Preservation*, *17*(3), 370-375.

Hussain, S. Z., Naseer, B., Qadri, T., Fatima, T., & Bhat, T. A. (2021). Cherries (Prunus Cerasus)—Morphology, Taxonomy, Composition and Health Benefits. In *Fruits Grown in Highland Regions of the Himalayas: Nutritional and Health Benefits* (pp. 63-75). Cham: Springer International Publishing.

Ivanova, I., Serdiuk, M., Malkina, V., Bandura, I., Kovalenko, I., Tymoshchuk, T., Tonkha, O., Tsyz, O., Mushtruk, M., & Omelian, A. (2021). The study of soluble solids content accumulation dynamics under the influence of weather factors in the fruits of cherries. Potravinarstvo Slovak Journal of Food Sciences, 15, 350–359. https://doi.org/10.5219/1554.

La Gioia, A., Porter, E., Merunka, I., Shahzad, A., Salahuddin, S., Jones, M., & O'Halloran, M. (2018). Open-ended coaxial probe technique for dielectric measurement of biological tissues: Challenges and common practices. *Diagnostics*, *8*(40). https://doi.org/10.3390/diagnostics8020040.

Li, X., Wei, Y., Xu, J., Feng, X., Wu, F., Zhou, R., Jin, J., Xu, K., Yu, X., & He, Y. (2018). SSC and pH for sweet assessment and maturity classification of harvested cherry fruit based on NIR hyperspectral imaging technology. *Postharvest Biology and Technology, 143,* 112-118. https://doi.org/10.1016/j.postharvbio.2018.05.003.

Lin, F., Chen, D., Liu, C., & He, J. (2024). Non-destructive detection of golden passion fruit quality based on dielectric characteristics. *Applied Sciences, 14*(5), 2200.

Liu, D., & Guo, W. (2017). Nondestructive determination of soluble solids content of persimmons by using dielectric spectroscopy. *International Journal of Food Properties, 20*(sup3), S2596-S2611. https://doi.org/10.1080/10942912.2017.1381114.

Liu, D., Wang, E., Wang, G., Wang, P., Wang, C., & Wang, Z. (2021). Non-destructive sugar content assessment of multiple cultivars of melons by dielectric properties. *Journal of the Science of Food and Agriculture, 101*(10), 4308-4314. https://doi.org/10.1002/jsfa.11070.

Lleo, L., Ruiz-Altisent, M., Hernández, N., & Gutiérrez, P. (2007). Application of microwave return loss for sensing internal quality of peaches. Biosystems Engineering, 96(4), 525–539. https://doi.org/10.1016/j.biosystemseng.2006.11.010.

Navarkhele, V. V., Kapre, A. K., & Shaikh, A. A. (2015). Dielectric properties of black soil with chemical fertilizers at X-band. *Indian Journal of Radio and Space Physics*, *44*, 102e105.

Nelson, S. O., Guo, W., Trabelsi, S., & Kays, S. J. (2007). Dielectric properties of watermelons for quality sensing. *Measurement Science and Technology, 18*(7), 1887–1892. https://doi.org/10.1088/0957-0233/18/7/022.

Nelson, S. (2015). *Dielectric properties of agricultural materials and their applications*. Academic Press. https://doi.org/10.1016/B978-0-12-802305-1.00001-4.

Pozar, D. M. (2021). *Microwave Engineering: Theory and Techniques*. John Wiley & Sons.

Qin, Y., Wei, B., Liu, S., Li, M., Cai, C., & Hou, T. (2021, April). A new method for measuring complex relative permittivity of dielectric material by reflection method. In *Journal of Physics: Conference Series* (Vol. 1871, No. 1, p. 012002). IOP Publishing. https://doi.org/10.1088/1742-6596/1871/1/012002.

Queb-González, D. B., Lopez-Malo, A., Sosa-Morales, M. E., & Villa-Rojas, R. (2020). Postharvest heat treatments to inhibit *Penicillium digitatum* growth and maintain quality of Mandarin (*Citrus reticulata* Blanco). *Heliyon*, *6*(1), e03166. https://doi.org/10.1016/j.heliyon.2020.e03166.

Ropelewska, E., Çetin, N., & Günaydın, S. (2023). Non-destructive discrimination of vacuum-dried banana using image processing operation and machine learning approach. *Food and Bioproducts Processing*, *141*, 36-48.

Sanchez, P. D. C., Hashim, N., Shamsudin, R., & Mohd Nor, M. Z. (2020). Applications of imaging and spectroscopy techniques for non-destructive quality evaluation of potatoes and sweet potatoes: A review. *Trends in Food Science & Technology*, *96*, 208-221. https://doi.org/10.1016/j.tifs.2019.12.027.

Shivamurthy, H. T., Mascolo, V., Romano, R., Neto, A., & Spirito, M. (2018, June). Complex permittivity extraction of layered biological samples. In *2018 IEEE/MTT-S International Microwave Symposium-IMS* (pp. 1573-1576). IEEE. https://doi.org/10.1109/MWSYM.2018.8439281.

Skierucha, W., Wilczek, A., & Szyplowska, A. (2012). Dielectric spectroscopy in agrophysics. *International Agrophysics*, *26*(2), 187-197.

Sosa-Morales, M. E., Valerio-Junco, L., Lopez-Malo, A., & Garcia, H. S. (2010). Dielectric properties of foods: Reported data in the 21st Century and their potential applications. *LWT Food Science and Technology*, *43*(8), 1169e1179.

Tang, Y., Zhang, H., Liang, Q., Xia, Y., Che, J., & Liu, Y. (2024). Non-destructive testing of the internal quality of Korla fragrant pears based on dielectric properties. *Horticulturae*, *10*(6), 572. https://doi.org/10.3390/horticulturae10060572.

Wang, S., Bingke, Y., Zhou, Y., Qu, Y., Zhang, D., Chen, J., Russel, M., & Biswas, K. (2024). Dielectric spectroscopy with machine learning integrated approach to analyze the influence of N:P ratio on algae growth modeling for forecasting. *Microchemical Journal*, *200*, 110254. https://doi.org/10.1016/j.microc.2024.110254.

Witten, I. H., Frank, E., Hall, M. A., Pal, C. J., & Data, M. (2005). Practical machine learning tools and techniques. In *Data mining* (Vol. 2, No. 4, pp. 403-413). Amsterdam, The Netherlands: Elsevier.

Yu, S., Liu, Y., Tang, Y., Li, X., Li, W., Li, C., ... & Lan, H. (2022). Non-destructive quality assessment method for Korla fragrant pears based on electrical properties and adaptive neural-fuzzy inference system. *Computers and Electronics in Agriculture*, *203*, 107492.

Zadeh, V.M., Afrooz, K., Shamsi, M., & Rostami, M. A. (2019). Measuring the dielectric properties of date palm fruit, date palm leaflet, and Dubas bug at radio and microwave frequency using two-port coaxial transmission/reflection line technique. *Biosystems Engineering*, *181*, 73-85. https://doi.org/10.1016/j.biosystemseng.2019.03.003.

Zhu, X., Fang, L., Gu, J., & Guo, W. (2016). Feasibility investigation on determining soluble solids content of peaches using dielectric spectra. *Food Analytical Methods, 9*(6), 1789–1798. https://doi.org/10.1007/s12161-015-0348-7

Zhu, Z., Zhu, X., Kong, F., & Guo, W. (2019). Quantitatively determining the total bacterial count of raw goat milk using dielectric spectra. *Journal of Dairy Science*, *102*(9), 7895-7903.

Table captions

Table 1. Downsampling intervals and the number of frequency

Table 1. Downsampling intervals and the number of frequency.

Sampling frequency interval (MHz)	Total frequency	Percentage of Information
15 (original)	201	100.00%
30	101	50.25%
60	51	25.37%
90	34	16.92%
120	26	12.94%
150	21	10.44%
180	17	8.45%
210	15	7.46%
240	13	6.47%
270	12	5.97%
300	11	5.47%

Figure captions

Fig 1. A schematic diagram of the dielectric measurement system

Fig 2. The results of the *SSC* (%Brix) of sweet cherry samples with means and errors

Fig 3. Resulting dielectric properties from sweet cherry samples at different frequencies

Fig 4. The correlation coefficient results on the heat map of the *ML* models at different frequency intervals

Fig. 5. The *RMSE* results on the radar chart of the *ML* models at different frequency intervals

Fig. 6. The *MAE* results on the vertical drop line of the *ML* models at different frequency intervals

Fig. 7. The time taken to build the *ML* models for combined dielectric properties

Fig 1. A schematic diagram of the dielectric measurement system.
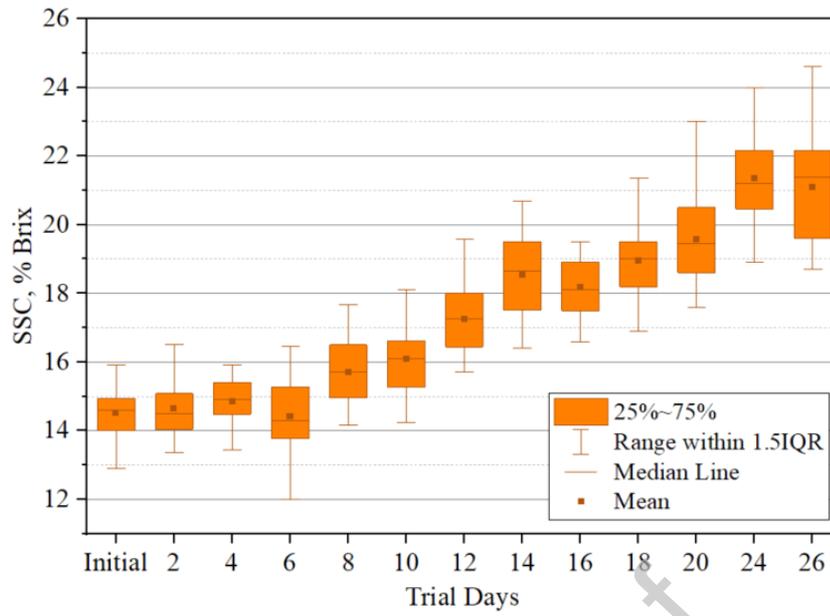
Fig 2. The results of the *SSC* (%Brix) of sweet cherry samples with means and error.

Fig 3. Resulting dielectric properties from sweet cherry samples at different frequencies.
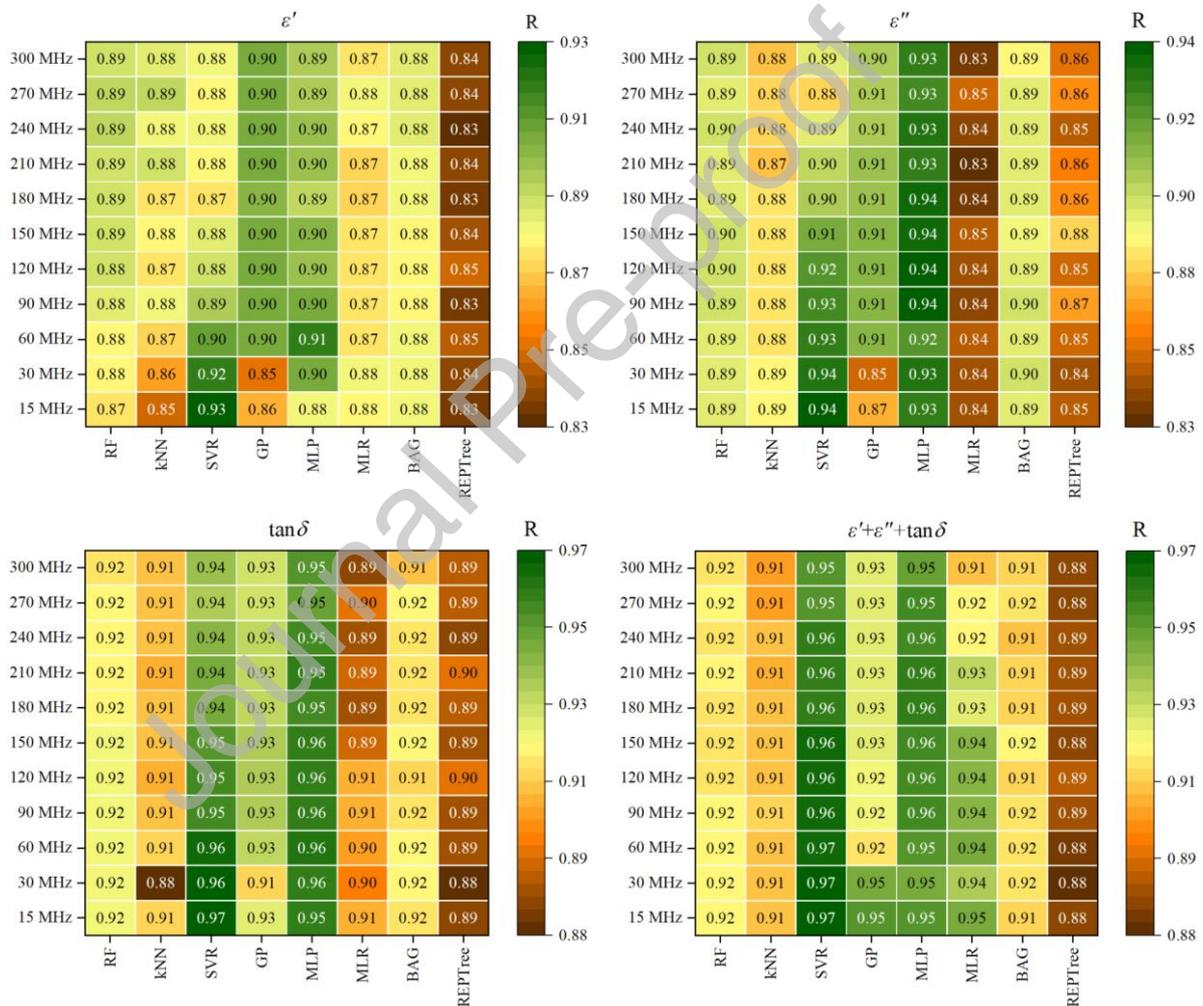
Fig 4. The correlation coefficient results on the heat map of the *ML* models at different frequency intervals.
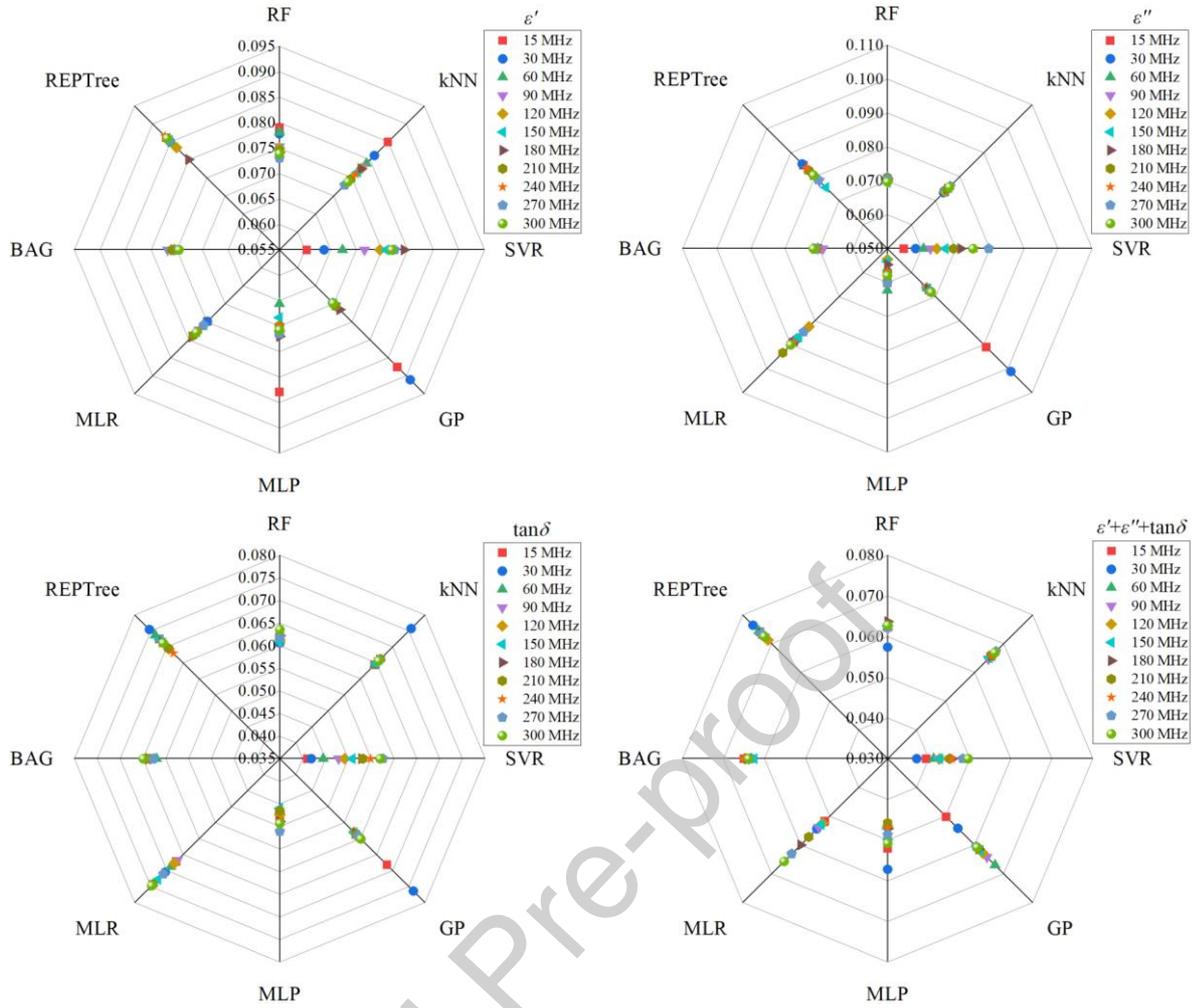
Fig. 5. The *RMSE* results on the radar chart of the *ML* models at different frequency intervals.
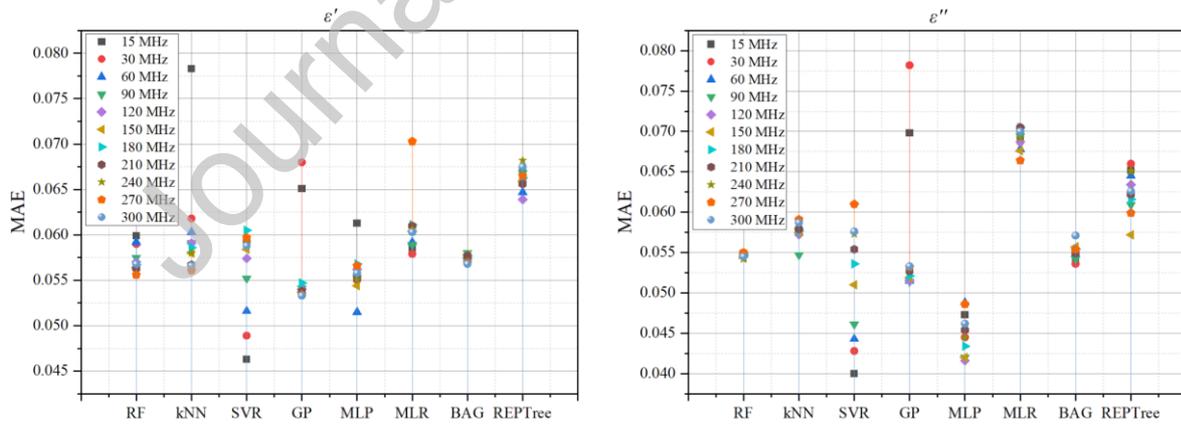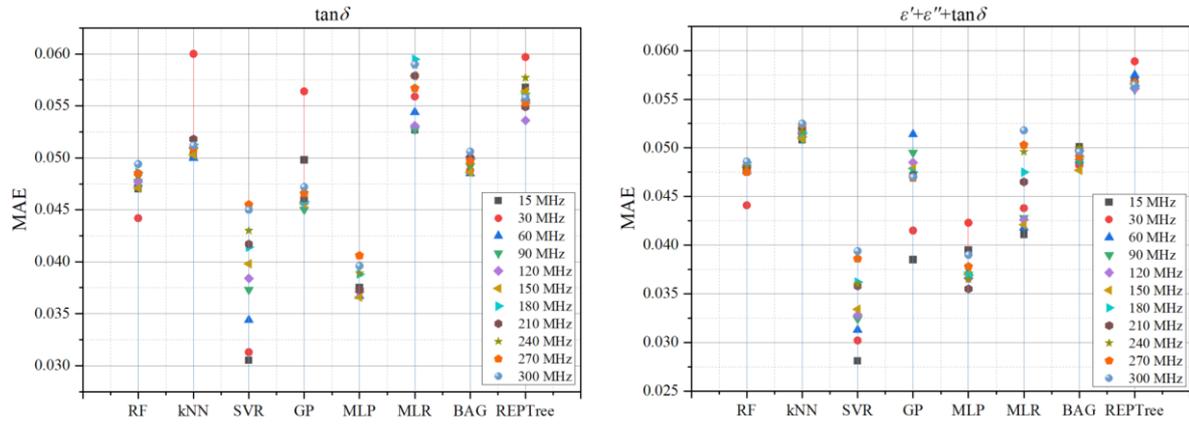
Fig. 6. The *MAE* results on the vertical drop line of the *ML* models at different frequency intervals.
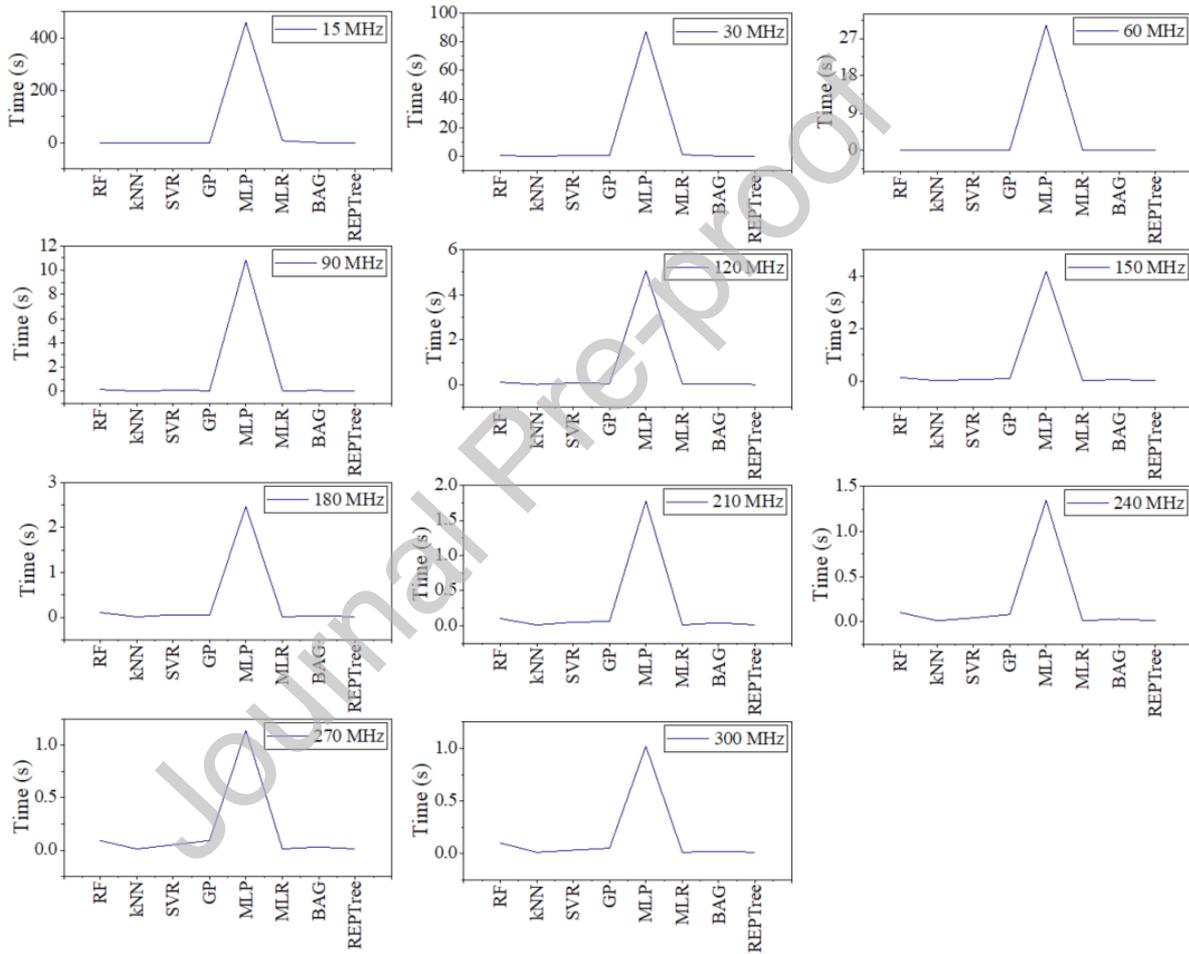
Fig. 7. The time taken to build the *ML* models for combined dielectric properties.

Appendix A.: Data tables and figures

This appendix contains detailed data tables for the down-sampled frequency ranges and ML models. As stated in the main text, sampled data at frequency intervals of 15, 30, 60, 90, 120, 150, 180, 210, 240, 270, and 300 MHz were used to predict SSC using eight machine learning algorithms. This analysis resulted in an additional 17 tables and 3 figures, presented below.

Table A.1. Statistical analysis of SSC (°Brix) of sweet cherry samples

Harvest day	Mean	Standard deviation	Standard error	Median	Minimum	Maximum
Initial	14.53	0.7940	0.1449	14.59	12.90	15.92
2	14.63	0.7847	0.1432	14.49	13.36	16.50
4	14.85	0.7559	0.1380	14.92	13.43	15.91
6	14.41	1.1101	0.2027	14.30	12.00	16.46
8	15.70	0.9297	0.1697	15.71	14.15	17.67
10	16.08	0.9581	0.1749	16.09	14.23	18.11
12	17.24	0.9062	0.1654	17.26	15.72	19.58
14	18.53	1.1932	0.2178	18.64	16.40	20.70
16	18.17	0.8320	0.1518	18.10	16.60	19.51
18	18.95	1.1290	0.2061	19.02	16.90	21.36
20	19.58	1.3471	0.2459	19.44	17.60	23.00
22	20.77	1.1475	0.2095	20.89	18.10	22.83
24	21.34	1.3153	0.2401	21.21	18.90	23.98
26	21.09	1.4400	0.2629	21.39	18.70	24.60

Table A.2. The correlation coefficient results of the ML models for dielectric constant (ε')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.8730	0.8766	0.8763	0.8834	0.8845	0.8864	0.8857	0.8859	0.8888	0.8900	0.8869
kNN	0.8471	0.8609	0.8704	0.8782	0.8732	0.8772	0.8740	0.8840	0.8792	0.8884	0.8838
SVR	0.9254	0.9151	0.9041	0.8915	0.8843	0.8807	0.8722	0.8772	0.8775	0.8759	0.8778
GP	0.8626	0.8504	0.9009	0.9010	0.9007	0.9014	0.8971	0.9001	0.9011	0.9015	0.9015
MLP	0.8785	0.9017	0.9141	0.9003	0.8960	0.8995	0.8893	0.8965	0.8959	0.8908	0.8931
MLR	0.8768	0.8795	0.8739	0.8736	0.8713	0.8714	0.8682	0.8692	0.8747	0.8778	0.8733
BAG	0.8802	0.8789	0.8784	0.8751	0.8766	0.8763	0.8790	0.8783	0.8813	0.8817	0.8815
REPTree	0.8338	0.8367	0.8450	0.8340	0.8476	0.8400	0.8337	0.8379	0.8296	0.8377	0.8362

Table A.3. The correlation coefficient results of the ML models for loss factor (ε'')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.8943	0.8931	0.8932	0.8947	0.8959	0.8970	0.8941	0.8935	0.8961	0.8943	0.8949
kNN	0.8851	0.8851	0.8796	0.8803	0.8822	0.8806	0.8792	0.8737	0.8766	0.8772	0.8750
SVR	0.9441	0.9365	0.9316	0.9274	0.9232	0.9144	0.9044	0.8984	0.8916	0.8782	0.8911
GP	0.8709	0.8509	0.9114	0.9119	0.9113	0.9099	0.9089	0.9066	0.9056	0.9051	0.9042
MLP	0.9293	0.9346	0.9246	0.9414	0.9427	0.9403	0.9389	0.9327	0.9347	0.9256	0.9315
MLR	0.8409	0.8435	0.8444	0.8384	0.8407	0.8453	0.8360	0.8294	0.8383	0.8500	0.8332
BAG	0.8935	0.8960	0.8931	0.8978	0.8922	0.8902	0.8918	0.8929	0.8914	0.8924	0.8867
REPTree	0.8462	0.8421	0.8516	0.8656	0.8511	0.8782	0.8617	0.8561	0.8469	0.8630	0.8552

Table A.4. The correlation coefficient results of the ML models for loss tangent ($\tan\delta$)

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.9240	0.9240	0.9225	0.9229	0.9232	0.9241	0.9192	0.9198	0.9196	0.9212	0.9158
kNN	0.9136	0.8793	0.9131	0.9102	0.9095	0.9117	0.9097	0.9094	0.9105	0.9103	0.9110
SVR	0.9659	0.9649	0.9604	0.9541	0.9518	0.9484	0.9449	0.9442	0.9408	0.9353	0.9367

	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
GP	0.9262	0.9128	0.9325	0.9336	0.9339	0.9336	0.9323	0.9318	0.9304	0.9302	0.9277
MLP	0.9548	0.9570	0.9574	0.9559	0.9555	0.9565	0.9516	0.9545	0.9528	0.9469	0.9501
MLR	0.9089	0.8996	0.9042	0.9087	0.9075	0.8931	0.8885	0.8921	0.8877	0.8974	0.8855
BAG	0.9182	0.9187	0.9208	0.9179	0.9145	0.9192	0.9159	0.9163	0.9158	0.9198	0.9123
REPTree	0.8857	0.8771	0.8865	0.8871	0.8953	0.8874	0.8907	0.8952	0.8850	0.8901	0.8896

Table A.5. The correlation coefficient results of the ML models for combined dielectric properties

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.9201	0.9209	0.9199	0.9205	0.9187	0.9187	0.9164	0.9189	0.9183	0.9198	0.9176
kNN	0.9117	0.9113	0.9109	0.9135	0.9113	0.9114	0.9103	0.9089	0.9110	0.9059	0.9059
SVR	0.9684	0.9677	0.9651	0.9632	0.9627	0.9627	0.9579	0.9586	0.9576	0.9528	0.9507
GP	0.9523	0.9464	0.9170	0.9220	0.9245	0.9260	0.9268	0.9273	0.9280	0.9281	0.9271
MLP	0.9529	0.9454	0.9529	0.9557	0.9560	0.9556	0.9565	0.9579	0.9557	0.9542	0.9496
MLR	0.9450	0.9374	0.9437	0.9412	0.9418	0.9426	0.9273	0.9325	0.9190	0.9190	0.9107
BAG	0.9132	0.9185	0.9172	0.9158	0.9138	0.9199	0.9142	0.9148	0.9126	0.9174	0.9142
REPTree	0.8817	0.8759	0.8808	0.8857	0.8909	0.8840	0.8875	0.8869	0.8872	0.8828	0.8837

Table A.6. The RMSE results of the ML models for dielectric constant (ε')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0790	0.0778	0.0781	0.0753	0.0750	0.0746	0.0742	0.0745	0.0731	0.0730	0.0739
kNN	0.0849	0.0812	0.0790	0.0780	0.0776	0.0766	0.0775	0.0747	0.0760	0.0730	0.0739
SVR	0.0603	0.0637	0.0673	0.0716	0.0746	0.0758	0.0793	0.0772	0.0769	0.0775	0.0769
GP	0.0875	0.0911	0.0709	0.0705	0.0703	0.0700	0.0716	0.0705	0.0700	0.0697	0.0698
MLP	0.0829	0.0717	0.0656	0.0702	0.0709	0.0683	0.0720	0.0699	0.0698	0.0715	0.0707
MLR	0.0749	0.0749	0.0764	0.0772	0.0780	0.0778	0.0792	0.0788	0.0771	0.0759	0.0779
BAG	0.0749	0.0751	0.0755	0.0768	0.0755	0.0762	0.0757	0.0760	0.0746	0.0747	0.0746
REPTree	0.0862	0.0851	0.0845	0.0859	0.0834	0.0854	0.0801	0.0856	0.0866	0.0855	0.0860

Table A.7. The RMSE results of the ML models for loss factor (ε'')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0705	0.0709	0.0709	0.0702	0.0698	0.0695	0.0706	0.0709	0.0698	0.0705	0.0697
kNN	0.0735	0.0733	0.0747	0.0747	0.0738	0.0743	0.0748	0.0749	0.0762	0.0760	0.0757
SVR	0.0548	0.0583	0.0606	0.0626	0.0645	0.0670	0.0715	0.0694	0.0756	0.0797	0.0751
GP	0.0910	0.1012	0.0665	0.0661	0.0663	0.0667	0.0668	0.0676	0.0678	0.0679	0.0682
MLP	0.0601	0.0568	0.0623	0.0536	0.0532	0.0538	0.0547	0.0573	0.0566	0.0601	0.0579
MLR	0.0888	0.0879	0.0877	0.0896	0.0825	0.0867	0.0897	0.0934	0.0889	0.0847	0.0901
BAG	0.0708	0.0699	0.0704	0.0689	0.0705	0.0712	0.0711	0.0711	0.0704	0.0713	0.0717
REPTree	0.0848	0.0853	0.0823	0.0782	0.0819	0.0755	0.0808	0.0826	0.0836	0.0787	0.0806

Table A.8. The RMSE results of the ML models for loss tangent (tanδ)

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0608	0.0606	0.0613	0.0609	0.0611	0.0609	0.0622	0.0622	0.0622	0.0619	0.0636
kNN	0.0643	0.0757	0.0644	0.0657	0.0659	0.0649	0.0659	0.0662	0.0660	0.0658	0.0656
SVR	0.0410	0.0419	0.0445	0.0479	0.0492	0.0509	0.0529	0.0532	0.0549	0.0576	0.0570
GP	0.0682	0.0764	0.0588	0.0582	0.0580	0.0581	0.0584	0.0587	0.0593	0.0592	0.0601
MLP	0.0480	0.0472	0.0466	0.0472	0.0474	0.0460	0.0489	0.0465	0.0487	0.0511	0.0493
MLR	0.0671	0.0705	0.0688	0.0667	0.0677	0.0728	0.0746	0.0743	0.0747	0.0711	0.0747
BAG	0.0633	0.0629	0.0620	0.0634	0.0643	0.0630	0.0638	0.0636	0.0637	0.0627	0.0649
REPTree	0.0724	0.0754	0.0736	0.0719	0.0694	0.0722	0.0715	0.0695	0.0680	0.0713	0.0712

Table A.9. The RMSE results of the ML models for combined dielectric properties

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0623	0.0574	0.0624	0.0620	0.0627	0.0627	0.0637	0.0627	0.0628	0.0623	0.0628
kNN	0.0654	0.0654	0.0654	0.0647	0.0652	0.0651	0.0657	0.0661	0.0658	0.0673	0.0669
SVR	0.0394	0.0371	0.0412	0.0424	0.0427	0.0427	0.0457	0.0451	0.0457	0.0483	0.0496

GP	0.0502	0.0542	0.0670	0.0642	0.0629	0.0620	0.0614	0.0611	0.0607	0.0605	0.0608
MLP	0.0521	0.0572	0.0493	0.0472	0.0468	0.0472	0.0472	0.0458	0.0472	0.0487	0.0508
MLR	0.0517	0.0544	0.0524	0.0538	0.0524	0.0529	0.0600	0.0572	0.0629	0.0631	0.0657
BAG	0.0650	0.0632	0.0632	0.0638	0.0645	0.0625	0.0643	0.0640	0.0649	0.0636	0.0639
REPTree	0.0741	0.0764	0.0744	0.0723	0.0712	0.0737	0.0736	0.0724	0.0731	0.0734	0.0725

Table A.10. The MAE results of the ML models for dielectric constant (ε')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0599	0.0590	0.0592	0.0575	0.0570	0.0564	0.0567	0.0563	0.0556	0.0556	0.0568
kNN	0.0783	0.0618	0.0603	0.0591	0.0591	0.0580	0.0586	0.0567	0.0580	0.0561	0.0566
SVR	0.0463	0.0489	0.0516	0.0552	0.0574	0.0584	0.0605	0.0593	0.0591	0.0597	0.0588
GP	0.0651	0.0680	0.0543	0.0541	0.0540	0.0537	0.0547	0.0538	0.0535	0.0534	0.0533
MLP	0.0613	0.0550	0.0515	0.0553	0.0562	0.0544	0.0568	0.0552	0.0551	0.0566	0.0558
MLR	0.0584	0.0579	0.0592	0.0589	0.0604	0.0604	0.0611	0.0610	0.0606	0.0703	0.0603
BAG	0.0572	0.0570	0.0576	0.0580	0.0575	0.0578	0.0577	0.0577	0.0571	0.0569	0.0568
REPTree	0.0671	0.0660	0.0647	0.0668	0.0639	0.0659	0.0663	0.0656	0.0682	0.0665	0.0675

Table A.11. The MAE results of the ML models for loss factor (ε'')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0549	0.0550	0.0548	0.0546	0.0546	0.0543	0.0547	0.0548	0.0542	0.0550	0.0545
kNN	0.0578	0.0574	0.0582	0.0547	0.0572	0.0575	0.0577	0.0579	0.0588	0.0591	0.0587
SVR	0.0400	0.0428	0.0443	0.0461	0.0610	0.0510	0.0536	0.0554	0.0573	0.0610	0.0576
GP	0.0698	0.0782	0.0515	0.0514	0.0515	0.0519	0.0521	0.0527	0.0531	0.0532	0.0533
MLP	0.0473	0.0445	0.0488	0.0420	0.0416	0.0420	0.0434	0.0454	0.0445	0.0486	0.0462
MLR	0.0689	0.0687	0.0678	0.0694	0.0687	0.0676	0.0705	0.0705	0.0692	0.0664	0.0700
BAG	0.0542	0.0536	0.0548	0.0542	0.0554	0.0557	0.0554	0.0549	0.0553	0.0554	0.0571
REPTree	0.0652	0.0660	0.0645	0.0609	0.0634	0.0572	0.0616	0.0622	0.0651	0.0599	0.0626

Table A.12. The MAE results of the ML models for loss tangent (tanδ)

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0470	0.0442	0.0475	0.0477	0.0477	0.0471	0.0486	0.0485	0.0483	0.0485	0.0494
kNN	0.0503	0.0600	0.0500	0.0507	0.0510	0.0503	0.0513	0.0518	0.0509	0.0509	0.0512
SVR	0.0305	0.0313	0.0344	0.0373	0.0384	0.0398	0.0414	0.0417	0.0430	0.0455	0.0450
GP	0.0498	0.0564	0.0460	0.0450	0.0456	0.0456	0.0458	0.0461	0.0465	0.0466	0.0472
MLP	0.0375	0.0372	0.0367	0.0372	0.0371	0.0366	0.0388	0.0373	0.0389	0.0406	0.0396
MLR	0.0527	0.0559	0.0544	0.0527	0.0531	0.0579	0.0595	0.0579	0.0589	0.0567	0.0590
BAG	0.0497	0.0488	0.0485	0.0492	0.0501	0.0486	0.0499	0.0500	0.0494	0.0497	0.0506
REPTree	0.0568	0.0597	0.0558	0.0560	0.0536	0.0564	0.0553	0.0549	0.0577	0.0553	0.0558

Table A.13. The MAE results of the ML models for combined dielectric properties

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.0477	0.0441	0.0476	0.0477	0.0482	0.0480	0.0485	0.0479	0.0481	0.0475	0.0486
kNN	0.0508	0.0514	0.0517	0.0509	0.0516	0.0510	0.0516	0.0520	0.0517	0.0524	0.0525
SVR	0.0281	0.0302	0.0313	0.0324	0.0328	0.0334	0.0362	0.0358	0.0360	0.0386	0.0394
GP	0.0385	0.0415	0.0514	0.0495	0.0485	0.0479	0.0475	0.0472	0.0470	0.0469	0.0470
MLP	0.0395	0.0423	0.0372	0.0370	0.0366	0.0370	0.0369	0.0355	0.0365	0.0378	0.0390
MLR	0.0411	0.0438	0.0418	0.0428	0.0426	0.0421	0.0475	0.0465	0.0496	0.0503	0.0518
BAG	0.0501	0.0482	0.0487	0.0486	0.0497	0.0477	0.0491	0.0497	0.0500	0.0491	0.0496
REPTree	0.0568	0.0589	0.0575	0.0560	0.0561	0.0566	0.0564	0.0568	0.0565	0.0567	0.0565

Table A.14. The time taken to build the ML models for dielectric constant (ε')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.21	0.10	0.09	0.11	0.07	0.08	0.08	0.07	0.07	0.09	0.06
kNN	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
SVR	0.13	0.05	0.03	0.02	0.03	0.03	0.04	0.04	0.03	0.03	0.03
GP	0.17	0.07	0.07	0.07	0.06	0.06	0.06	0.08	0.09	0.06	0.06

MLP	50.37	10.23	2.27	1.12	0.58	0.43	0.29	0.23	0.18	0.18	0.14
MLR	0.33	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
BAG	0.10	0.05	0.04	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.01
REPTree	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Table A.15. The time taken to build the ML models for loss factor (ε'')

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.25	0.09	0.12	0.09	0.07	0.07	0.10	0.09	0.08	0.07	0.08
kNN	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
SVR	0.17	0.07	0.03	0.03	0.03	0.04	0.04	0.05	0.04	0.04	0.04
GP	0.15	0.04	0.04	0.08	0.04	0.05	0.06	0.05	0.04	0.08	0.04
MLP	36.24	9.63	2.23	0.96	0.63	0.45	0.33	0.28	0.20	0.16	0.14
MLR	0.30	0.06	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
BAG	0.11	0.07	0.04	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01
REPTree	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Table A.16. The time taken to build the ML models for loss tangent ($\tan\delta$)

Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.14	0.09	0.10	0.10	0.08	0.07	0.07	0.07	0.06	0.06	0.06
kNN	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
SVR	0.14	0.05	0.03	0.06	0.03	0.03	0.05	0.05	0.04	0.03	0.03
GP	0.14	0.04	0.07	0.04	0.07	0.06	0.06	0.08	0.04	0.04	0.05
MLP	48.12	8.59	2.32	0.96	0.64	0.46	0.29	0.29	0.18	0.14	0.13
MLR	0.34	0.06	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
BAG	0.10	0.06	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.01	0.01
REPTree	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Table A.17. The time taken to build the ML models for combined dielectric properties

| Models	15 MHz	30 MHz	60 MHz	90 MHz	120 MHz	150 MHz	180 MHz	210 MHz	240 MHz	270 MHz	300 MHz
RF	0.26	0.34	0.15	0.16	0.11	0.13	0.11	0.10	0.10	0.09	0.10
kNN	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
SVR	0.38	0.32	0.10	0.09	0.07	0.06	0.06	0.05	0.04	0.05	0.03
GP	0.15	0.23	0.09	0.06	0.05	0.08	0.05	0.06	0.08	0.09	0.05
MLP	458.6	87.11	30.2	10.85	5.06	4.18	2.46	1.78	1.35	1.14	1.02
MLR	8.47	1.04	0.14	0.06	0.03	0.02	0.01	0.01	0.01	0.01	0.01
BAG	0.29	0.15	0.09	0.07	0.04	0.05	0.04	0.04	0.03	0.03	0.02
REPTree	0.05	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
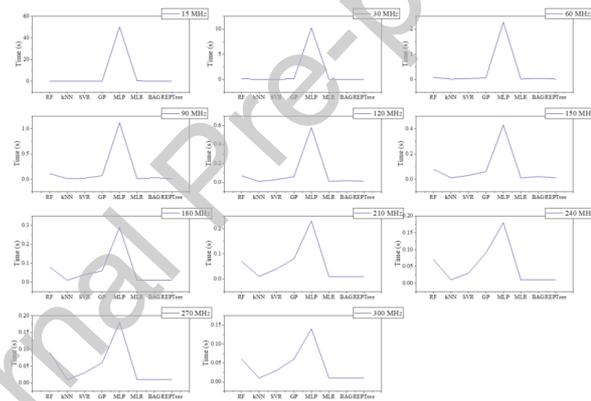
Fig. A.1. The time taken to build the ML models for dielectric constant (ε')
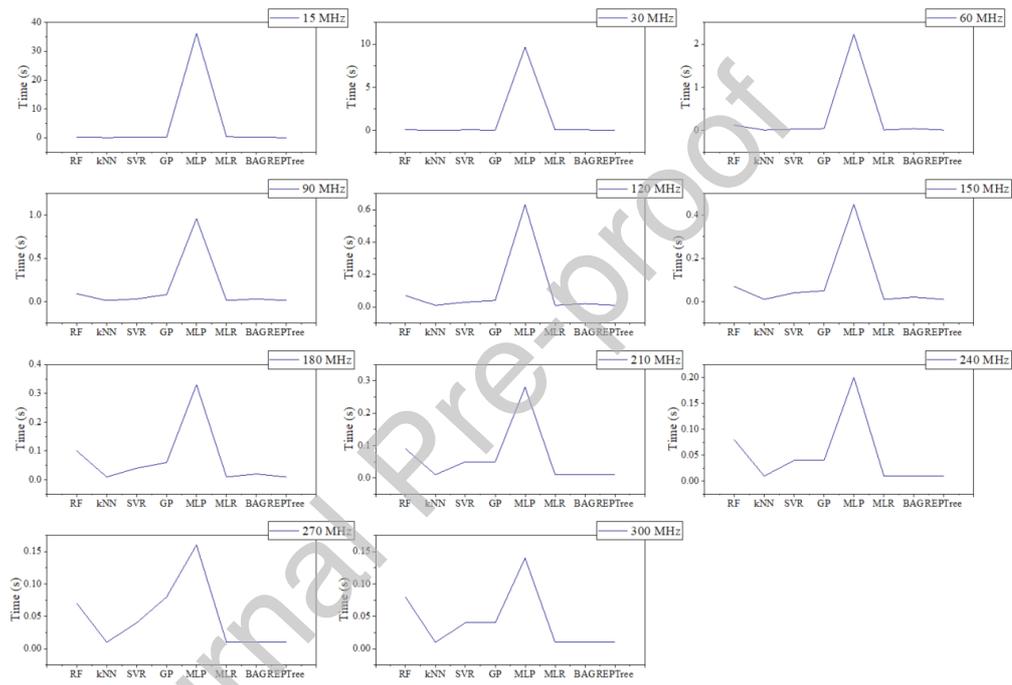
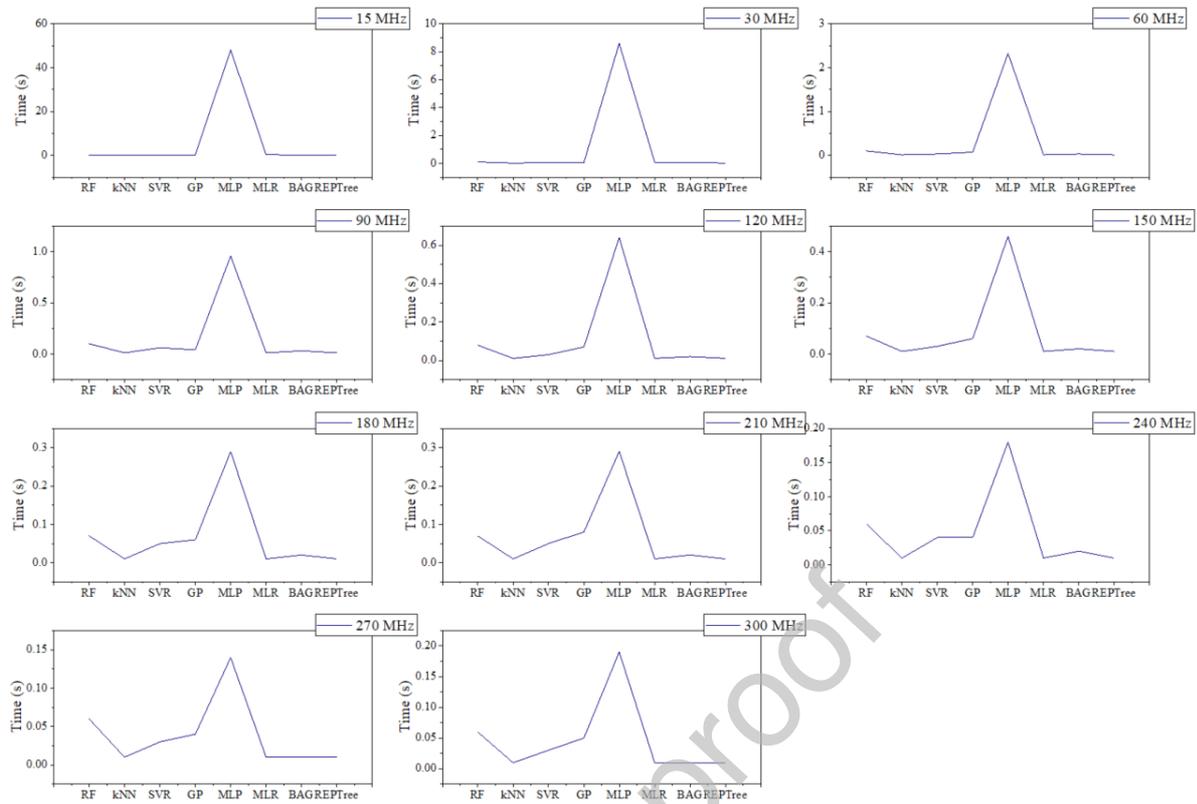Fig. A.2. The time taken to build the ML models for loss factor (ε'')

Fig. A.3. The time taken to build the ML models for loss tangent (tanδ)

■ **Ethics Statement**

X Not applicable: This manuscript does not include human or animal research.

Declaration of interests

⊠ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: