

Extreme Gradient Boosting Model to Predict Antioxidant Activity of Extract from *Ainsliaea acerifolia*

Hyeon Cheol Kim , Woo Seok Lim , Si Young Ha ,
and Jae-Kyung Yang ,*

A machine learning (ML)-based framework was developed for predicting and optimizing the antioxidant activity of *Ainsliaea acerifolia* water extracts. While the response surface methodology (RSM) is deficient in modeling nonlinear interactions. In this study, three machine learning (ML) algorithms, Extreme Gradient Boosting (XGB), Random Forest (RF), and Support Vector Machine (SVM), were evaluated using extraction variables (temperature, time, and solvent-to-solid ratio) along with flavonoid and polyphenol content as input features. Among the models evaluated, the XGB model showed the most advanced antioxidant prediction capabilities, as evidenced by its R^2 of 0.9835 and RMSE of 2.52 on the test data set. The biological significance of the features was explored using SHAP analysis, revealing flavonoid content and extraction temperature as key contributors. A graphical user interface (GUI) was developed to facilitate real-time prediction, enhancing accessibility for researchers and industrial users. This approach improves operational efficiency by optimizing extraction conditions, predicting antioxidant activity from data including flavonoids and polyphenols, and reducing reagent usage. This study highlights the potential of ML as a sustainable alternative for natural product optimization and lays the groundwork for future research that integrates bioactivity prediction with formulation design.

DOI: 10.15376/biores.20.4.9103-9126

Keywords: Antioxidant; *Ainsliaea acerifolia*; Extreme gradient boosting; Flavonoids; Machine learning; Water extraction

Contact information: Department of Environmental Materials Science/Institute of Agriculture and Life Science, Gyeongsang National University, Jinju, 52828, Republic of Korea;

* Corresponding author: jkyang@gnu.ac.kr

INTRODUCTION

Antioxidant activity has been defined as the activity of removing reactive oxygen species (ROS) produced by the metabolic processes of the body (Ifeanyi 2018). Production of ROS during normal metabolism has been implicated in electron transport, gene expression, and antimicrobial activity (Pham-Huy *et al.* 2008). However, ROS can be increased through UV stress and food intake (Golovynska *et al.* 2023). Exposure to excessive ROS has been implicated as a cause of chronic diseases and conditions. These include inflammation, diabetes, and cancer (Wang *et al.* 2021). In this context, many synthetic antioxidants have been developed, such as butylated hydroxytoluene (BHT) and butylated hydroxyanisole (BHA), but their use has been increasingly restricted due to controversies about their safety (Mizobuchi *et al.* 2022). In addition, increasing consumer awareness of antioxidant safety and a growing preference for natural products has motivated research to identify sources of natural antioxidants.

Ainsliaea acerifolia, a perennial herb belonging to the Compositae family, is distributed in the mountainous regions of Korea (Jung *et al.* 2000). The extract of *A.*

acerifolia was found to contain abundant polyphenols, including major secondary metabolites such as quinic acid derivatives, sesquiterpene lactones, and lignans (Choi *et al.* 2006). These polyphenols were found to possess various biological activities, including antioxidant activity (Park 2010). In addition, various polar compounds, such as polyphenols, were complexly mixed in the extracts (Macías *et al.* 1999). Therefore, researchers are focusing on the antioxidant activity of plant extracts rather than individual compounds (Ed-Dahmani *et al.* 2024; Kongolo Kalembe *et al.* 2024). Therefore, there is reason to propose *A. acerifolia* extract as a natural source of antioxidants. Studying the extraction process is needed to obtain extracts with high antioxidant activity from *A. acerifolia* and to identify the influencing factors.

Water extraction is an environmentally friendly and safe solvent to produce extracts from plants such as *A. acerifolia* (Plotka-Wasyłka *et al.* 2017). Although many plant extracts have been obtained using organic solvents, water extraction is highly attractive from a cost perspective. Additionally, the antioxidant activity of the extract has been found to be sensitively influenced by extraction conditions, including extraction temperature and solvent-to-solid ratio (Belwal *et al.* 2016). High thermal temperatures in the extraction process could alter the phenolic extractability due to the destruction of cell walls, but it could also cause partial degradation of phenolic compounds that affect antioxidant activity (Barreira *et al.* 2009; Choi *et al.* 2006). Generally, a higher solvent ratio resulted in higher total phenolic content and antioxidant capacity (Cacace and Mazza 2003). However, the solvent-to-solid ratio varied according to the type of plant, requiring individual assessment (Michiels *et al.* 2012). To obtain *A. acerifolia* extracts with high antioxidant activity, it was essential to understand the above extraction processes and identify the major influencing factors.

Various methods have been developed to assess the antioxidant capacity of compounds, with *in vitro* chemical assays being the most used. Among these, the 2,2'-azino-bis-3-ethylbenzothiazoline-6-sulphonic acid (ABTS) radical scavenging assay has been widely used (Dong *et al.* 2015). This test measures the amount of ABTS that changes from a relatively stable turquoise color to colorless using a spectrophotometer. However, these methods were influenced by environmental factors, were time consuming and costly, and the results varied depending on the skill of the experimenter. Researchers have traditionally used response surface methodology (RSM) to predict and optimize antioxidant activity. It is essentially a collection of mathematical and statistical methods useful for experimental design, model development and process optimization, considering parameter interactions (Khedmati *et al.* 2017). However, RSM can only use standardized quadratic equations within the experimental range. Relationships between data that contain curvature are not always well accommodated by quadratic equations. To overcome this problem, quadratic models can be transformed using logarithms or exponential functions. However, transforming responses or inputs is time-consuming and sometimes it is difficult to know what form of transformation is best (Baş and Boyacı 2007). Furthermore, if discrete variables are chosen as part of the experiment, RSM can result in a continuous approximation of the discrete design, which can lead to significant inaccuracy (Karimifard and Alavi Moghaddam 2018).

To overcome the statistical limitations of RSM, researchers have introduced machine learning (ML) into their studies, which is a field of artificial intelligence that uses computer algorithms to derive mathematical models capable of making predictions directly from trainable data. The ML methods are useful for inferring outcomes in complex non-linear relationships between variables and outcomes (Ryo and Rillig 2017). ML has the potential to predict outcomes from trainable data without the need to explicitly understand the mechanisms of variable interactions. Through ML,

experimental results could be predicted from variables, allowing rapid determination of results and independence from the skill of the experimenter.

Compared to traditional chemical analysis, machine learning offers a complementary approach that can improve efficiency and reduce the need for extensive chemical reagents at certain stages of prediction or screening. However, it is not a replacement for experimental validation, but rather a tool to guide and streamline empirical studies. Recent studies have compared RSM and ML and reported improvements in ML for optimizing process variables and predicting output. Yikmis *et al.* (2024) showed the highest R^2 values of 0.99, 0.98, and 0.99 for ML compared to RSM in predicting TPC content, TAC content, and DPPH antioxidant activity of extracts of *Viburnum opulus* L., respectively. Chen *et al.* (2024) found that ML had an R^2 value of 0.97 in predicting the antioxidant activity of *Salvia miltiorrhiza* extract-derived constituents. Li *et al.* (2022) demonstrated the higher performance of ML compared to RSM in optimizing the ultrasonic extraction of *Polygala tenuifolia* and process parameters. It is evident that the utilisation of machine learning (ML) for the enhancement or forecasting of the antioxidant activity of aqueous *A. acerifolia* extracts remains unreported. The authors' hypothesis is that ML algorithms can be used to predict and optimize antioxidant activity.

Among ML models, the extreme gradient boosting (XGB) algorithm employs gradient boosting techniques with regularization to prevent model overfitting and to enhance generalization performance. It has been documented that this model exhibits superior performance in comparison to other machine learning models. Lee and Aan (2024) introduced the XGB model for predicting antioxidant activity from the spectroscopic data of fruit juices. They reported an R^2 value of 0.980, which outperformed the multiple linear regression model and the random forest model. Nashi *et al.* (2025) employed the XGB model to predict antioxidant activity from the polyphenolic composition of extracts of date palm seeds. This approach yielded an R^2 value of 0.92, indicating a high degree of accuracy. In their 2023 study, Fujimoto and Gotoh (2023) employed the XGB model to predict the antioxidant activity of compounds with analogous structures derived from plant phenolic compounds. The XGB model demonstrated consistent prediction accuracy, as evidenced by an RMSE value of 0.1939.

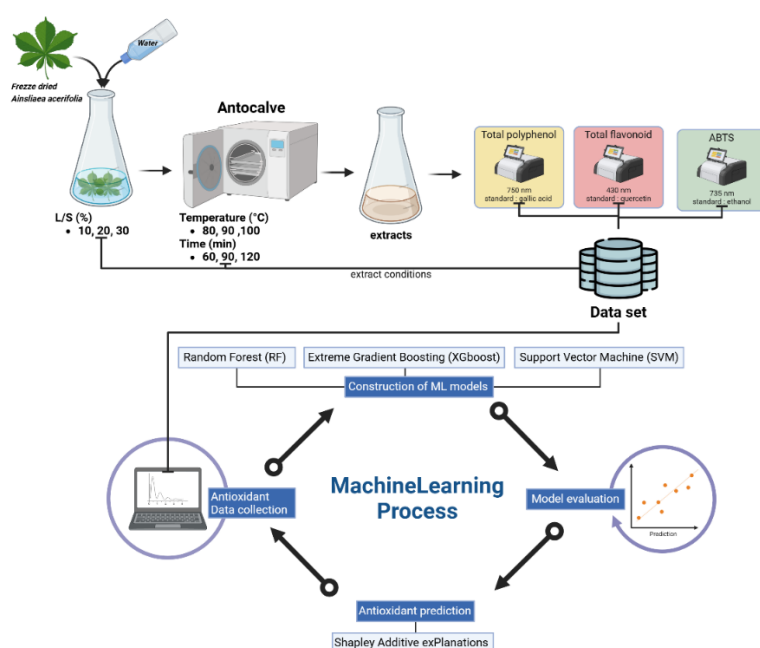


Fig. 1. Machine learning process for the prediction and optimization of production of

extract

The aim of this paper is to apply ML algorithms to predict and optimize the antioxidant activity of *A. acerifolia* water extract (Fig. 1). The water extracts of *A. acerifolia* leaves were obtained by varying the temperature, time, and liquid/solid ratio. The ML model was trained with data that included the polyphenol and flavonoid content of the extracts along with extraction variables and predicted antioxidant activity. Three ML algorithms, extreme gradient boosting (XGB), random forest (RF), and support vector machine (SVM), were used to develop the ML model. The relative importance and impact of each input variable were investigated. The Shapley Additive Explanations (SHAP) approach was used to interpret the developed ML model. The ML model was applied to a graphical user interface, allowing researchers to quickly and easily predict the antioxidant activity of the extracts. The results obtained from this study could help in understanding and improving the antioxidant activity of *A. acerifolia*.

EXPERIMENTAL

Material

The *A. acerifolia* seedlings were obtained from a native farm in South Korea (Yeoju, Gyeonggi-do) and subsequently transplanted into pots with a 5-cm separation between each seedling. The seedlings were irrigated at an interval of 12 h. Following the harvesting of the maple, the leaves were collected, thoroughly cleaned to remove any soil debris, and then subjected to a freeze-drying process. The freeze-dried leaves were then ground and passed through a 40-mesh wire sieve. Samples were stored at 4 °C in tightly sealed bags until use.

Water extraction of *A. acerifolia*

Water extraction of *A. acerifolia* was performed by varying the temperature, time, and S/L ratio according to the conditions in Table 1. A sample of the powder was placed in a 300-mL flask with different liquid/solid ratios, capped, and extracted with an autoclave (ST-65G, JEIO Tech, Korea) under different temperature and time conditions. The autoclave was set to melting mode. The temperature was raised at a rate of 4.2 °C /min. The flask was left at room temperature to cool. After the end of the extraction, it was gravity filtered using Whatman filter paper No. 2 and the extract was subjected to antioxidant activity assay. All experiments were measured in triplicate.

Table 1. Conditions for Water Extraction of *Ainsliaea acerifolia*

Variable	Symbol	Conditions		
Temperature (°C)	Temp	80	90	100
Time (min)	Time	60	90	120
Liquid/Solid ratio	L/S	10	20	30

Determination of total polyphenol content

Total polyphenol content (TPC) was measured using a slightly modified version of a standardized protocol (Singleton and Rossi 1965). Briefly, a mixture containing 100 µL of *A. acerifolia* extract and 100 µL of Folin-Ciocalteu colorimetric reagent solution was incubated with 100 µL of 2% Na₂CO₃ (sodium carbonate) solution for 30 min at room temperature conditions. The resulting assay mixture was measured calorimetrically at 750 nm using a UV spectrophotometer (SpectraMax 190, Molecular Devices LLC, San Jose, CA, USA). A calibration curve was generated using gallic acid

as a control standard. The TPC was obtained from a standard curve with gallic acid as the standard and expressed as mg gallic acid equivalent (GAE) per g of sample.

Determination of total flavonoid content

Total flavonoid content (TFC) was determined following a standardized protocol (Lee *et al.* 2017). The *A. acerifolia* extract (100 μ L) were combined with 100 μ L of 2% aluminum chloride solution and the mixture was allowed to react at room temperature for 10 min. The mixture was measuring absorbance at 430 nm using a UV-spectrophotometer (SpectraMax 190, Molecular Devices LLC, San Jose, CA, USA). The TFC was calculated from a standard curve with quercetin as the standard and expressed as mg quercetin equivalent (QE) per g of sample.

ABTS radical scavenging assay

The antioxidant capacity of the extracts, assessed by the ABTS (2,2'-azinobis(3-ethylbenzothiazoline-6-sulphonic acid) radical scavenging assay, was evaluated according to the method described by Ha *et al.* (2024). The ABTS working solution was prepared by combining equal volumes of 7.4 mM ABTS and 2.6 mM potassium peroxydisulfate solutions, which were then allowed to react for 24 h in the dark at room temperature. The solution was then diluted with ethanol to an absorbance of 0.7 ± 0.02 at 735 nm. A total of 190 μ L of the prepared ABTS solution was mixed with 10 μ L of *A. acerifolia* extract and incubated for 6 min at room temperature. Absorbance at 735 nm was recorded using a UV spectrophotometer (SpectraMax 190, Molecular Devices LLC, San Jose, CA, USA) with 98% ethanol as a control.

Two-factor interaction model

The statistical software Design-Expert (version 13, State-Ease Inc., Minneapolis, MN, USA) was used to construct a two-factor interaction (2FI) model. In the analysis, extraction process variables, polyphenols, and flavonoids were included as influencing factors, while ABTS antioxidant activity was considered as the response variable. The significance of these variables within the model was assessed using an analysis of variance (ANOVA). An equation reflecting the contribution of the effectors was derived to estimate the ABTS antioxidant activity.

Machine learning model

The authors used 81 data points to train and evaluate the machine learning model.

Training and test data were randomly split 8:2. The test data was used to evaluate the model without being involved in training the model. This ensures that we get objective predictions rather than predictions from overfitted models. The coefficient of determination (R^2) and root mean squared error (RMSE) was then used to assess the performance of the model (Renaud and Victoria-Feser 2010). R^2 is a metric that indicates the extent to which a prediction accurately represents the true value. It provides a standardized measure of fit when comparing multiple models trained on the same dataset. RMSE, conversely, offers an interpretive approach to model prediction accuracy by providing the mean size of the prediction error in the same units as the outcome variable. The utilization of these two metrics in regression problems is pervasive, as they offer a complementary array of information. The R^2 and RMSE were calculated according to Eqs. 1 and 2,

$$\text{Coefficient of determination}(R^2) = 1 - \frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{\sum_{n=1}^N (y_n - \bar{y})^2} \quad (1)$$

where N is the number of observations, y_n is the actual value corresponding to the n -

th data point, \widehat{y}_n is the predicted value for the n-th data point, and y is the average value for the N observations. Equation 2 is as follows,

$$\text{Root Mean Squared Error(RMSE)} = 1 - \sqrt{\frac{1}{n} \sum_{n=1}^N (y_n - \widehat{y}_n)^2} \quad (2)$$

where N is the number of observations, y_n is the actual value corresponding to the n-th data point, and \widehat{y}_n is the predicted value for the n-th data point.

Feature selection

Feature selection was performed based on domain knowledge and existing experimental results related to the antioxidant activity of *A. acerifolia* extracts. Five variables were selected as input variables: extraction temperature, extraction time, solvent/solids ratio, total flavonoid content, and total polyphenol content. The withering variable has been reported in several studies to have a significant effect on the antioxidant activity properties of plant extracts (Abeysinghe *et al.* 2021; Antony and Farid 2022; Camel 2000; Pan *et al.* 2000). In this work, the SHAP (SHapley Additive exPlanations) value to evaluate the importance of variables in the optimized model. SHAP value leverages game theory concepts to provide insight into how much each feature contributes to model predictions. This technique provides a better understanding of the model behavior that gives features their importance (Li *et al.* 2024).

Extreme gradient boosting model

Extreme gradient boosting is one of the gradient boosting-based supervised learning algorithms that support preventing overfitting and parallel processing. Gradient boosting is an algorithm that sequentially adds new learning models with weights in the direction of minimizing the learning error of several weak decision tree (Zhang and Haghani 2015). A new learning model is created at every step instead of modifying the existing learning models. The model's error is reduced using gradient descent. The XGB applies a penalty to the loss function to prevent overfitting to the training data. Furthermore, the drawback of consuming significant learning resources due to the sequential data learning characteristic of the gradient model was resolved through parallel processing (Chen *et al.* 2015).

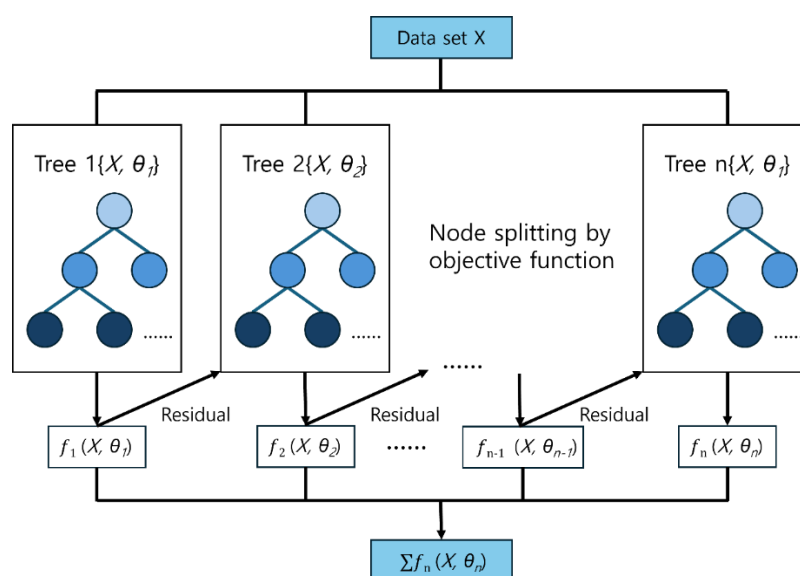


Fig. 2. Schematic diagram of the XGB model

Random forest model

Random forest (RF) is one of the supervised learning algorithms used for various classification and regression problems (Biau and Scornet 2016). The RF integrates several DT to form an ensemble regressor and predicts the outcome by averaging the output values of each DT. If the number of DT is sufficient, RF reduces the overall variance and prediction error by averaging uncorrelated trees, thus not causing model to overfit. RF having the characteristics of bagging, can maintain accuracy even if some data are missing.

Support vector machine model

The SVM was one of the algorithms used for various classification and regression problems (Suthaharan 2016). A major advantage of SVM is the adoption of the structural risk minimization principle, proven superior to the empirical risk minimization principle used in conventional neural network structures. Therefore, SVM is generally less vulnerable to overfitting issues. It also demonstrates robustness against outliers, performing well in predictions for data with values that differ from the general pattern.

Random sample consensus model

The Random Sample Consensus (RANSAC) algorithm is a regression algorithm used when dealing with data with many outliers. This is because the algorithm can effectively identify and discard data containing outliers to obtain an accurate model. The RANSAC algorithm uses a heuristic approach and is effective at finding a satisfactory model with limited data and in a relatively short time.

Optimize hyper parameters

The RandomizedSearchCV method from the scikit-learn library was applied for the random search to find the optimal hyperparameters. The randomized search combines K-fold CVs to determine the given parameter values. After randomly exploring the parameter combinations of the fitting, the optimal parameter combination is returned (Bergstra and Bengio 2012). In this study, k is set to 5 and RMSE is chosen as the loss function. The optimal parameter combination corresponded to the lowest RMSE value. The search range of different hyperparameters and the optimal hyperparameter combination for all models are summarized in Table 2.

Table 2. The Optimized Hyper Parameters in Models Built in this Study

Algorithm	Hyper-parameter	Range of Parameters	Optimized Value
SVM	c	0.1 to 100	57.49
	epsilon	0.01 to 0.5	0.23
	kernel	linear, rbf	rbf
	gamma	scale, auto, 0.01, 0.1, 1	0.01
RF	n_estimators	50 to 500	435
	max_depth	None, 10, 20, 30, 40	30
	min_sample_split	2, 5, 10	5
	min_sample_leaf	1, 2, 4	1
	max_features	None, sqrt, log2	Sqrt
XGB	n_estimators	50 to 500	297
	learning_rate	0.01 to 0.2	0.19

	max_depth	3 to 13	10
	subsample	0.4 to 0.6	0.65
	gamma	0 to 0.3	0.29
RANSAC	min_sample	0.2 to 0.6	0.78
	residual_threshold	1, 5, 10	10.0
	max_trails	100 to 1000	600

Optimization extract process

The water extraction process of *A. acerifolia* was optimized to find the conditions with maximum ABTS antioxidant activity. The factors that were investigated included extraction temperature, extraction time, solvent-to-solid ratio, total polyphenol content, and total flavonoid content. Among the optimized machine learning (ML) models, the model with the highest R^2 value was selected. The GridSearchCV method from the scikit-learn library was applied for the grid search to ascertain the optimal extract condition. The visualization was implemented using Plotly (6.0.1), incorporating axis labels, color bars, and contour labels to effectively convey the optimal extraction condition.

Graphical user interface

A graphical user interface (GUI) was implemented using the PyQt5 library (version 5.15.10) in Python (version 3.10.14), taking advantage of the optimized structure of the XGB model (Meier 2019). This interface displays the status of the application on the monitor and allows user interaction *via* mouse and keyboard. Through adjusting conditions *via* buttons, users can efficiently obtain and validate accurate predictions generated by the trained model. Therefore, the authors applied the XGB model to the GUI for the purpose of predicting antioxidant activity. Kumar *et al.* (2022) adopted a GUI to display the results of machine learning models and reduce the repetition of code execution. The GUI improves accessibility for researchers by providing visualization and insights into model performance and simplifying code execution for predictive applications.

RESULTS AND DISCUSSION

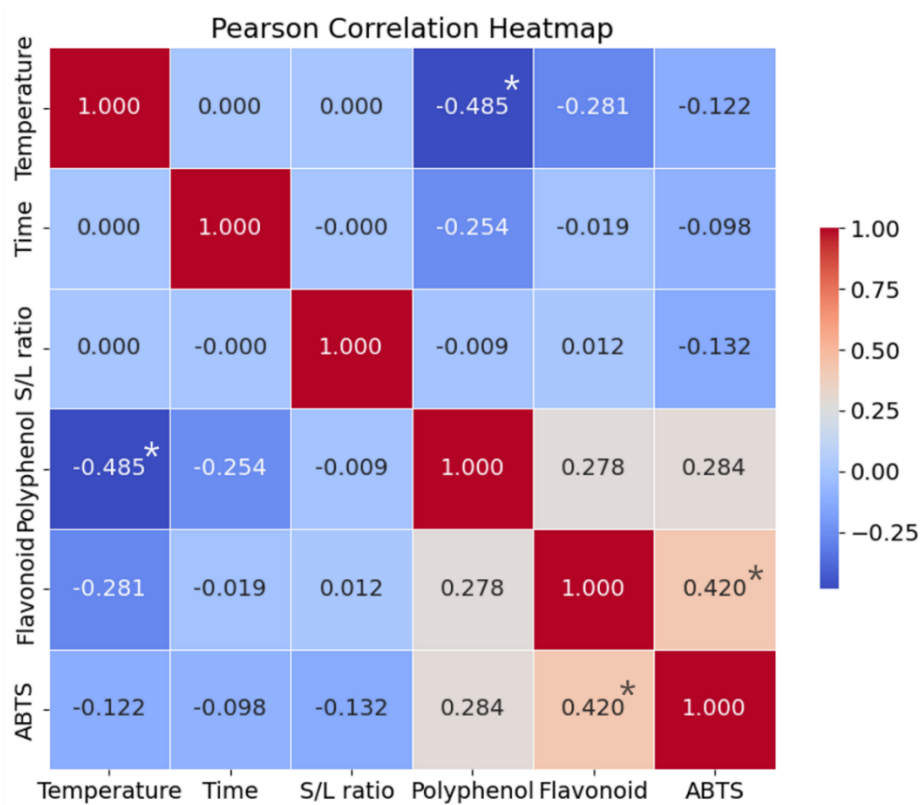
Collection of the Dataset

The input variables and predictors are shown in Table S1. Extracts were collected from *A. acerifolia* using the water extraction method.

Analysis of Pearson Correlation

To further explore the linear relationship between antioxidant activity and input characteristics, the Pearson correlation coefficient matrix is shown in Fig. 3, and only values with a p-value of 0.05 or less to test the hypothesis are marked with “*”. The gradient of elemental colors from blue to red in the matrix plot represents the logarithmic increase of the Pearson correlation coefficient from -1 to 1. For the antioxidant activity, there was a positive correlation with the flavonoids. Flavonoids are major secondary metabolites produced by plants and exhibit antioxidant activity through scavenging reactive oxygen species and antioxidant enzyme activity (Williamson *et al.* 2018). This is due to the contribution of the hydroxy group of ring B in the flavonoid structure to antioxidant activity (Rice-Evans *et al.* 1996). Nagarajan *et al.* (2020) reported that flavonoid polymers are excellent antioxidants due to the presence of many water-level hydroxyls in their molecules. Polyphenols showed a

lower correlation coefficient (0.284) for homeostatic activity compared to flavonoids. This indicates that when the degree of polymerization of polyphenols exceeded a threshold, the complexity of the molecule reduced the availability of hydroxyl groups, which negatively affected their antioxidant activity (Espín and Wichers 2000). Polyphenols and flavonoids were negatively correlated with temperature, with correlation coefficients of -0.485 and -0.281, respectively. Elevated extraction temperatures result in the degradation or loss of some heat-sensitive volatile phenolic/flavonoid compounds (Xiao *et al.* 2008).



*=significant at $P < 0.05$, respectively.

Fig. 3. Pearson correlation heatmap of temperature, time, S/L ratio, polyphenol, flavonoid, and ABTS

Predictive Performance of 2FI Model

To build a prediction model for antioxidant activity, the 2FI model and three ML algorithms, namely XGB, RF, and SVM, were used. The results of the ANOVA analysis of the 2FI model and the coefficient analysis values of the variables are shown in Table 3. The lower the p-value of a factor, the more it influences the 2FI model. The 2FI model for ABTS antioxidant activity (%) to predict the relationship between independent and dependent variables can be expressed as follows:

$$\text{ABTS} = 264.45086 - 1.32937A - 2.36092B - 9.06582C - 36.27082D + 3.39669E + 0.022547AB + 0.053713AC - 0.255629AD + 0.048973AE + 0.012333BC + 0.134263BD - 0.009934BE + 1.2921CD + 0.059328CE + 0.928443DE$$

where A is the temperature, B is the time, C is the liquid /solid ratio, D is the polyphenol content, and E is the flavonoid content.

The effect of independent variables on antioxidant activity was tested for adequacy and goodness of fit by ANOVA.

Table 3. ANOVA and Coefficients in Coded Factors For Two Factor Interaction Model

Source	Sum of Squares	df	Mean Square	F-Value	p-Value	Coefficient Estimate
Model	18436.47	15	1229.1	9.21	< 0.0001	
A ¹⁾	4.1	1	4.1	0.0307	0.8614	0.5878
B ²⁾	226.23	1	226.23	1.7	0.1974	-3.83
C ³⁾	886.14	1	886.14	6.64	0.0122	5.67
D ⁴⁾	85.43	1	85.43	0.6404	0.4265	4.12
E ⁵⁾	1892.31	1	1892.31	14.18	0.0004	14.5
AB	1115.59	1	1115.59	8.36	0.0052	6.76
AC	717.5	1	717.5	5.38	0.0235	5.37
AD	26.99	1	26.99	0.2023	0.6543	-3.27
AE	423.24	1	423.24	3.17	0.0796	-8.83
BC	253.96	1	253.96	1.9	0.1724	3.7
BD	77.71	1	77.71	0.5825	0.4481	5.16
BE	148.69	1	148.69	1.11	0.295	-5.37
CD	1164.55	1	1164.55	8.73	0.0044	16.54
CE	770.94	1	770.94	5.78	0.0191	10.7
DE	1062.89	1	1062.89	7.97	0.0063	21.43
Residual	8671.42	65	133.41			
Cor Total	27107.89	80				
R ²	0.6063					

¹⁾ Temperature (°C)

²⁾ Time (h)

³⁾ Liquid/Solid ratio (%)

⁴⁾ Polyphenol content (mgQue/g)

⁵⁾ Flavonoid content (mgGAE/g)

Table 3 summarizes the results of goodness of fit, variance, model adequacy, and coefficients of determination. Statistical analysis showed that the 2FI model had a very low p-value ($p < 0.0001$), which was highly significant. However, the coefficient of determination (R^2) indicated that 60.63% of the variation could be explained by the fitted model. The independent variables (Flavonoid content, Liquid/Solid ratio) and the interactions (Temperature-Time, Liquid/Solid ratio- Polyphenol content, Polyphenol content-Flavonoid content) influence antioxidant activity. The 2FI model had a low coefficient of determination for predicting the antioxidant activity of *A. acerifolia*, suggesting that it is not a suitable method for predicting antioxidant activity.

Predictive Performance of Machine Learning Model

Four ML algorithms were used to build the ML models: XGB, SVM, RF, and RANSAC. The dataset was randomly partitioned into training and testing sets, with a ratio of 80:20. Independent validation using external datasets was not performed. It is acknowledged that this may result in a restriction of the generalizability of the model's predictive performance, such that it remains consistent under entirely new experimental conditions. Each model used K-fold validation and RandomizedSearchCV for hyper-

parameter optimization. The optimized values of the hyper parameters are shown in Table 2. R^2 specifies the correlation between the predicted value and the target value. It is one of the commonly preferred parameters to measure the performance of a model. A comparison of the target values is shown in Fig. 4 and Fig. 5. The performance of the trained ML model was evaluated using training and test sets. The RMSE and R^2 values of the models on the train and test sets are shown in Table 4. The best performing model on the training set was the SVM model (R^2 : 0.9656, RMSE: 9.5996), but the best performing model on the test set was the XGB model (R^2 : 0.9835, RMSE: 2.5182).

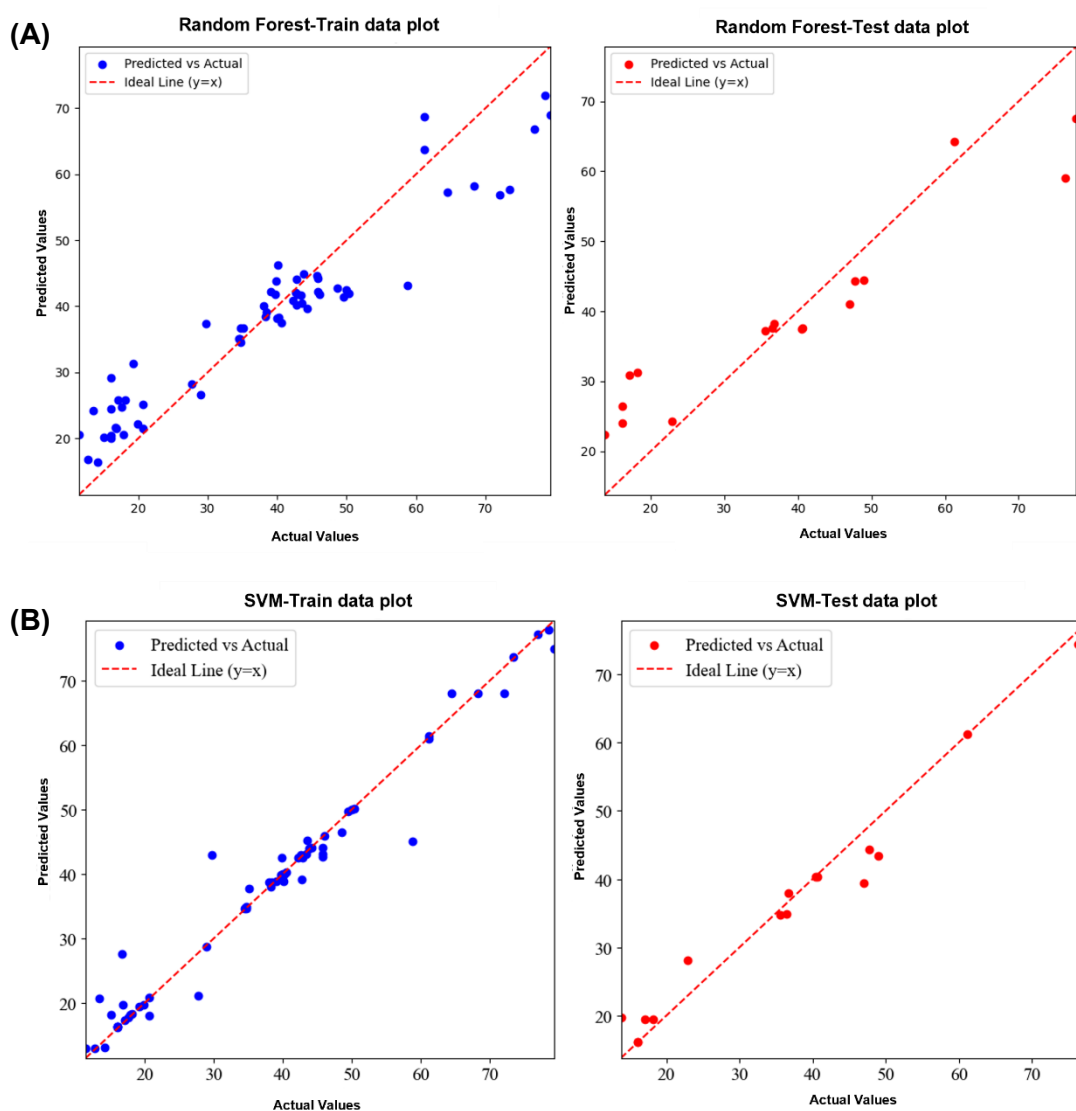


Fig. 4. Scatter plot of actual versus RF (A) and SVM(B) model predicted values for ABTS radical scavenging activity

The 2FI model (R^2 :0.6063), a mathematical model, exhibited a substantially lower predictive capacity compared to other machine learning models. This finding suggests that traditional mathematical models are effective in capturing linear relationships between input variables but are limited in capturing non-linear relationships. Zhu *et al.* (2024) and Alqahtani *et al.* (2025) found that the model based on XGB outperformed the traditional multiple linear regression model and other machine learning models in predicting the activity of extracts. The findings indicate that boosting-based machine learning (ML) models demonstrate superior efficacy in predicting the activity of extracts from variables (Temperature, Time, Liquid /Solid

ratio, Polyphenol content, and Flavonoid content) when compared to conventional regression models.

Table 4. R^2 and RMSE Results for Each Machine Learning Model on the Train Set and Test Set

	Train RMSE	Train R^2	Test RMSE	Test R^2
RF	13.2732	0.8741	8.0924	0.8292
SVM	9.5996	0.9656	3.2435	0.9726
XGB	12.8292	0.9992	2.5182	0.9835
RANSAC	18.3341	0.106	16.6802	0.2743

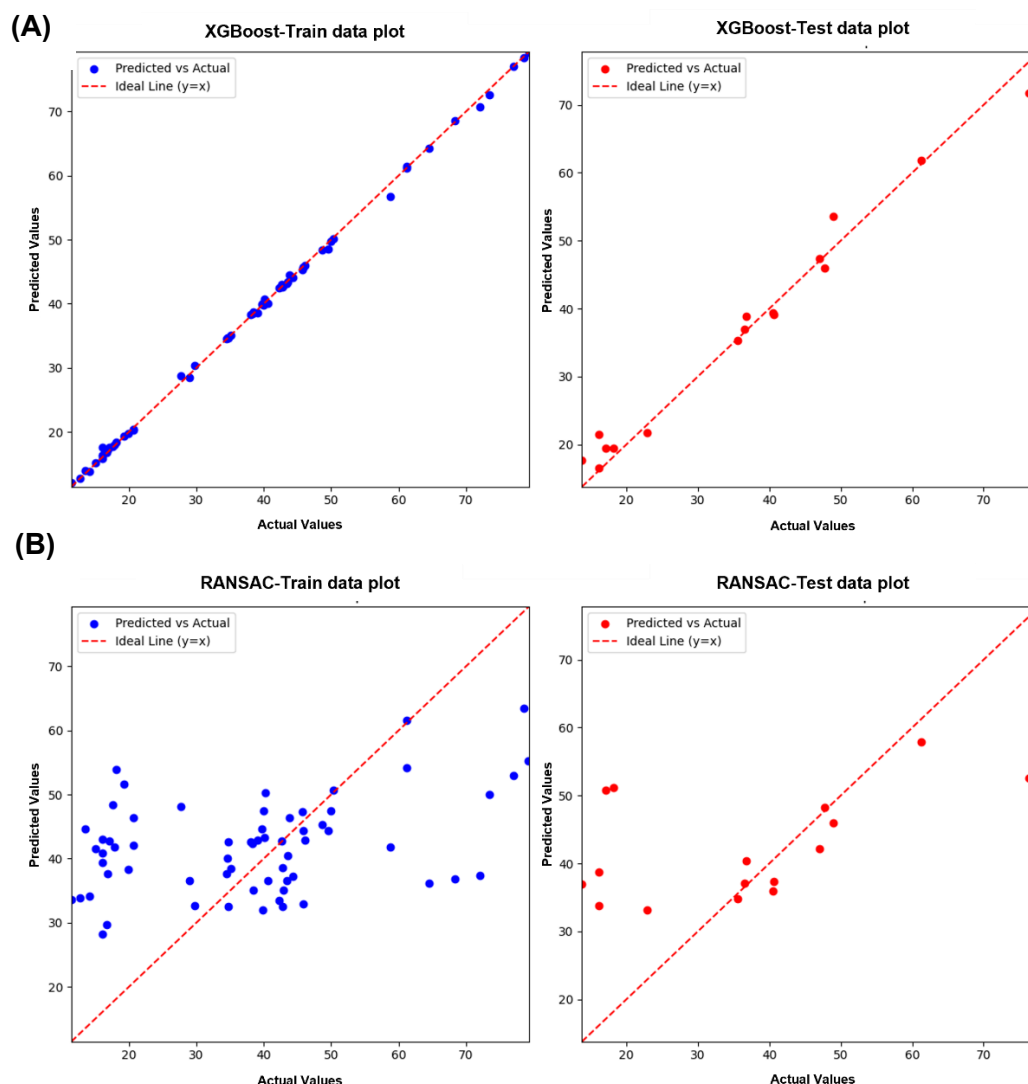


Fig. 5. Scatter plot of actual versus XGB (A) and RANSAC(B) model predicted values for ABTS radical scavenging activity

Evaluating Features of Importance in ML Models

The ML-based feature analysis was performed to assess the importance of the input features. The model was selected based on its prediction performance to generate feature importance. For antioxidant activity, flavonoids are the most important feature (~7) (Fig. 6), which is consistent with previous studies that flavonoids are an important factor influencing antioxidant activity (Abeyasinghe *et al.* 2021). Flavonoids are renowned for their antioxidant activity, which is attributed to their capacity to stabilize free radicals by donating hydrogen. Flavonoids are a class of phytonutrients that have

been shown to possess significant antioxidant properties. Their structural characteristics, particularly the presence of multiple hydroxy groups on the B ring, facilitate the donation of hydrogen to free radicals, thereby neutralizing them (Sekher Pannala *et al.* 2001; Wolfe and Liu 2008). Extraction temperature is the second important feature; the actual temperature is a factor that can determine the content of both flavonoids and polyphenols (Antony and Farid 2022). If the temperature of the extraction increases, the cell matrix opens up and consequently increases the availability of flavonoids for extraction. In addition, at higher temperatures, solvent viscosity decreases, and diffusivity increases, which increases extraction efficiency (Camel 2000; Pan *et al.* 2000). Considering the above features, the most important way to increase antioxidant activity is to optimize the extraction temperature.

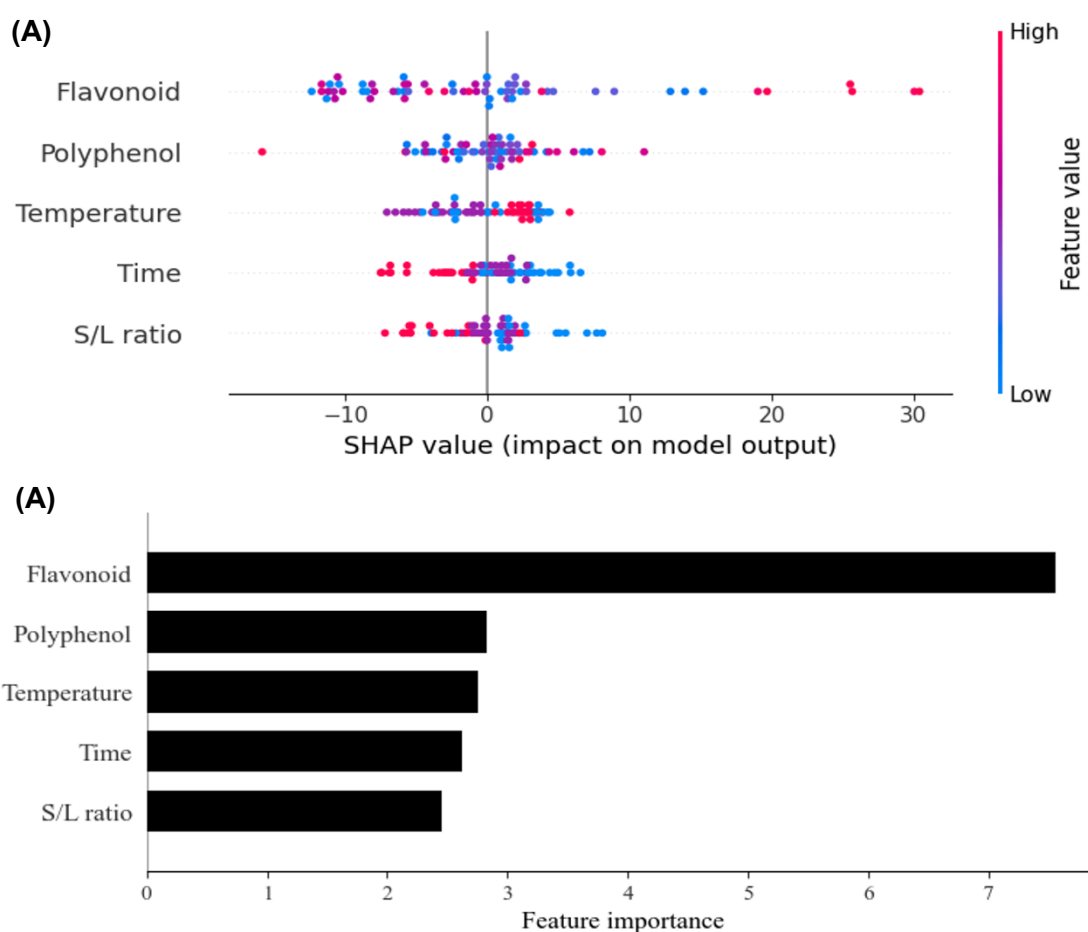


Fig. 6. SHAP value (A) and feature importance (B) of each input feature for predicting ABTS antioxidant activity

Extraction Process Optimization

The XGB model, which demonstrated the highest R^2 value among the various machine learning models, was selected to optimize the extraction process. Figure 7 illustrates the multivariate relationships among process variables (Temperature, Time, S/L Ratio, Polyphenol, and Flavonoid), as well as the model's predicted ABTS radical-scavenging activity (%). This was achieved by employing parallel coordinates visualization of 100 randomly sampled grid-search experiments. Each polyline in the graph represents an individual experimental run, with its color on the "Plasma" scale corresponding to the predicted ABTS inhibition (10% to 90%). This perspective underscores the pivotal role that combinations of extraction conditions play in

determining antioxidant performance. The process variable values that were optimized with the XGB model are as follows: The experiment yielded the following results: temperature of 80 degrees Celsius, time of 75 minutes, S/L ratio of 25, polyphenol of 2.0, flavonoid of 39.4, and maximum ABTS antioxidant activity of 84.31%.

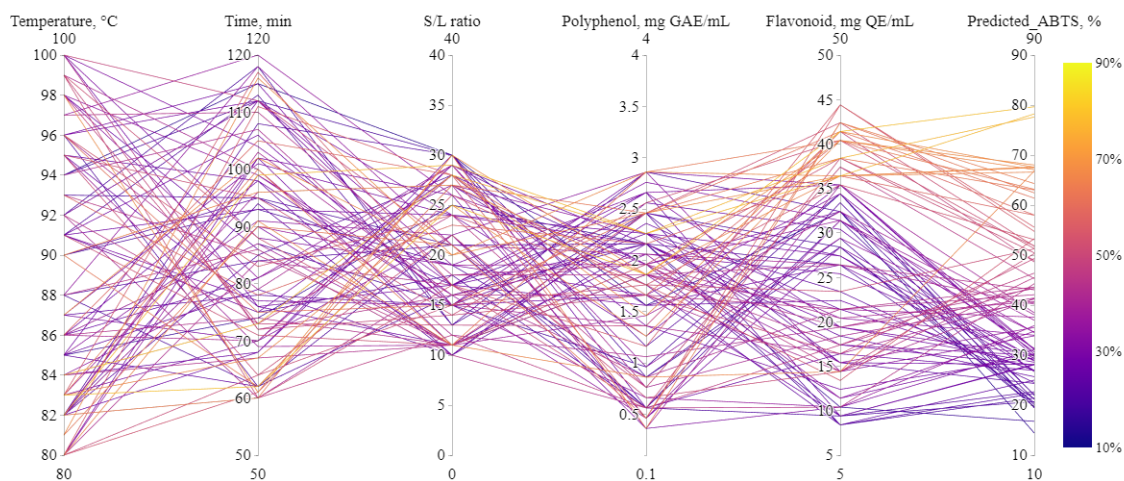


Fig. 7. Parallel coordinates mapping of experimental runs: correlating Temperature, Time, S/L ratio, Polyphenol/Flavonoid contents with predicted ABTS activity

Comparison of 2FI and ML Results

Data on the antioxidant activity of *A. acerifolia* extract were used to optimize and predict results using the XGB model. This model was trained to learn from the antioxidant properties of the extract and the interactions between extraction temperature, time, liquid/solid ratio, polyphenols, and flavonoids. Among the machine learning models evaluated, XGB showed superior performance compared to the two-factor interaction (2FI) model, achieving higher prediction accuracy. The improved performance of XGB can be attributed to its advanced data processing capabilities. While 2FI models are commonly used for numerical optimization of individual variables, they are limited to quadratic regression, which limits their predictive power. In contrast, XGB models effectively capture and learn the non-linear interactions between process variables within each tree, making them highly adaptable to different applications. Figure 7 shows a comparison of the predicted antioxidant activity of maple aroma extract using 2FI and ML. The R^2 values of the RSM (2FI) model and the ML (XGB) model are 0.60 and 0.98, respectively. Therefore, it can be concluded that XGB efficiently optimized and predicted the antioxidant activity of the extract as a function of temperature, time, and solvent concentration. Kunjiappan *et al.* (2024) compared the use of 2FI and machine learning models to predict the bioactivity of *Vitis vinifera* extracts and reported that the machine learning model achieved higher prediction accuracy. Kabilan *et al.* used RSM to find the optimal conditions for extracts of *Boerhavia diffusa* Linn and combined it with machine learning to predict bioactivity, showing high reliability (0.957).

While the present study achieved a high degree of success in accurately predicting the antioxidant activity of *A. acerifolia* extracts using XGB model, it is important to note that its limitations are primarily associated with its exclusive reliance on a particular plant species (*A. acerifolia*) and a single antioxidant activity assay. Consequently, the study's generalizability to other plant materials or antioxidant activity mechanisms may be constrained. Moreover, the XGB model is constrained by the absence of additional external validation through independent datasets.

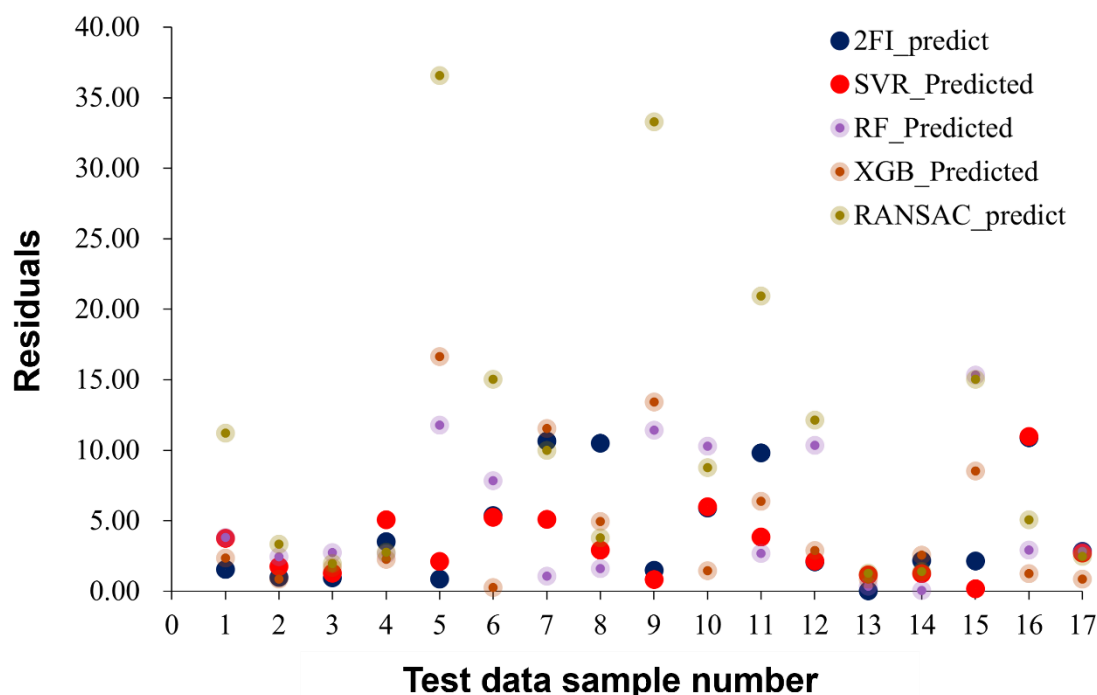


Fig. 8. Scatter plot of residuals of predicted data (2FI, SVR, RF, XGB, RANSAC) against actual data for ABTS antioxidant activity

Development of a Graphical User Interface

To predict the antioxidant activity of an extract, the authors developed a user-friendly graphical user interface (GUI). Figure 8 shows the schematic of the GUI. In the GUI, users can predict the antioxidant activity of an extract by entering the extraction temperature, time, liquid/solid ratio, and extraction factors of polyphenols and flavonoids. The GUI efficiently generated predictions in less than 0.4 s. Leveraging the high predictive accuracy of the XGB model, the system provides reliable results. This framework increases the efficiency of analysis and decision making for both researchers and industry professionals.

In preliminary experiments, the GUI-integrated antioxidant activity model demonstrated a strong prediction probability of 0.98. In addition, the performance of the model was further evaluated using an external dataset. At the following link the GUI program can be downloaded ([Antioxidant prediction.zip](#)).

The XGB-based prediction model proposed in this study has the potential to enhance the efficiency of developing antioxidant functional materials by deriving optimal extraction conditions without the necessity of repeated experiments. Furthermore, it can be utilized to forecast the antioxidant activity of extracts through the implementation of measurement methods such as DPPH and FRAP. Also, the GUI has the capacity to simulate and predict an array of extraction conditions in real time within R&D environments. This capability enhances the efficiency and precision of decision-making processes.

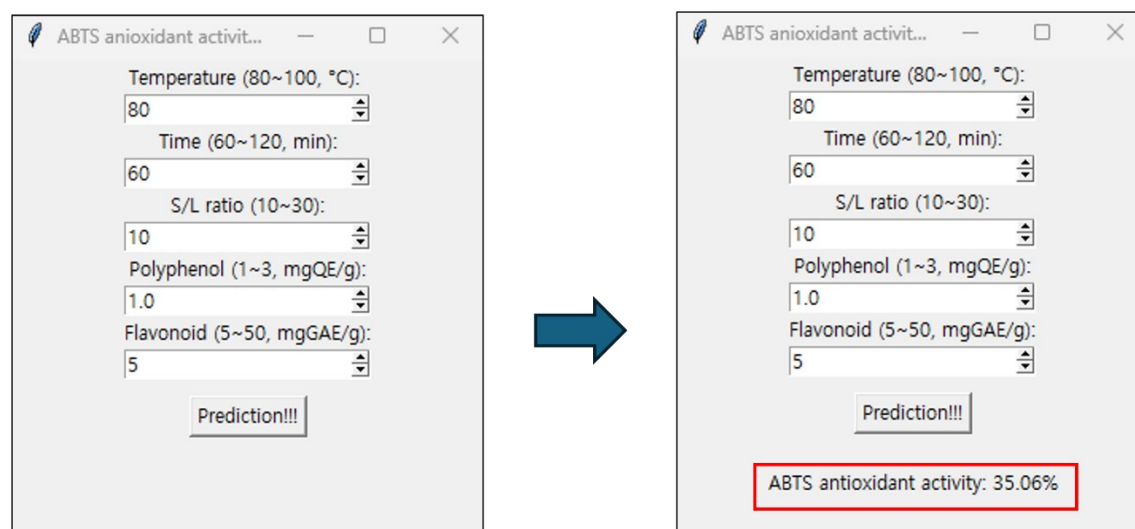


Fig. 9. Schematic representation of the GUI from the ABTS antioxidant activity prediction model

CONCLUSIONS

1. Based on the analysis and results presented, the study successfully developed a predictive model for antioxidant activity using various machine learning techniques. The extraction of antioxidant compounds from *A. acerifolia* was thoroughly analyzed, with a particular focus on the relationship between input variables and antioxidant activity.
2. The study highlighted the significant positive correlation between flavonoid content and antioxidant activity, emphasizing the role of extraction temperature on flavonoid preservation. The XGB model emerged as the most effective predictive tool, surpassing both the 2FI model and other machine learning algorithms such as RF and SVM in terms of prediction accuracy. The XGB model's ability to handle nonlinear relationships and its high R^2 value of 0.9805 demonstrate its robustness in predicting antioxidant activity.
3. This is particularly important for optimizing extraction processes and enhancing the quality of the extracts. Moreover, the development of a user-friendly graphical user interface (GUI) based on the XGB model allows for rapid and accurate predictions of antioxidant activity.
4. This tool simplifies the decision-making process for researchers and industry professionals, offering a practical application of the study's findings. In conclusion, the study provides a comprehensive framework for predicting and optimizing antioxidant activity in plant extracts, with the XGB model playing a central role in advancing the analytical capabilities in this domain. Future research could expand on this work by exploring additional variables and refining the model further to enhance its applicability across different types of plant extracts.

ACKNOWLEDGMENTS

This study was carried out with the support of R&D Program for Forest Science Technology (Project No. "RS-2023-KF00245261382116530003") provided by Korea Forest Service (Korea Forestry Promotion Institute).

DATA AVAILABILITY

All datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

REFERENCE CITED

- Abeyasinghe, D.T., Kumara, K. A. H., Kaushalya, K. A. D., Chandrika, U. G., and Alwis, D. D. D. H. (2021). "Phytochemical screening, total polyphenol, flavonoid content, in vitro antioxidant and antibacterial activities of Sri Lankan varieties of *Murraya koenigii* and *Micromelum minutum* leaves," *Heliyon* 7(7), article e07449. DOI: 10.1016/j.heliyon.2021.e07449
- Alqahtani, N. K., Alnemr, T. M., Farag, H. A. S., Ismail, R., and Habib, H. M. (2025). "Machine learning insights into the antioxidant and biomolecular shielding effects of polyphenol-rich 18 date palm pit extracts," *Food Chem. X* 27, article 102480. DOI: 10.1016/j.fochx.2025.102480
- Antony, A., and Farid, M. (2022). "Effect of temperatures on polyphenols during extraction," *Appl. Sci.* 12(4), article 2107. DOI: 10.3390/app12042107
- Barreira, J. C., Alves, R. C., Casal, S., Ferreira, I. C., Oliveira, M. B. P., and Pereira, J. A. (2009). "Vitamin E profile as a reliable authenticity discrimination factor between chestnut (*Castanea sativa* Mill.) cultivars," *J. Agric. Food. Chem.* 57(12), 5524-5528.
- Baş, D., and Boyacı, İ. H. (2007). "Modeling and optimization I: Usability of response surface methodology," *J. Food. Eng.* 78(3), 836-845. DOI: 10.1016/j.jfoodeng.2005.11.024
- Belwal, T., Dhyani, P., Bhatt, I. D., Rawal, R. S., and Pande, V. (2016). "Optimization extraction conditions for improving phenolic content and antioxidant activity in *Berberis asiatica* fruits using response surface methodology (RSM)," *Food Chem.* 207, 115-124. DOI: 10.1016/j.foodchem.2016.03.081
- Bergstra, J., and Bengio, Y. (2012). "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.* 13(2), article 2188395. DOI: 10.5555/2503308.2188395
- Biau, G., and Scornet, E. (2016). "A random forest guided tour," *Test* 25, 197-227. DOI: 10.1007/s11749-016-0481-7
- Cacace, J. E., and Mazza, G. (2003). "Optimization of extraction of anthocyanins from black currants with aqueous ethanol," *J. Food Sci.* 68(1), 240-248. DOI: 10.1111/j.1365-2621.2003.tb14146.x
- Camel, V. (2000). "Microwave-assisted solvent extraction of environmental samples," *Trends Anal. Chem.* 19(4), 229-248. DOI: 10.1016/S0165-9936(99)00185-5

- Chen, B., Zhao, Y., Yu, D., Lin, F., Xu, Z., Song, J., and Li, X. (2024). "Optimizing the extraction of active components from *Salvia miltiorrhiza* by combination of machine learning models and intelligent optimization algorithms and its correlation analysis of antioxidant activity," *Prep. Biochem. Biotechnol.* 54(3), 358-373. DOI: 10.1080/10826068.2023.2243493
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., and Zhou, T. (2015). "Xgboost: Extreme gradient boosting," *R Package Version 0.4-2* 1(4), 1–4.
- Choi, S. Z., Yang, M. C., Choi, S. U., and Lee, K. R. (2006a). "Cytotoxic terpenes and lignans from the roots of *Ainsliaea acerifolia*," *Arch. Pharm. Res.* 29, 203-208. DOI: 10.1007/BF02969394
- Choi, Y., Lee, S. M., Chun, J., Lee, H. B., and Lee, J. (2006b). "Influence of heat treatment on the antioxidant activities and polyphenolic compounds of shiitake (*Lentinus edodes*) mushroom," *Food Chem.* 99(2), 381-387. DOI: 10.1016/j.foodchem.2005.08.004
- Dong, J., Cai, L., Xing, Y., Yu, J., and Ding, Z. (2015). "Re-evaluation of ABTS• assay for total antioxidant capacity of natural products," *Nat. Prod. Commun.* 10(12), 1934578X1501001239. DOI: 10.1177/1934578X1501001239
- Ed-Dahmani, I., El fadili, M., Kandsi, F., Conte, R., El Atki, Y., Kara, M., Assouguem, A., Touijer, H., Lfitat, A., Nouioura, G., *et al.* (2024). "Phytochemical, antioxidant activity, and toxicity of wild medicinal plant of *Melilotus albus* extracts, *in vitro* and *in silico* approaches," *ACS Omega* 9(8), 9236-9246. DOI: 10.1021/acsomega.3c08314
- Espín, J. C., and Wichers, H. J. (2000). "Study of the oxidation of resveratrol catalyzed by polyphenol oxtase. Effect of polyphenol oxidase, laccase and peroxidase on the antiradical capacity of resveratrol," *J. Food Biochem.* 24(3), 225-250. DOI: 10.1111/j.1745-4514.2000.tb00698.x
- Fujimoto, T., and Gotoh, H. (2023). "Feature selection for the interpretation of antioxidant mechanisms in plant phenolics," *Molecules* 28(3), article 1454. DOI: 10.3390/molecules28031454
- Golovynska, I., Golovynskyi, S., and Qu, J. (2023). "Comparing the impact of NIR, visible and UV light on ROS upregulation *via* photoacceptors of mitochondrial complexes in normal, immune and cancer cells," *Photochem. Photobiol.* 99(1), 106-119. DOI: 10.1111/php.13661
- Ha, S. Y., Jung, J. Y., Kim, H. C., and Yang, J. (2024). "Optimization of antioxidant activity and phenolic extraction from *Ainsliaea acerifolia* stem using ultrasound-assisted extraction technology," *BioResources* 19(3), 6325-6338. DOI: 10.15376/biores.19.3.6325-6338
- Ifeanyi, O. E. (2018). "A review on free radicals and antioxidants." *Int. J. Curr. Res. Med. Sci.* 4(2), 123-133. DOI: 10.22192/ijcrms.2018.04.02.019
- Jung, C., Kwon, H., Choi, S., Lee, J., Lee, D., Ryu, S., and Lee, K. (2000). "Phytochemical constituents of *Ainsliaea acerifolia*," *Korean Journal of Pharmacognosy* 31(2), 125-129.
- Kabilan, S. J., Sivakumar, O., Sumanth, G. B., Kannan, S., Kunjiappan, S., and Sundar, K. (2024). "Optimization and analysis of ultrasound-assisted solvent extraction of bioactive compounds from *Boerhavia diffusa* Linn. using RSM, ANFIS and machine learning algorithm," *Journal of Food Measurement and Characterization* 18(6), 4204-4220. DOI: 10.1007/s11694-024-02487-w
- Karimifard, S., and Alavi Moghaddam, M. R. (2018). "Application of response surface methodology in physicochemical removal of dyes from wastewater: A critical review," *Sci. Total Environ.* 640-641, 772-797. DOI:

- 10.1016/j.scitotenv.2018.05.355
- Khedmati, M., Khodaii, A., and Haghshenas, H. F. (2017). "A study on moisture susceptibility of stone matrix warm mix asphalt," *Constr. Build. Mater.* 144, 42-49. DOI: 10.1016/j.conbuildmat.2017.03.121
- Kongolo Kalembe, M. R., Makhuvele, R., and Njobeh, P. B. (2024). "Phytochemical screening, antioxidant activity of selected methanolic plant extracts and their detoxification capabilities against AFB1 toxicity," *Heliyon* 10(2), article e24435. DOI: 10.1016/j.heliyon.2024.e24435
- Kumar, K. S., Sai Sathya, M., Nadeem, A., and Rajesh, S. (2022). "Diseases prediction based on symptoms using database and GUI," in: *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, pp. 1353-1357. DOI: 10.1109/ICCMC53470.2022.9753707
- Kunjiappan, S., Ramasamy, L. K., Kannan, S., Pavadai, P., Theivendren, P., and Palanisamy, P. (2024). "Optimization of ultrasound-aided extraction of bioactive ingredients from *Vitis vinifera* seeds using RSM and ANFIS modeling with machine learning algorithm," *Sci. Rep.* 14(1), article 1219. DOI: 10.1038/s41598-023-49839-y
- Lee, K. J., Ma, K., Cho, Y., Lee, J., Chung, J., and Lee, G. (2017). "Phytochemical distribution and antioxidant activities of Korean adzuki bean (*Vigna angularis*) landraces," *J. Crop Sci. Biotechnol.* 20, 205-212. DOI: 10.1007/s12892-017-0056-0
- Li, M., Sun, H., Huang, Y., and Chen, H. (2024). "Shapley value: From cooperative game to explainable artificial intelligence," *Auton. Intell. Syst.* 4(1), 2. DOI: 10.1007/s43684-023-00060-8
- Li, X., Chen, S., Zhang, J., Yu, L., Chen, W., and Zhang, Y. (2022). "Optimization of ultrasonic-assisted extraction of active components and antioxidant activity from *polygala tenuifolia*: A comparative study of the response surface methodology and least squares support vector machine," *Molecules* 27(10), article 3069. DOI: 10.3390/molecules27103069
- Macías, F. A., Simonet, A. M., Galindo, J. C. G., and Castellano, D. (1999). "Bioactive phenolics and polar compounds from *Melilotus messanensis*. Part 1," *Phytochem.* 50(1), 35-46. DOI: 10.1016/S0031-9422(98)00453-1
- Meier, B. (2019). *Python GUI Programming Cookbook: Develop Functional and Responsive User Interfaces with Tkinter and PyQt5*, Packt Publishing Ltd., Birmingham, England.
- Michiels, J. A., Kevers, C., Pincemail, J., Defraigne, J. O., and Dommes, J. (2012). "Extraction conditions can greatly influence antioxidant capacity assays in plant food matrices," *Food Chem.* 130(4), 986-993. DOI: 10.1016/j.foodchem.2011.07.117
- Mizobuchi, M., Ishidoh, K., and Kamemura, N. (2022). "A comparison of cell death mechanisms of antioxidants, butylated hydroxyanisole and butylated hydroxytoluene," *Drug Chem. Toxicol.* 45(4), 1899-1906. DOI: 10.1080/01480545.2021.1894701
- Nagarajan, S., Nagarajan, R., Kumar, J., Salemm, A., Togna, A. R., Saso, L., and Bruno, F. (2020). "Antioxidant activity of synthetic polymers of phenolic compounds," *Polymers* 12(8), article 1646. DOI: 10.3390/polym12081646
- Pan, X., Liu, H., Jia, G., and Shu, Y. Y. (2000). "Microwave-assisted extraction of glycyrrhizic acid from licorice root," *Biochem. Eng. J.* 5(3), 173-177. DOI: 10.1016/S1369-703X(00)00057-7

- Park, H. (2010). "Chemistry and pharmacological action of caffeoylquinic acid derivatives and pharmaceutical utilization of chwinamul (Korean mountainous vegetable)," *Arch. Pharm. Res.* 33, 1703-1720. DOI: 10.1007/s12272-010-1101-9
- Pham-Huy, L. A., He, H., and Pham-Huy, C. (2008). "Free radicals, antioxidants in disease and health," *Int. J. Biomed. Sci.* 4(2), 89.
- Płotka-Wasyłka, J., Rutkowska, M., Owczarek, K., Tobiszewski, M., and Namieśnik, J. (2017). "Extraction with environmentally friendly solvents," *Trends Analyt. Chem.* 91, 12-25. DOI: 10.1016/j.trac.2017.03.006
- Renaud, O., and Victoria-Feser, M. (2010). "A robust coefficient of determination for regression," *J. Stat. Plan. Inference.* 140(7), 1852-1862. DOI: 10.1016/j.jspi.2010.01.008
- Rice-Evans, C. A., Miller, N. J., and Paganga, G. (1996). "Structure-antioxidant activity relationships of flavonoids and phenolic acids," *Free Radic. Biol. Med.* 20(7), 933-956. DOI: 10.1016/0891-5849(95)02227-9
- Ryo, M., and Rillig, M. C. (2017). "Statistically reinforced machine learning for nonlinear patterns and variable interactions," *Ecosphere* 8(11), article e01976. DOI: 10.1002/ecs2.1976
- Sekher Pannala, A., Chan, T. S., O'Brien, P. J., and Rice-Evans, C. A. (2001). "Flavonoid B-ring chemistry and antioxidant activity: Fast reaction kinetics," *Biochem. Biophys. Res. Commun.* 282(5), 1161-1168. DOI: 10.1006/bbrc.2001.4705
- Singleton, V. L., and Rossi, J. A. (1965). "Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents," *Am. J. Enol. Vitic.* 16(3), 144-158. DOI: 10.5344/ajev.1965.16.3.144
- Suthaharan, S. (2016). "Support vector machine," in: *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*, Springer, New York, NY, USA, pp. 207-235.
- Wang, P., Gong, Q., Hu, J., Li, X., and Zhang, X. (2021). "Reactive oxygen species (ROS)-responsive prodrugs, probes, and theranostic prodrugs: Applications in the ROS-related diseases," *J. Med. Chem.* 64(1), 298-325. DOI: 10.1021/acs.jmedchem.0c01704
- Williamson, G., Kay, C. D., and Crozier, A. (2018). "The bioavailability, transport, and bioactivity of dietary flavonoids: A review from a historical perspective," *Compr. Rev. Food Sci. Food Saf.* 17(5), 1054-1112. DOI: 10.1111/1541-4337.12351
- Wolfe, K. L., and Liu, R. H. (2008). "Structure-activity relationships of flavonoids in the cellular antioxidant activity assay," *J. Agric. Food. Chem.* 56(18), 8404-8411. DOI: 10.1021/jf8013074
- Xiao, W., Han, L., and Shi, B. (2008). "Microwave-assisted extraction of flavonoids from radix astragali," *Sep. Purif. Technol.* 62(3), 614-618. DOI: 10.1016/j.seppur.2008.03.025
- Yıkımsı, S., Duman Altan, A., Türkol, M., Gezer, G. E., Ganimet, Ş., Abdi, G., Hussain, S., and Aadil, R. M. (2024). "Effects on quality characteristics of ultrasound-treated gilaburu juice using RSM and ANFIS modeling with machine learning algorithm," *Ultrason. Sonochem.* 107, article 106922. DOI: 10.1016/j.ultsonch.2024.106922
- Zhang, Y., and Haghani, A. (2015). "A gradient boosting method to improve travel time prediction," *Transp. Res. Part C- Emerg. Technol.* 58, 308-324. DOI: 10.1016/j.trc.2015.02.019

Zhu, P., Li, R., and Lu, A. (2024). "Electrode impedance modeling based on XGboost algorithm for analyzing the antioxidant properties of juice," *Journal of Food Measurement and Characterization* 18(6), 5031-5042. DOI: 10.1007/s11694-024-02553-3

Article submitted: March 11, 2025; Peer review completed: April 6, 2025; Revised version received: May 7, 2025; Accepted: May 20, 2025; Published: August 27, 2025
DOI: 10.15376/biores.20.4.9103-9126

APPENDIX

Table S1. Classifying Training and Test Data for Model Training

No.	Class	Temperature (°C)	Time (min)	Liquid/Solid Ratio	Polyphenol (mgQE/g)	Flavonoid (mgGAE/g)	ABTS (%)
1	Train	80	120	20	1.13	30.68	16.09
2	Train	90	120	30	1.87	9.28	28.98
3	Train	90	60	30	0.92	9.49	12.8
4	Train	100	60	30	1.65	17.35	46.1
5	Train	80	120	20	1.37	30.92	16.1
6	Train	100	60	10	1.09	28.56	40.01
7	Train	90	90	20	1.5	25.42	13.48
8	Train	80	90	20	1.72	43.3	76.99
9	Train	90	120	20	1.28	17.55	44.28
10	Train	100	60	30	1.93	17.63	48.61
11	Train	90	60	20	2.21	27.8	18.12
12	Train	80	60	20	1.13	26.37	49.54
13	Train	100	120	20	1.10	11.28	42.28
14	Train	80	60	20	1.85	27.09	50.36
15	Train	90	120	30	1.09	8.5	16.72
16	Train	90	120	10	0.80	9.75	45.8
17	Train	100	60	10	0.77	28.24	39.76
18	Train	90	60	30	0.89	9.46	11.45
19	Train	80	60	10	1.50	19.14	45.75
20	Train	90	90	30	1.58	11.85	16.84
21	Train	80	90	10	2.55	40.66	61.22
22	Train	100	90	30	0.36	25.50	34.70
23	Train	90	60	30	0.95	9.52	14.15
24	Train	90	120	10	0.70	9.65	39.88
25	Train	90	90	20	1.70	25.62	20.62
26	Train	80	90	30	2.61	36.01	79.29
27	Train	80	120	30	1.85	17.32	16.10
28	Train	90	60	20	1.57	27.16	17.6
29	Train	80	60	30	1.62	22.19	45.84
30	Train	100	60	30	1.37	17.07	43.59
31	Train	80	90	30	2.01	35.41	73.47
32	Train	80	120	10	1.05	27.60	15.10
33	Train	100	90	20	0.86	35.40	39.07
34	Train	80	60	30	1.43	22.00	42.70
35	Train	90	120	20	1.80	18.07	58.78
36	Train	100	120	20	1.46	11.64	43.44

37	Train	100	120	10	1.00	20.30	35.13
38	Train	80	60	20	1.49	26.73	49.95
39	Train	90	60	10	2.30	12.61	19.22
40	Train	80	120	30	0.57	16.04	16.08
41	Train	100	90	10	1.12	8.43	34.45
42	Train	80	120	10	1.11	27.66	20.66
43	Train	100	60	20	0.78	18.72	42.76
44	Train	80	90	10	1.71	39.82	61.2
45	Train	100	120	30	0.75	34.57	40.57
46	Train	90	120	10	0.75	9.70	42.84
47	Train	90	90	30	1.66	11.93	19.91
48	Train	100	90	10	1.68	8.99	34.77
49	Train	100	90	10	1.40	8.71	34.61
50	Train	90	120	20	0.76	17.03	29.78
51	Train	90	60	20	0.93	26.52	17.08
52	Train	90	90	10	0.75	12.76	64.53
53	Train	100	90	20	0.90	35.44	40.12
54	Train	90	90	10	0.82	12.83	68.33
55	Train	100	60	10	1.41	28.88	40.26
56	Train	80	120	10	1.08	27.63	17.88
57	Train	100	120	10	1.44	20.74	38.33
58	Train	100	120	20	1.28	11.46	42.86
59	Train	90	90	10	0.89	12.90	72.13
60	Train	100	90	20	0.82	35.36	38.02
61	Train	80	90	20	2.92	44.50	78.59
62	Train	80	60	10	1.39	19.03	43.80
63	Train	100	60	20	0.37	18.31	38.49
64	Train	90	90	20	1.90	25.82	27.76
65	Test	90	120	30	1.48	8.89	22.85
66	Test	100	90	30	0.62	25.76	35.60
67	Test	100	120	30	0.83	34.65	40.68
68	Test	100	90	30	0.88	26.02	36.50
69	Test	90	60	10	2.25	12.56	18.17
70	Test	80	120	20	0.89	30.44	16.08
71	Test	80	60	30	1.81	22.38	48.98
72	Test	100	60	20	1.19	19.13	47.03
73	Test	90	60	10	2.20	12.51	17.12
74	Test	80	120	30	1.21	16.68	16.09
75	Test	90	90	30	1.50	11.77	13.77
76	Test	80	90	20	2.32	43.90	77.79

77	Test	100	120	10	1.22	20.52	36.73
78	Test	80	60	10	1.61	19.25	47.70
79	Test	80	90	30	2.31	35.71	76.38
80	Test	80	90	10	2.13	40.24	61.21
81	Test	100	120	30	0.67	34.49	40.46