

MACHINE LEARNING BASED ESTIMATION OF DRYING CHARACTERISTICS OF APPLE SLICES

Necati Çetin ^{1,*}

¹Erciyes University, Faculty of Agriculture, Department of Biosystems Engineering, 38039, Kayseri, Turkey



Abstract

Machine learning algorithms have been usually used in food drying. These models are also effectively used for nonlinear processes such as heat and mass transfer. Estimation of drying characteristics is also important for optimizing drying conditions. Estimating of moisture rate and drying rate ensures accurate and high quality drying of the product under air-convective drying conditions. In this study, drying rate (DR) and moisture ratio (MR) were estimated in air-convective conditions with the use of drying time, moisture content (d.b.), and effective moisture diffusivity as input. In addition, two different validation methodology was performed as *k*-fold cross validation and train test split. In the present study random forest-RF; multilayer perceptron-MLP; and *k*-nearest neighbor-kNN were performed to estimate of drying rate and moisture ratio. As a result, correlation coefficients were found above 0.8500 for moisture ratio and 0.8722 for drying rate. The findings show that algorithms could be successfully applied for the estimation of drying rate and moisture ratio.

Keywords: Apple, random forest, neural network, nearest neighbors, moisture content

1. INTRODUCTION

Apple (*Malus communis* L.) is the second most produced fruit worldwide after banana. Apple, which has the ability to adapt to various ecological environments, is grown in a wide geographical area (Spengler, 2019). The world's annual apple production is about 87 million tons. Turkey, which produces approximately 3 million tons per year, ranks third after China and USA (FAO, 2021). Drying is one of the most common methods used in the processing and preservation of food products (Deng et al., 2019). Air-convective method, which is also called hot air drying, hot air is passed over the product layer. Evaporating moisture creates a thin boundary layer saturated with water on the product surface. The high partial vapor pressure of the saturated layer creates an inhibiting effect on drying. For this reason, temperature and air velocity are important as they determine the rate of removal of the moisture evaporated from the product (Cemeroğlu, 2011).

Mathematical modeling is among the most important stages of drying technologies. Proper drying kinetics need to be known for estimation of drying rate, optimization of drying parameters and design of dryers (Cruz et al., 2015; Naderinezhad et al., 2016). Machine learning offers linear and nonlinear models that can estimate past and future values in an input-output link (Zhang et al., 2012). Machine learning-based estimation applications are effective tools used in the design of accurate and reliable estimators. These models are generally used for the correct selection of descriptive features in the quality assessment of agricultural products (Omid et al., 2010; Mollazade

et al., 2012). Recently, machine learning technologies have been developed and applied for the evaluation of agricultural products. These techniques offer solution-oriented approaches through rapid and simple simulations (Soares et al., 2013). However, these algorithms can provide unsatisfactory performances on highly nonlinear datasets (Song et al., 2018; Zou et al., 2010). The aim of this study; i) drying of apples under air-convective drying, ii) determine the drying characteristics such as moisture content, moisture ratio, drying rate, drying time, and effective moisture diffusivity, iii) estimate the moisture ratio and drying rate.

2. MATERIALS AND METHODS

2.1. Materials

Two common apple cultivars Granny Smith and Red Delicious were used as material in the study. Apples were supplied from a local farmer in Yeşilhisar district of Kayseri. Fresh products were brought to the drying laboratory in the portable type of refrigerator in order to prevent moisture loss and were kept in the refrigerator of +4 ° C during the drying period.

2.2. Methods

2.2.1. Drying Processes and Drying Characteristics

Tests were performed in convective drying cabinet (ETHK-20M, TR). The temperature could be set between 40 and 280 °C and the air velocity was adjusted with the value of 0.5 m s⁻¹. The moisture loss of the samples was recorded by measuring with a precision balance (±0.001 g) every 60 minutes. Initial moisture content was determined as 85.17% and 86.37% (wet basis) for Granny Smith and Red Delicious, respectively, and 50±5 g samples were used in drying processes. To determine the initial moisture content, the samples were dried in an oven at 105 °C for 24 hours. The drying process was continued according to the response surface method until the samples reached the equilibrium moisture.

The moisture content of the material before and after the studies were calculated by using the following equation (Eq.1), respectively (Yağcıoğlu, 1999).

$$M_0 = \frac{(W_0 - W_d)}{W_d} = \frac{W_w}{W_d} \quad (1)$$

Where: M_0 is moisture content (kg water. kg drymatter⁻¹); W_0 is initial weight (kg); W_d is final weight (kg) and W_w is wet weight (kg).

The MR of apple slices during drying was determined using the following equation (Pinar et al., 2021):

$$MR = \frac{M_t - M_e}{M_0 - M_e} \quad (2)$$

where M_t (kg kg⁻¹) is moisture content at time t (dry basis), M_0 (kg kg⁻¹) initial moisture content (on a dry basis), M_e (kg kg⁻¹) is moisture content (dry basis) at equilibrium.

The DR was determined using following equation (Therdthai and Zhou, 2009):

$$DR = \frac{M_{t+dt} - M_t}{dt} \quad (3)$$

where DR is the drying rate (g water g dry matter⁻¹.min), Mt+dt is the moisture content at time t dt (g water g dry matter⁻¹), dt is the drying time (min).

2.2.2. Effective Moisture Diffusivity

Effective moisture diffusivity (D_{eff}) values were calculated using Fick's second law of diffusion. The general solution of the equation for slab geometry (spherical) is showed as follow (Crank, 1975; Çetin et al., 2022):

$$MR = \frac{M_t - M_e}{M_o - M_e} = \frac{8}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{(2n-1)^2} \exp\left(-\frac{(2n-1)^2 \pi^2 D_{eff} t}{4L^2}\right) \quad (4)$$

where D_{eff} is effective moisture diffusivity (m²s⁻¹), L is half thickness of slice (m), t is drying duration. For longer drying times, the first term of the above equation is used. The slope of the drying time plot against ln(MR) gives k₀ in the equation below (Çetin, 2022a).

$$\ln(MR) = \frac{8}{\pi^2} \exp\left(-\frac{\pi^2 D_{eff} t}{4L^2}\right) \quad (5)$$

$$k_0 = \frac{D_{eff} \pi^2}{4L^2} \quad (6)$$

2.2.3. Response Surface Methodology

The response surface method is a statistical approach used for the optimization (Kaur et al., 2009). Box-Behnken experimental design (Box and Behnken, 1960) was used with three repetitions, 3-factor and 3-level at the center point was used to develop prediction models based on a quadratic polynomial equation for the parameters examined (Karaman and Sağdıç, 2019). Box-Behnken design was given in Table 1. The regression coefficients were obtained using the Design-Expert® software (Design Expert, 2021).

Table 1. Box–Behnken design and variables

Variables	Unit	Symbol	Level 1 Low, (-1)	Level 2 Mid, (0)	Level 3 High, (1)
Temperature	°C	T	50	55	60
Thickness	mm	ST	5	7	9
Drying time	h	DT	8	9	10
Design					
No	Temperature		Thickness	Drying time	
1	50		5	9	
2	50		7	8	
3	50		7	10	
4	50		9	9	
5	55		9	10	
6	55		5	8	
7	55		7	9	
8	55		7	9	

9	55	7	9
10	55	9	8
11	55	5	10
12	60	7	8
13	60	7	10
14	60	5	9
15	60	9	9

2.2.4. Estimation of the Drying Characteristics

In the present study, three different algorithms were utilized. Results were evaluated with the R, MAE, RMSE, RRSE and RAE. Five features (drying time, moisture ratio, moisture content (d.b.), drying rate, effective moisture diffusivity) were selected as input for prediction. Moreover, two different validation methods were performed as k-fold cross validation and train-test split. In the k-fold cross validation, k value was selected as 10. In the train test split method, train and test values were selected as 66% and 34%, respectively. In the present study, random forest, multilayer perceptron, and k-nearest neighbor algorithms were used. MLP model structure are given in Fig 1.

In the MLP prediction, number of epochs, learning ratio, momentum coefficient, and activation function were selected as 1000, 0.3, 0.2, and sigmoid, respectively. The correlation-based feature selection was used (Çetin, 2022b). In the k-NN prediction Euclidean distance rule was performed and k value was 1.

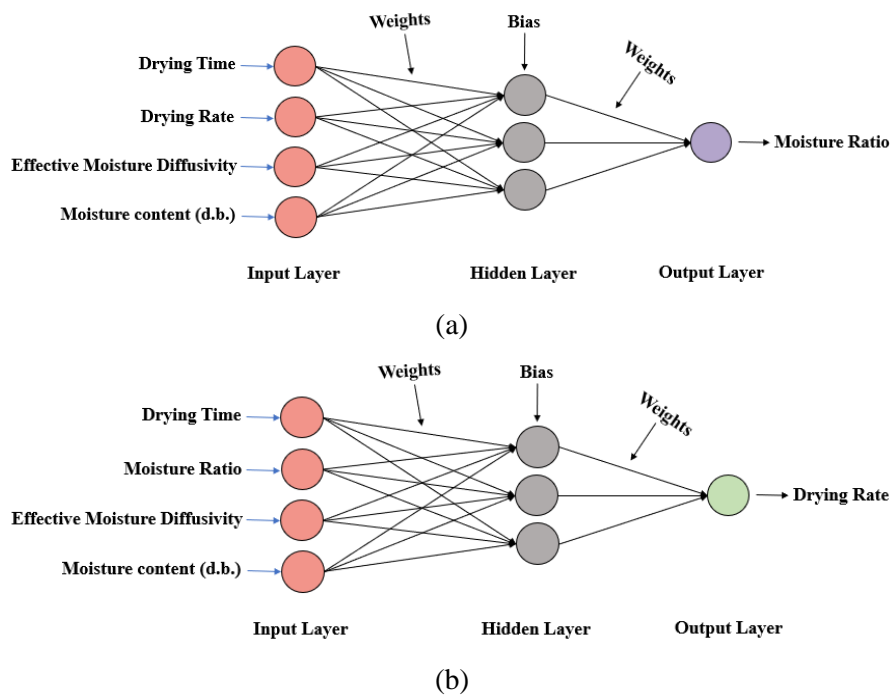


Figure 1. MLP structure of moisture ratio (a) and drying rate (b)

2.2.5. Model Performances

Model performance was evaluated by correlation coefficient (R), mean absolute error (MAE), relative absolute error (RAE) and root mean square error (RMSE), and root relative squared error (RRSE), Equations (8-10) (Parker, 2001).

$$R = \frac{1}{n-1} \sum_{i=1}^n \frac{(M_i - \bar{M})(E_i - \bar{E})}{S_M S_E} \quad (8)$$

$$MAE = \sum_{i=1}^n \frac{|E_i - M_i|}{n} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (E_i - M_i)^2}{n}} \quad (10)$$

$$RAE = \frac{\sum_{i=1}^n |E_i - M_i|}{\sum_{i=1}^n |M - M_i|} \times 100 \quad (11)$$

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (E_i - M_i)^2}{\sum_{i=1}^n (M - M_i)^2}} \times 100 \quad (12)$$

Where; E_i : Estimated target value, M_i : Measured target value, \bar{E} : mean of estimated target values, \bar{M} : mean of measured target values, S_E : Sum of estimated target values, S_M : Sum of measured target values and n : Number of data. R were analyzed to evaluate success of predictions in related with principles specified in Colton (1974). As the R values get closer to 1, the estimation success increases.

3. RESULTS AND DISCUSSIONS

3.1. Moisture ratio estimation

Performance results are given in Table 2 and 3. In the k-fold cross validation, the highest R (0.9950) and lowest RMSE (0.0007) were determined in k-NN algorithm for Granny Smith cultivar. The lowest MAE (0.0007), RRSE (9.66%), and RAE (2.99%) values were also found in k-NN algorithm. For Red Delicious, MLP had the greatest R and lowest RMSE with the values of 0.9901 and 0.0014, respectively. However, the lowest MAE (0.0003) was obtained from k-NN.

In the train-test split method, the highest R was found as 0.9999 in k-NN for Granny Smith. In addition, the lowest RRSE (2.30%), RAE (1.12%), RMSE (0.0002), and MAE (0.0001) were determined in the k-NN. For Red Delicious, MLP had the highest R and lowest RRSE as 0.9995 and 5.99%. This algorithm was followed by RF with the R value of 0.9979 and an RMSE value of 0.0007. The lowest R with the value of 0.9750 found in k-NN algorithm.

Table 2. Performance results of moisture ratio (k-fold cross validation)

<i>Granny Smith</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.8500	0.0024	0.0037	%47.58	%51.20
k-NN	0.9950	0.0001	0.0007	%2.99	%9.66
RF	0.8561	0.0022	0.0038	%44.00	%52.73
<i>Red Delicious</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9901	0.0010	0.0014	%12.58	%13.68
k-NN	0.9849	0.0003	0.0017	%4.19	%16.71
RF	0.9858	0.0010	0.0017	%12.72	%17.01

Table 3. Performance results of moisture ratio (train-test split)

<i>Granny Smith</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9229	0.0023	0.0032	%43.28	%39.23
k-NN	0.9999	0.0001	0.0002	%1.12	%2.30
RF	0.9061	0.0027	0.0045	%50.86	%55.73
<i>Red Delicious</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9995	0.0027	0.0033	%11.30	%5.99
k-NN	0.9750	0.0011	0.0024	%14.82	%24.95
RF	0.9979	0.0006	0.0007	%6.85	%7.10

Onwude et al. (2016) by using artificial neural network algorithms predicted moisture ratio of zucchini and the R^2 values were indicated between 0.9590 and 0.9920. Çetin (2022b), by using random forest (RF), multilayer perceptron (MLP), gaussian processes (GP), support vector regression (SVR), and k-nearest neighbor (k-NN) to estimate moisture ratio of orange. For different algorithms greatest R values were reported as 0.9944 and 0.9898, for Washington Navel and Valencia cultivars respectively. Sağlam and Çetin (2022) estimated moisture ratio by using machine learning algorithms. For Granny Smith Oregon Spur, and Golden Delicious cultivars, Random Forest was most successful algorithm with R value of 0.9841, 0.9873, and 0.9800 respectively.

3.2. Drying rate estimation

Model performance results are presented in Table 4 and 5. For Granny Smith, the greatest R (0.9999) was found in k-NN, and the lowest R (0.9672) was found in RF in k-fold cross validation method. The lowest MAE (0.0001), RMSE (0.0002), RAE (1.12%) and RRSE (2.41%) were found in k-NN.

For Red Delicious, the highest R values determined as 0.9886 and 0.9882 for MLP and k-NN, respectively. In addition, the lowest RAE, RRSE, and MAE were found in k-NN with the values of 4.26%, 15.02%, and 0.0001 respectively.

In the train-test method, the greatest R values were determined from k-NN (0.9999) and MLP (0.9995). The k-NN had the lowest MAE and RMSE as 0.0001 for Granny Smith cultivar. The highest R obtained from MLP and k-NN as 0.9879 and 0.9814, respectively. The lowest MAE (0.0002) and RAE (8.61%) values were obtained from k-NN algorithm.

Table 4. Performance results of drying rate (k-fold cross validation)

<i>Granny Smith</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9988	0.0015	0.0020	%7.53	%5.02
k-NN	0.9999	0.0001	0.0002	%1.12	%2.41
RF	0.9672	0.0056	0.0120	%28.61	%30.74
<i>Red Delicious</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9886	0.0003	0.0004	%13.11	%15.58
k-NN	0.9882	0.0001	0.0004	%4.26	%15.02
RF	0.9753	0.0005	0.0006	%20.23	%22.67

Table 5. Performance results of drying rate (train-test split)

<i>Granny Smith</i>					
Models	R	MAE	RMSE	RAE	RRSE
MLP	0.9995	0.0027	0.0033	%11.30	%5.99
k-NN	0.9999	0.0001	0.0001	%1.01	%1.22
RF	0.8722	0.0178	0.0373	%74.30	%68.58
<i>Red Delicious</i>					
Algorithms	R	MAE	RMSE	RAE	RRSE
MLP	0.9879	0.0004	0.0006	%17.11	%21.71
k-NN	0.9814	0.0002	0.0006	%8.61	%21.01
RF	0.9740	0.0006	0.0008	%25.19	%28.68

Menlik et al. (2010) estimated apple drying rate by using ANN and the authors were found the R² and RMSE values as 0.9990 and 0.0000008, respectively. Sağlam and Çetin (2022) in the air-convective drying used k-NN, GP, MLP, SVR, and RF to estimate apple drying rates. The greatest R value was found with the values of 0.9995 in RF algorithm for Golden Delicious cultivar. Çetin (2022b) for drying rate estimation used five different algorithms (RF, MLP, GP, SVR and k-NN). The author reported that the R values ranged from 0.8618 and 1.0000 for air-convective dried orange slices.

4. CONCLUSIONS

In this study, drying rate and moisture ratio were estimated by three machine learning algorithms. According to results developed algorithms were appropriate for estimations. The k-NN algorithm was the most successful algorithm for Granny Smith cultivar, while the MLP was the most successful algorithm for Red Delicious. Generally, RF algorithm performance was lower than the other models in estimation of drying rate and moisture content. The results of MLP and k-NN algorithms were close to each other for drying rate estimation. As a result, the findings show that algorithms could be successfully applied for the estimation of drying rate and moisture ratio. Machine learning method could be used to drying control systems for the optimization.

5. REFERENCES

- Box, G.E.P., Behnken, D.W. (1960). Some new three level designs for the study of quantitative variables. *Technometrics*, 2, 455-475.
- Cemeroğlu B. (2011). Meyve ve Sebze İşleme Teknolojisi [Fruit and Vegetable Processing Technology], 2.Cilt, Nobel Akademik Yayıncılık Eğitim Danışmanlık Tic. Lmt. Şti. Yayın No: 191 (In Turkish).
- Colton, T. (1974). Statistics in medicine (p. 179). Little Brown and Co.
- Crank, J. (1975). Mathematics of Diffusion, 2nd ed.; Oxford University Press: London, UK,; p. 414.
- Cruz, A. C., Guiné, R. P., Gonçalves, J. C. (2015). Drying kinetics and product quality for convective drying of apples (cvs. Golden Delicious and Granny Smith). *International Journal of Fruit Science*, 15(1), 54-78.
- Çetin, N., Ciftci, B., Kara, K., & Kaplan, M. (2022). Effects of gradually increasing drying temperatures on energy aspects, fatty acids, chemical composition, and in vitro ruminal fermentation of acorn. *Environmental Science and Pollution Research*, 1-17.
- Çetin, N., Sağlam, C. (2022). Rapid detection of total phenolics, antioxidant activity and ascorbic acid of dried apples by chemometric algorithms. *Food Bioscience*, 47, 101670.
- Çetin, N. (2022a). Comparative assessment of energy analysis, drying kinetics, and biochemical composition of tomato waste under different drying conditions. *Scientia Horticulturae*, 305, 111405.
- Çetin, N. (2022b). Prediction of moisture ratio and drying rate of orange slices using machine learning approaches. *Journal of Food Processing and Preservation*, e17011.
- Deng, L. Z., Mujumdar, A. S., Zhang, Q., Yang, X. H., Wang, J., Zheng, Z. A., Gao, Z.J., Xiao, H. W. (2019). Chemical and physical pretreatments of fruits and vegetables: Effects on drying characteristics and quality attributes—a comprehensive review. *Critical reviews in Food Science and Nutrition*, 59(9), 1408-1432.
- Design Expert 1305, 2021. Software for design of experiments Stat-Ease Inc Minneapolis USA.
- Food and Agriculture Organization (FAO) FAOSTAT.<http://www.fao.org/faostat/en/#data2022>.
- Karaman, K., Sagdic, O. (2019). Zygosaccharomyces bailii and Z rouxii induced ethanol formation in apple juice supplemented with different natural preservatives: A response surface methodology approach. *Journal of Microbiological Methods*, 163: 105659.
- Kaur, S., Sarkar, B. C., Sharma, H. K. Singh, C. (2009). Optimization of enzymatic hydrolysis pretreatment conditions for enhanced juice recovery from guava fruit using response surface methodology. *Food Bioprocess Technology*, 2, 96-100.
- Menlik, T., Özdemir, M. B., Kirmaci, V. (2010). Determination of freeze-drying behaviors of apples by artificial neural network. *Expert Systems with Applications*, 37(12), 7669-7677.
- Mollazade, K., Omid, M., Arefi, A. (2012). Comparing data mining classifiers for grading raisins based on visual features. *Computers and electronics in agriculture*, 84, 124-131.
- Naderinezhad, S., Etesami, N., Poormalek Najafabady, A., Ghasemi Falavarjani, M. (2016). Mathematical modeling of drying of potato slices in a forced convective dryer based on important parameters. *Food Science & Nutrition*, 4(1), 110-118.

- Omid, M., Khojastehnazhand, M., Tabatabaeefar, A. (2010). Estimating volume and mass of citrus fruits by image processing technique. *Journal of food Engineering*, 100(2), 315-321.
- Onwude, D. I., Hashim, N., Janius, R. B., Nawi, N., Abdan, K. (2016). Modelling the convective drying process of pumpkin (*Cucurbita moschata*) using an artificial neural network. *International Food Research Journal*, 23, s237.
- Parker, J. R. (2001). Rank and response combination from confusion matrix data. *Information Fusion*, 2(2), 113-120.
- Pinar, H., Çetin, N., Ciftci, B., Karaman, K., Kaplan, M. (2021). Biochemical composition, drying kinetics and chromatic parameters of red pepper as affected by cultivars and drying methods. *Journal of Food Composition and Analysis*, 102, 103976.
- Sağlam, C., Çetin, N. (2022) Machine learning algorithms to estimate drying characteristics of apples slices dried with different methods. *Journal of Food Processing and Preservation*, e16496.
- Soares, J. D. R., Pasqual, M., Lacerda, W. S., Silva, S. O., Donato, S. L. R. (2013). Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. *Scientia Horticulturae*, 155, 24-29.
- Higashi, H., ElMasry, G. M., Nakauchi, S. (2016). Sparse regression for selecting fluorescence wavelengths for accurate prediction of food properties. *Chemometrics and Intelligent Laboratory Systems*, 154, 29-37.
- Spengler, R. N. (2019). Origins of the apple: the role of megafaunal mutualism in the domestication of *Malus* and rosaceous trees. *Frontiers in plant science*, 10, 617.
- Therdthai, N., Zhou, W. (2009). Characterization of microwave vacuum drying and hot air drying of mint leaves (*Mentha cordifolia* Opiz ex Fresen). *Journal of Food Engineering*, 91(3), 482-489.
- Yagcioglu, A. (1999). Tarımsal ürünleri kurutma tekniği [Drying techniques of agricultural products]. Ege Üniversitesi, Ziraat Fakültesi, Yayın No: 536 (in Turkish).
- Zhang, H., Song, T. Q., Wang, K. L., Wang, G. X., Hu, H., Zeng, F. P. (2012). Prediction of crude protein content in rice grain with canopy spectral reflectance. *Plant, Soil and Environment*, 58(11), 514-520.
- Zou, H. Y., Wu, H. L., Fu, H. Y., Tang, L. J., Xu, L., Nie, J. F., Yu, R. Q. (2010). Variable-weighted least-squares support vector machine for multivariate spectral analysis. *Talanta*, 80(5), 1698-1701.