



Using an Adaptive Neuro-fuzzy Interface System (ANFIS) to Estimate Walnut Kernel Quality and Percentage from the Morphological Features of Leaves and Nuts

Mehdi Rezaei¹ · Abbas Rohani² · Shaneka S. Lawson³

Received: 2 October 2020 / Accepted: 26 May 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2022

Abstract

Walnut genetic improvement and orchard management would significantly benefit from accurate prediction of critical yield-related traits. In this study, an adaptive neuro-fuzzy interface system (ANFIS) was used to predict walnut kernel percentage and kernel quality. ANFIS uses principles of artificial neural network (ANN) learning as well as fuzzy principles. A total of 14 morphological characteristics of 100 walnut genotypes from Golestan province in Iran were used as model inputs. Correlation analysis and principal component analysis (PCA) were tested for their ability to reduce the model input numbers needed for accurate output. Eight features (four nut-related traits, four leaf characteristics) were revealed to be the most useful ANFIS input variables. Modeling data revealed ANFIS could predict walnut kernel percentage with a coefficient of determination (R^2) of 99%. Accuracy in detection of kernel quality was also 99%. These results indicated that a combination of the fuzzy c-means (FCM) method with the hybrid training algorithm were the most useful when designing the ANFIS model. Therefore, ANFIS is a highly recommended tool for modeling walnut yield traits.

Keywords ANFIS · Classification · Fuzzy inference system · *Juglans regia* · Neural network

Abbreviations

ANFIS	Adaptive neuro-fuzzy interface system
ANN	Artificial neural network
F1	Full leaf length
F2	Length of the largest leaflet
F3	Width of the largest leaflet
F4	Number of leaflets
F5	Leaf shape
F6	Leaf color
F7	Tree growing habit
F8	Nut shape
F9	Nut diameter

F10	Nut length
F11	Shell thickness
F12	Shell seal diameter
F13	Fruit width
F14	Fruit length
FIS	Fuzzy inference system
FCM	Fuzzy c-means
FG	Fuzzy grids
FN	Number of positive samples rejected incorrectly
FP	Number of negative samples identified incorrectly
GP	Grid partitioning
IR	Influence radius
$K_e P_e$	Kernel percentage
MF	Membership function
PCA	Principal component analysis
R^2	Coefficient of determination
RMSE	Root mean square error
SC	Subtractive clustering
SF	Squash factor
TN	Number of positive samples rejected correctly
TP	Number of positive samples identified correctly

✉ Abbas Rohani
arohani@um.ac.ir

¹ Department of Horticultural Sciences, Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran

² Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

³ Department of Forestry and Natural Resources, USDA Forest Service, Northern Research Station, Hardwood Tree Improvement and Regeneration Center (HTIRC), Purdue University, 715 West State Street, West Lafayette, USA

Introduction

Persian walnut (*Juglans regia* L.) is one of the most important nuts in the world. Domestication of walnut trees began in Iran and Afghanistan (Arzani et al. 2008; Khadivi-Khub and Ebrahimi 2015). Iran, with a production exceeding 0.41 million tons of dried walnut, is the third-largest producer of walnut after China and the United States (Faostat 2018). Traditional walnut seed orchards in Iran are considered a precious genetic resource. Longstanding orchards were developed from large non-grafted seedling trees to ensure genetic diversity, which is easily observable in the morphological traits of individual trees (Arzani et al. 2008). Walnut sexual reproduction has also contributed to the substantial genetic variations in tree growing habits and the qualitative and quantitative characteristics of nuts (Khadivi-Khub and Ebrahimi 2015; Shamlu et al. 2018; Khadivi et al. 2019). Today, with the development of new data analysis methods and the combination of biological data and mathematical modeling (i.e., machine learning), more information is known about nonlinear relationships between biological system traits and their interacting factors (Motta and Pappalardo 2013; Almudevar et al. 2020). Data obtained from walnut orchards, with their segregating populations, could be useful for evaluating new computational methods. These original approaches can provide a new viewpoint from which to survey walnut trait relationships. Computational models can be used universally by growers, the processing industry, and walnut tree breeders to aid future genetic diversity and management decisions.

Walnut breeding programs typically seek to confirm proper overlapping of flowering and fruit harvesting dates, increase nut yield, certify an easy separation of kernel and shell during nut-cracking, enhance kernel flavour quality, and improve resistance to important diseases such as anthracnose and blight. Nut-related quantitative traits are also important considerations for walnut breeders (Cosmulescu and Botu 2012). For example, consumers expect large, light amber-coloured kernels that are easily removed from the shell and are appetizing. An ideal walnut should weigh 12–18 g and have a kernel percentage greater than 50% (Germain et al. 1997; Cosmulescu and Botu 2012; Ebrahimi Aziz et al. 2015). Improvement of walnut kernel percentage is a primary walnut breeding goal (Khadivi et al. 2019). Kernel size directly influences walnut kernel percentages, with a positive correlation between kernel dimensions and kernel percentage being reported (Ebrahimi Aziz et al. 2015; Khadivi-Khub and Ebrahimi 2015). Walnut trees mature slowly and produce fruit later in life. Early prediction of nut and kernel properties based on vegetative features is invaluable for the selection of genotypes for classical breeding programs. Estimation of walnut kernel

percentage prior to cracking the nut can also be a significant achievement for use in the processing industry.

Machine learning techniques such as predictive modeling and neural networking have many applications in various aspects of biological systems (Almudevar et al. 2020). Several modern studies have employed computational software models to predict plant yield efficiency and genetic value based on vegetative properties observed in early growth stages (Azevedo et al. 2015; Rad et al. 2015; Chen et al. 2018; Mochida et al. 2019; Ali and Deo 2020). Most recently, scientists have introduced numerous artificial intelligence methods to model complex and nonlinear phenomena and analyze these processes (Almudevar et al. 2020). The adaptive neuro-fuzzy interface system (ANFIS) is one of the most important of these intelligence methods and is a neural network and fuzzy inference combination (Jang 1993). An ANFIS is a specific type of artificial neural network (ANN), based on the Takagi–Sugeno fuzzy inference system (FIS) (Jang 1993). The Takagi–Sugeno FIS model applies a rule base, a database, and a reasoning mechanism to the weighted averages of fuzzy inputs to generate crisp outputs. In other words, the primary network components, namely the premise and consequence parts, are identified with the ANFIS model. The successful use of this method was previously reported by Eski et al. (2018) for the prediction of almond (*Amygdalus communis* L.) nut physical properties. The ANFIS model was also used to sort walnut based on acoustic emissions analysis (Khalifa and Komarizadeh 2012) and kernel color (DemİR 2018).

Although walnut sorting was modeled by machine learning and evidenced in several studies, no reports were found regarding the use of this or other soft computing methods for walnut genetic improvement. In this study, we attempt to predict walnut kernel percentage from data collected on leaf characteristics and nut traits aided by an ANN, specifically ANFIS. The second objective is to evaluate the capability of ANFIS to classify the four defined classes of walnut kernel quality, thus allowing prediction of genotypes that produce desirable, consumer-preferred kernels from characteristics identifiable in early growth stages.

Material and Methods

Data Set

This study was conducted on 100 walnut genotypes from North-eastern Iran, a primary walnut production site within the Golestan province (37.2898°N, 55.1376°E). Trees were randomly selected from commercial orchards and individual wild-grown trees. The studied genotypes were numbered from 1–100 and labeled according to site and sample tree number. A total of 14 morphological traits were mea-

sured based on the International Plant Genetic Resources Institute (IPGRI) walnut descriptor for each genotype (McGranahan et al. 1994). Between 80 and 100 walnut fruits and 15–20 leaves were randomly selected from each tree for use in data analysis. Walnut fruits were harvested at the husk cracking stage and after removing the outer husk by hand, were dried at room temperature for 30 days before kernel and nut traits were measured.

Measured Traits

Leaf morphological traits and tree growing habits of the selected walnut genotypes were given designations between F1 and F7 while the seven nut-related features were given designations between F8 and F14. Full leaf length (F1), and length (F2) and width (F3) of the largest leaflet were measured in centimeters with a digital caliper. Number of leaflets (F4), along with leaf shape (F5), color (F6), and tree growing habit (F7) were measured and coded according to published IPGRI descriptions (McGranahan et al. 1994). Continuously variable characteristics such as nut shape (F8) and shell texture were recorded on a scale of 1–9, following IPGRI guidelines. Diameter (F9) and nut length (F10), shell thickness (F11), shell seal diameter (F12), fruit width (F13), and length (F14) were also measured in centimeters with a digital caliper. Nut and kernel weights were recorded with a digital scale with ± 0.01 g accuracy. Kernel fill percentages (K_eP_e) resulted from the average of 100 nuts per genotype. Moreover, genotypes were sorted into kernel quality grades (1–4) based on kernel characteristics (weight, color, ease of shell separation, kernel fill percentage) (Table 1).

Modeling and Classification Using an Adaptive Neuro-fuzzy Inference System (ANFIS)

An ANFIS is a multilayer feed-forward neural network with fuzzy parameters that was used to predict K_eP_e and kernel quality classifications based on data from Table 1. This network type is used as a confident option when relationships between system variables are very complicated. The ANFIS model is based on the first-order Sugeno inference system and results from the merger of an ANN with an FIS. An ANFIS model with two input variables consists of five computational layers (Fig. 1). These layers are defined as: 1. Fuzzification, 2. product, 3. strength normalization,

4. defuzzifier, and 5. summation. Fuzzification is the first ANFIS step in which membership function (MF) values are calculated and where the relationship between model inputs and outputs is established through a set of logical rules. Defuzzification is final calculation step in the ANFIS modeling process. In this study, three optimization methods were used to create the FIS, including grid partitioning (GP), subtractive clustering (SC), and fuzzy c-means clustering (FCM). These three methods are used to optimize fuzzy system parameters. GP is most often used for ANFIS design; however, when model input numbers are high, the number of rules increases exponentially. The number of rules equals the number of MFs to the power of input variable numbers. GP serves to divide input data sets into high-dimensional fuzzy grids (FG). Compared to GP, SC produces fewer fuzzy rules. The SC data set is divided into clusters. Fuzzy IF-THEN Statement rules are designed based on cluster centers. The value of the cluster radius influences the number of fuzzy IF-THEN rules. The FCM method calculates cluster centers based on minimizing ANFIS training phase error by the K-means method. Additional details on these methods can be found in (Jang 1993).

Fig. 2 provides the formulae required to calculate each ANFIS layer when assuming two input variables (x_1, x_2). Initially, input variables are converted to linguistic variables (A, B) using membership functions (μ). Output of the second layer is called firing strength, which is obtained through logical operators (AND/OR). Firing strength is the weight of each rule. The output of the third layer is called normalized firing strength. In this layer, normalization operations are performed. In the fourth layer, which are adaptive nodes, the total weight of IF-THEN rules (r_i) are calculated. Each rule can be stated as IF (x_1 is A AND x_2 is B), THEN ($f = p x_1 + q x_2 + c$). The symbols p , q , and c represent the following parameters. Finally, in the fifth layer, the sum of the output of the fourth layer is presented as the crisp output value or predicted K_eP_e . The nodes of the fifth layer are called fixed nodes. The output of each layer is displayed with O_L .

Optimal values of MF and consequent parameters for ANFIS model experiences during the training phase are recorded as a training data set and training algorithms. In this study, two backpropagation and hybrid algorithms were used in training the ANFIS model. In the backpropagation

Table 1 Walnut kernel quality characteristics using weight, color, ease of shell separation, and fill percentage

Kernel quality (grade)	Kernel weight (g)	Kernel color (shade)	Shell separation (easy to difficult)	Kernel fill (%)
1	>5.5	Extra light, light	Very easy to easy	>50
2	4.5–6	Extra light, light, light amber	Easy to difficult	45–55
3	3.5–5.5	Light, light amber	Easy to difficult	40–50
4	<4.5	Light amber, amber	Easy to difficult	<45

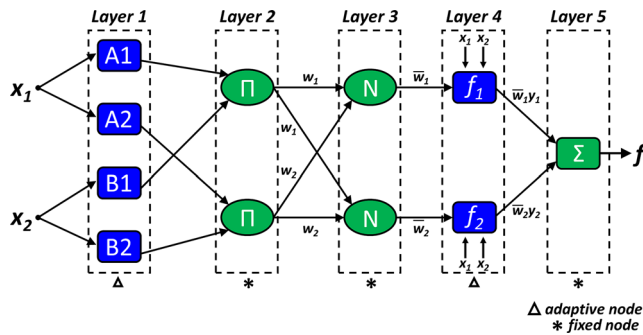


Fig. 1 Adaptive neuro-fuzzy interface system structure

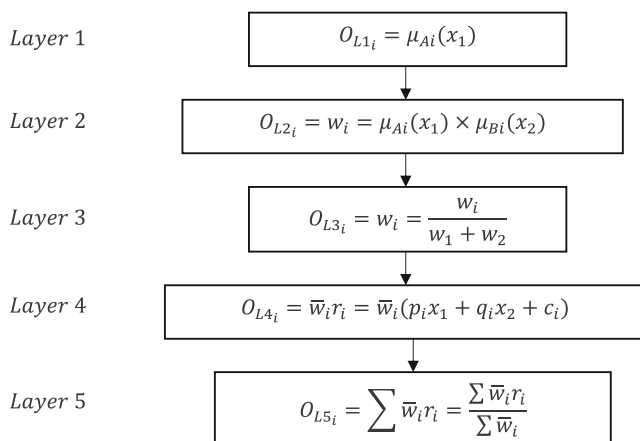


Fig. 2 Formulae indicating the computations required for each adaptive neuro-fuzzy interface system layer

algorithm, output errors are calculated for each layer and used to update layer parameters. The slowness of the training process and the potential for becoming mired in local minima are detractors of this training method. The hybrid training algorithm, named for the use of two gradient descent and least-squares optimization methods, does not have these two problems.

Evaluation of Model Capability

The ANFIS model capability was evaluated by root means square error (RMSE) and coefficient of determination (R^2) with a regression line between the real ($KePe_a$) and predicted ($KePe_p$) data sets. The closer R^2 is to 1 and RMSE to zero, the greater likelihood the model will accurately predict and model $KePe$. The objective is to classify walnut kernel quality into four grading classes based on data from Table 1 using ANFIS. The accuracy criteria were used to evaluate this classification (Eq. 3) and optimal results were achieved at 100% accuracy.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (KePe_{ai} - KePe_{pi})^2}{n}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (KePe_{ai} - KePe_{pi})^2}{\sum_{i=1}^n (KePe_{ai} - \overline{KePe_a})^2} \quad (2)$$

$$Accuracy (\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (3)$$

where TP and TN are the number of positive and negative samples correctly identified and correctly rejected, respectively. FP and FN are the number of negative and positive samples incorrectly identified and incorrectly rejected.

The existing data set (all 100 genotypes) was randomly divided into two training data sets, one 80 genotype data and testing set and one 20 genotype data set. The training data set was used to adjust and optimize ANFIS model parameters. Testing data sets were used to assess the model's functional capability in new situations. Model generalizability was also assessed by testing data sets of multiple sizes. MATLAB R2020a software (MathWorks, Inc., Nat-

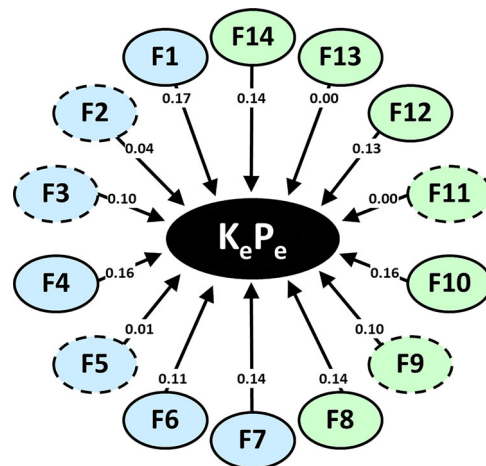


Fig. 3 Correlation values between walnut kernel fill percentage ($KePe$) and vegetative growth (blue boxes) and nut features (green boxes). Dotted outlines emphasize features with low correlation coefficients

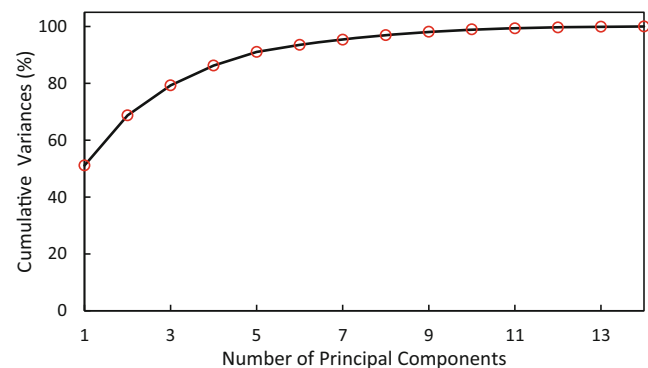


Fig. 4 Principal component analysis requires fewer nut and vegetative growth characteristics to explain input variance

Table 2 Model prediction error values (root mean square error) for fuzzy inference system (FIS) construction methods used in K_eP_e estimation

FIS generation methods	Optimization method	PCA	Train	Test	Total
Grid partition (GP)	Backpropagation	No	3.15	4.88	3.89
		Yes	5.14	6.47	5.55
	Hybrid	No	2.05	5.16	3.14
		Yes	5.56	6.76	5.23
Subtractive clustering (SC)	Backpropagation	No	0.3	0.11	0.27
		Yes	4.87	4.79	4.85
	Hybrid	No	0.01	7.51	3.36
		Yes	4.51	8.31	5.49
FCM clustering (FCM)	Backpropagation	No	4.71	4.98	4.76
		Yes	5.3	5.3	5.3
	Hybrid	No	1.66	12.25	8.44
		Yes	4.91	6.54	5.27

PCA principal component analysis

ick, MA, USA) was used to run the ANFIS model of this study.

Results and Discussion

Modeling Kernel Percentage

Feature Selection and Reduction

Associations between walnut kernel fill percentage (K_eP_e) and leaf, nut, and fruit characteristics were determined (Fig. 3). Correlation to K_eP_e resulted in values <0.1 for three growth characteristics (F2, F3, F5) and three nut and fruit traits (F9, F11, F13). Previous studies of walnut morphological diversity correlating kernel fill percentage to leaf shape and leaf length and width also noted a lack of significant correlation (Ebrahimi Aziz et al. 2015; Shamlu et al. 2018). Similar to our results, Khadivi et al. (2019) reported a correlation coefficient of $r=0.04$ between kernel fill percentage and fruit diameter in walnut. A significant negative correlation has been described between kernel fill percentage and fruit diameter (Khadivi-Khub and Ebrahimi 2015). Additionally, kernel fill percentage has no significant correlation with fruit weight and size in walnut (Ebrahimi Aziz et al. 2015), although a significant positive correlation exists between kernel fill percentage and kernel weight (Sharma and Sharma 2001; Cosmulescu and Botu 2012; Ebrahimi Aziz et al. 2015; Khadivi-Khub and Ebrahimi 2015). Therefore, for modeling purposes, removing these six features reduced the total input number for the ANFIS model from 14 features to eight (Fig. 3).

As evidenced in Fig. 4, principal component analysis (PCA) reduced model input numbers from 14 features to nine principal components and could explain 98% of the total variance of the original inputs. Decreasing the number

of model inputs can reduce the required computing volume and construct a simpler model.

This study utilized three distinct methods (grid partition, GP; subtractive clustering, SC; FCM clustering, FCM) to create the FIS and MF parameters, and rules were adjusted using backpropagation and hybrid training. Application of a variety of ANFIS methods with RMSE criteria to the training, testing, and total stages highlighted marked differences (Table 2). Features derived from PCA failed to reduce RMSE; thus, despite the ability to simplify the model and reduce required inputs, it is necessary to use all 14 features

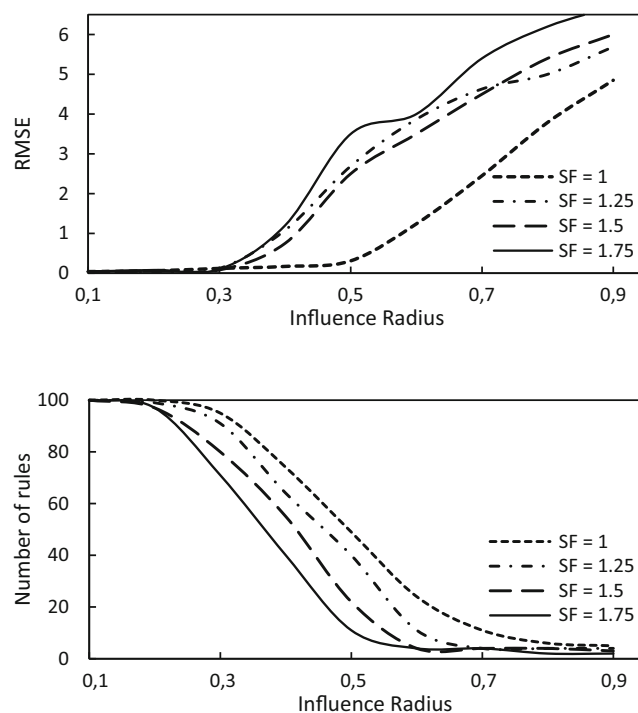


Fig. 5 Influence radius and squash factor (SF) effects on adaptive neuro-fuzzy interface system prediction error and number of rules in K_eP_e estimation

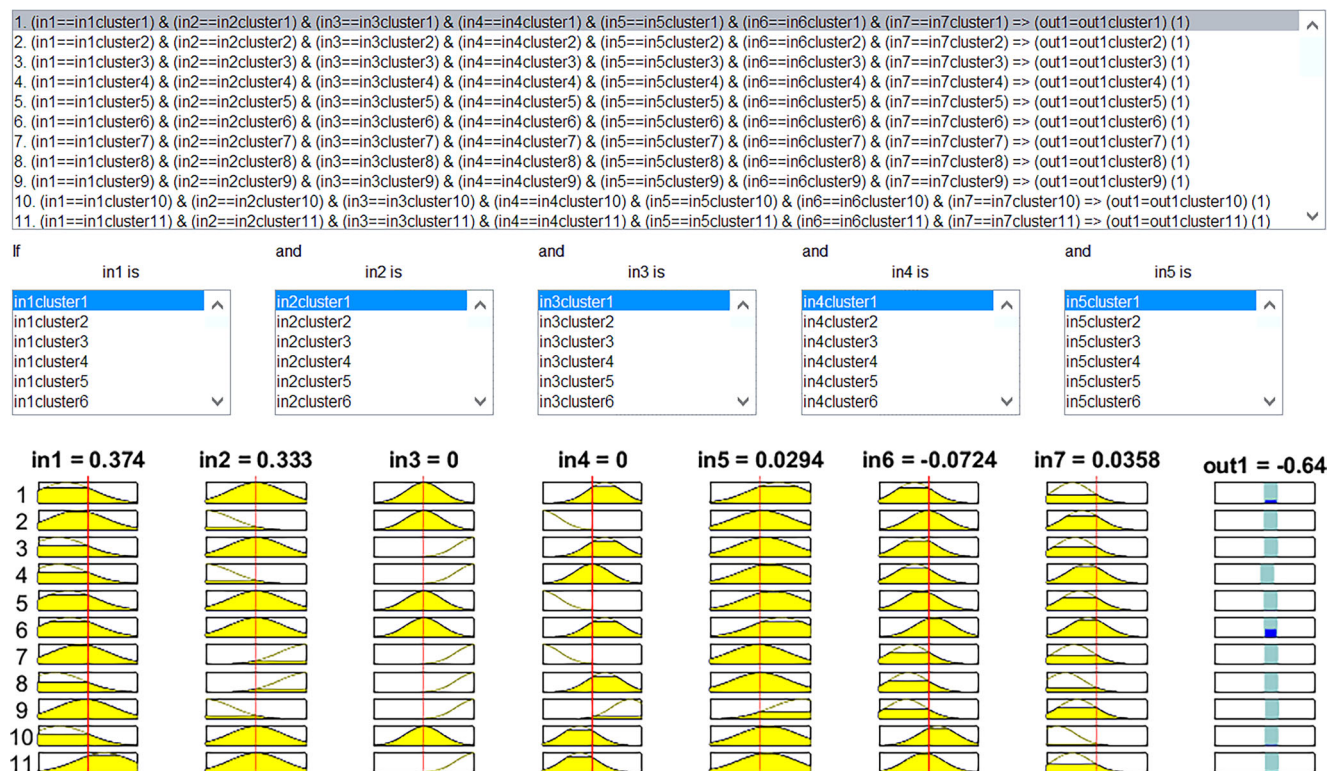


Fig. 6 Table of rules and working steps for the fuzzy inference system of our adaptive neuro-fuzzy interface system model

to generate the least amount of error. The lowest error rates were obtained when the FIS was generated with the SC method and optimized using backpropagation. The resultant RMSE value for $K_e P_e$ estimation was 0.27. Therefore, with a mean value for $K_e P_e$ of 45.66, a RMSE of 0.27 is acceptable.

Use of the ANFIS model necessitates finding optimal values for two critical parameters, influence radius (IR) and squash factor (SF). An increased IR leads to increased RMSE, while increased SF exacerbates the issue (Fig. 5). Lowering IR to 0.1 decreased RMSE but led to an increase in the number of rules to 100, increased computations, and altered graphed output from a smooth to a rough response curve. A combination where IR = 0.5 and SF = 1 with 11 rules exhibited the best fit with the least RMSE (Fig. 5). The working steps of the FIS system along with the MF of the rules and their relationship to each other are presented

to show the effect on output (Fig. 6). The settings used in the design of the ANFIS model, and the rules designed between input variables and their classification position are based on Gaussian membership function (Table 3).

Our results have shown that the ANFIS model can predict the walnut kernel fill percentage with an error rate of less than 1%. However, the amount of data used in the training phase can be influenced by model training performance and testing stage error as a generalization capability. Our study reported variation in RMSE with changes in training data set size from 50–90%, but the amount was minimal (Table 4). Thus, the ANFIS model is a reliable predictor of walnut kernel fill percentage based on vegetative growth characteristics and nut features as inputs.

Regression equations for actual and predicted ($y = a + bx$, R^2 , [$R^2 = 0.99$]) $K_e P_e$ values at the 50% training data size indicated a slope and y-intercept close to one and

Table 3 Adaptive neuro-fuzzy interface system model features and parameters

Feature	Descriptor
AND method	Prod
OR method	Probor
Implication method	Prod
Aggregation method	Sum
Defuzzification method	Wtaver
Rules between input variables	45

Table 4 Adaptive neuro-fuzzy interface system root mean square error response to deviations in training size during the train, test, and total phases

Training sizes	Train	Test	Total
90/10	0.28	0.09	0.27
80/20	0.3	0.11	0.27
70/30	0.35	0.15	0.21
60/40	0.6	0.37	0.52
50/50	0.67	0.38	0.55

Fig. 7 Comparison of actual and predicted K_eP_e values for the 50% training set size

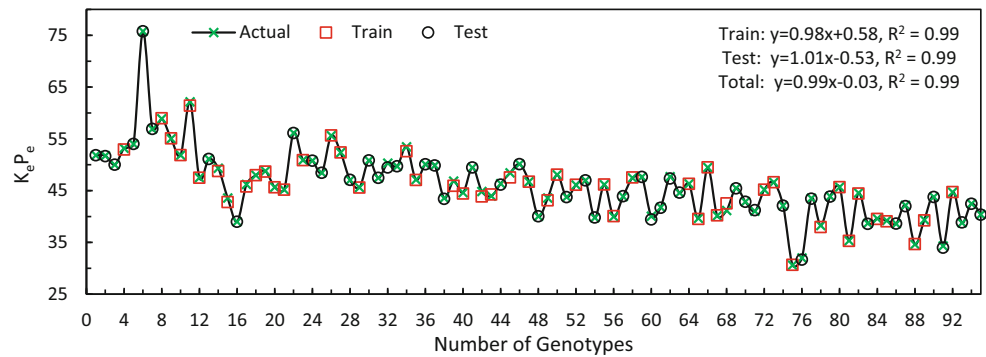
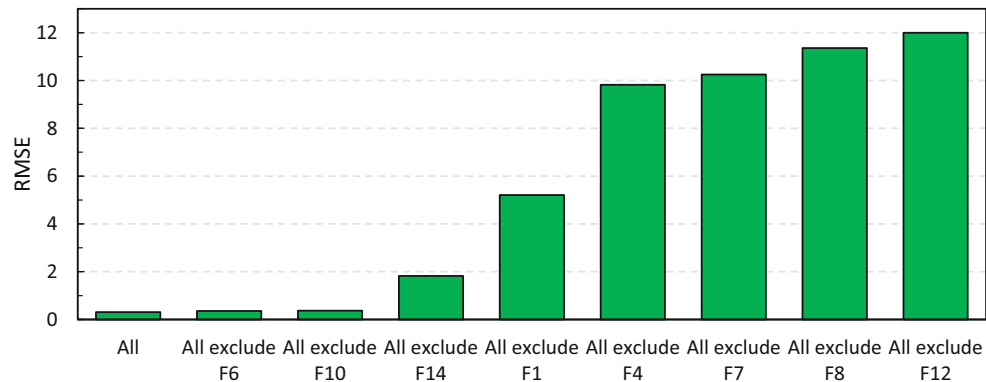


Fig. 8 Effectiveness of K_eP_e estimation by the adaptive neuro-fuzzy interface system model after exclusion of selected features



zero, respectively (Fig. 7). The ANFIS model facilitated the highly accurate prediction of K_eP_e values in the test stage using only 50% of the total training stage dataset.

A sensitivity analysis was used to measure the effects of each of the eight features with the higher correlation values and revealed that removal of any one of these features from the dataset increased prediction error. The impact of each trait on RMSE revealed that F12 (shell seal diameter) had the greatest impact, while F6 (leaf color) was least responsible for error (Fig. 8). Ordering traits from lowest to highest showed leaf color, nut length, fruit length, full leaf length, number of leaflets, tree growing habit, nut shape,

and shell seal diameter (F6, F10, F14, F1, F4, F7, F8, and F12, respectively) showed RMSE of the final four traits to be above 9.8 and significantly different from the other traits in the study (Fig. 8). The substantial difference in RMSE highlights the considerable contributions of these traits in predicting the walnut kernel fill percentage. The highest RMSE, 12, was obtained from the exclusion of shell seal diameter, while an RMSE of 0.36 was determined from the exclusion of leaf color.

Table 5 Model prediction error values (root mean square error) for fuzzy inference system (FIS) construction methods used in kernel quality classification

Generate FIS	Optimization method	PCA	Train	Test	Total
Grid partition (GP)	Backpropagation	No	56.25	54.16	55.24
		Yes	45.91	42.88	44.75
	Hybrid	No	49.11	48.24	48.85
		Yes	47.25	43.14	46.66
Subtractive clustering (SC)	Backpropagation	No	95.1	89.5	94
		Yes	89.95	88.91	89
	Hybrid	No	100	36.74	88.02
		Yes	100	42.1	89
FCM clustering (FCM)	Backpropagation	No	40.7	52.61	43.05
		Yes	40.36	31.62	39
	Hybrid	No	90.5	10.1	75
		Yes	100	31.6	87

PCA principal component analysis

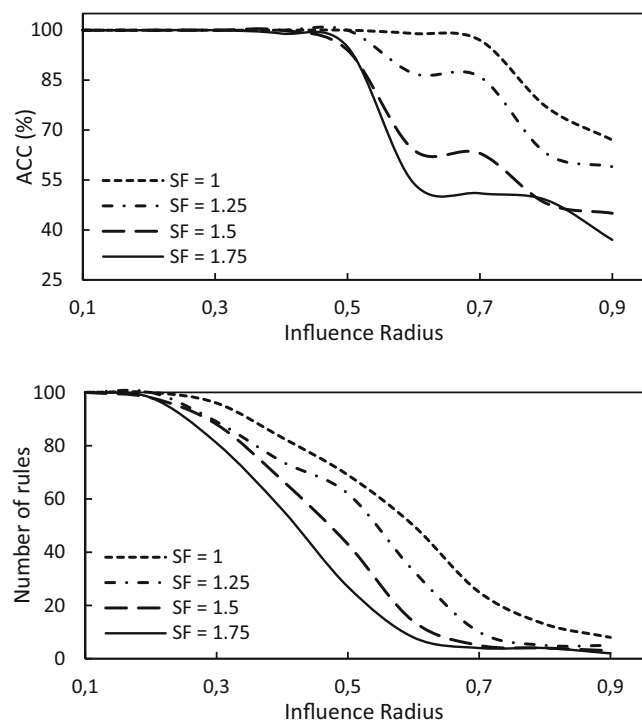


Fig. 9 Influence radius and squash factor (*SF*) effects on adaptive neuro-fuzzy interface system prediction error and number of rules in kernel quality classification

Classification

Walnut kernel quality was determined using a classifier. Similar to K_eP_e modeling, classifier prediction accuracy is dependent on the FIS construction method, the model training algorithm, and the nature of features. Consumers and the kernel packaging industry desire large, brightly colored kernels that are appetizing and separate easily from the nut shell. The ANFIS classifier was used to predict walnut kernel quality and divide them into groups designated 1–4. Greater classification accuracy was reported for the training (95%) and testing (89%) stages using the SC method for FIS construction with a backpropagation training algorithm and eight selected features as ANFIS classifier inputs (Ta-

ble 5). A similar improvement in accuracy was discovered when modeling K_eP_e using these methods.

ANFIS classification accuracy is contingent upon IR and SF (See Fig. 7 results). A total of 14 walnut genotype traits were evaluated, and eight features, four vegetative growth traits, and four nut related traits were used as model variables similar to those observed during K_eP_e modeling.

The parameters producing the least RMSE and lowest number of rules was the combination of IR = 0.6 and SF = 1 with 50 rules (Fig. 9). Kernel quality classification accuracy was 99% based on eight selected traits. Therefore, the ANFIS model can discern kernel quality based on the changes in inputs.

Eight selected features were examined using a sensitivity analysis to evaluate how the exclusion of each influenced the accuracy of the ANFIS model (Fig. 10). If the exclusion of a feature decreases the accuracy of the model classifier, then that feature is more significant than the combination of the other traits. Presenting the traits in order from those deemed least to most important for accuracy of the ANFIS model, it was established that nut shape (F8) was the primary influence on model accuracy. The remaining traits, tree growing habit, number of leaflets, nut length, shell seal diameter, full leaf length, fruit length, and leaf color (F7, F4, F10, F12, F1, F14, and F6, respectively) influenced accuracy to a continuously lesser degree with leaf color (F6) again having the least influence on the model.

Comparison of sensitivity analyses for K_eP_e modeling and kernel quality classification indicated selected traits affected the model differently depending on ANFIS model purpose. While shell seal diameter (F12) was the most important for walnut kernel fill percentage, nut shape (F8) was the most important in kernel grading. As such, the RMSE associated with the prediction of kernel fill percentage increased from 0.31–12 after exclusion of shell seal diameter (F12), while the exclusion of nut shape (F8) reduced classifier accuracy > 50% (Fig. 10).

Research studies of walnut morphological diversity have shown nut shape to be a highly diverse trait (Arzani et al. 2008; Shamlu et al. 2018; Khadivi et al. 2019). Ebrahimi A. et al. (2009). Ebrahimi et al. (2009), in a study on the

Fig. 10 Effectiveness of kernel color quality classification by the adaptive neuro-fuzzy interface system model after exclusion of selected features

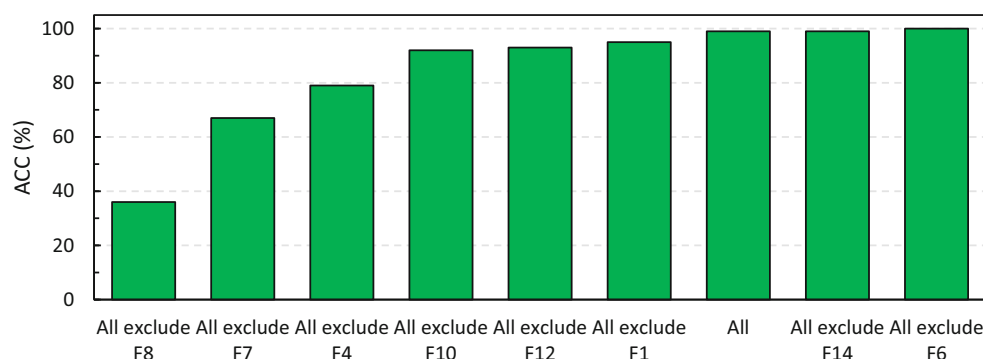


Table 6 Assessment of accuracy of kernel quality classification to adaptive neuro-fuzzy interface system model deviations in training size during the train, test, and total phases

Training sizes	Train	Test	Total
90/10	98.89	100.00	99.00
80/20	98.70	100.00	99.00
70/30	93.33	95.00	94.00
60/40	92.86	93.33	93.00
50/50	82.40	93.90	88.00

physical characteristics of walnut genotypes reported that a coefficient of variance in the nut shape was > 51%. Most walnuts produced in Iran are from orchards with seedling trees, where each tree has a particular nut shape. This allows walnut shape to be a distinctive physical trait that the packaging and grading industry can use for screening.

To further examine the ability of ANFIS to be used as an accurate classifier for walnut kernel grading based on growth and nut characteristics, the size of the model training set was reduced from 90 to 50% of the total patterns. Results indicated accuracy in kernel quality classification decreased as training data set size decreased. Thus, these data show that ANFIS has less generalizability in the kernel quality classification than K_cP_c modeling (see Table 6).

Conclusion

This study sought to present an accurate prediction model for walnut kernel fill percentage and classification of kernel quality using ANN principles and fuzzy system benefits acquired by using ANFIS. Seven growth- and seven nut-related traits were selected as independent variables for the ANFIS model. Correlation analyses confirmed that a fuzzy neural relationship existed between eight of the 14 features (F1, F4, F6, F7, F8, F10, F12, and F14) and that they could reliably be used in the ANFIS model to predict kernel fill percentage and kernel quality. PCA was used in an attempt to decrease the required number of input variables; however, model accuracy decreased, and the method was used only for comparison. Use of the ANFIS model to estimate kernel fill percentage resulted in 0.27% RMSE with a 99% correlation between actual and predicted data. The percent accuracy of the model in kernel classification was also 99%. Both uses for the ANFIS model were promising; however, the generalizability of ANFIS in modeling was superior to classification. A global model of ANFIS structure, or any model based on soft computing, can be developed with the addition of more complete and universal data sets. At that point, proposed models can be applied universally for walnut genetic improvement or other walnut orchard management issues.

Conflict of interest M. Rezaei, A. Rohani, and S.S. Lawson declare that they have no competing interests.

References

- Ali M, Deo RC (2020) Modeling wheat yield with data-intelligent algorithms: artificial neural network versus genetic programming and minimax probability machine regression. In: Handbook of Probabilistic models. Elsevier, Amsterdam, pp 37–87
- Almudevar A, Oakes D, Hall J (2020) Statistical modeling for biological systems: in memory of Andrei Yakovlev. Springer Nature, Heidelberg, Berlin, New York
- Arzani K, Mansouri-Ardakan H, Vezvaei A, Roozban MR (2008) Morphological variation among Persian walnut (*Juglans regia*) genotypes from central Iran. *N Z J Crop Hortic Sci* 36(3):159–168
- Azevedo AM, de Andrade Júnior VC, Pedrosa CE, de Oliveira CM, Dornas MFS, Cruz CD, Valadares NR (2015) Application of artificial neural networks in indirect selection: a case study on the breeding of lettuce. *Bragantia*. <https://doi.org/10.1590/1678-4499.0088>
- Chen D, Shi R, Pape JM, Neumann K, Arend D, Graner A, Chen M, Klukas C (2018) Predicting plant biomass accumulation from image-derived parameters. *GigaSci* 7(2):giy1
- Cosmulescu S, Botu M (2012) Walnut biodiversity in south-western Romania resource for perspective cultivars. *Pak J Bot* 44(1):307–311
- Demir B (2018) Application of data mining and adaptive neuro-fuzzy structure to predict color parameters of walnuts (*Juglans regia* L.). *Türk J Agric For* 42(3):216–225
- Ebrahimi A, Zarei A, Fatahi R, Varnamkhasti MG (2009) Study on some morphological and physical attributes of walnut used in mass models. *Sci Hortic* 121(4):490–494
- Ebrahimi A, Khadivi-Khub A, Nosrati Z, Karimi R (2015) Identification of superior walnut (*Juglans regia*) genotypes with late leafing and high kernel quality in Iran. *Sci Hortic* 193:195–201
- Eski İ, Demir B, Gürbüz F, Kuş ZA, Yılmaz KU, Uzun M, Ercişli S (2018) Design of neural network predictor for the physical properties of almond nuts. *Erwerbs-Obstbau* 60(2):153–160
- Faostat (2018) FAO web page. <http://www.fao.org/faostat>. Accessed 22 July 2018
- Germain E, Charlot G, Prunet JP (1997) Fernor and Fernet, two new walnut cultivars registered in France. *Riv Frutticolt Ortofrutticolt* 59:86–87
- Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst, Man, Cybern* 23(3):665–685
- Khadivi A, Montazeran A, Rezaei M, Ebrahimi A (2019) The pomological characterization of walnut (*Juglans regia* L.) to select the superior genotypes—an opportunity for genetic improvement. *Sci Hortic* 248:29–33
- Khadivi-Khub A, Ebrahimi A (2015) The variability in walnut (*Juglans regia* L.) germplasm from different regions in Iran. *Acta Physiol Plantarum* 37(3):57
- Khalifa S, Komarizadeh MH (2012) An intelligent approach based on adaptive neuro-fuzzy inference systems (ANFIS) for walnut sorting. *Aust J Crop Sci* 6(2):183–187
- McGranahan GH, Germain E, Ramos DE, Riggert K (1994) Descriptor list for walnut (*Juglans* spp.). International Plant Genetic Resources Institute, Rome
- Mochida K, Koda S, Inoue K, Hirayama T, Tanaka S, Nishii R, Melgani F (2019) Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective. *GigaSci* 8(1):giy153
- Motta S, Pappalardo F (2013) Mathematical modeling of biological systems. *Brief Bioinform* 14(4):411–422
- Rad MRN, Fanaei HR, Rad MRP (2015) Application of artificial neural networks to predict the final fruit weight and random forest to

- select important variables in native population of melon (*Cucumis melo* L.). *Sci Hortic* 181:108–112
- Shamlu F, Rezaei M, Lawson S, Ebrahimi A, Biabani A, Khan-Ahmadi A (2018) Genetic diversity of superior Persian walnut genotypes in Azadshahr, Iran. *Physiol Mol Biol Plants* 24(5):939–949
- Sharma OC, Sharma SD (2001) Genetic divergence in seedling trees of Persian walnut (*Juglans regia* L.) for various metric nut and kernel characters in Himachal Pradesh. *Sci Hortic* 88(2):163–171