DOI: http://dx.doi.org/10.22092/cbj.2012.100427

Application of unsupervised weighting algorithms for identifying important attributes and factors contributing to grain and biological yields of wheat

E. Bijanzadeh*, Y. Emam, E. Ebrahimi, and M. Ebrahimi

Faculty of Agriculture, Shiraz University, Shiraz, Iran.
* Corresponding author's e-mail address: ebijanzadeh@gmail.com

Received: September 2012
Accepted: December 2012

ABSTRACT

Bijanzadeh, E., Y. Emam, E. Ebrahimi, and M. Ebrahimi. 2012. Application of unsupervised weighting algorithms for identifying important attributes and factors contributing to grain and biological yields of wheat. **Crop Breeding Journal 2(2): 111-117.**

To identify important attributes/factors that contribute to grain and biological yields of wheat, 9912 sets of diverse data from field studies were extracted, and supervised attribute-weighting models were employed. Results showed that when biological yield was the output, grain yield, nitrogen applied, rainfall, irrigation regime, and organic content were the most important factors/attributes, highlighted by 9, 7, 5, 3 and 3 weighting models, respectively. In contrast, when grain yield was the output, biological yield, location, and genotype were identified by 8, 6, and 5 weighting models, respectively. Also, five other features (cropping system, organic content, 1000-grain weight, spike number m-2 and soil texture) were selected by three models as the most important factors/attributes. Field water status, such as the irrigation regime or the amount of rainfall, was another important factor related to the biological or grain yield of wheat (weight ≥ 0.5). Our results showed that attribute/factor classification by unsupervised attribute-weighting models can provide a comprehensive view of the important distinguishing attributes/factors that contribute to wheat grain or biological yield. This is the first report on identifying the most important factors/attributes contributing to wheat grain and biological yields-using attribute-weighting algorithms. This study opened a new horizon in wheat production using data mining.

Keywords: attribute weighting, data mining, unsupervised model, wheat

INTRODUCTION

Data mining using various methodologies has been developed at both commercial and research centers (Ebrahimi and Ebrahimi, 2009; Ebrahimi *et al.,* 2008). Recently, agricultural and biological research studies have used various data mining techniques for analyzing large data sets and establishing useful classification patterns in data sets (Bijanzadeh *et al.,* 2010). However, considering its novelty and diversity, data mining methods are expected to produce even more fruitful results (Hsiao *et al.,* 2006).

Not all attributes in a data set are important; some are redundant or irrelevant. Data sets that include irrelevant attributes can misguide the clustering results and make them hard to explain (Liu and Motoda, 2008). In data mining, screening, clustering and decision tree algorithms, factor selection and attribute-weighting algorithms are useful for identifying irrelevant attributes to be excluded from the data set (Lakizadeh *et al.,* 2011; Bijanzadeh *et al.,* 2010; Ebrahimi and Ebrahimi, 2010; Ebrahimi *et al.,* 2009, Liu and Motoda, 2001).

Until now, researchers have only considered a restricted number of characteristics under field conditions that contribute to crop yield and yield components. It has now become obvious that analyzing a large number of factors under different field conditions can provide a comprehensive overview of important features responsible for yield improvement (Shekofa *et al.,* 2011).

Bijanzadeh *et al.* (2010) reported that based on a supervised factor selection model, the type of cultivation affected wheat grain yield, whereas soil pH had a marginal effect on wheat grain yield. They also demonstrated that factor classification using factor selection algorithms may be a suitable option for determining the important factors contributing to wheat grain yield, and for providing a comprehensive view of different traits.

Recently, there has been great interest in employing various attribute-weighting algorithms to identify the critical factors involved in different phenomena (Lakizadeh *et al.,* 2011). Attribute-weighting (or factor-selection) models reduce the size of attributes, creating a more manageable set of attributes for modelling. The main idea of attribute-weighting is to choose a subset of input variables by eliminating attributes with little or no predictive information (Ashrafi *et al.,* 2011).

Given that wheat (*Triticum aestivum* L.) is the most important food crop in the world (Emam, 2007), in this study various weighting algorithms were employed to determine the most important attributes contributing to the biological yield and grain yield of wheat. It is expected that an

Table 1. Articles in the literature that were used for different attribute weighting models of wheat grain yield in Iran.

Authors	Province-Location	Type of treatment
Abhari *et al*., 2008	Golestan- Gorgan	Drought stress
Afiuni *et al*., 2006	Isfahan-Rod Dasht	Different genotypes, Salt stress
Akbari *et al*., 2006	Khorasan-e-Razavi-Mashhad	Different genotypes
Dastfal *et al*., 2008	Fars-Darab	Different genotypes, Drought stress
Emam *et al*., 2000	Fars- Badjgah	Different nitrogen levels
Emam *et al*., 2009	Fars- Badjgah	Different nitrogen levels, Different genotypes
Emam *et al*., 2007	Fars- Badjgah	Different genotypes, Drought stress
Farahani and Arzani, 2007	Isfahan- Najaf Abad	Different genotypes
Faraji *et al*., 2006	Khozestan- Ramin	Different nitrogen levels, Drought stress
Ghodsi *et al*., 2005	Khorasan-e-Razavi-Mashhad	Drought stress
Kiani *et al*., 2004	Golestan-Agh Ghala	Salt stress, Drought stress
Modhej *et al*., 2008	Khozestan-Ahvaz	Different nitrogen levels
Momtazi and Emam, 2006	Fars- Badjgah	Length of growing season, Plant density
Moussavi Nik *et al*., 2006	Sistan and Balochestan-Zabol	Different phosphorus and zinc levels
Roustaii *et al*., 2003	Western Azarbaijan-Maragheh	Different genotypes
Sadegh Zadeh Ahari *et al*., 2006	Kohkiloyeh -Gachsaran	Different genotypes
Sadegh Zadeh Ahari *et al*., 2005	Western Azarbaijan-Maragheh	Different genotypes

Table 2. Identifying the most important attributes (weights ≥ 0.5) for determining wheat biological yield by different weighting algorithms (values closer to 1 show greater effectiveness of the attribute in determining biological yield).

Weighting algorithm	Attribute	Weight
Information gain	Grain yield	1.0
	Nitrogen applied	0.5
	Rainfall	0.5
Information gain ratio	Grain yield	1.0
	Organic content	0.5
	Plant height	0.5
	Irrigation regime	0.5
	Nitrogen applied	0.5
	Rainfall	0.5
	1000-kernel weight	0.5
Rule	Nitrogen applied	1.0
	Grain yield	0.7
	Length of growing season	0.6
	Plant height	0.5
	Irrigation regime	0.5
	Grain number per spike	0.5
Deviation	Grain yield	1.0
Chi-squared statistic	Grain yield	1.0
	Nitrogen applied	0.6
	Organic content	0.5
	Phosphorus applied	0.5
Gini Index	Grain yield	1.0
	Nitrogen applied	0.6
	Rainfall	0.5
Uncertainty	Grain yield	1.0
	Phosphorus applied	0.6
	Nitrogen applied	0.6
	Organic content	0.5
Relief	Length of growing season	1.0
	Soil pH	0.9
	Harvest index	0.8
	Irrigation regime	0.8
	Plant density	0.8
	Nitrogen applied	0.7
	Rainfall	0.6
Support vector machine (SVM)	Grain yield	1.0
	1000-kernel weight	0.6
	Rainfall	0.5
Principal component analysis (PCA)	Grain yield	1.0

intelligent agricultural information system will be built to assist experts in the field increase wheat grain yield and production.

MATERIALS AND METHODS

Data collection

Data from field studies describing the effects of field conditions (attributes or factors) on wheat grain or biological yields in Iran were extracted from the literature (Table 1). As a result, 472 records with 21 factors (9912 data) including test site, cropping system (dryland or irrigated), rainfall (mm), soil texture, soil pH, irrigation water EC (dS m^{-2}), nitrogen, phosphorus and potassium applied to the soil (kg ha^{-1}), soil organic content (%), length of growing season (days), plant height (cm), irrigation regime (according to available water), genotype, 1000-kernel weight (g), spike number m^{-2}, plant density (plant m^{-2}), harvest index (%), grain number per spike and grain and biological yields (kg ha^{-2}) of wheat were prepared and a data set formed.

Statistical analysis

The data set was imported into Rapid Miner (RapidMiner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44227 Dortmund, Germany) software; biological and grain yields were set as output variables, and the rest as input variables. Factors such as grain yield, biological yield, rainfall and plant height were classified as continuous variables, while others such as location, genotype and soil texture were classified as categorical variables.

Attribute-weighting algorithms

The following supervised attribute-weighting algorithms were applied to the data set:

- Weighting by information gain: This operator calculated the relevance of a factor by computing the information gain in class distribution.
- Weighting by the information gain ratio: This algorithm calculated the relevance of a feature by computing the information gain ratio for class distribution.
- Weighting by rule:_This operator calculated the relevance of a factor by computing the error rate of a model on the sample data set without the factor.
- Weighting by deviation: The operator created weights from the standard deviations of all attributes. The values were normalized by the average, minimum or maximum of the attribute.
- Weighting by the chi-squared statistic: This operator calculated the relevance of a factor by computing, for each attribute in the input sample data set, the value of the chi-squared statistic with respect to the class attribute.
- Weighting by the Gini Index: This operator calculated the relevance of a factor by computing the Gini Index of the class distribution, if the given sample data set would have been split according to the factor in question.
- Weighting by uncertainty: This operator calculated the relevance of an attribute by measuring the symmetrical uncertainty with respect to the class.
- Weighting by Relief: This operator measured the relevance of a factor by sampling the examples and comparing the value of the current factor for the nearest example of the same, and of a different, class. This version also worked with multiple classes and regression data sets. The resulting weights were normalized into the interval between 0 and 1.
- Weighting by Support Vector Machine (SVM): This operator used the coefficients of the normal vector of a linear SVM as feature weights.
- Weighting by Principal Component Analysis (PCA): This operator used the factors of the first principal component as feature weights.

Data were normalized before running the models, so it is reasonable to expect that all weights will be presented as a digit between 0 and 1; showing the importance of each attribute for the target attribute (biological or grain yield).

RESULTS

In the Rapid Miner software, biological yield and grain yield of wheat were each the output once; the other variables were the inputs. Attributes with a weight of 0.5 or higher were regarded as important attributes contributing to biological yield or grain yield (Tables 2 and 3).

1. Biological yield as output

- Weighting by information gain

Grain yield was the sole attribute whose weight was equal to 1.0; nitrogen applied and rainfall had a weight of 0.5 (Table 2).

- Weighting by the information gain ratio

When this model was applied to the data set, just grain yield had a weight equal to 1.0 (Table 2); six other attributes, including organic content, plant height, irrigation regime, nitrogen applied, rainfall and 1000-kernel weight had weights equal to 0.5.

- Weighting by rule

Three important attributes were nitrogen applied, grain yield, and growing season length, which had weights of 1.0, 0.7, and 0.6, respectively (Table 2). Plant height, irrigation regime, and grain number per spike were the other three attributes with weights equal to 0.5 by this model.

•Weighting by deviation

Only grain yield, selected as the most important attribute, had a weight equal to 1.0 (Table 2).

•Weighting by chi-squared statistic

Grain yield was weighted at 1.0 (Table 2), and nitrogen applied, organic content and phosphorus applied had weights of 0.6, 0.6, and 0.5, respectively.

•Weighting by the Gini Index

In the Gini Index model, grain yield, nitrogen applied and rainfall were the most important attributes, with values of 1.0, 0.6, and 0.5, respectively (Table 2).

•Weighting by uncertainty

In the uncertainty model, which is similar to the Gini Index algorithm, grain yield, phosphorus applied, nitrogen applied and organic content were the four important attributes, with values of 1.0, 0.6, 0.6, and 0.5, respectively (Table 2).

• Weighting by Relief

Seven attributes had weights equal to or higher than 0.6, including growing season length, soil pH, harvest index, irrigation regime, plant density, nitrogen and rainfall by the Relief model (Table 2).

• Weighting by Support Vector Machine (SVM)

When this model was applied, grain yield, 1000-kernel weight and rainfall had weights equal to or higher than 0.5 (Table 2).

• Weighting by Principal Component Analysis (PCA)

Similar to the deviation model, with the PCA model, grain yield was the sole attribute with a value of 1.0 (Table 2).

2. Grain yield as output

•Weighting by information gain

As was shown in Table 3, genotype, biological yield and location had weights ≥ 0.5.

•Weighting by the information gain ratio

When the information gain ratio was applied to the data set, biological yield had a weight equal to 1.0 (Table 3). Four other attributes, including organic content, culture type, 1000-kernel weight and spike number m^{-2}, were had weights equal to 0.5.

•Weighting by rule

As presented in Table 3, just grain yield had a weight equal to 1.0. Seven other attributes, including rainfall, location, biological yield, soil texture, growing season length, nitrogen applied and grain number per spike, showed weights ≥ 0.5.

•Weighting by deviation

The biological yield with a weight equal to 1.0 was the sole attribute selected by the deviation model (Table 3).

•Weighting by chi-squared statistic

Just two attributes-genotype and location-had weights of 1.0 and 0.5, respectively by the chi-squared statistic model (Table 3).

•Weighting by the Gini Index

Using the Gini Index, three attributes-genotype, biological yield and location-had weights higher than 0.5 (Table 3).

• Weighting by uncertainty

Of 20 attributes used as inputs in the uncertainty algorithm, in addition to biological yield with value of 1.0, 13 attributes had values ≥ 0.5 (Table 3).

• Weighting by Relief

Soil texture, location and harvest index had weights of 1.0, 1.0, and 0.5, respectively (Table 3).

• Weighting by Support Vector Machine (SVM)

In the SVM model, the highest weight (1.0) was calculated for 1000-kernel weight (Table 3). Culture type, plant height, irrigation regime, spike number/m^2, rainfall, biological yield and organic content were seven important attributes with weights ≥ 0.5.

• Weighting by Principal Component Analysis (PCA)

In the PCA model, similar to the deviation model, just the biological yield attribute was assigned a value of 1.0 (Table 3).

Overall, when biological yield was the output, 12 attributes with values ≥ 0.5 were calculated by various weighting models (Table 1). In contrast, when grain yield was applied as output in the weighting models, 16 attributes had weights ≥ 0.5 (Table 3).

The attribute-weighting algorithms that selected the most important attributes (features) are shown in Table 4. When biological yield was the output, grain yield, nitrogen applied, rainfall, irrigation regime and organic content were the most important attributes highlighted (repeated) by 9, 7, 5, 3 and 3 weighting models, respectively. In contrast, when grain yield was the output, biological yield, location and genotype were repeated by 8, 6, and 5 weighting models (Table 4). Five attributes, including organic content, culture type, 1000-kernel weight, spike number m^{-2} and soil texture, were selected by three models as the most important attributes.

DISCUSSION

In this study, attribute-weighting algorithms were useful for identifying the most important attributes to be excluded from the data set. Analyzing a large number of features under different field conditions (9912 data) could provide a comprehensive overview of the most important features responsible for grain yield improvement by 10 attribute-weighting models (Tables 2 and 3). Biological yield and spike number m^{-2} had important effects on wheat grain yield, and 8 and 3 models selected these attributes when grain yield was the output (Table 4). Farahani and Arzani (2007) found that biological yield and spike number m^{-2} were strongly correlated to wheat grain yield. Ghodsi *et al.* (2005) also reported a significant relationship between spike number m^{-2} and grain yield.

Table 3. Identifying the most important attributes (weights ≥ 0.5) for determining wheat grain yield by different weighting algorithms (values closer to 1 show greater effectiveness of the attribute in determining wheat grain yield).

Weighting algorithm	Attribute	Weight
Information gain	Genotype	1.0
	Biological yield	0.7
	Location	0.5
Information gain ratio	Biological yield	1.0
	Organic content	0.5
	Cropping system	0.5
	1000-kernel weight	0.5
	Spike number/m^2	0.5
Rule	Genotype	1.0
	Rainfall	0.8
	Location	0.7
	Biological yield	0.7
	Soil texture	0.6
	Length of growing season	0.6
	Nitrogen applied	0.6
	Grain number per spike	0.5
Deviation	Biological yield	1.0
Chi-squared statistic	Genotype	1.0
	Location	0.5
Gini Index	Genotype	1.0
	Biological yield	0.8
	Location	0.5
Uncertainty	Biological yield	1.0
	Cropping system	0.9
	Location	0.8
	Phosphorus applied	0.8
	Genotype	0.8
	Organic content	0.7
	Soil texture	0.7
	Plant height	0.6
	1000-kernel weight	0.6
	Irrigation regime	0.6
	Nitrogen applied	0.6
	Soil pH	0.5
	Length of growing season	0.5
	Spike number/m^2	0.5
Relief	Soil texture	1.0
	Location	1.0
	Harvest index	0.5
Support vector machine (SVM)	1000-kernel weight	1.0
	Cropping system	0.7
	Plant height	0.7
	Irrigation regime	0.7
	Spike number/m^2	0.6
	Rainfall	0.6
	Biological yield	0.6
	Organic content	0.5
Principal component analysis (PCA)	Biological yield	1.0

When biological yield was the output, nitrogen and grain yield had a strong relationship with biological yield, with values of 0.5 to 1.0 in various attribute-weighting algorithms (Tables 2 and 4). Emam *et al.* (2009) showed that nitrogen applied to the soil, a key element in crop nutrition, had an important role in increasing biological yield and wheat grain yield. Furthermore, Abhari *et al.* (2008) found a positive relationship between biological yield and wheat grain yield.

Using a supervised feature selection algorithm, Bijanzadeh *et al.* (2011) found that culture types such as dryland or irrigated farming severely affected wheat grain yield, and mean grain yield under dryland (1966 kg/ha) and irrigated conditions (4000 kg/ha) was significantly different ($P \leq 0.01$). Furthermore, the relationship of culture type and genotype with wheat grain yield has been reported by researchers under water stress conditions (e.g., Ahmadi and Sio Se Mardeh, 2003; Emam *et al.,* 2007). In a field study with 42 durum and bread wheat genotypes and F1 hybrids, there was a significant difference between genotypes for yield and yield components, and genotype was considered to be an important factor in determining final grain yield (Farahani and Arzani, 2007). In the present study, models 5 and 3 indicated that genotype and culture type were related to grain yield (Table 4).

Ten different attribute-weighting models showed that while biological yield was an important attribute for improving wheat grain yield, harvest index was less important in modern wheat genotypes in Iran, and was only selected by the Relief model when biological yield or grain yield were the outputs (Tables 2 and 3). Sharma and Smith (1996) found that wheat grain yield may be increased by improving biomass at a given level of harvest index in three winter wheat populations. Austin (1984) showed that an alternative for increasing grain yield is to increase wheat biomass.

Farid *et al.* (1996) reported that improving harvest index appears to be difficult, and recent increases in wheat grain yield have been attributed to increases in biomass production. Recently, Bijanzadeh *et al.* (2010) reported that modern wheat cultivars grown in Iran show variation in biomass production, and that it might be possible to improve wheat grain yield by selecting cultivars with higher biomass.

Field water status, such as irrigation regime or rainfall, was another important attribute related to biological yield or wheat grain yield (Table 4). Bijanzadeh *et al.* (2011) reported that water stress decreased grain yield of five wheat cultivars by decreasing 1000-kernel weight. In the present study, 1000-kernel weight was another important attribute selected by models 3 and 2, when biological yield and grain yield were the outputs, respectively (Table 4).

Potassium applied to soil (0.133 value) was not found to be important (value ≤0.5) using all attribute-weighting models. Malakoti (2003) found that soils in western and southern Iran were rich in available potassium ions, and that farmers often did not apply potassium fertilizer in these areas.

For the first time, our results showed that attribute classification by unsupervised attribute-weighting models can provide a comprehensive view of important distinguishing attributes such as biological yield, location, and genotype, which contribute to wheat grain yield. This study opened a new vista in wheat production using attribute-

Table 4. The number of attribute-weighting algorithms that selected the most important attributes related to biological or grain yield of wheat. # is the number of algorithms that selected the attribute.

Output	Attribute	# Repeat
Biological yield	**Grain yield**	**9**
	Nitrogen applied	**7**
	Rainfall	**5**
	Irrigation regime	**3**
	Organic content	**3**
	1000-kernel weight	**2**
	Length of growing season	**2**
	Plant height	**2**
	Phosphorus applied	**2**
Grain yield	**Biological yield**	**8**
	Location	**6**
	Genotype	**5**
	Organic content	**3**
	Cropping system	**3**
	1000-kernel weight	**3**
	Spike number/m^2	**3**
	Soil texture	**3**
	Nitrogen applied	**2**
	Rainfall	**2**
	Irrigation regime	**2**
	Length of growing season	**2**
	Plant height	**2**

weighting and data mining methods that would benefit newcomers in this field.

ACKNOWLEDGEMENTS

We would like to thank the Department of Crop Production and Plant Breeding of Shiraz University, Shiraz, Iran, for financial support of this research. We also greatly appreciate and acknowledge the technical support of the Bioinformatics Research Group, Green Research Centre, Qom University.

REFERENCES

Abhari, A., S. Galeshi, N. Latifi, and N. Kalateh. 2008. The effects of some growth parameters on grain yield of wheat genotypes yield under drought stress conditions. J. Agric. Nat. Res. 14: 104-112.

Ahmadi, A., and A. Sio Se Mardeh. 2003. Relationships among growth indices, drought resistance and yield in wheat cultivars of different climates of Iran under stress and non stress conditions. Iran. J. Agric. Sci. 34: 667-679.

Ashrafi, E. A. Alemzadeh, M. Ebrahimi, E. Ebrahimie, N. Dadkhodaei, and M. Ebrahimi. 2011. Determining specific amino acid features in P1B-ATPase heavy metal transporters which provides a unique ability in small number of organisms to cope with heavy metal pollution. Bioinfo. Biol. Insights 5: 59-82.

Austin, R. B. 1984. Crop characteristics and the potential of wheat yield. J. Agric. Sci. Camb. 98: 447-453.

Bijanzadeh, E., and Y. Emam. 2011. Evaluation of assimilate remobilization and yield of wheat cultivars under different irrigation regimes in arid climate. Arch. Agron. Soil Sci. 58(11) 1243-1259.

Bijanzadeh, E., Y. Emam, and E. Ebrahimie. 2010. Determining the most important features contributing to wheat grain yield using supervised feature selection model. Aust. J. Crop Sci. 4: 402-407.

Ebrahimi, M., and E. Ebrahimie. 2009. Application of neural networks methods to define the most important features contributing to xylanase enzyme thermostability. Pp. 2885-2891. In: Proceedings of the Eleventh Congress on Evolutionary Computation.

Ebrahimi, M., and E. Ebrahimi. 2010. Sequence-based prediction of enzyme thermostability through bioinformatics algorithms. Current Bioinfo. 5: 195-203.

Ebrahimi, M., Ebrahimi E., and Ebrahimi, M. 2009. Searching for patterns of thermostability in proteins and defining the main features contributing to enzyme thermostability through screening, clustering, and decision tree algorithms. EXCLI J. 8: 218-233.

Ebrahimie, E., Ebrahimi, M. Ebrahimi M., Zinati, Z. Delavar, A. and Mohammadi-Dehcheshmeh, M. 2008 Determining the main factors contributing to xylanase resistance to different pH by applying the asymptotic distribution of a transformation on the Pearson correlation coefficient. Proceedings of the 3rd International Conference on Pattern Recognition in Bioinformatics.

Emam, Y., Ranjbari, A. M., and Bahrani, M. 2007. Evaluation of yield and yield components in wheat genotypes under post-anthesis drought stress. J. Sci. Technol. Nat. Res. 11: 317-328.

Emam, Y. 2007. Cereal production. 3rd Edition, Shiraz University Press, Shiraz. 190 pp.

Emam, Y., S. Salimi, and A. Shekofa. 2009. Effect of nitrogen level on grain yield and yield components of wheat (*Triticum aestivum* L.) under irrigation and rainfed conditions. Iranian J. Field Crops Res. 7: 323-334.

Farahani, E., and A. Arzani. 2007. Study of genetic diversity of cultivars and F1 hybrids of wheat using agronomic and morphological traits. J. Sci. Techn. and Nat. Res. 10: 341-355.

Farid, M., M. Niknejad, and Y. Emam. 1996. Analysis of phytomass in 36 Iranian winter wheat (*Triticum aestivum* L.) genotypes. Iranian Agric. Res. 15: 69-77.

Ghodsi, M., M. R. Chai Chi, D. Mazaheri, and M. R. Jalal Kamali. 2005. Determination of susceptibility of developmental stages in bread wheat to water stress and its effect on yield and yield components. Seed and Plant 20: 489-509.

Hsiao, H.W., M. S. Tasi, and S. C. Wang. 2006. Spatial data mining of collocation patterns for decision support in agriculture. Asian J. Health Infor. Sci. 1: 61-72.

Lakizadeh, A., P. Agha-Golzadeh, M. Ebrahimi, and E. Ebrahimie. 2011. Engineering thermostable enzymes: Application of unsupervised clustering algorithms. Adv. Stud. Biol. 3: 63-78.

Liu, H., and H. Motoda. 2001. Feature extraction, construction and selection: a data mining perspective. Boston, Kluwer Academic Publishers. 434pp.

Liu, H., and H. Motoda. 2008. Computational Methods of Feature Selection. Chapman and Hall/CRC Taylor and Francis Group 6000 Broken Sound Parkway NW.

Malakoti, M. J. 2003. Recommendations and applications of chemical fertilizers in fields of Iran. Tarbiat Modarres University Press.

Sharma, R. C., and E. L. Smith. 1996. Selection for high and low harvest index in three winter wheat populations. Crop Sci. 26: 1147-1150.

Shekoofa, A., Y. Emam, M. Ebrahimi, and E. Ebrahimie. 2011. Application of supervised feature selection methods to define the most important traits affecting maximum kernel water content in maize. Aus. J. Crop Sci. 5: 162-168.