Trakia Journal of Sciences, No 1, pp 47-54, 2014

Available online at:
http://www.uni-sz.bg

ISSN 1313-7050 (print)
ISSN 1313-3551 (online)

Original Contribution

POTATO MASS MODELING WITH DIMENSIONAL ATTRIBUTES USING REGRESSION AND ARTIFICIAL NEURAL NETWORKS

R. Farhadi*, D. Ghanbarian

Agricultural Machinery Department, Faculty of Agriculture, University of Shahrekord, Shahrekord, Iran,

ABSTRACT
In this research ten ordinary regression methods, quadratic response surface model and two artificial neural networks were applied for finding relations between potato tuber dimensions (length, width and thickness) as input data and mass of tuber as output data. This relation is useful for designing potato sorters. Four hundred and seventy five potatoes of Morfana variety randomly selected from farms of Hamedan province in Iran. Two thirds of data was used for relation calculation and one third for examination and evaluation. Results showed that smoothing spline technique had the best outcomes for mass prediction in regression methods while generalized regression neural network conduced the best results between artificial neural networks. Comparison of the best regression and artificial neural network showed that, in most cases, artificial neural networks had less errors and more precision. At last, generalized regression neural network with three inputs and mean of absolute value of error =9.1 gram is suggested to be the best method for predicting mass of potato tubers.

Key words: Sorter, Artificial neural networks, Back propagation model, Generalized regression neural network, Quadratic response surface model, smoothing spline technique

INTRODUCTION

Potato is one of the most important agricultural products in Iran and other countries of the world. Grading seed tubers is a basic post harvest operation of this product. The best method of potato sorting is gradation based on mass according to the studies carried out by Goryachkin [1], Shaym [2], Butler et al. [3] and Ghanbarian et al. [4]. Although weight of tubers is the best criterion for separating seed tubers, weight sizing mechanisms are not customary because of being slow and costly. It is usually performed by dimensional attributes (length, width and thickness of tubers). Therefore, determining the best relationship between mass and dimensions is one of the researchers' activities. The study of research papers in the potato gradation field show that most researchers have used ordinary regression methods for finding relation of potato and its dimensions (Shaym [2], Butler et al. [3], Ghanbarian et al. [4], Dalvand [5]). Recently, artificial neural networks as powerful implement for modeling complex relation between inputs and outputs have been used by agricultural researchers. Rios-Cabrera et al. [6] used image processing by three models of Back propagation, Perceptron and Fuzzy ARTMAP for finding misshapen and defected potatoes. Amiryousefi & Mohebbi [7] applied multilayer feed forward neural network to model mass transfer during osmotic dehydration of potato slices. Zangeneh et al. [8] compared application results of two different approaches, parametric model and artificial neural networks for assessing economical productivity, total costs of production and benefit to cost ratio of potato crop. Amiryousefi et al. [9] evaluated changes in physical characteristics of pomegranate using multilayer feed forward neural network.

***Correspondence to**: *Rouhollah Farhadi, Agricultural Machinery Department, Faculty of Agriculture, University of Shahrekord, P.O.Box 115, Shahrekord, Iran, Telephone: +989133718720, Fax: +983814424412, E-mail: Farhadi.Roohollah@gmail.com.*

This study aims at investigating the efficiency of artificial neural networks and making a comparison between conventional regression methods and artificial neural networks for potato mass modeling with dimensional attributes.

MATERIAL AND METHODS

Four hundred and seventy five potatoes of Morfana variety randomly selected from a farm in Hamedan province of Iran. Morfana variety is one of the most important potato varieties in Iran [4]. After cleaning them, dimensions and mass were measured by caliper and digital balance respectively with precision of 0.1 mm and 0.1 gram. Data characteristics, after being transferred to spreadsheet Excel software, are presented in **Table 1**.

Table 1. *Statistical characteristics of measured data.*

	Mass	Height	Width	length
	g	mm	mm	mm
Mean	155.8	51.4	60.5	73.3
Standard deviation	128.8	25.6	16.3	22.7
Min	11.1	21.1	26.3	30.1
Max	833.4	112.0	109.0	170.0
Coefficient of variation (%)	82.7	49.8	26.9	31.0

The difference between minimum and maximum was measured for the study of data distribution status and it was divided by three for each characteristic (length, width, height and mass) and then appeared three sorts of: small, mean and large. Circumstance of data dispensation is shown in **Table 2**.

Table 2. *Data distribution at three sorts of small, mean and large.*

			Mass	Height	Width	length
			g	mm	mm	mm
Difference of min and max			822.3	90.9	82.7	139.9
Sorts distances			274.1	30.3	27.6	46.6
distribution status	Small	numbers	405	16	3	48
		%	85.26	3.37	0.63	10.1
	Mean	numbers	63	354	198	338
		%	13.26	74.53	41.68	71.16
	large	numbers	7	105	274	89
		%	1.48	22.1	57.68	18.74

Two thirds of data was used for relation calculation (317 data) and one third for examination and evaluation (158 data). Whereas mass criterion contains more importance at gradation and 158 selective data for test should have same dispensation of original data until all input range is tested and examined at relation, distribution percentage of **Table 2** – mass part, was exerted for data. Hence, according **Table 3**, 158 test data should be chosen from total input data.

Table 3. *Distributions of test data.*

	Small	Mean	Large	Sum
Percentage	85.26	13.26	1.48	100
Numbers	135	21	2	158

A special program was written with FORTRAN language for the selection of random data with no repetition according to the pattern presented in **Table 3**.

Now, conventional regression method is investigated. The following three methods are generally used for mass prediction according tuber dimensions:

1. One variable regression consists of relation between mass and each dimension separately: M=f (L), M=f (W), M=f (H).
2. Two variable regressions consist of relation among mass and dual components of dimensions: M=f (L, W), M=f (L, H), M=f (H, W).
3. Three variable regressions consist of relation among mass and all dimensions: M=f (L, W, H).

MATLAB software was used for regression calculations. This software contains complete and comprehensive collection of regression methods and graphical capabilities. Ten methods of curve fitting toolbox were applied for one variable and Quadratic response surface model for others. Criterion of the best choice was based on maximum of determination coefficient (R^2) and minimum of SSE (Sum of squares due to error) and RSE (Regression standard error) in one variable state. Quadratic response surface model utilizes following equation for fitting and MATLAB software presents related coefficients (b_i, b_{ij}):

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^3 \quad \text{Eq. (1)}$$

Now, artificial neural network method briefly is introduced. ANN is formed by some neurons placed together with special order. Each net contains minimum of (or at least) two layers, input and output. Moreover, several hidden layers can be placed between the input and output layers **(Figure 1).**

More ANN architectures and algorithms change by alteration of neuron model, relation among neurons and weight function. Two ANN types of Generalized Regression Neural Networks (GRNN) and Back propagation (BP) were selected at MATLAB neural network toolbox. These two models have been applied in many research papers for function approximation and relation creation between inputs and outputs (Wang et al. [10], Effendi et al. [11], Ghamari et al. [12], Heidari et al. [13] and Ghamari [14]).

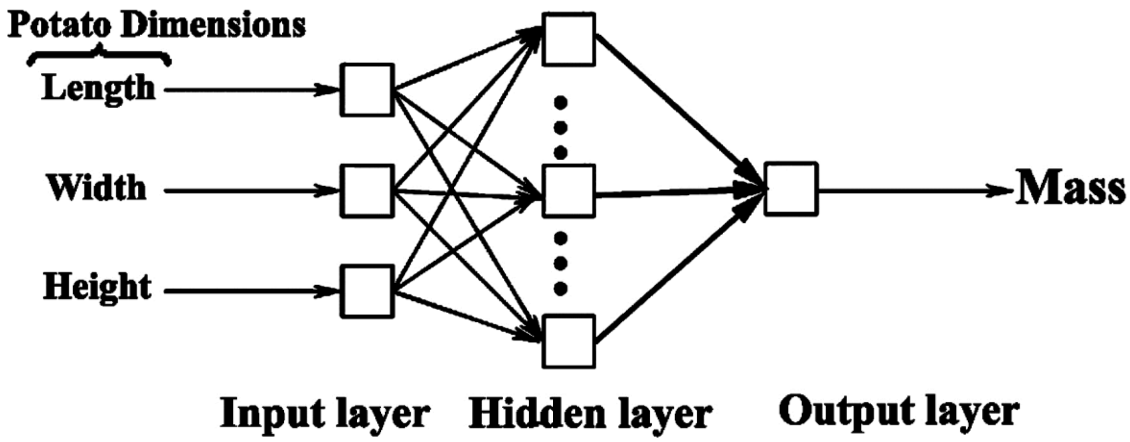

Figure 1. Different layers of ANN.

The BP model contains different training methods that Levenberg-Marquardt (trainlm) and quasi-Newton (trainbfg) were chosen because they had the best results considering examined various tests. Other points that should determine in the BP model are layers numbers and neurons of each layer with applied functions. This part majorly depends on experience and demands examination of different phases. The model was selected with three layers of 1-500-1 (first layer may be two or three based on input condition) and transfer functions of Logsig and Purelin.

In GRNN model, the number of neurons and input layers are equal. The best result was selected with examination of various spread values. Spread value exhibits band width that neurons respond at input distance. Spread value should be large enough to cover input distance strongly and presents proper respond. If the spread value increases highly each neuron effectively responds at wide zone and no desirable result is achieved.

158 test-data were used for evaluation and comparison of regression methods. Results of regression calculation and actual data were compared; then the mean and standard deviation of absolute value of error were computed as is shown in the following formulas:

$$MAE = \frac{\sum_{i=1}^{n} \left| m_{ai} - m_{ei} \right|}{n} \quad \text{Eq.}$$

$$SDAE = \sqrt{\frac{\sum_{i=1}^{n} \left(\left| m_{a_i} - m_{e_i} \right| - MAE \right)^2}{(n-1)}} \quad \text{Eq. (3)}$$

Where:
MAE: mean of absolute value of error
Ma: actual mass
M_e: estimated mass
N: number of inputs
SDAE: standard deviation of absolute value of error

Results were calculated in ANN after defining and training the model. Similar to regression methods, the mean and standard deviation of absolute value of error were computed. Finally outcomes of ANN and regression methods based on mean of absolute value of error were compared with the test of two mean comparisons (t-test).

RESULTS AND DISCUSSION

Among conventional regression methods (one variable), smoothing spline had the best results for mass prediction with regard to value comparison of R^2, SSE and RSE. **Table 4** and **Figure 2** show calculation results of W-M state.

Table 4. *Results of ten regression methods for w-m state.*

Methods	Equations	Regression results		
		RSE	SSE	R^2
Exponential	$ae^{bx} + ce^{dx}$	34.894	381103.08	0.926
Logarithmic	$a\ln x + b$	61.335	1185038.3	0.770
Fourier	$a_0 + (a_1 \cos x + b_1 \sin x) + \ldots + (a_8 \cos 8x + b_8 \sin 8x)$	123.768	4580227.6	0.110
Gaussian	$a_1 e^{\frac{x \; b_1}{c_1}^{2}} + \ldots + a_8 e^{\frac{x \; b_8}{c_8}^{2}}$	29.178	249444.4	0.952
Polynomial	$a_0 + a_1 x + \ldots. + a_9 x^9$	34.810	372004.41	0.928
Power	$ax^b + c$	34.713	378364.69	0.926
Rational	$\frac{a_0 + a_1 x + \ldots. + a_5 x^5}{a_0 + a_1 x + \ldots. + a_4 x^4}$	33.060	335544.6	0.935
Smoothing spline	Variable polynomial oscillating 1 to 3 degrees(Linear, Quadratic and cubic equations)	23.195	90158.643	0.982
Sum of Sin functions	$a_1 \sin(b_1 x + c_1) + \ldots + a_8 \sin(b_8 x + c_8)$	208.659	1.28 E7	-1.479
Weibull	$abx^{(b\text{-}1)} e^{(\text{-}ax^b)}$	202.232	1.29 E7	-1.503

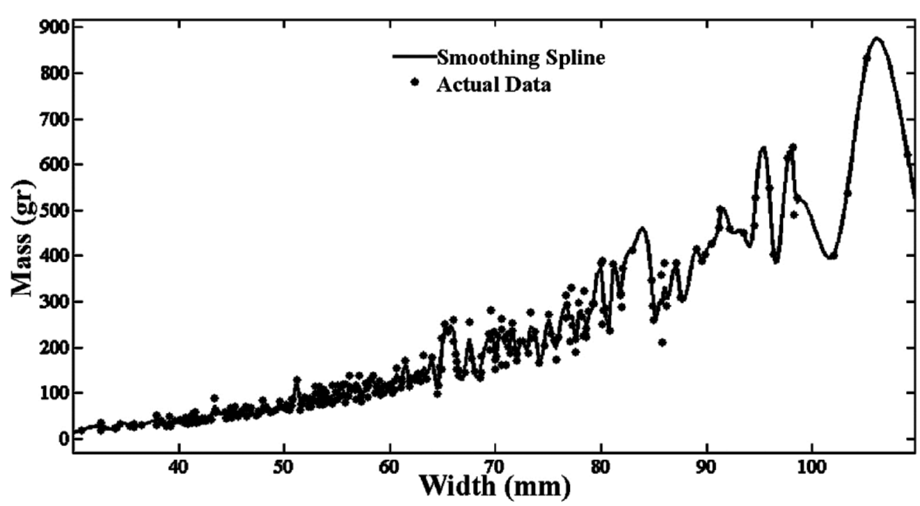

Figure 2. The best regression methods in one variable state (R^2=0.982). – Smoothing Spline ● Actual data

Calculation summary concerning ANNs is presented in **Table 5**. The GRNN model had better conditions and lower error than the BP model. **Figure 3** shows GRNN outcome with three inputs and mean of absolute value of error =9.1 gram.

Table 5. *Comparison results of GRNN and BP model.*

	Models						
	GRNN			BP			
	GRNN results		Spread value	trainlm		trainbfg	
	Mean	SD		Mean	SD	Mean	SD
M=f(L)	26.636	28.152	5	31.550	33.980	28.920	34.916
M=f(W)	21.076	29.582	2	23.449	30.702	23.100	31.817
M=f(H)	27.848	37.434	5	31.585	44.826	33.471	48.251
M=f(L,W)	12.820	14.643	3	16.382	18.690	15.695	20.867
M=f(L,H)	14.646	20.502	5	18.206	21.580	16.832	23.794
M=f(W,H)	19.992	28.641	6	34.464	56.607	23.557	31.195
M=f(L,W,H)	9.144	16.247	3	16.484	28.340	20.993	28.382

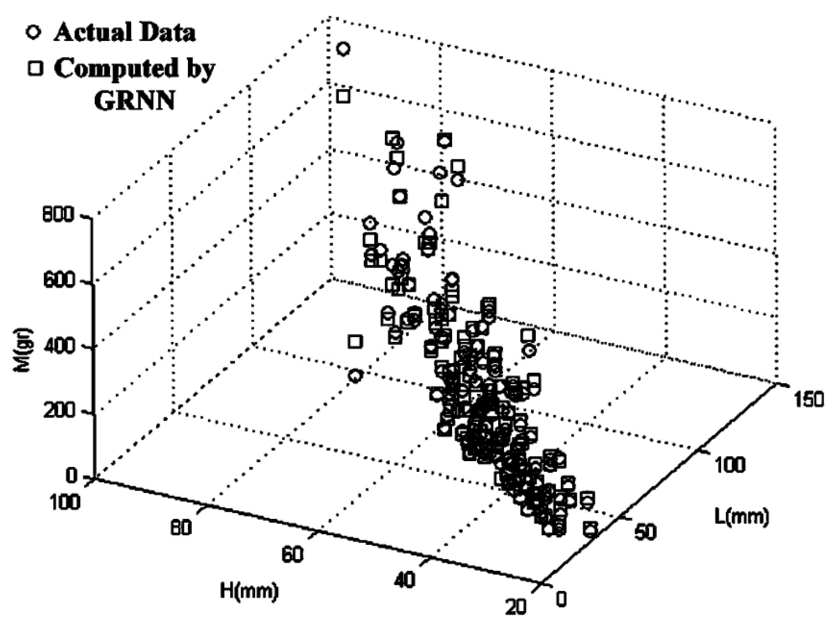

Figure 3. GRNN predicted mass versus actual data.
○ Actual data
□ Computed by GRNN

The best values of regression methods (smoothing spline and multiple quadratic regression) and artificial neural network (GRNN) are presented in **Table 6** with results of t comparative test for discernment of differences significance.

Table 6. *Absolute values of error for comparison of the best regression methods and ANNs.*

	Regression		Neural network		Comparison		
	Smoothing spline / multiple quadratic regression		GRNN		t	$t_{5\%}$	Status
	Mean	SD	Mean	SD			
M=f(L)	35.37	43.65	24.64	28.15	2.11	1.969	*
M=f(W)	27.02	35.56	21.08	29.58	1.62	1.968	-
M=f(H)	35.5	54.41	27.85	37.43	1.64	1.969	-
M=f(L,W)	12.19	16.09	12.82	14.64	0.36	1.968	-
M=f(L,H)	12.58	16.41	14.65	20.5	0.99	1.968	-
M=f(W,H)	20.35	22.87	19.99	28.64	0.12	1.968	-
M=f(L,W,H)	8.86	15.67	9.14	16.25	0.16	1.968	-

* Significant difference at 5% level
\- Non significant

Artificial neural network method contained lower absolute value of error in many cases than regression methods. Although artificial neural network primacies did not have significant difference apart from one item, it shows that ANN has competition power and seems to be equally efficient comparing to the best conventional regression methods. Results of the applied regression in the studies of Shaym [2], Butler et al. [3], Ghanbarian et al. [4] and Dalvand [5] were similar to mentioned regressions, therefor artificial neural network can improve potato mass modeling.

CONCLUSIONS

The best regression method was smoothing spline with comparison of results at whole cases (singly order state). Minimum error values were obtained at perfect state then two variables and finally at a single order. Proper architectures of 1-500-1, 2-500-1 and 3-500-1 were acquired for the BP model. Training methods of Levenberg-Marquardt (trainlm) and quasi-Newton (trainbfg) in the BP model and transfer functions of Logsig and Purelin had better results. Between two neural network models of GRNN and BP, GRNN model worked more efficiently and had lower error in most cases. Regression method had less error just in two cases, although those differences with artificial neural network were not significant.

The GRNN model functions as proper, competitive and coequal method to the best regression techniques. The generalized regression neural networks model with three inputs and mean of absolute value of error =9.1

gram is suggested to be the best method for predicting mass of potato tubers.

Abbreviations: ANN- Artificial neural networks, BP- Back propagation, GRNN- Generalized regression neural network, H- Height, L- Length, M- Mass, Ma- Actual mass, MAE- Mean of absolute value of error, Me- Estimated mass, N- Number of inputs, RSE- Regression standard error, SD- Standard deviation, SDAE- Standard deviation of absolute value of error, SSE- Sum of squares due to error, W- Width.

REFERENCES

1. Goryachkin, V.P., About potato grading. Collected works, Kolos, Moscow, 1968.
2. Shaym, M., Criteria for separating potato tubers in to different size- grades. *Journal of Agricultural Engineering*, 19 (4): 77-83, 1982.
3. Butler, G.P., Bernet, T. and Manrique, K., Mechanization of potato grading on small-scale farms: A case study from Peru. *Experimental Agriculture*, 41:81-92, 2005.
4. Ghanbarian, D., Kolchin, N.N, Hasan Beigi, S.R. and Ebrahimi, R., Design and development of a small potato-grading machine using Capron net. *Journal of Food Process Engineering,* 33(6): 1148-1158, 2010.
5. Dalvand, M.J., Mechanical properties of potato tuber cv. analytic cultivated in Iran, electronic *Journal of Environmental, Agricultural and Food Chemistry*, 11(4): 424-435, 2012.
6. Rios-Cabrera, R., Lopez-Juarez, I. and Sheng-Jen, H., Ann analysis in a vision approach for potato inspection. *Journal of Applied Research and Technology,* 6(2): 106-119, 2008.
7. Amiryousefi, M.R. and Mohebbi, M., Neural network approach for modeling the mass transfer of potato slices during osmotic dehydration using genetic algorithm. African *Journal of Agricultural Research*, 5(1):70-77, 2010.
8. Zangeneh, M., Omid, M. and Akram, A., A comparative study between parametric and artificial neural networks approaches for economical assessment of potato production in Iran. *Spanish Journal of Agricultural Research*, 9(3): 661-671, 2011.
9. Amiryousefi, M.R., Zarei, M., Azizi, M. and Mohebbi, M., Modeling Some Physical Characteristics of Pomegranate (Punica granatum L.) Fruit during Ripening Using Artificial Neural Network. *Journal of Agricultural Science and Technology*. 14: 857-867, 2012.
10. Wang, Y. M., Traore, S. and Kerh, T., Computational performance of reference evapotranspiration in semiarid zone of Africa. *Scientific Research and Essay,* 4 (6): 577-583, 2009.
11. Effendi, Z., Ramli, R. and Ghani, J.A., A Back Propagation Neural Networks for Grading Jatropha curcas Fruits Maturity. *American Journal of Applied Sciences*, 7 (3): 390-394, 2010.
12. Ghamari, S., Borghei, A.M., Rabbani, H., Khazaei, J. and Basati, F., Modeling the terminal velocity of agricultural seeds with artificial neural networks. *African Journal of Agricultural Research,* 5(5): 389-398, 2010.
13. Heidari, M.D., Omid, M. and Akram, A., Application of artificial neural network for modeling benefit to cost ratio of broiler farms in tropical regions of Iran. *Research Journal of Applied Sciences, Engineering and Technology* 3(6): 546-552, 2011.
14. Ghamari, S., Classification of chickpea seeds using supervised and unsupervised artificial neural networks. *African Journal of Agricultural Research*, 7(21): 3193-3201, 2012.