

Comparative Study of Soft Computing Methodologies for Energy Input–Output Analysis to Predict Potato Production

Sara Rajabi Hamedani¹ · Misbah Liaqat² · Shahaboddin Shamshirband² ·
Othman Saleh Al-Razgan³ · Eiman Tamah Al-Shammari⁴ · Dalibor Petković⁵

Published online: 28 April 2015
© The Potato Association of America 2015

Abstract In this study, an adaptive neuro-fuzzy inference system (ANFIS) was developed to predict potato production in Iran. Data related to potato yield from 2010 to 2011 was collected from 50 potato producers in Hamedan, Iran. The resulting ANFIS network has an input layer with eight neurons and an output layer with a single neuron (potato yield). The energy inputs were manual labor, diesel, chemical fertilizers, and manure from farm animals, chemicals, machinery, water, and seed. The most significant and influential inputs were selected from the eight initial inputs and the ANFIS network was used to choose the parameters that have the most influence on potato yield. A new ANFIS model was created after the three most influential parameters were selected. The new ANFIS model was then utilized to estimate yield using the three energy inputs. Next, the ANFIS model results were compared with the results from the support vector regression (SVR) technique. The end results revealed that ANFIS provided more accurate predictions and had the capacity to generalize. The Pearson correlation coefficient (r) for ANFIS

potato yield prediction was 0.9999 in the training and testing phases, while the SVR model had a correlation coefficient of 0.8484 in training and 0.9984 in testing.

Resumen En este estudio se desarrolló un sistema de inferencia adaptativa de lógica difusa (ANFIS) para predecir la producción de papa en Irán. Se colectaron datos relacionados con el rendimiento de papa de 2010 a 2011 de 50 productores en Hamedan, Irán. La red ANFIS resultante tiene una capa de insumos con ocho neuronas y una capa de salidas con una única neurona (rendimiento de papa). Los insumos de energía fueron mano de obra, diésel, fertilizantes químicos y estiércol de animales de granja, químicos, maquinaria, agua y semilla. Se seleccionaron los insumos más significativos y de influencia de los ocho insumos iniciales, y se usó la red ANFIS para escoger los parámetros que tienen la mayor influencia en el rendimiento de papa. Se creó un nuevo modelo ANFIS después que se seleccionaron los tres parámetros de mayor influencia. Entonces se utilizó el nuevo modelo ANFIS para estimar rendimiento usando los tres insumos de energía. Después, los resultados del modelo ANFIS se compararon con los resultados de la técnica de regresión de vector de respaldo (SVR). Los resultados finales revelaron que ANFIS suministró predicciones más precisas y tuvo la capacidad de generalizar. El coeficiente de correlación de Pearson (r) para la predicción del rendimiento de papa por ANFIS fue 0.9999 en las fases de formación y de prueba, e el modelo SVR tuvo un coeficiente de correlación de 0.8484 en formación y 0.9984 en prueba.

✉ Shahaboddin Shamshirband
shahab1396@gmail.com

¹ Department of Agricultural Science, Payame Noor University, Tehran, Iran

² Department of Computer System and Information Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

³ Kuwait Institute for Scientific Research (KISR), Kuwait City, Kuwait

⁴ Department of Information Science, College of Computing Sciences and Engineering, Kuwait University, Kuwait City, Kuwait

⁵ Faculty of Mechanical Engineering, Department for Mechatronics, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Serbia

Keywords Potato yield · Input energy · Prediction · ANFIS · SVR

Introduction

The potato (*Solanum tuberosum* L.) is an important source of several nutrients including vitamins, such as Vitamin C and B6, and minerals, such as iron and potassium. It is a significant crop in Iran, grown in almost every province. The provinces that produce the largest potato crops are Hamedan, Tabriz and Mashad. In Iran, potatoes were grown on approximately 185,000 hectares in 2012 (Anonymous 2012).

In any form of agriculture, energy plays a fundamental role. All phases of agriculture, from tillage to transporting the harvest to the market, require energy (Rykaczewska 2015; Larkin and Halloran 2010). The types of energy consumption in agriculture can be defined as direct or indirect. Direct energy consumption occurs when energy is used to operate machinery, control the temperature in farm buildings, and provide artificial light. Examples of indirect energy consumption include the energy used to fabricate fertilizers, pesticides, herbicides, and farm machinery, and to produce seeds (Ozkan et al. 2004). Population growth, the decreasing availability of arable land, and the rising standard of living have all increased the demand for energy by agriculture (Esengun et al. 2007). Discovering the relationship between energy inputs and output affords the data required to create policies that would optimize resources, encourage renewable energy use, and expand sustainable farming (Brown et al. 2014). The relationship between energy input and output in the production of several crops including cucumbers (Bolandnazar et al. 2014), basil (Pahlavan et al. 2012), chickpeas (Salami and Ahmadi 2010), grapes (Rajabi-Hamedani et al. 2011), canola (Taherri-Garavand et al. 2010), barley (Mobtaker et al. 2010), and garlic (Samavatean et al. 2011) has been investigated in several studies. Energy input and output prediction benefits farmers, governments, and related agricultural industries.

Food production is extremely important to governments due to the role it plays in national security. Internal prognosticators allow governments to create policies that can benefit agriculture in terms of technical and market assistance (Geoffrey 1994). Adaptive neuro-fuzzy inference systems (ANFIS) are often used by researchers. For example, Akbarzadeh et al. (2009) used ANFIS to develop a soil erosion model. Krueger et al. (2011) also used ANFIS to model root distribution patterns, while Khoshnevisan et al. (2014) assessed the ability of ANFIS models to predict wheat yield based on energy inputs. There is a need for a system that can help analyze energy input and output in predicting potato production.

In this study, ANFIS is used to examine energy input and output for predicting potato yield. The inputs of human labor, diesel, fertilizers (manmade and manure from farm animals), chemical additives, machinery, water and seeds are analyzed separately. There are currently no studies in which the relationship between energy input, output and potato yield is modeled with support vector regression (SVR). Thus, one of the primary

goals of this study is to determine whether ANFIS or SVR is more effective for predicting yield based on energy inputs.

Initially, an ANFIS network was used to find the parameter with the strongest influence on yield. This process is known as “variable selection” and serves to find subsets of the full set of variables that are potentially good predictors (Castellano and Fanelli 2000; Dieterle et al. 2003; Cibas et al. 1996; Andersson et al. 2000). One variable selection method is to use previous knowledge to eliminate variables that are not relevant. Another, more advanced technique is to view variable selection as an optimization procedure where genetic algorithms can be exploited (Sofge 2002). In these cases, the goal is to select input variables that can reduce the error between the predicted output variables. ANFIS, among the most powerful neural network systems (Gocić et al. 2015a; Kwong et al. 2009), is employed for variable selection in this study.

ANFIS was used to monitor the variables and identify the significant energy inputs with respect to potato production. The three most significant variables were selected and applied to develop an ANFIS model. ANFIS is a robust tool favored by researchers for modeling (Al-Ghandoor and Samhoury 2009; Petković et al. 2012a, b; Petković and Čojbašić 2012), making predictions (Hosoz et al. 2011; Gocić et al. 2015b; Sivakumar and Balu 2010) and control in engineering systems (Kurnaz et al. 2010; Ravi et al. 2011; Khoshnevisan et al. 2015; Petković et al. 2012a, b; Tian and Collins 2005). ANFIS facilitates a fuzzy modeling procedure to gather data (Aldair and Wang 2011) and it can also be used to organize fuzzy inference systems using input/output data pairs. The next step in our study was to compare ANFIS and SVR results (Sivapragasam et al. 2001). SVR algorithms work well for regression problems but do not account for error approximation in the data or the generalized model. These algorithms rely on a structural risk minimization principle and statistical learning theory (Yang et al. 2009; Zhang et al. 2013). Three kernel functions were used in this work in addition to the SVR scheme. The radial basis (SVR_rbf), polynomial (SVR_poly), and linear (SVR_linear) functions were used to develop a function that can estimate potato production.

Materials and Methods

Data Collection

Fifty potato producers from the province of Hamedan participated in this study. Hamedan is located at 36° 40' latitude and 48° 31' longitude. It receives average rainfall of 317.7 mm and has mean average temperature of 11.3 °C.

The 50 farmers were randomly selected to take part in a face-to-face questionnaire, in which questions related to potato cultivation from 2010 to 2011 were asked. Sample size was determined with the Neyman technique (Zangeneh et al. 2011). The farmers were asked about their practices in terms

of manual labor, diesel, fertilizers, chemicals, water, machinery, and manure from farm animals (energy inputs) to produce a potato crop (energy output). The energy input was quantified per hectare, and this value was multiplied by the coefficient of the energy equivalent. The energy equivalent for both the inputs and output was changed into energy per unit area (Table 1). For example, the energy equivalent for machinery was found using the following equation:

$$ME = E \frac{G}{T}$$

where ME indicates energy used by machines (MJh^{-1}), E denotes the equivalent energy used in the machinery production (MJkg^{-1}), G is the machine weight in kg, and T signifies the economic life of the machine (h). In this study, the energy input was $92225.11 \text{ MJha}^{-1}$ and the output was $103009.2 \text{ MJha}^{-1}$. Renewable energy comprised $34561.59 \text{ MJha}^{-1}$ compared to $57663.52 \text{ MJha}^{-1}$ of non-renewable energy. Table 2 summarizes the energy inputs.

Neuro-Fuzzy Computing

Soft computing is an innovative approach to constructing computationally intelligent systems that possess humanlike expertise within a specific domain. These systems are supposed to adapt in changing environments, learn to do better and explain their decision-making process. It is usually more beneficial to employ several computing methods in a synergistic way rather than building a system based exclusively on one technique. This is useful in confronting real-world computing problems. The result of such synergistic use of computing techniques is the

development of complementary hybrid intelligent systems. The epitome of designing and constructing intelligent systems of this kind is neuro-fuzzy computing: first, neural networks recognize patterns and adapt to cope with evolving environments; second, fuzzy inference systems include human knowledge to implement decision-making and differentiation. Combining and integrating these two complementary methodologies produces a novel discipline called neuro-fuzzy computing.

Adaptive Neuro-Fuzzy Inference Systems

ANFIS (adaptive neuro-fuzzy inference systems) is a class of adaptive networks functionally equivalent to fuzzy inference systems. In this study, the fuzzy inference system utilized has three inputs, x , y and z , and one output, f . The first-order Sugeno fuzzy model with two fuzzy if-then rules was applied as follows:

$$\begin{aligned} \text{Rule 1 : if } x \text{ is } A \text{ and } y \text{ is } C \text{ and } z \text{ is } E \text{ then } f_1 &= p_1x + q_1y + r_1z + s \\ \text{Rule 2 : if } x \text{ is } B \text{ and } y \text{ is } D \text{ and } z \text{ is } E \text{ then } f_2 &= p_2x + q_2y + r_2z + s \end{aligned} \quad (1)$$

The ANFIS architecture for three inputs (x , y and z) is shown in Fig. 1. Nodes in the same layer have similar functions. The output of the i th node in layer l is denoted as $O_{l,i}$.

The first layer consists of input variable membership functions (MFs) and it supplies the input values to the next layer. Every node i is an adaptive node with a node function.

$$\begin{aligned} O_{l,i} &= \mu_{A_i}(x), \text{ for } i=1, 2, \text{ or} \\ O_{l,i} &= \mu_{C_{i-2}}(y), \text{ for } i=3, 4, \text{ or} \\ O_{l,i} &= \mu_{E_{i-4}}(z), \text{ for } i=5, 6 \end{aligned} \quad (2)$$

Table 1 Energy equivalent of inputs and output in agricultural production

Inputs and output	Unit	Energy equivalent (MJh^{-1})	References
Inputs			
1. Human labor	h	1.96	Ozkan et al. 2004
2. Machinery	h	13.06	
3. Diesel fuel	L	47.8	
4. Chemical fertilizers	kg		
Nitrogen		78.1	
Phosphorus		17.4	
Potassium		13.7	
5. Farmyard manure	ton	303.1	
6. Chemicals	kg		Esengun et al. 2007
Insecticides		120	
Herbicides		120	
7. Water	m^3	1.02	
8. Seed (potato)	kg	3.6	
Potato	kg	3.6	
Outputs			
Potato	kg	3.6	Esengun K, et al., 2007

Table 2 The input energy values and their percentage of the total energy input

Input (unit)	Total energy equivalent (MJha ⁻¹)	Percentage (%)
1. Human labor (h)	571.91	62
2. Chemical fertilizer(kg)		
a. Nitrogen	32881.46	35.65
b. Phosphorus	2451.71	2.65
c. Potassium	392.25	0.42
3. Farnyard manure (ton)	6612	7.16
4. Chemicals (kg)		
a. Insecticide	264.6	0.28
b. Herbicide	102.3	0.11
5. Machinery (h)	2256.57	2.44
6. Water for irrigation (m ³)	13550.07	14.69
7. Diesel fuel (L)	19314.63	20.94
8. Seed (potato)	13827.61	14.99
Total energy input (MJha ⁻¹)	92225.11	100
Total output energy (MJha ⁻¹)	103009.2	
Renewable energy(MJha ⁻¹)	34561.59	37.47
Non-renewable energy(MJha ⁻¹)	57663.52	62.52

where x or y or z is the input to node i and A_i or B_{i-2} or E_{i-4} is an associated linguistic label (such as ‘small’ or ‘large’). In other words, $O_{i,i}$ is the membership grade of a fuzzy set A, C and $E = A_1, A_2, C_1, C_2, E_1, E_2$, which stipulates the extent to which the specified input x or y or z satisfies the quantifier A, C or E . In this instance, the membership function can be any suitable parameterized membership function. Membership functions are represented by $\mu_{A_i}(x), \mu_{C_{i-2}}(y), \mu_{E_{i-4}}(z)$. The generalized bell function is used here, as it is best able to generalize non-linear parameters:

$$\mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x - C_i}{a_i}\right)^{2b_i}} \tag{3}$$

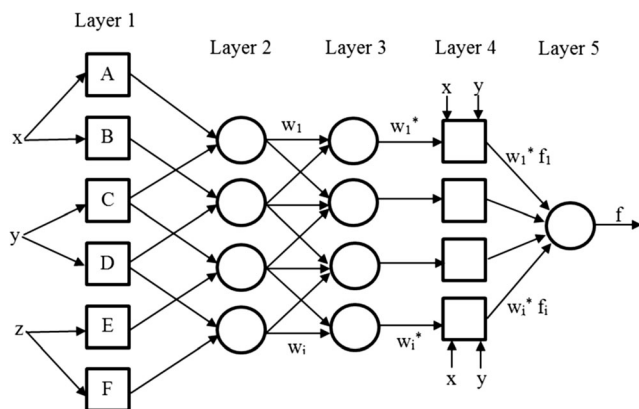


Fig. 1 ANFIS Structure

where $\{a_i, b_i, c_i\}$ is the variable set. The bell-shaped function varies accordingly as the values of the variables change, therefore manifesting different types of membership functions for fuzzy set A . Variables in the first layer are called premise variables.

The second layer (membership layer) multiplies incoming signals from the first layer and sends the product out. Each node in the 2nd layer is a fixed node and its output is the resultant of all incoming signals:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{C_{i-2}}(y)\mu_{E_{i-4}}(z), \quad i = 1, 2 \tag{4}$$

The third layer (i.e., rule layer) is non-adaptive, where every node i calculates the ratio of the rule’s firing strength to the sum of all rules’ firing strengths as:

$$O_{3,i} = w_i^* = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2. \tag{5}$$

The outputs of this layer are called normalized firing strengths or normalized weights.

The fourth layer (i.e., defuzzification layer) provides the output values resulting from the inference of rules, where every node i is an adaptive node with the node function:

$$O_{4,i} = w_i^* \cdot f_i = w_i^* (p_i x + q_i y + r_i) \tag{6}$$

where $\{p_i, q_i, r_i\}$ are the consequent parameters.

The fifth layer sums up all inputs from the fourth layer and converts the fuzzy classification results into a crisp output. The node in the fifth layer is not adaptive and computes the overall output of all incoming signals:

$$O_{5,i} = \sum_i w_i^* \cdot f_i = \frac{\sum_i w_i^* f_i}{\sum_i w_i} \tag{7}$$

The parameters in the ANFIS architecture were identified by applying the hybrid learning algorithm. In the forward pass of this algorithm, functional signals go forward until Layer 4. Consequent parameters are identified by the least squares estimate. In the backward pass, the error rates propagate backwards and premise parameters are updated by gradient descent.

Supervised Vector Machines (SVM)

An SVM is a learning technique from the linear classifier family. The formula includes structural risk minimization (SRM) logic, which is very different from the empirical risk minimization (ERM) method used in statistical learning techniques. For instance, SRM eases the maximum bound on a generalization error but ERM creates the lowest error in the training data. SVMs are more likely to be generalized and

their global optimum is assured (Chen et al. 2010; Yang et al. 2014; Raghavendra and Deka 2014).

A Radial Basis Function as a Kernel

SVM uses kernel functions to chart data to a higher dimensional feature space – a characteristic that accounts for the SVM flexibility (Wang et al. 2010; Ayat et al. 2005). In higher dimensional feature spaces, linear solutions are comparable to non-linear solutions in the initial, lower dimensional input space. This attribute makes SVM an appropriate choice for examining issues related to hydrology that are non-linear in nature. A few methods use non-linear kernels for working with regression problems while using SVMs. The radial basis function (RBF) is a compressed supported kernel, appropriate for restricting the computation training process, thus allowing SVM to be compared with other probable kernel features. In this study, RBF with a σ parameter was used.

Support Vector Regression (SVR)

SVR was designed to discover which function has the highest number of ϵ deviations from the actual destination vector for flat training data (Yousefi et al. 2015; Ju and Hong 2013). Non-linear support vector regression is discovered by the kernel function. SVR requires the user to define the kernel-specific parameters. SVR also requires for the size error in ϵ and the best values of the legalization argument C to be established. SVR has the advantage of an algorithm that incorporates the outcome of the quadratic programming function, resulting in unique, superior, and complete solutions.

The training data is expressed as $\{x_i, y_i\}_{i=1}^N$, where $x_i \in \mathcal{R}^p$ identifies the p -dimensional input vector and $y_i \in \mathcal{R}$ denotes a scalar measured output that indicates the system’s output. The purpose here is to establish a function that indicates the degree of dependency of y_i on the input x_i . This function can be stated as $y=f(x)$ and the linear function y can be described as:

$$y = w^T \varphi(x) + b \tag{8}$$

where w represents the weight vector and b signifies the bias. A non-linear mapping function is $\phi(.)$. The mapping function takes the original input data and places it in a high-dimensional space that moves non-linear separable problems into linearly separable space. The non-linear function that maps data to a higher dimensional feature space is $\phi(.) = \mathcal{R}^p \rightarrow \mathcal{R}^h$. Equations 9 and 10 describe the optimization problem and equality constraints:

$$\min j(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 \tag{9}$$

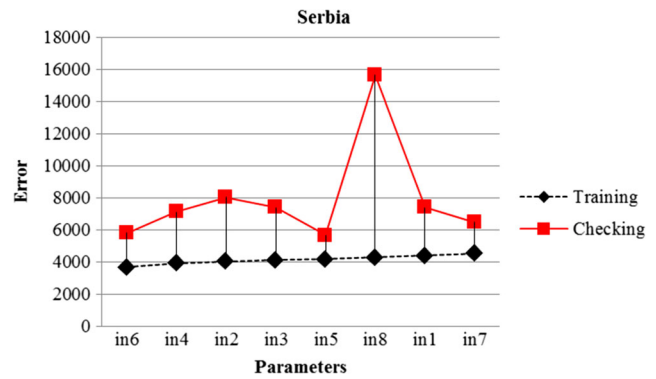


Fig. 2 Every input parameter’s influence on production yield

subject to

$$y_i = w^T \phi(x_i) + b + e_i, i = 1, \dots, N \tag{10}$$

In these equations, e_i represents the random error and $\gamma \in \mathcal{R}^+$ signifies the regularization parameter for optimizing the compromise between minimizing training errors and the model’s complexity. The purpose of seeking optimal parameters is to decrease the regression model’s prediction errors. The optimal model was selected by decreasing the cost function where e_i is minimized. This formula relates to the regression high-dimensional feature space. The possibly infinite nature of this space means there is no easy solution to this problem. To address this issue, the Lagrange function is expressed as follows:

$$L(w, b, e; \alpha) = J(w, e) - \sum_{i=1}^N \alpha_i \{w^T \phi(x_i) + b + e_i - y_i\} \tag{11}$$

The answer to Eq. 9 can be obtained by making partial differentiation considering $w, b, e; \alpha$, as demonstrated below:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^N \alpha_i \phi(x_i) \tag{12}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow b = \sum_{i=1}^N \alpha_i = 0 \tag{13}$$

$$\frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = y - e_i, i = 1, \dots, N \tag{14}$$

$$\frac{\partial L}{\partial x_i} = 0 \rightarrow w^T \phi(x_i) + e_i - y_i = 0, i = 1, \dots, N \tag{15}$$

Table 3 ANFIS regression errors for the potato production prediction

	Training error	Checking error
Input 6	3672.5187	5807.7052
Input4/Input6	1758.0953	6651.6758
Input3/Inpu5/Input6	2.7419	148672.4173

The assumed values for b and α_i (b^a, α_i^a) can be found by answering the linear system. The resulting model is described as:

$$y = f(x) = \sum_{i=1}^N \alpha_i^a K(x, x_i) + b^a \tag{16}$$

where $K(x, x_i)$ defines the kernel function. In this study, the non-linear RBF kernel is written as:

$$K(x, x_i) = \exp\left(-\frac{1}{\sigma^2} \|x-x_i\|^2\right) \tag{17}$$

In Eq. 16, σ indicates the kernel function parameter of the RBF kernel.

The regularization parameter is important to the model and it establishes a compromise between minimizing the fitting error and the smoothness of the estimated function. While running the calibration model, the kernel function value decreased. Using SVR was proposed in this work to create a model for predicting potato production. To develop the proposed SVR model, care was taken to establish the SVR parameters. Rather than decreasing observable training errors, SVR minimizes generalization errors, which is bound to improve the generalized performance.

Results and Discussion

Energy Use Patterns

The energy use patterns and final potato yield amounts are summarized in Table 2. The average values for manual labor, diesel, chemical fertilizers and manure from farm animals, chemicals and water were 571.91, 19314.63, 35725.38, 6612, 366.92 and 13550.06 MJ^{ha}⁻¹, respectively. The energy use pattern shows that nitrogen fertilizer (35.62 %), diesel

(20.94 %) and water (14.69 %) comprised most of the energy for potato production in this area. The high fertilizer use is attributed to the mistaken belief held by many farmers that crop yield will increase in proportion to the amount of fertilizer used. Additionally, the high diesel and water consumption can be attributed to using the wrong type or inefficient machinery relative to farm size.

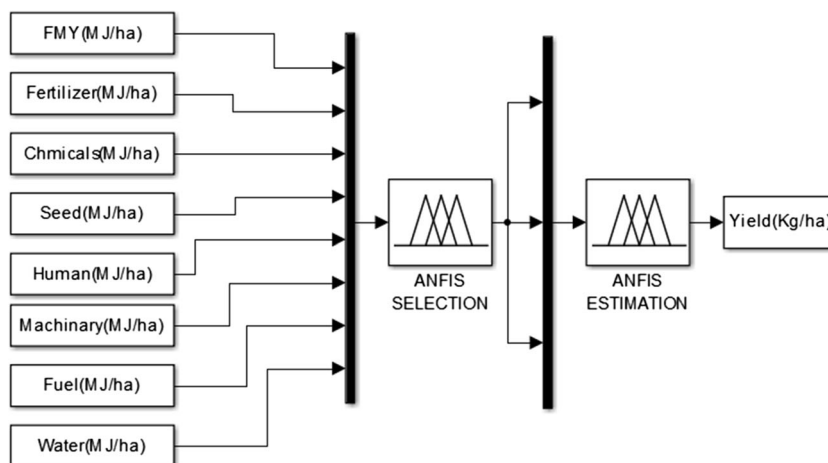
The total energy input needed for potato production in this study reached 92225.11 MJ^{ha}⁻¹ and the output was 103009.2 MJ^{ha}⁻¹. The energy input applied in the current study can be classified as either renewable or non-renewable. The non-renewable energy inputs include diesel, chemicals, chemical fertilizers, and machinery, which account for 62.52 % of the total energy input. The renewable energy input accounted for only 37.48 % of the total energy input. These results are in agreement with the outcome of previous, similar studies. For example, Khoshnevisan et al. (2014) found that 85 % of the energy consumed in wheat production is from non-renewable resources. Additionally, Samavatean et al. (2011) discovered that 63.26 % of the energy input for garlic production was from non-renewable resources.

In this study, the available inputs were scrutinized to determine the optimal combination of energy inputs with the most influence on output. To accomplish this, ANFIS was employed with functions to build a model for each combination of input functions with a training period of one epoch, after which the performance results were reported (Fig. 2). Input 6 (Machinery) had the most influence on potato yield. Figure 2 also shows that the least error was the most relevant with respect to output.

The graph and results in Fig. 2 indicate that input variable 6 is the most influential parameter on predicting potato production.

Similar training and checking error indicate there was no over-fitting, meaning that the ANFIS model can be built using more than one input parameter. To verify this claim, a search was performed to look for the optimal combinations of 2 and 3 input parameters as shown in

Fig. 3 ANFIS selection and detection of potato production yield



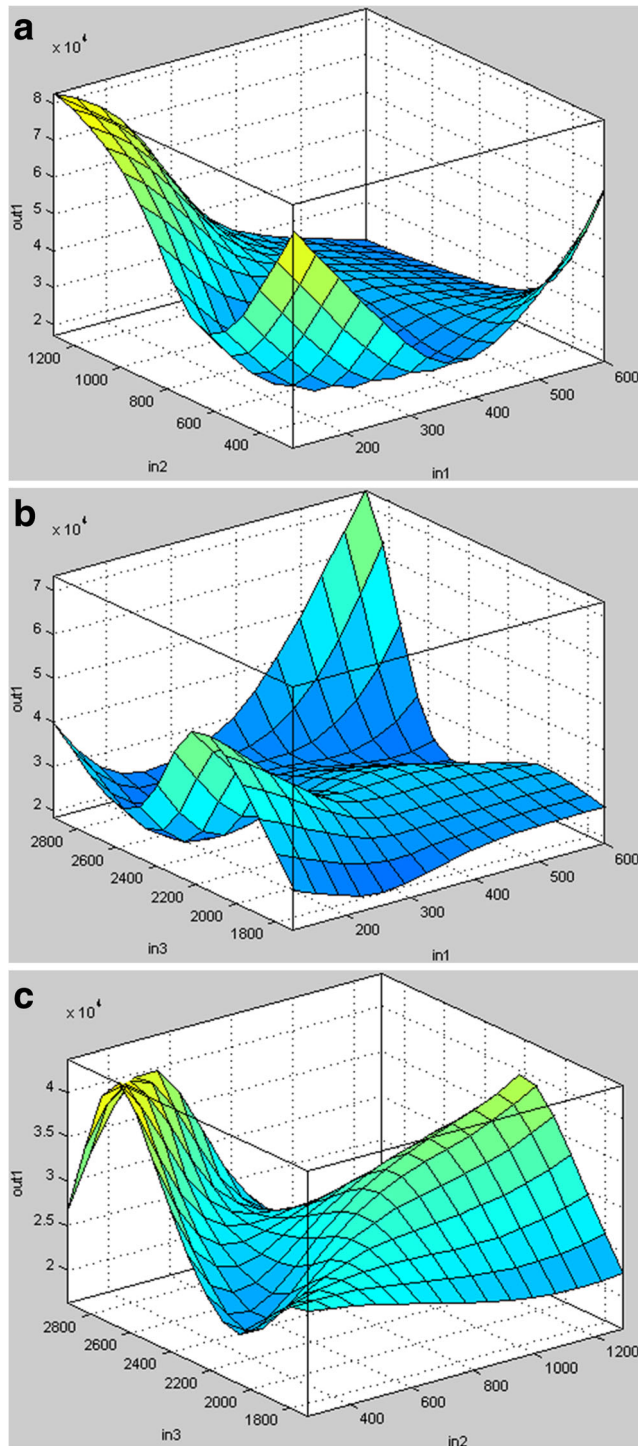


Fig. 4 ANFIS decision surfaces for potato production yield estimation: input1: Chemicals, input2: Human labor and input3: Machinery

Table 3. According to Table 3, input combinations 4 & 6 and 3, 5 & 6 were the optimal combinations for potato production. Simple models with simple structures are always preferable, thus avoiding using more than two inputs in the model. Chemicals, manual labor, and machinery served as input parameters for the original training and checking datasets.

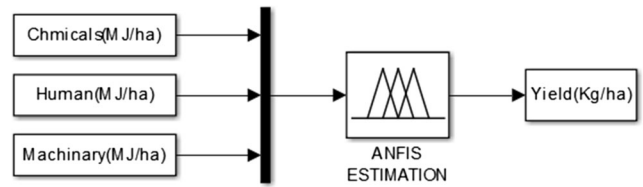


Fig. 5 SIMULINK block diagram for ANFIS prediction of potato yield

An ANFIS model was developed for potato yield prediction after choosing the three most optimal parameters. The scheme used by the algorithm for selection and estimation is shown in Fig. 3. The ANFIS selection and estimation are also illustrated in Fig. 3.

The ANFIS decision surfaces for estimating production with the three optimal parameters are represented in Fig. 4.

Finally, a SIMULINK block diagram was generated to illustrate the ANFIS estimation of potato yield (Fig. 5).

Model Performance Evaluation

In this study, ANFIS and SVR models were compared. Linear, polynomial, and radial basis kernel functions were used with the SVR model. To compare the ANFIS, SVR_rbf, SVR_poly, and SVR_linear models, the statistical indicators shown below were employed:

- 1) root-mean-square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}}, \tag{18}$$

- 2) Pearson correlation coefficient (r)

$$r = \frac{n \left(\sum_{i=1}^n O_i \cdot P_i \right) - \left(\sum_{i=1}^n O_i \right) \cdot \left(\sum_{i=1}^n P_i \right)}{\sqrt{\left(n \sum_{i=1}^n O_i^2 - \left(\sum_{i=1}^n O_i \right)^2 \right) \cdot \left(n \sum_{i=1}^n P_i^2 - \left(\sum_{i=1}^n P_i \right)^2 \right)}} \tag{19}$$

where P_i represents the experimental function and O_i signifies the forecast values. The amount of test data is denoted by n .

Table 4 User-defined parameters for SVR

Drought estimation										
SVR_rbf			SVR_poly				SVR_linear			
<i>C</i>	<i>t</i>	<i>e</i>	<i>C</i>	<i>t</i>	<i>e</i>	γ	<i>d</i>	<i>C</i>	<i>t</i>	<i>e</i>
1000	0.001	0.01	1000	0.001	0.01	2	1.1	1000	0.001	0.01

Table 5 Performance indices of various approaches for potato production estimation

Method	Training		Testing	
	RMSE	r	RMSE	r
ANFIS	9.76	0.9999	12.9	0.9999
SVR_rbf	3027	0.8484	411	0.9984
SVR_poly	4665	0.2077	6472	0.6130
SVR_linear	4500	0.2564	6396	0.6328

The SVR kernel functions used to predict potato production are RBF, a polynomial function and a linear function. The parameters associated with these kernels are C , e , γ , d and t . The accuracy of the SVM model relies on model parameter selection. The user-defined parameters (C , e , γ , d and t) were selected after several trials using different combinations of polynomial kernels (C and d) and radial basis function kernels (C and t). The optimal user-defined parameter values are summarized in Table 4.

Performance Analysis

The expected and actual values for each SVR kernel function and ANFIS were compared using root-mean squared error (RMSE) and the Pearson correlation coefficient. The performance indices for different techniques used to estimate potato production are shown in Table 5. According to this table, ANFIS is the most effective in estimating potato yield.

Conclusion

This study revealed that an ANFIS network can be used to model the energy inputs used for potato production in Iran. The Pearson correlation coefficient (r) for ANFIS-predicted potato yield was 0.9999 in the training and testing phases. The SVR model had a correlation coefficient of 0.8484 in training and 0.9984 in testing. The conclusions drawn from the results can be summarized as follows.

- An analysis of the effect of input parameters on the output revealed that manual labor, chemicals, and machinery have the greatest impact on output.
- The results indicate that ANFIS is a valuable tool for predicting potato production using certain input energies.
- The RMSE and Pearson correlation coefficient indicated that the ANFIS predictions were more accurate than the SVR predictions.
- A sensitivity analysis of the effect of input parameters on output showed that chemical fertilizers and manure from farm animals, diesel, and chemicals have a significant

effect on potato production, while electricity, manual labor and transportation have significantly less effect on output.

The soft computing methods utilized in this study have superior learning and prediction abilities. This study also revealed that the proposed prediction model has the capacity to overcome the lack of artificial neural networks without defining network structure, thus avoiding trapping in the local optimum.

Acknowledgments This work is supported by the Bright Spark Unit, University of Malaya, Malaysia.

References

- Akbarzadeh, A., R.T. Mehrjardi, H. Rouhipour, M. Gorji, and H.G. Rahimi. 2009. Estimating of soil erosion covered with rolled erosion control systems using rainfall simulator (neuro-fuzzy and artificial neural network approaches). *Journal Applied Science Research* 5: 505–514.
- Aldair, A.A., and W.J. Wang. 2011. Design an intelligent controller for full vehicle nonlinear active suspension systems. *International Journal on Smart Sensing and Intelligent Systems* 4: 224–243.
- Al-Ghandoor, A., and M. Samhouri. 2009. Electricity consumption in the industrial sector of Jordan: application of multivariate linear regression and adaptive neuro-fuzzy techniques. *Jordan Journal of Mechanical and Industrial Engineering* 3: 69–76.
- Andersson, F.O., M. Åberg, and S.P. Jacobsson. 2000. Algorithmic approaches for studies of variable influence, contribution and selection in neural networks. *Chemometrics and Intelligent Laboratory Systems* 51: 61–72.
- Anonymous. 2012. Annual agricultural statistics. Ministry of Jihad-e-Agricultural of Iran.
- Ayat, N.E., M. Cheriet, and C.Y. Suen. 2005. Automatic model selection for the optimization of SVM kernels. *Pattern Recognition* 38: 1733–1745.
- Bolandnazar, E., A. Keyhani, and M. Omid. 2014. Determination of efficient and inefficient greenhouse cucumber producers using Data Envelopment Analysis approach, a case study: Jiroft city in Iran. *Journal of Cleaner Production* 5: 1–8.
- Brown, C.R., K.G. Haynes, M. Moore, M.J. Pavek, D.C. Hane, and S.L. Love. 2014. Stability and broad-sense heritability of mineral content in potato: copper and sulfur. *American Journal of Potato Research* 91: 618–624.
- Castellano, G., and A.M. Fanelli. 2000. Variable selection using neural-network models. *Neurocomputing* 31: 1–13.
- Chen, S.-T., P.-S. Yu, and Y.-H. Tang. 2010. Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology* 385: 13–22.
- Cibas, T., F.F. Soulié, P. Gallinari, and S. Raudys. 1996. Variable selection with neural networks. *Neurocomputing* 12: 223–248.
- Dieterle, F., S. Busche, and G. Gauglitz. 2003. Growing neural networks for a multivariate calibration and variable selection of time-resolved measurements. *Analytica Chimica Acta* 490: 71–83.
- Esengun, K., O. Gunduz, and G. Erdal. 2007. Input-output energy analysis in dry apricot production of Turkey. *Energy Conversion and Management* 48: 592–598.
- Geoffrey, A.P. 1994. Economic forecasting in agriculture. *International Journal of Forecasting* 10: 81–135.
- Gocić, M., S. Motamedi, S. Shamshirband, D. Petković, S. Ch, R. Hashim, and M. Arif. 2015a. Soft computing approaches for

- forecasting reference evapotranspiration. *Computers and Electronics in Agriculture* 113: 164–173.
- Gocić M, S. Motamedi, S. Shamsirband, D. Petković, and R. Hashim. 2015b. Potential of adaptive neuro-fuzzy inference system for evaluation of drought indices. *Stochastic Environmental Research and Risk Assessment*. doi:10.1007/s00477-014-0972-6
- Hosoz, M., H.M. Ertunc, and H. Bulgurcu. 2011. An adaptive neuro-fuzzy inference system model for predicting the performance of a refrigeration system with a cooling tower. *Expert Systems with Applications* 38: 14148–14155.
- Ju, F.-Y., and W.-C. Hong. 2013. Application of seasonal SVR with chaotic gravitational search algorithm in electricity forecasting. *Applied Mathematical Modelling* 37: 9643–9651.
- Khoshnevisan, B., S.H. Rafiee, M. Omid, and H. Mousazade. 2014. Development of intelligent based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Information processing in agriculture*. online first.
- Khoshnevisan, B., S.H. Rafiee, M. Omid, H. Mousazadeh, S. Shamsirband, and S.H. Ab Hamid. 2015. Developing a fuzzy clustering model for better energy use in farm management systems. *Renewable and Sustainable Energy Reviews* 48: 27–34.
- Krueger, E., S.A. Prior, D. Kurtener, H.H. Rogers, and G.B. Runion. 2011. Characterizing root distribution with adaptive neuro-fuzzy analysis. *International Agrophysics* 25: 93–96.
- Kurnaz, S., O. Cetin, and O. Kaynak. 2010. Adaptive neuro-fuzzy inference system based autonomous flight control of unmanned air vehicles. *Expert Systems with Applications* 37: 1229–1234.
- Kwong, C.K., T.C. Wong, and K.Y. Chan. 2009. A methodology of generating customer satisfaction models for new product development using a neuro-fuzzy approach. *Expert Systems with Applications* 36: 11262–11270.
- Larkin, R.P., and J.M. Halloran. 2010. Management effects of disease-suppressive rotation crops on potato yield and soilborne disease and their economic implications in potato production. *American Journal of Potato Research* 91: 429–439.
- Mobtaker, H.G., A. Keyhani, A. Mohammadi, S.H. Rafiee, and A. Akram. 2010. Sensitivity analysis of energy inputs for barely production in Hamedan Province of Iran. *Agriculture Ecosystems and Environment* 137: 367–372.
- Ozkan, B., H. Akcaoz, and C. Fert. 2004. Energy input-output analysis in Turkish agriculture. *Renewable Energy* 29: 39–57.
- Pahlavan, R., M. Omid, and A. Akram. 2012. Energy input-output analysis and application of artificial neural networks for predicting greenhouse basil production. *Energy* 37: 171–176.
- Petković, D., and Ž. Čojbašić. 2012. Adaptive neuro-fuzzy estimation of autonomic nervous system parameters effect on heart rate variability. *Neural Computing and Applications* 21: 2065–2070.
- Raghavendra, S.N., and P.C. Deka. 2014. Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing* 19: 372–386.
- Rajabi-Hamedani, S., A. Keyhani, and R. Alimardani. 2011. Energy use patterns and econometric models of grape production in Hamedan province of Iran. *Energy* 36: 6345–6351.
- Ravi, S., M. Sudha, and P.A. Balakrishnan. 2011. Design of intelligent self-tuning GA ANFIS temperature controller for plastic extrusion system. *Modelling and Simulation in Engineering* 2011: 1–8.
- Rykczevska, K. 2015. The Effect of High Temperature Occurring in Subsequent Stages of Plant Development on Potato Yield and Tuber Physiological Defects. *American Journal of Potato Research*, online first.
- Salami, P., and H. Ahmadi. 2010. Energy inputs and outputs in a chickpea production system in Kurdistan, Iran. *African Crop Science Journal* 18: 51–57.
- Samavatean, N., S.H. Rafiee, H. Mobli, and A. Mohamadi. 2011. An analysis of energy use and relation between energy inputs and yield, cost and income of garlic production in Iran. *Renewable Energy* 36: 1808–1813.
- Sivakumar, R., and K. Balu. 2010. ANFIS based distillation column control. *IJCA Special Issue on Evolutionary Computation* 2: 67–73.
- Sivapragasam, C., S.-Y. Liong, and M.F.K. Pasha. 2001. Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics* 3(3): 141–152.
- Sofge, D. 2002. *Using genetic algorithm based variable selection to improve neural network models for real-world systems*, 16–19. Las Vegas: Proceedings of the International Conference on Machine Learning and Applications.
- Taherri-Garavand, A., A. Asakereh, and K. Haghani. 2010. Energy elevation and economic analysis of canola production in Iran a case study: Mazandaran province. *International Journal of Environmental Sciences* 1: 1–10.
- Tian, L., and C. Collins. 2005. Adaptive neuro-fuzzy control of a flexible manipulator. *Mechatronics* 15: 1305–1320.
- Wang, T., H. Huang, S. Tian, and J. Xu. 2010. Feature selection for SVM via optimization of kernel polarization with Gaussian ARD kernels. *Expert Systems with Applications* 37: 6663–6668.
- Yang, H., K. Huang, I. King, and M.R. Lyu. 2009. Localized support vector regression for time series prediction. *Neurocomputing* 72: 2659–2669.
- Yang, X., L. Tan, and L. He. 2014. A robust least squares support vector machine for regression and classification with noise. *Neurocomputing* 140: 41–52.
- Yousefi, M., B. Khoshnevisan, S. Shamsirband, S. Motamedi, M.H.N. Md. Nasir, M. Arif, and R. Ahmad. Support vector regression methodology for prediction of output energy in rice production. *Stochastic Environmental Research and Risk Assessment*. doi:10.1007/s00477-015-1055-z
- Zangeneh, M., Omid, M., 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: 661–671
- Zhang, L., W.-D. Zhou, P.-C. Chang, J.-W. Yang, and F.-Z. Li. 2013. Iterated time series prediction with multiple support vector regression models. *Neurocomputing* 99: 411–422.