Journal of Mechanical Engineering Research and Developments
ISSN: 1024-1752
CODEN: JERDFO
Vol. 43, No. 5, pp. 373-383
Published Year 2020

Plural Genetic Algorithms Approach to Control Agricultural Mechanization and Wheat Production

Yousif Y. Hilal*, Raqeeb Hummadi Rajab, Arkan M. A. Seddiq

Department of Agricultural Machines and Equipment, College of Agriculture and Forestry, University of Mosul, Iraq

*Corresponding Author Email: yousif.yakoub@uomosul.edu.iq

ABSTRACT: For many years the Iraqi wheat production has been facing the challenge of the reduced rate of grain yield caused by the drop in wheat production and high land usage. In the wheat production, the presence of many variables, selecting critical input energy play a crucial role in apprehending different issues, i.e., optimization yield and decision making. The study presents a genetic algorithm program developed to identify the relevant variables affecting the wheat grain yield and straw. In this method, a subset of variables was obtained from a large set. Under a given set of assumptions of active interest in wheat grain yield and straw production predictions in Iraq. Research findings used three years of the wheat mechanization operation that includes tillage, seedbed preparation implements, seeding implements, fertilizing, and pesticide implements, and harvesting. P-fertilizer, seed, and N-fertilizer consumption are considered the most important variables in wheat farm operations, its importance being the relative values of 0.431,0.327 and 0.273. These variables impacted wheat operation during the three years at 19007.1, 28985.607 and 6788.8275 MJ /ha for Bakrajo, Ranya and Chamchamal, respectively. The research concludes that the genetic algorithm method is a user-friendly variable selection tool with excellent results because it can choose variables correctly.

KEYWORDS: Genetic algorithms; agricultural mechanization; wheat production; fertilizers.

INTRODUCTION

Wheat crop (*Triticum*) is a member of the *Poaceae* crop and one of the most staple foods worldwide. One hectare of wheat can produce approximately 1.4 to 4.6 tonnes times compared with those of other major grassy crops. The average wheat yield for the farms from 2013 to 2016 was 4.49 metric tonnes per hectare. The most efficient producers may yield about eight tonnes per hectare [1].

The wheat production in Iraq is considerably competitive and one of the major economic sectors contributing to the total revenue of the country. Currently, wheat is one of the most profitable world crops because of its biomass production [2]. Over the last ten years, the Iraqi wheat industry has encountered many challenges. One of these challenges is the drop in the production of grain yield, which leads to a decrease in the production rate due to the increase in land usage. Eight million hectares are available; only 4 to 5 million hectares being cultivated of the available [3,4]. The wheat yield has a different amount of products in multiple areas of Iraq distributed between high yield, medium and worst, which have significantly affected the efficiency of production. In Iraq, plantation materials are believed to be capable of achieving a yield of four tons of grain yield per hectare per year. However, in fact, average yields are only 50–60 percent of this potential [5,6].

Various parameters, such as kind of wheat areas, climate change, air pollution, and agricultural mechanization can affect the yield of wheat, either positively or negatively. moreover, production systems in agricultural fields from diverse ecosystems are quite different given their energy utilities and energy provision potential [7]. For each energy source and field operation, there is a corresponding norm, which is called a conversion coefficient or energy equivalent. Conversion coefficients help to standardize the unit of all inputs to MJ/ha [8].

FAO [9], demonstrates that there is an increment in the global demand and consumption of wheat crop products. Therefore, factors that can lead to maximum production should be identified and established. A conventional approach is testing a set of affected markers serially by a single-effect marker using traditional procedures [10]. Although widely understood to compute, these approaches consider the removal or addition of one factor once as a step, depending on selected variables [11]. According to Ficken [12], establishing specific parameters that exert considerable effect collectively would be difficult because linear relationships or linear correlations consider only one parameter at a time. Another tactic is examining the most of potential models provided with u factors or parameters to select a model.

Genetic algorithms (GAs) are methods of optimization designed based on natural selection and heredity science [13]. Variables being optimized are epitomized as genes on a chromosome. The most widely used chromosome is a string consisting of 1s and 0s. In natural selection, 1 denotes that the corresponding variable is selected, and 0 denotes otherwise [14]. The fitness of a chromosome is specified by computing the response function score. For most production in wheat, it is generally assumed that there are a few factors with large effects and many with small effects. Therefore, the main objectives of the research are to develop a genetic algorithm as input variable selection models to identify the most significant variables that affect wheat production based on large-scale energy consumption. It also, explore the effects of the selected significant variables in the production of wheat and energy and propose improvements to reduce energy consumption that will lead to the maximum wheat production in Iraq.

MATERIAL

Description of Data Used and the Sites Analysis

The models were developed to select the most significant variables in wheat yield based on prior values of energy consumption. The study was carried out in three locations (Ranya, Bakrajo, Chamchamal) of the north and center of Sulaimani, Iraq, as are stated in Figure (1). These data were taken from the years 2016 to 2019. Table 1 presents data collected with 19 variables to input and two Variables to output. The average high temperatures are between 7,9°C to 38,9°C and average low Temperatures: -0,2°C to 24,1°C. The average annual rainfall in most areas is between 202 to 370 mm.

METHODS

Proposed Method and Details of GA program

GAs can be grouped under many different versions that perform the basic steps, differently. The GA is developed based on correlation analysis CA, the programme presented in this study is developed for the selection of the variable method. CA is applied as the fitness function to develop the GA for model selection. The GA method is used as an optimization tool that reduces

the number of parameters and selects the optimal set with a more effective role compared with other sets to obtain a highly accurate result with the lowest error. All the steps are introduced as follows:

RANYA: 36.255816, 44.883399

Bakrajo: 35.5537,45.3602

Chamchamal: 35.5333312, 44.83333

Figure 1. wheat plantations at Sulaimani

Table 1. Input and Output Variables.

A. Inputs energy	**Unit**	**Energy equivalent (MJ ha-1)**	**Reference**
1-Human labor	H	1.96	[15]
Tillage			
Secondary tillage			
Sowing and covering the seeds			
Fertilization			
Pest control			
Harvesting			
2-Diesel fuel	L	47.3	[16]
Tillage			
Secondary tillage			
Sowing and covering the seeds			
Fertilization			
Pest control			
Harvesting			
3- Fertilizers	Kg		
Farmyard manure		0.3	
Nitrogen (N)		75.46	[17]
Phosphate (P2O2)		13.07	[17]
Potassium (K2O)		11.15	[17]
4- Pest control	Kg		[18]
Pesticide		115	[19]
Herbicides		295	
5-Seed	Kg	25	[20]
B. Output	Kg		[21]
Grain yield		14.7	
Straw		12.5	

Step1. Description of parameters and reading the data of the CA-GA: The default parameters set in GA were the crossovers: the one-point crossover. mutation: bit-flip mutation; selection: random selection. population size: 100; Number of runs: 100 and read the input and output data by selecting the range of data.

Step 2. The population's initiation: a chromosome contains genes; each gene being coded as one if the corresponding variable is selected and 0 if not.

Step 3. Evaluation: CA method is applied to each subset to assess the response. CA is a data processing technique used to calculate the relationship between two variables. One of the approaches uses correlation analysis as the means of estimating the fitness values [22]. The CA in the form of a matrix with the corresponding minimum value for every attribute. The minimum is computed by a magnitude only [23,24]. Independent variables that have strong correlation cannot be a part of predicting dependent variables together, only independent variables that have lower correlation coefficients are fit to take part in the operations [25]. The fitness function taken over here is [23]:

$$f(x) = 1 - \min(\, R_X \,) \ldots \ldots \qquad (1)$$

where:

min (Rx): is the minimum value of correlation coefficient corresponding to any attribute of X.

Step 4. Validation**:** comparing the correlation of the two new chromosomes with one of the existing chromosomes of the current population. As a guide to optimization, the programme uses CA for calculating fitness values. Our fitness function comprises two objective tasks because we considered the optimality of selecting subset regarding two objectives. In the filter phase, the first objective is to maximize the inter-correlation (i.e., the high correlations between input variables, and output variables). Therefore, output variables are given by [26]:

$$R(S,y) = \frac{1}{/S/} \sum_{k=1}^{/S/} /CORR2(x,y)/ \ \ldots \ldots \ldots \qquad (2)$$

Where: R (S,y): is the overall correlation among the selected input variables (S), the corresponding output variables (y), and the minimized intra-correlation, i.e., the low mutual inter-correlations among the input variables, result in a high correlation between the input and output variables. R(S) is the overall correlation between the selected variable subset (i.e., intra-correlation) and defined by [27] as:

$$R(S) = \frac{1}{C(S,i)} \sum_{K=1}^{/S/} \sum_{L=K+1}^{/S/} /Corr2(X_K, X_L)/ \ \ldots \ldots \ldots \qquad (3)$$

Where: C (|S|,i): is the number of i− combinations from the selected variable subset S. Thus, the fitness function concerning the first objective was obtained. For the second objective of the fitness function, CA is used for the evaluation of chromosomes. CA results are fed back to the GA programme to guide the search.

Step 5. This step in this process is called modeling selection when the final criterion is reached. A variable is a characteristic peak, where a set of variables is represented as a chromosome in the GA. The GA evolution process ends. If not, go back to Step 6.

Step 6. in this step, start again to selection is conducted when the estimates do not satisfy the criterions

Step 7. Crossover and mutation: in this step, two new chromosomes must be verified to avoid containing the same variables.

Step 8. Loop the steps 8–3 until a necessary termination criterion is satisfied. The proposed CA-GA method for selection of the optimal subset of variables as shown in Figure (2).

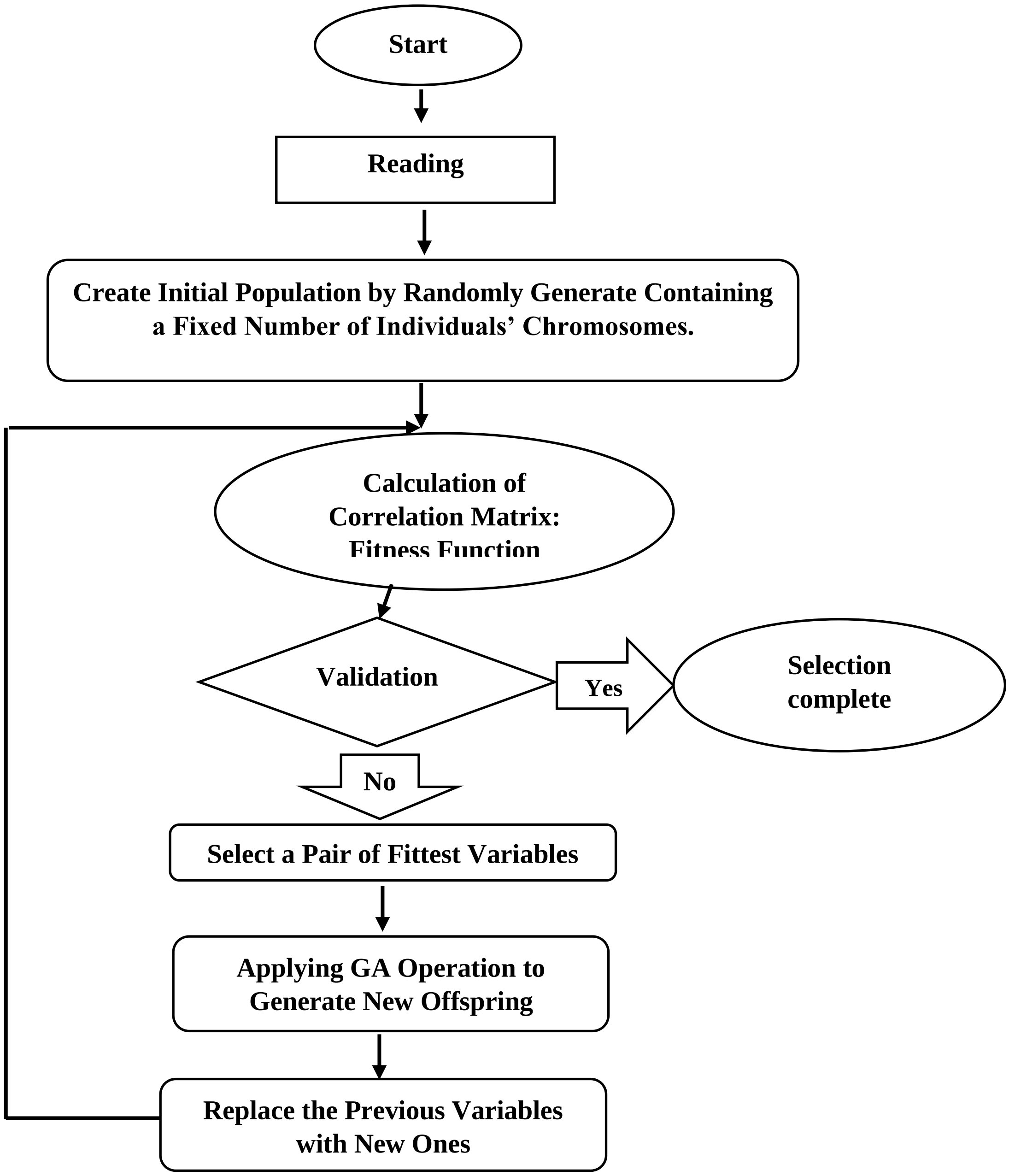

Figure 2. The flow chart of the Genetic algorithm program.

Sensitivity test

A sensitivity method is utilized to identify how different values of an input variable affect a particular output variable under a specified set of assumptions [28]. Pianosi et al. [29] state that the mathematical models that involve many input sensitivity tests are considered essential elements for building the model and quality assurance. In this study, a sensitivity analysis is performed to determine the effectiveness of a variable using the suggested model. The performances of the different possible interactions of variables are also evaluated. Consequently, the performances of the selected group's variables are investigated by the optimal GA model using the correlation.

RESULTS AND DISCUSSION

The productivity of grain yield and straw per hectare of a wheat plantation is the most critical indicator in measuring the efficiency and effectiveness of the farms. The grain yield and straw of the wheat plantation achieved at the same time can also be an indicator of whether the cultivated plantation industry is economically viable. It is impossible to find the same environment and kind of wheat area in each province; they are not the same everywhere. Therefore, in this study, each province is treated separately.

The optimal selection of variables:

The proposed method applies GA to construct an optimal selection of variables. This method is a filter algorithm that ranks variable subsets according to a correlation-based optimizing evaluation function. The selection of evaluation function was toward subsets that contain variables that are highly correlated with the wheat yield and uncorrelated with one another. Irrelevant variables were disregarded because they present a low correlation with the wheat yield. Redundant variables are examined because they are highly correlated with one or more of the remaining variables.

The results are displayed as in Figures 3, 4 and 5 to see how the GA performs with the best and mean values of the population in every generation in Ranya, Bakrajo and Chamchamal for output wheat energy. From these figures, we can see GA converges quickly to the solution until optimization terminated: average change in the penalty fitness value less than options. The function tolerance and constraint violation are lower than the constraint tolerance options. The best value found to fitness value was - 0.2113, - 0.1831 and -0.2663 for Ranya, Bakrajo and Chamchamal, respectively.

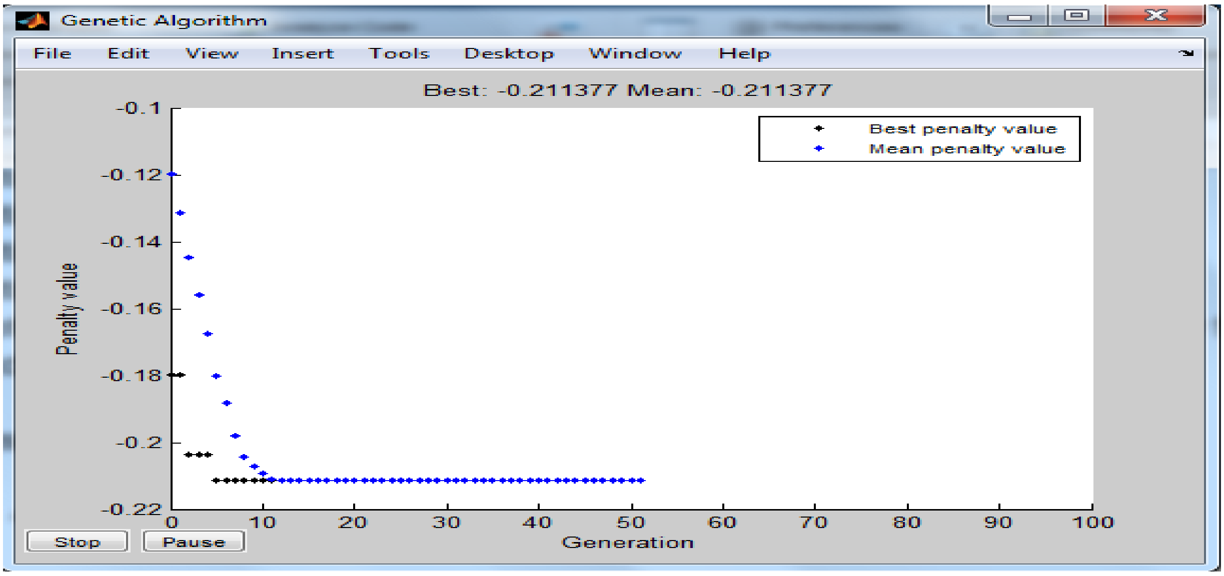

Figure 3. The best fitness value of genetic algorithm performance in Ranya.

The function was automatically called with the best point found by GA. The best solution to variables selection associated with fitness value that the smallest one, it is very close to zero. As a result, the best fitness plot is level and the algorithm stalls at start generation number 10 until 50.

The search started with the minimum correlation set of variables, which contains zero. Every single-variable addition to the minimum set is evaluated, and the selected variable is added to the subset because it exhibits the highest correlation. Subsequently, each of the remaining variables is tested with the selected variables, and the most suitable ones are chosen. Similarly, in the next stage, the following variable is added to the subset. The final step is testing the single remaining variable with the current subset. The most suitable subset found is returned. The variable selection exercise uses the CA as the objective function for GA. A total of 19 variables are available, and each chromosome is 19 bits long. Each of these bits is called a gene, and each of these 19 genes consists of either a randomly generated 0, which indicates the variable that will not be selected, or 1, which states the variable that will be chosen. Variables that appear in at least 50 runs are chosen to ensure that they are the most informative ones and not just selected by chance.

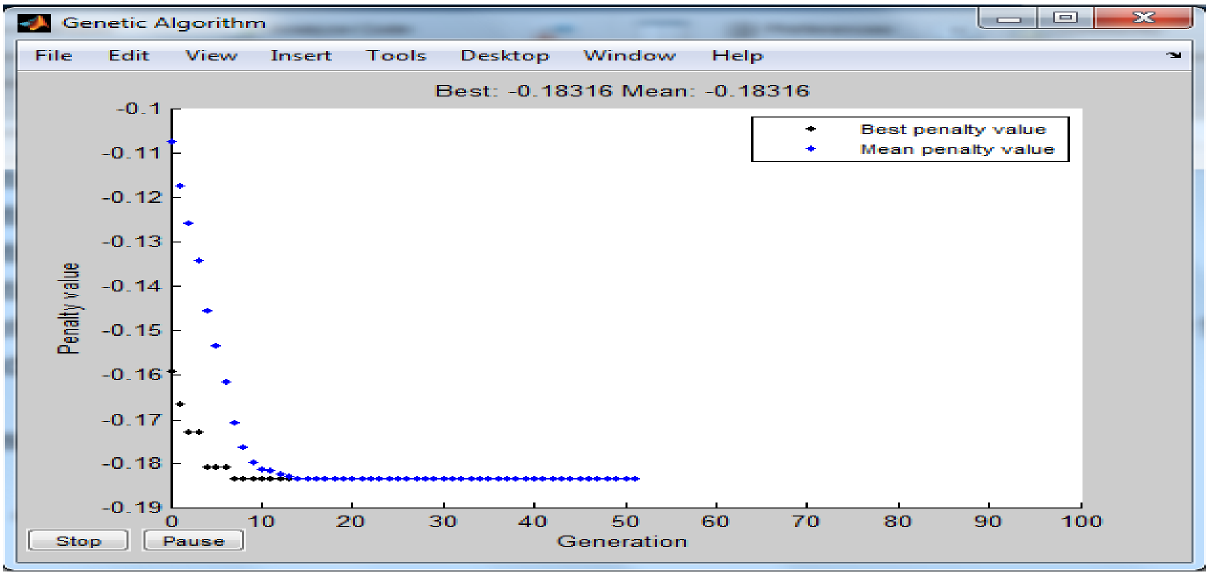

Figure 4. The best fitness value of genetic algorithm performance in Bakrajo.

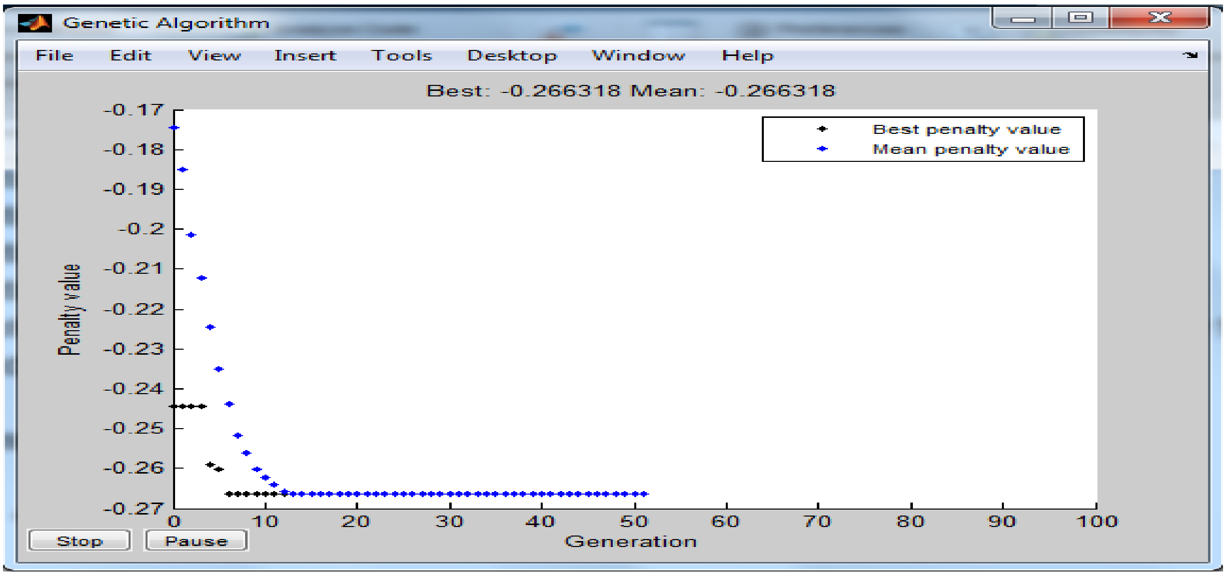

Figure 5. The best fitness value of genetic algorithm performance in Chamchamal.

Tables 2 showed that GA successfully determined the selected variables. A model selection devoid of a number of spurious variables has been created. As the results presented in this Table, the proposed method selected four input energy, namely, seed, P-fertilizer, human labor, and diesel fuel, which were separated from the remaining variables. The selection was based on the highest affected relation with energy produced by wheat in Ranya. Moreover, the method selected four inputs of energy consumption in Bakrajo, including P-fertilizer, N-fertilizer, seed, and diesel fuel. Whereas, the proposed method chose five variables that including N-fertilizer, herbicides, P-fertilizer, diesel fuel and pesticide in Chamchamal.

The wheat industry has faced and is facing regular strong issues its uses, the environmental change and social damages caused by its cultivation. Over the past ten years, climate change had a profoundly negative effect on the cultivation of wheat, which will have a clear impact on the future of wheat growing [5]. One of the causes of the production gap is climate change, which has prompted farmers and investors to find solutions to the problem. Through excessive use of fertilizers, and an increase of labor forces (due to the reality of agricultural mechanization in wheat plantations). Farmers and investors used planting material as a key input for solving the gap of wheat production [30].

Table 2. The Model Selection for Ranya, Bakrajo and Chamchamal.

	Ranya	Bakrajo	Chamchamal
1	Seed	P-fertilizer	N-fertilizer
2	P-fertilizer	N-fertilizer	Herbicides
3	Human labor	Seed	P-fertilizer
4	Diesel fuel	Diesel fuel	Diesel fuel
5			Pesticide

Sensitivity Results

The simulation results demonstrated a successful GA in finding the selected variables. GA finds a model selection that includes a number of extra spurious variables. We explored the strength of the energy output of the wheat production from the perspective of selection models, which were selected by GA due to they proved a clear improvement in the models. The sensitivity analysis was used to assess the IVI. Sensitivity can be determined with the distinction between the highest and lowest values for each scenario. The analysis was conducted by changing independent variable indicators for different possible changes in Ranya, Bakrajo, and Chamchamal.

Figure (6) showed that sensitivity analysis of the data sets in a wheat farm, it appears that the use of seed is significant regarding the impact categories in Ranya. It was considered the most important variable effectively on output energy because its relative importance value was 0.327. This value was followed by those of P-fertilizer, human labor and diesel fuel with relative importance value were 0.268,0.188 and 0.143, respectively.

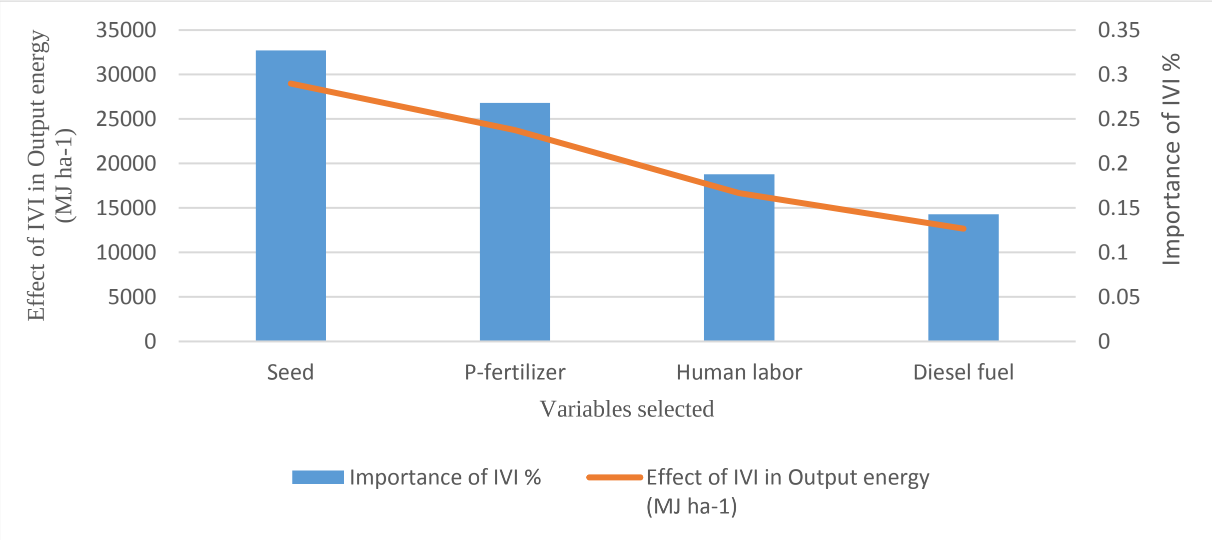

Figure 6. The most significant independent variables for Ranya.

According to the data collected, wheat plantation operations in the farms were positively sensitive to seed, P-fertilizer, human labor, and diesel fuel. For example, seed impact on the output energy in wheat operations was about 28985.607 MJ/ha. The reason being that during recent years, energy use in wheat agriculture was developed in response to a limited supply of arable land, increasing populations and a desire for higher standards of living. P-fertilizer and human labor have negatively affected output energy which can thereby reduce the production of wheat. For example, P-fertilizer affected on the output energy by 23755.788 MJ/ha.

As shown in Figures 7 and 8, P-fertilizer and N-fertilizer consumption was considered the most important variable with its importance relative value of 0.431 and 0.273 in wheat plantation operations in Bakrajo and Chamchamal, respectively. According to the data collected. P-fertilizer and N-fertilizer were 100% of normalized importance that scored the highest impact on the energy of wheat yield. It impacts on the output energy in wheat operations was about 19007.1 and 6788.8275 MJ/ha during the period of study. Sensitivity analysis of the inputs energy consumption presented in Figure (7) showed that N-fertilizer, seed, and diesel fuel high correlated with wheat operations and productions in Bakrajo.

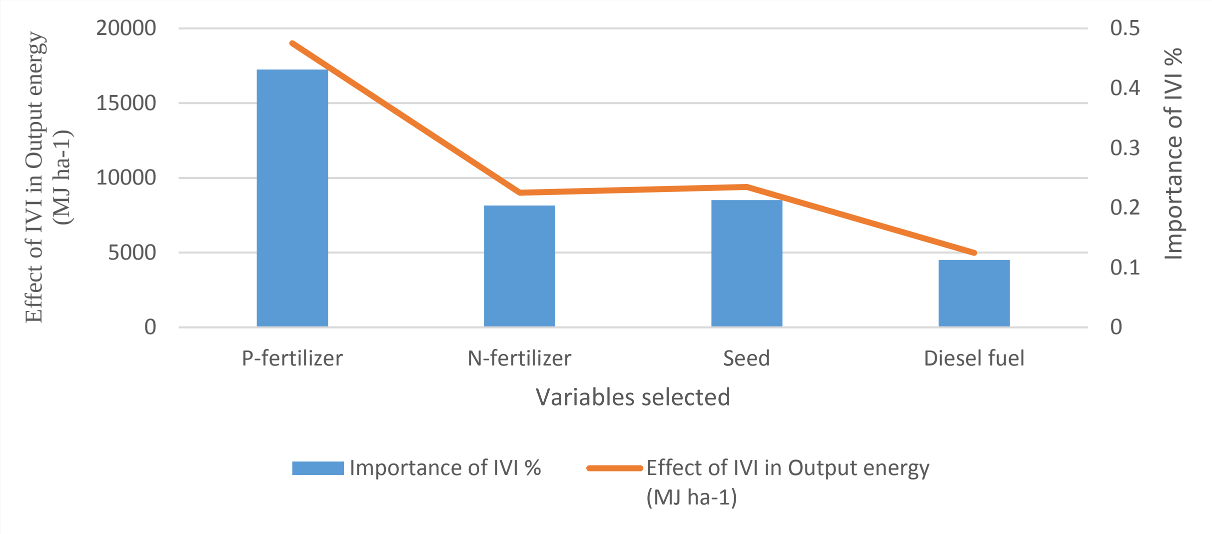

Figure 7. The most significant independent variables for Bakrajo.

As the results in Figure (8) in Chamchamal, the herbicides, P-fertilizer, diesel fuel, and pesticide high correlated with the productivity of wheat plantations, respectively. The results agreed with [31] referring to that wheat requires several types of nutrients in the form of fertilizers to achieve significant growth rates. The results agreed with the study carried out by [32], the rates of application differ depending on the age of the plants. However, climate change has resulted in an increase in demand for input energy to compensate for the shortage of wheat production.

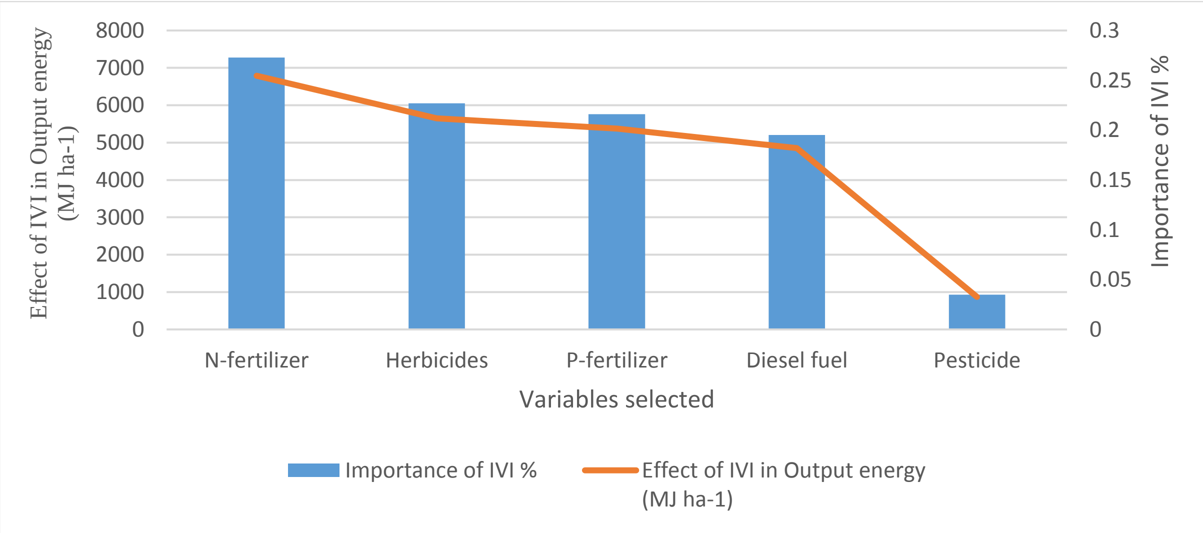

Figure 8. The most significant independent variables for Chamchamal.

CONCLUSIONS

This research mainly aims to clarify how the proposed method can be used to implement a GA in variable selection and optimization, which are essential for understanding the potential of wheat production. The effectiveness of GA produces fast and efficient solutions for incorrect time. Experimental results show that GA can be successfully used to select variables from large variables, allowing a remarkably comprehensive search of the solution space. Sensitivity analysis for variable selections simultaneously determines the effect of input energy on several elements on land productivity. GAs with CA can be readily applied to a wide range to obtain an optimization selection model.

ACKNOWLEDGMENTS

The authors would like to thank Mosul University for support during the time of the study. The Central Statistical Organization CSO and KRSO for supplying the information used in this research.

REFERENCES

[1] *FAO STAT*. *Database*. Retrieved online. http://faostat3.fao.org/home/E. 2018.

[2] S. H. Ewaid, S.A. Abed and N. Al-Ansari. “Water footprint of wheat in Iraq”. *Water*. Vol.11. No. 3, Pp.535-547. 2019.

[3] N. Adamo, N. Al-Ansari, V. Sissakian, J. Laue and S. Knutsson. “The Future of the Tigris and Euphrates Water Resources in view of Climate Change”. *Journal of Earth Sciences and Geotechnical Engineering*. Vol.*8*. No.3, Pp.59-74.2018.

[4] A.J. Alwan, A. J. “Effects of variety productivity on the profitability and comparative advantage for Iraqi wheat production in the year 2017”. *Plant Archives*. Vol.19. No.2, Pp.1194-1198. 2019.

[5] Central Statistical Organization CSO. “Wheat and barley production in 2018”. *In Arabic*.2019.

[6] KRSO. Winter Crops Planted Area Survey in Kurdistan Region (*Area - Yield - Production - Expenditure*) 2016-2017.2019.

[7] H. Ghasemi-Mobtaker, A. Kaab, and S. Rafiee. “Application of life cycle analysis to assess environmental sustainability of wheat cultivation in the west of Iran”. *Energy*. Vol.193. No.15, Pp.116-128. 2020.

[8] R. Fadavi, N. Samavatean, A. Keyhani and S. Saied. “An analysis of improving energy use with data envelopment analysis in apple orchard”. *Asian journal of agriculture and rural development*.Vol.*2*. No.2, Pp 277-286. 2012.

[9] FAO. “FAOSTAT statistics database”. Available at: 2019. http://www.fao. org/faostat/. Last access: 01.01.2017.

[10] M.C. Rodriguez,L. Dupont-Courtade and W.Oueslati. “Air pollution and urban structure linkages: Evidence from European cities”. *Renewable and Sustainable Energy Reviews*. Vol.53, Pp.1–9.2016.

[11] S.Szymczak, E. Holzinger,A. Dasgupta, J.D. Malley, A.M. Molloy, J.L. Mills, and J.E. Bailey-Wilson.”r2VIM: A new variable selection method for random forests in genome-wide association studies”. *BioData mining*. Vol. 9. No.7. 2016. https://doi.org/10.1007/978-3-319-07692-8_32

[12] F.A. Ficken. “The simplex method of linear programming”. *Courier Dover Publications*. New York, USA. Pp.1-64.2015.

[13] J.Trejos, M.A. Villalobos-Arias and J.L. Espinoza. “Variable Selection in Multiple Linear Regression Using a Genetic Algorithm”. *Handbook of Research on Modern Optimization Algorithms and Applications in Engineering and Economics*, Pp.133-160. 2016.

[14] Q. Zhu and A.T. Azar. *Complex system modeling and control through intelligent soft computations*, Springer.London, UK.Pp.1-44. 2015.

[15] F. Kardoni, S. Parande, K. Jassemi and S. Karami. “Energy input-output relationship and economical analysis of wheat product ion in Khuzestan province of Iran”. *International Journal of Agronomy and Plant Production*. Vol. 4. No.9, Pp.2187–2193.2013.

[16] O. Kitani, and T. Jungbluth. *CIGR handbook of agricultural engineering*. American Society of Agricultural Engineers Publication. USA.Pp.30-47.1999.

[17] A. Mohammadi, A. Tabatabaeefar, S. Shahin, S. Rafiee and A. Keyhani. “Energy use and economical analysis of potato production in Iran a case study: ardabil province”. *Energy Conversion and Management*. Vol. 49. No.12, Pp. 3566-3570. 2008.

[18] K. Esengun, O. Gunduz and G. Erdal. "Input-output energy analysis in dry apricot production of Turkey". *Energy Conversion and Management.* Vol. 48, No.2, Pp.592-598.2007.

[19] Z. Hussain, M. Azam Khan and M. Irfan. "Water energy and economic analysis of wheat production under raised bed and conventional irrigation systems: A case study from a semi-arid area of Pakistan". *Soil Tillage research*. Vol.109. No.2, Pp. 61-67. 2010.

[20] S. Khan, M.A. Khan and N. Ltife. "Energy requirements and economic analysis of wheat, rice and barley production in Australia". *Soil and Environment*. Vol.29, No.1, Pp. 61-68.2010.

[21] O. Kitani. (1999). *CIGR handbook of agricultural engineering*. American Society of Agricultural Engineers Publication. USA. Vol. 5. Pp.14-27.1999.

[22] S. Jungjit and A.A. Freitas. "A New Genetic Algorithm for Multi-Label Correlation-Based Feature Selection". *In Proceedings*. Pp. 285-290. 2015. Available from: https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2015-48.pdf

[23] R. Tiwari and M.P. Singh. "Correlation-based attribute selection using a genetic algorithm". *International Journal of Computer Applications*. Vol.4.No.8,Pp. 28–34.2010.

[24] Y.F. Jin, Z.Y. Yin, S.L. Shen and P.Y. Hicher. "Investigation into MOGA for identifying parameters of a critical-state-based sand model and parameters correlation by factor analysis". *Acta Geotechnica*.Vol.11.No.5,Pp.1131–1145.2016.

[25] S. Goswami, S. Saha, S. Chakravorty, A. Chakrabarti and B. Chakraborty. "A new evaluation measure for feature subset selection with a genetic algorithm". *International Journal of Intelligent Systems and Applications*. Vol. 7.No.10,Pp. 28-36. 2015.

[26] M. Kantardzic. *Data mining: concepts, models, methods, and algorithms*. 2nd edition. John Wiley & Sons. USA.Pp.140-144.2011.

[27] A.A. Freitas. *Data mining and knowledge discovery with evolutionary algorithms*. 2nd edition. Springer Science & Business Media. New York, USA.Pp.107-163.2013.

[28] M. Paruggia. "Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models". *Journal of the American Statistical Association*.Vol.101.No.473,Pp. 398-399.2012.

[29] F.Pianosi, K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson and T. Wagener. "Sensitivity analysis of environmental models: A systematic review with practical workflow". *Environmental Modelling and Software*.Vol. 79.No.3,Pp.214-232. 2016.

[30] N. Raheem Lahmod, J. Talib Alkooranee, A.A. Gatea Alshammary and J. Rodrigo-Comino. "Effect of Wheat Straw as a Cover Crop on the Chlorophyll, Seed, and Oilseed Yield of Trigonella foeunm graecum L under Water Deficiency and Weed Competition". *Plants*. Vol. 8. No.11, Pp. 503-519. 2019.

[31] S.H. Pishgar-Komleh, T. Zylowski, S. Rozakis and J. Kozyra. "Efficiency under different methods for incorporating undesirable outputs in an LCA+ DEA framework: A case study of winter wheat production in Poland". *Journal of environmental management*. Vol. 260. No.4, Pp.110-138.2020.

[32] R.P. Lollato, B.M. Figueiredo, J. S. Dhillon, D.B. Arnall and W.R. Raun. "Wheat grain yield and grain-nitrogen relationships as affected by N, P, and K fertilization: A synthesis of long-term experiments". *Field Crops Research*. Vol. 236.No. 4, 42-57.2019.