Optimization of sequence-dependent harvesting time at sugarcane farms using meta-heuristic methods

Negar Hafezi[1*], Mohammad Javad Sheikhdavood[1], Houshang Bahrami[1], Seyed Enayatallah Alavi[2]

(1. Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, 61357-83151, Iran; 2. Department of Computer Engineering, Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, 61357-83151, Iran)

Abstract: The framework of the present problem is based on predicting the recoverable sugar percentage, which has a fundamental role in sugarcane harvesting time. After forecasting this parameter, two objectives are pursued one tries to maximize the sugar production quantity according to the proper sugarcane age and variety, and the other tries to minimize the completion time of harvesting operations in a specified sequence of sugarcane farms. To estimate the recoverable sugar percentage parameter was used an Elman Neural Network (ENN).Then, the sequence-dependent harvesting time problem was formulated by a hybrid model called the Travelling Thief Problem (TTP). To solve this bi-objective problem has been used two meta-heuristic algorithms, called NSGA-II and SPEA2. Results indicate that the bi-objective optimization problem can be increased the sugar production quantity by 32.93% and be decreased the completion time of harvesting operations via finding optimal routes by 57.7% compared to the actual harvesting sequence. The statistical testing results show that the NSGA-II is superior to the SPEA2 in terms of achieving better convergence, generating more non-dominated solutions, improving the distribution of solutions, and shortening the running time.

Keywords: Sugarcane harvesting; Operations Research; Meta-heuristic Methods; Travelling Thief Problem

Citation: Hafezi, N., M. J. Sheikhdavoodi, H. Bahrami and S. E. Alavi. 2024. Optimization of sequence-dependent harvesting time at sugarcane farms using meta-heuristic methods. Agricultural Engineering International: CIGR Journal, 26(4): 217-233.

1 Introduction

Sugarcane (*Saccharum officinarum* L.) is a tropical perennial grass of the Gramineae family that is used for sugar production. The plant is also grown for biofuel production. In Iran, sugarcane is a valuable commodity. Recently, due to the increasing population growth rate and sugar consumption, plenty of sugar has been imported from abroad (Kaab et al., 2019). In 2020, sugarcane production in Iran was 7.83 million tons. Sugarcane production of Iran increased from 578,000 tons in 1971 to 7.83 million tons in 2020 and with an average annual growth rate of 7.66% (Knoema, 2022). In harvest planning, sugarcane should be cropped at the appropriate time and maturity to maximize sugar recovery from a specific variety and prevent cane deterioration and quality problems (Junqueira and Morabito, 2017). The harvesting date of a farm is not usually decided upon very far ahead, but many thumb rules are currently used to make scheduling decisions (Stray, 2010). Sugarcane harvesting at farms is usually done with regard to variety and crop age. The purpose of the operational planning of sugarcane harvesting is to determine the best harvest time and route so that the

Receive date: 2023-06-30 **Accepted date:** 2024-10-04

***Corresponding author: Negar Hafezi.** Ph.D graduate, Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, Ahvaz, 61357-83151, Iran. Email: hafezi.negar@gmail.com.

total cost of harvesting and transportation operations decreases, and sugar yield is maximized (Kittilertpaisan and Pathumnakul, 2015). Knowledge of operations research, which is one of the sub-branches of applied mathematics, is intended to facilitate decision-making based on scientific standards due to the limitations. In the agricultural sector, there are various constraints such as weather conditions, economic and social issues, so benefiting from this knowledge can help to solve problems and make the right decisions. Glen (1987) was the first to use the decision support tools in the agricultural sector (Glen, 1987; Jena and Aragão, 2011). In recent studies, the use of operations research has played a significant role in the scheduling problem of sugarcane harvesting (Munoz and Lee, 2019).

One of the issues of operations research in sugarcane scope can be referred to as value chain optimization, cutting and transportation crew scheduling, and prediction of sugarcane yield indicators (Jena and Poggi, 2013). For example, Lopez Milan et al. (2006) presented a mixed-integer programming (MIP) model to solve the problem of cost minimization of sugarcane transportation from fields to the mill for a working day. Florentino and Pato (2014) presented a bi-objective genetic heuristic for minimizing collection and transportation costs and maximizing energy balance obtained from residues of the sugarcane harvesting. The authors concluded that the heuristic approach could increase solution quality and improve computing time. Additionally, Sungnul et al. (2017) studied a multi-objective optimization model to find the optimal harvest time for maximizing revenue and minimizing gathering cost. Many problems in real-world include the optimization of several objectives simultaneously.

To obtain the optimum solution there will be a set of optimal trade-offs between the conflicting objectives where these optimum solutions are called Pareto-optimal solutions (Abido and Bakhashwain, 2005). The proper time to harvest each farm is at maturity peak of the sugarcane (Jena and Poggi, 2013). Sugarcane has a certain age to harvest and can only be cropped at the interval of its age (Jena and Poggi, 2013). While harvesting, the manager must provide the sequence of sugarcane farms to be cropped. Generally, sequencing refers to a scheduling problem, so determining the harvest sequence of sugarcane farms is commonly seen as a general type of scheduling problem as well (Jena and Aragão, 2011).

The present study proposes a different method for harvest scheduling that deals with the sequencing problem of sugarcane farms by creating a bi-objective optimization model. In this study, there are two objective functions in which one tries to maximize the sugar production quantity according to the proper sugarcane age and variety, but the other tries to minimize the completion time of harvesting operations in a specified sequence of sugarcane farms. Hence, to solve this bi-objective problem whose complexity is NP-hard, has been employed two meta-heuristic algorithms, called NSGA-II and SPEA2. The proposed model has been formulated as a Travelling Thief Problem (TTP). The TTP was created in 2013 by Bonyadi et al. (2013).

TTP consists of two sub-problems, namely, the Travelling Salesman Problem (TSP) and Knapsack Problem (KP), both of which are among the most well-known issues of operations research and combined optimization (Mei et al., 2015). In TTP, there are a number of cities and items, a thief must visit each city once and pick some items from the cities to place in a rental knapsack. Each item includes a certain weight and profit. The knapsack has a certain weight and should not be the total weight of the picked items more than the knapsack weight (Mei et al., 2014). In the end, the cost of renting the knapsack is taken into account to estimate the total profit. The goal of TTP is to maximize profit from items picked and to minimize travel time or rent paid for the knapsack (Mei et al., 2014). In TTP, the travel speed of the thief depends on the current weight of the knapsack, so the heavier the knapsack, the slower the thief's speed and more time is needed to complete the tour (Bonyadi et al., 2013).

2 Materials and methods

2.1 Study location

In Iran, the sugarcane harvest performs mechanically, based on the scheduling of the harvest fleet that is allocated to farms. The harvest fleet includes crew and harvesting equipment, including mechanical harvesters, tractors, transportation means, and other service tools that are transferred from one farm to another when performing operations. In this country, the harvesting time horizon is split into a six-month period from November to April. The best month for sugarcane harvesting is in February when the sugar content of cane is at the highest level. This study is executed in the Debal Khazaei Agro-Industry Company. This site is located 20 kilometers south of Ahvaz city and east of the Karun River in the Khuzestan province of Iran, which is at 48º 35′ 08″ E and 31º 08′ 07″ N.

Figure 1 shows the location of the studied area. Over the past 50 years, the average temperature and annual precipitation have been 25°C and 213 mm in this area. The total area of this company is about 13,557 hectares, which is 11,956 hectares, and the rest of the canal, road, building, and a sugar factory. The size of each farm is 25 hectares. For simplifying harvest fleet relocation, the farming lands of the Company are divided into two sites, namely first production (right) and second production (left), so that the farms' ID is named accordingly.

This Company has about 390 green farms in each harvesting season that must be harvested within six months. General data was collected from the offices of Agriculture and Applied studies of the Company. The collected data included the farm sizes in hectares, amount of sugarcane cut from each farm in ton, recoverable sugar percentage (RS %) at each farm, sugarcane yield in ton per hectare, sugarcane variety (early, medium, and late-maturing), harvesting age (which is usually 10-13 months in ratoon crop and 14-18 months in plant crop), and distance between farms in kilometer.

In this paper, the harvesting sequence optimization of sugarcane farms has been proposed for a set of 62 farms around the sugar factory with four harvest fleets that each fleet consists of 6 harvesting machines. In addition, a hybrid binary integer programming model was used to formulate the present problem. Then, the model was solved by two approximation methods, including NSGA-II and SPEA2. MATLAB (R2017b) software was used to test and evaluate the proposed algorithms, which was installed on a personal computer with 4 GB RAM and 2.5 GHz processor speed.

Figure 1 A view of the studied area in the south of Khuzestan province, Iran

2.2 Elman neural network

Elman neural network (ENN) is a type of recurrent neural network (RNN) and subcategory of neural networks (NNs), which are formed by a significant number of neuronal cell models according to certain rules (Ren et al., 2018). The ENN possesses an input layer, a hidden layer, and an output layer which are connected in a feed-forward manner (Cruse, 2006). Elman networks often have two layers with one feedback of the output layer to the hidden layer entrance. This connection helps the network to detect and generate momentary and time-dependent patterns. The main difference between these networks and ordinary two-layer networks is its feedback, which the delay in this feedback provides information about the previous step in the current step. We have tried a single-hidden layer of 20 neurons in the proposed network. The number of epochs and the goal value of network training error are set to 2000 and 0.00001, respectively. The learning rate value is set to 0.05, the maximum value of validation failure is set to 10, and the amount of the network training time is set to 100 seconds. The network training was performed with back-propagation algorithm which updates the weights and bias of the network based on Levenberg-Marquardt optimization method. In hidden and output layers are used transfer functions of the hyperbolic tangent sigmoid and linear to transmit neurons' information, respectively. The mathematical definition of the *tansig* transfer function expressed in Equation 1:

$$tansig(n) = {}^{2}\!/_{(1 + e^{-2n})} - 1 \tag{1}$$

Where, n is the number of net input elements.

The input layer is made up of three quality indicators, namely cane fiber, cane moisture, and Brix. Brix is the total soluble solids content in syrup. Usually, before actual harvesting, twenty sugarcane stalks are cut from five locations within a farm representing the whole farm's crop and transferred to the laboratory to determine the qualitative properties. The output layer comprises one unit representing recoverable sugar percentage as the dependent variable. Figure 2 illustrates a two-layer ENN.

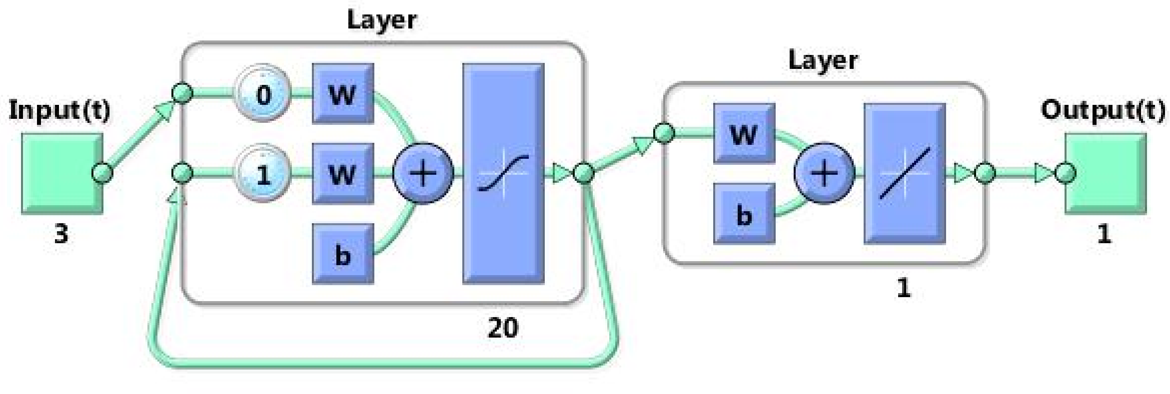


Figure 2 The layout of the ENN

The data spanned over the period 2011–2019 came from the sum of 2671 farms in the studied area which was used for four important sugarcane varieties CP57-614 (early maturing), SP70-1143 (early maturing), CP69-1062 (medium maturing), and CP48-103 (late maturing) in five classes, including plant crop and first to fourth ratoons. For supervised training of the NNs, the input and output samples are divided into training (60%), test (20%), and validation (20%) sets. In general, the data used in the proposed network needed normalization processing (Ren et al., 2018). In this model, the input data has been normalized into 0 and 1. The mathematical equation of normalization is expressed as follows:

$$X_n = {}^{(x - x_{min})}\!\Big/_{(x_{max} - x_{min})} \tag{2}$$

Where, x is the original data; X_n is the normalized data; x_{min} and x_{max} are the respective values of minimum and maximum original data. MATLAB (R2017b) software has been used to design and test the ENN model.

Two statistical metrics, namely, mean square error (MSE) and coefficient of determination (R^2) are used to assess proposed network performance. These statistical parameters are presented in Equations 3 and 4:

$$MSE = 1/_N \times \sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \quad (3)$$

$$R^2 = 1 - \sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \Big/ \sum_{i=1}^{N}(\bar{y} - y_i)^2 \quad (4)$$

Where, $\hat{y}_i$ is the *i-th* forecasted output, y_i is the *i-th* observed data, $\bar{y}$ is the mean of the observed data, and N is the number of observations. The R^2 is between 0 and 1. The "0" value indicates that the variables are incapable of predicting the dependent value, while the "1" value indicates the perfect prediction of the dependent variable by the independent variables (Nagelkerke, 1992).

2.3 Assumptions

To optimize the harvesting sequence problem the primary assumptions are taken into account, including all selected farms should be harvested at a given time horizon instance; One working day is equal to 14 hours, and time to use harvesters should not overcome the allowable working hours in a day; On the same variety, according to cane age, at first younger farms (plant crop) should be harvested, then older farms which are included first ratoon, second ratoon and so on; Crop harvesting priority belongs to the early maturing variety with high sugar content and then medium maturing and late maturing, respectively; For the sugarcane harvest, each farm is assigned only to one day and should not be harvested only part of it; The mechanical harvester is available when entering the farm until the harvesting operation is be completed at that farm; Weather and external disruptors are ignored at a given time horizon.

2.4 Mathematical model

The harvesting sequence problem was carried out to optimize the harvesting of 62 sugarcane farms within two weeks on February 5-19, 2020, when the farms' recoverable sugar percentage has reached maximum value according to harvesting age and sugarcane variety. The predicted amount of RS % is used as the main parameter in the sugarcane harvesting sequence problem. To optimize the farms harvesting sequence, we applied a bi-objective model that was formulated as a hybrid binary programming problem. The details of the mathematical model of the present problem are presented in Table 1. The model is expressed in Equations 5 - 10:

$$Maximize\ g(\bar{y}) = \sum_{l=1}^{p} RS_{lk} \cdot Q_{lk} \cdot y_{lk}, (k = 1, \ldots q) \quad (5)$$

$$Minimize\ f(\bar{x}, \bar{y}) = \sum_{i=1}^{n-1}(d_{x_i, x_{i+1}} / v_{x_i}) + (d_{x_n, x_1} / v_{x_n}) \quad (6)$$

Subject to

$$\sum_{l=1}^{p} Q_{lk} \cdot y_{lk} \leq C.T.H\ ,\ (k = 1, \ldots, q) \quad (7)$$

$$y_{lk} \in \{0, 1\}\ ,\ (l = 1, \ldots, p)\ , (k = 1, \ldots, q) \quad (8)$$

$$x \in P^n\ ,\ with\ \ x_1 = 1 \quad (9)$$

$$v_{x_i} = V_{max} - \left(\sum_{j=1}^{i} \left(\sum_{l=1}^{p} y_{lk}(x_j) \cdot Q_{lk} \right) \Big/ C.T.H \right) \cdot (V_{max} - V_{min}), (k = 1, \ldots, q) \quad (10)$$

Where, Equation 5 represents the first objective function which maximizes the sugar production (in ton) according to harvesting age and sugarcane variety. Equation 6 implies the second objective function, which minimizes the total completion time of sugarcane harvesting operations (in hour) at specified farms within a given time horizon. Constraint 7 ensures that the amount of sugarcane cut must not exceed the accumulated cutting capacity of harvesters on the same day. In Equation 8, $y_{lk}(x_j)$ is defined for harvesting each farm at the proper age, which returns "1" if the farm x_j having variety k can be cropped at age l and "0" otherwise. Equation 9 refers to all permutations of n farms in a specified variety for finding the harvesting sequence, which contains all farms exactly once are needed to calculate the second objective function. Finally, Equation 10 indicates that the harvesting operation speed is different, and it depends on sugarcane quantity, cutting capacity, and travel speed of the mechanical harvester.

Where, Equation 5 represents the first objective function which maximizes the sugar production (in ton) according to harvesting age and sugarcane vari-

ety. Equation 6 implies the second objective function, which minimizes the total completion time of sugarcane harvesting operations (in hour) at specified farms within a given time horizon. Constraint 7 ensures that the amount of sugarcane cut must not exceed the accumulated cutting capacity of harvesters on the same day. In Equation 8, $y_{lk}(x_j)$ is defined for harvesting each farm at the proper age, which returns "1" if the farm x_j having variety k can be cropped at age l and "0" otherwise. Equation 9 refers to all permutations of n farms in a specified variety for finding the harvesting sequence, which contains all farms exactly once are needed to calculate the second objective function. Finally, Equation 10 indicates that the harvesting operation speed is different, and it depends on sugarcane quantity, cutting capacity, and travel speed of the mechanical harvester.

Table 1 Indices, parameters, and decision variables in the harvesting sequence problem

Description	Type	Notation
Set of farms	Index	$i=\{1,\ldots,n\}$
Set of harvesting ages	Index	$l=\{1,\ldots,p\}$
Set of sugarcane varieties	Index	$k=\{1,\ldots,q\}$
Recoverable sugar predicted amount (in %) at farm *i* at age *l* for variety *k*	Parameter	RS_{lk}
Amount of sugarcane cut from farm *i* at age *l* for variety *k* (in ton)	Parameter	Q_{lk}
Cutting capacity of harvester (in ton/h)	Parameter	C
Working hours of harvesting machine per day (in h/day)	Parameter	T
Total number of harvesters available [1]	Parameter	H
The minimum speed of harvester during cutting operation (in km h^{-1})	Parameter	V_{min}
The maximum speed of harvester when moving between farms (in km h^{-1})	Parameter	V_{max}
Euclidean distance (in km) between farms $(i,i+1)$	Parameter	$d_{xi,xi+1}$
Decision vector: the farm is located in position *i*.	Real variable	x_i
Decision vector: if sugarcane variety *k*, at age *l* is proper for harvesting to take the value 1 and 0 otherwise.	Binary variable	y_{lk}
Decision variable: if farm x_j having variety *k* be harvested at the age *l* to take the value 1 and 0 otherwise.	Binary variable	$y_{lk}(x_j)$

Note: [1] The company has 24 active sugarcane harvesters of the Austoft 7000 series.

2.5 NSGA-II methodology

The non-dominated sorting genetic algorithm (NSGA) was presented by Srinivas and Deb (1994) in 1994. The NSGA has become a multi-objective algorithm by adding two essential operators to the simple genetic algorithm, which offers a set of the best solutions, known as the Pareto-optimal front, instead of finding the best solution. This algorithm is an efficient method for solving problems with several objective functions, but in order to select non-dominated individuals and computational complexity, it has some weaknesses. Hence, the second version of the NSGA algorithm called NSGA-II was introduced by Deb et al. (2000) in 2000. Generally, the steps of this algorithm include: (1) generating the initial population based on objectives and constraints of the problem, (2) evaluating the generated population according to the objective functions, (3) applying the non-dominated sorting method, (4) calculating the control parameter called crowding distance, (5) selecting the parent population for reproduction, (6) performing the recombination and mutation operators. Diagram of the NSGA-II is illustrated in Figure 3 (Left-side). The P_t and Q_t are parent and offspring populations at generation t, respectively. Both populations merge into a larger population called R_t with size $2N$. The F_1 is a set of the best non-dominated solutions from R_t. The F_2 is the second set of the best non-dominated solutions and so forth. As seen in this figure, the number of P_{t+1} from $P_t \cup Q_t$ members be selected based on their ranking, and the rest of them be removed to keep the number of the original population steady. The NSGA-II uses crowding distance to obtain a more uniform Pareto front than other algorithms and estimate the density of solutions enclosing a special solution i. The value of crowding distance for the solution I_j is calculated as follows:

$$\mathrm{d}_{I_j^m} = \mathrm{d}_{I_j^m} + \left(f_m^{(I_{j+1}^m)} - f_m^{(I_{j-1}^m)}\right) \Big/ \left(f_m^{max} - f_m^{min}\right) \quad (11)$$

Where, firstly, a solution is specified (i), then the difference in the objective function value of the next solution ($i+1$) from the previous solution ($i-1$) is divided by the difference of the maximum (f_m^{max}) and minimum (f_m^{min}) values of the objective function (Figure 3 (right-side)); this obtained ratio indicates the crowding distance (d_i). The m and j are the number of objective functions and objective function index, respectively (Deb, 2001).

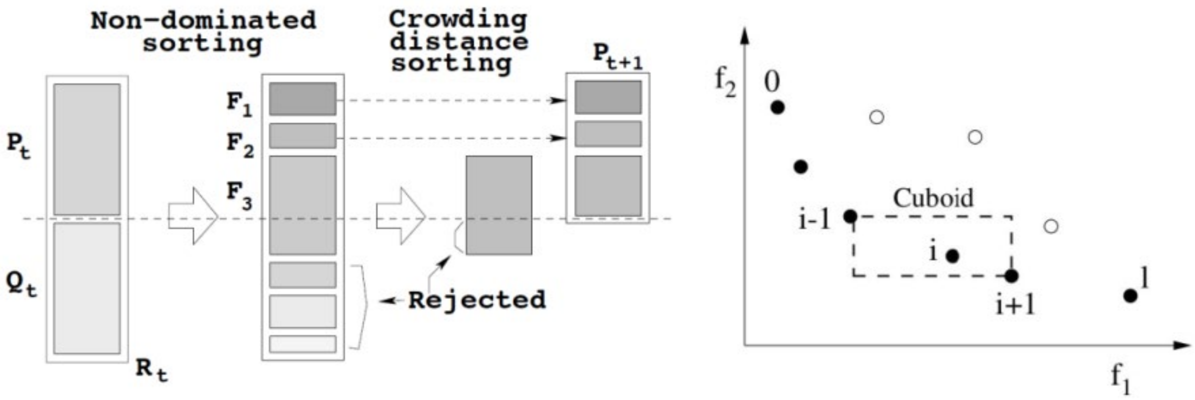


Figure 3 Schema of the NSGA-II method (left) and the crowding distance calculation (right) (Deb, 2001)

First, the solutions should be sorted using the rank and then by the crowding distance. Afterward, two parents are selected (Psychas et al., 2015). The parent solution is randomly selected in two ways. First, if the ranks are unequal, the solution is chosen that has a lower rank ($r_i<r_j$), then if the ranks are equal, the solution is chosen that has a greater crowding distance ($d_i>d_j$). The adjustable parameters of population size, number of generations, recombination and mutation probability, and mutation rate are tuned at 60, 100, 0.7, 0.3, and 0.02, respectively, to design the NSGA-II. This algorithm uses the binary tournament selection method, single-point recombination, and swap mutation operators.

2.6 SPEA2 methodology

The strength pareto evolutionary algorithm (SPEA) was introduced by Zitzler and Thiele (1999) in 1999. There is also a revised version of SPEA called SPEA2 (Zitzler et al., 2001). The main difference between SPEA and SPEA2 is population selection, which in SPEA is the choice based on individuals, while in SPEA2 the choice is based on region crowdedness. Basically, in SPEA2, there is a criterion to control the order of solutions, which is done by creating hyper-rectangle and hyper-grid in the objective function space. In the SPEA2 structure, a new approach is used to define fitness in which both the set of dominated solutions and non-dominated solutions have an effect, which implements the secondary factor, namely, the nearest neighbor approach to control the distribution and density estimation. In fact, it is a data distribution-based approach to remove additional solutions. The main loop of SPEA2 includes the number of populations, archive size, and the maximum number of generations as input and the set of non-dominated solutions as the output. Methods of the binary tournament, single-point, and normal distribution are used for selection, recombination, and mutation operators, respectively. The parameters of population size, number of generations, archive size, k-nearest neighbors, and recombination and mutation probability are adjusted 60, 100, 40, 10, 0.7, and 0.3, respectively, to design SPEA2.

2.7 Performance metrics

As regards, meta-heuristic approaches are approximate algorithms to solve optimization problems and have a random nature. Solving a problem in different ways may lead to various solutions, so evaluating algorithms and selecting the appropriate algorithm using various metrics are considered by researchers. But the convergence of Pareto solutions and providing density and diversity among the set of solutions are two distinct goals and somewhat conflicting in multi-objective evolutionary algorithms (Javid, 2021). Hence, there is no absolute criterion that measures the performance of algorithms. Thus, in this study, the following five metrics have been introduced as metrics of algorithms' performance evaluation. These metrics consist of the number of pareto solutions (NOP), generational distance (GD), spacing (S), maximum spread (MS), and the time index for running time of algorithms.

2.7.1 Number of pareto solutions metric

The NOP counts the number of non-dominated solutions by the algorithm, and the higher the number, the better the algorithm's performance.

2.7.2 Generational distance metric

The GD metric finds a mean distance of the solutions of Q from P^*. It is assumed that P^* is a known Pareto-optimal set (Van Veldhuizen, 1999; Gao et al., 2010). The GD metric can be calculated by Equation 12:

$$GD = \sqrt{\sum_{i=1}^{|Q|} {d_i}^2} \Big/ |Q| \tag{12}$$

For two objective problems, the parameter d_i is the minimum Euclidean distance between the solution $i \in Q$ and the nearest member of P^* (Gao et al., 2010), as follows:

$$d_i = \min_{k=1}^{|P^*|} \sqrt{\sum_{m=1}^{M} \left(f_m^i - f_m^{*(k)}\right)^2} \tag{13}$$

Where, $f_m^{*(k)}$ is the value of *m-th* objective function of the *k-th* member of P^* (Gao et al., 2010). In general, an algorithm with a smaller GD value is better.

2.7.3 Spacing metric

Schott (1995) proposed a metric for calculating the relative distance between successive points in set of the Pareto solutions (Gao et al., 2010; Schott, 1995) as follows:

$$S = \sqrt{1/|Q| \times \sum_{i=1}^{|Q|} (d_i - \bar{d})^2}$$

$$d_i = min_{k \in Q \wedge k \neq i} \sum_{m=1}^{M} \left| f_m^i - f_m^k \right|$$

$$\bar{d} = \sum_{i=1}^{|Q|} d_i \Big/ |Q| \tag{14}$$

Where, d_i is the value of the minimum sum of the absolute difference between *i-th* solution and other solutions of the objective function in set of the Pareto solutions. The $\bar{d}$ is the average value of the parameter d_i. Thus, an algorithm with a smaller spacing (S) is better efficient.

2.7.4 Maximum spread metric

The MS metric measures the diagonal length of the hyper-box made with the values of the furthest points of the objective function in set of the Pareto solutions (Zitzler, 1999; Sankararao and Gupta, 2006), as follows:

$$MS = \sqrt{\sum_{m=1}^{M} (max_{i=1:|Q|} f_m^i - min_{i=1:|Q|} f_m^i)^2} \tag{15}$$

The higher MS shows the larger area of the true Pareto front, which is covered by the obtained approximation front (Yen and He, 2014).

2.7.5 Algorithm running time metric

Running time is one of the most important metrics in evaluating algorithm efficiency. An algorithm that achieves the optimal solution in less time is more efficient (Salmasnia et al., 2018). To evaluate proposed algorithms' performance, we used statistical testing called independent sample t-test, which is done using statistical software of IBM SPSS 24.

3 Results and discussion

Figure 4 can be seen the network training process. This figure shows that the MSE diagram has the same behavior in all three data sets of the network. The best network's performance has been obtained at repetition 4 with error values 0.0025, 0.0023, and 0.0047 for training, validation, and test data sets, respectively. Besides, the network training has been stopped after ten consecutive increases in validation error and against 14 epochs.

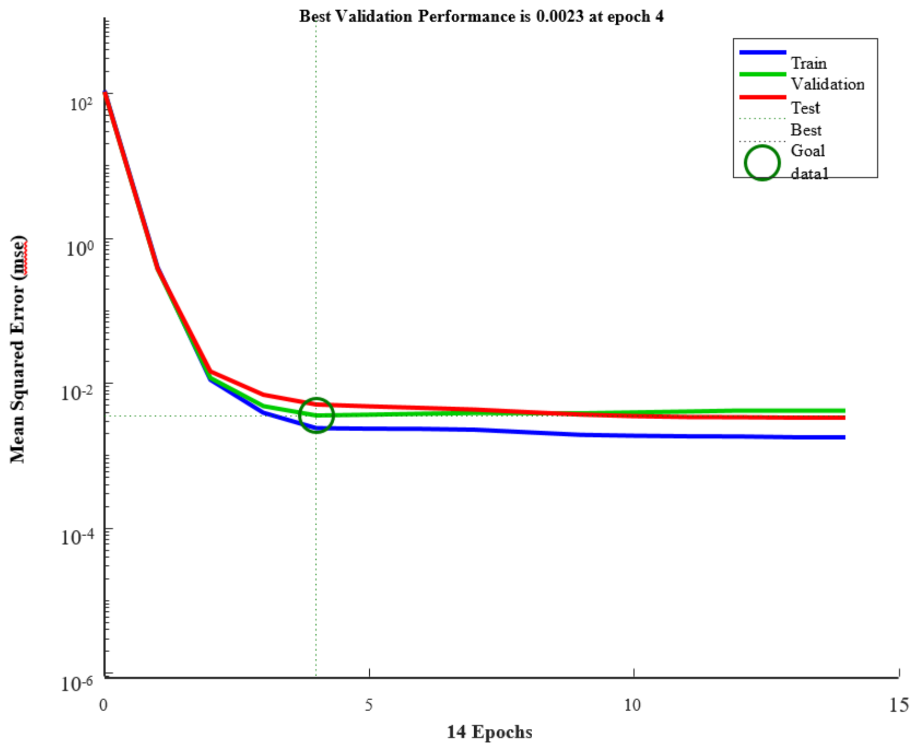


Figure 4 The training process of the designed ENN

Figure 5 (left side) shows the regression scatter plot in 3-20-1 topology for the network test data set. The network output, which is the percentage of recoverable sugar, is well matched on the target vector. The regression value is equal to 0.9995 in test data set which shows a desirable fitness between the actual data and predicted values by the network. The ENN model can well predict the output value with its target. Figure 5 (right side) illustrates that the predicted value of RS % for all three data sets of the designed network is approximately 10.6%. In addition, the minimum value is about 8%, and the maximum value is about 12.8%. Besides, results showed that the estimated values of RS % for the variety CP57-614 were 10.6% (PC), 10.7% (R01), 11% (R02), 10.7% (R03) and 10.4% (R04). For the variety SP70-1143, the values were 10% (PC), 10.3% (R01), 10.6% (R02), 10.8% (R03) and 10.9% (R04). Moreover, for the variety CP69-1062, recoverable sugar respective values were 10.2% (PC), 10.9% (R01), 10.5% (R02), 10.5% (R03) and 10.4% (R04). In relation to CP48-103, the obtained values were 10.4% (PC), 10.9% (R01), 10.8% (R02), 10.7% (R03) and 10.6% (R04). PC is termed as a plant crop that harvested for the first time; R01 is termed a crop harvested for the second time and so on.

Figure 6 demonstrates the optimal harvest routes of the selected farms in terms of the highest RS% in each sugarcane variety. The traveled distance between farms with the same variety in the actual harvesting sequence was equal to 43.23 km (CP57-614), 55.08 km (SP70-1143), 165.54 km (CP69-1062), and 53.15 km (CP48-103). While using the proposed optimization model, the traveled distance values for the above four varieties will be 23.21, 25.39, 49.86, and 28.10 km, respectively. This result shows, on average, a 60% reduction (47.60 km) in the traveled distance by mechanical harvesters. As seen in Figure 6, each farm has been harvested only once and completely at a 14-day time horizon instance. In this figure, “R” is an abbreviation of “Right” which means the first production, and “L” is an abbreviation of “Left” which means the second production that used to create the farm’s ID.

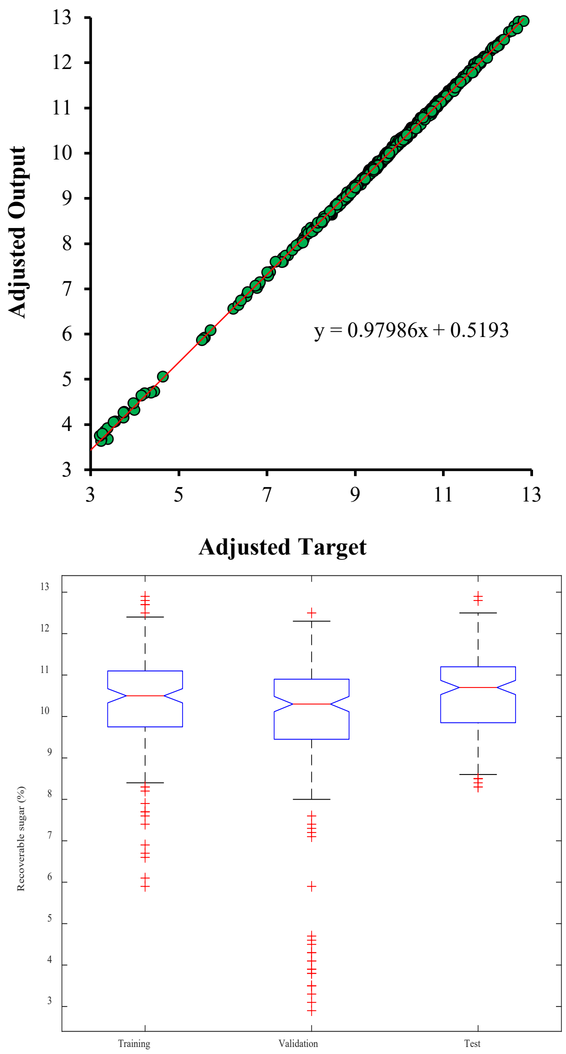
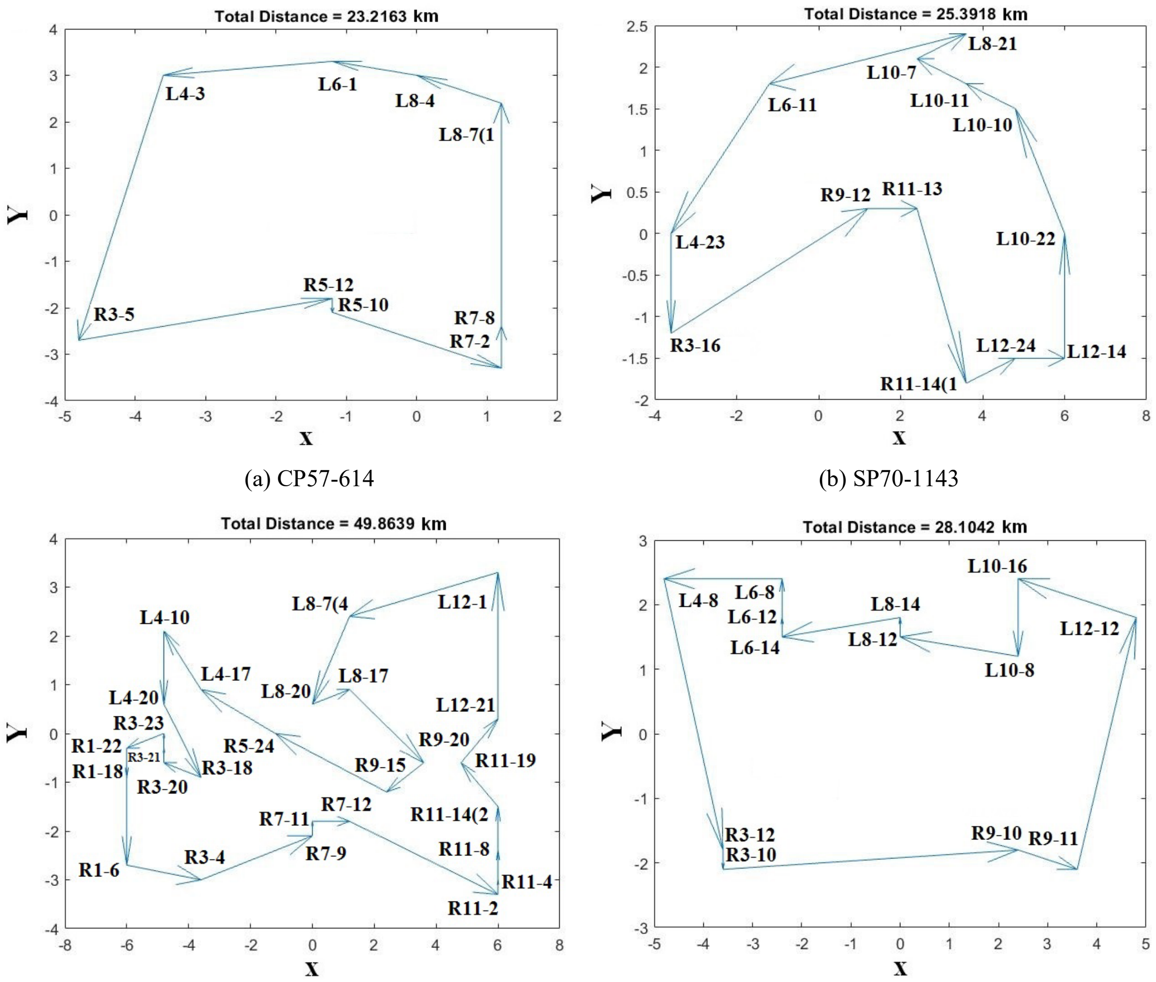


Figure 5 Showing of scatter plot (left) and box plot (right) of regression and predicted values of RS% by the ENN

Table 2 shows the optimal harvesting age in various crop classes of sugarcane varieties. As shown in this table, the proper age to harvest the variety CP57-614 in plant crop class is 14 months after planting which has the highest RS%. Then for the first to third ratoons is 11 months, and the fourth ratoon is 12 months after the previous harvest date. Afterward, in the variety SP70-1143, the proper age to harvest the farms is 15 months in plant crop class, and also ideal age for respective ratoon crops is 13 and 12 months. Also, for two varieties CP69-1062 and CP48-103, the proper harvesting age is obtained at 12 months in their different crop classes. Khan et al. (2017) found that the highest RS% in three genotypes early (CP69 1059), mid-early (B 4906, B 60267), and medium (B59 212) maturity were seen at 12 months of age. Furthermore, Hagos et al. (2014) found that the highest amount of sugar yield was obtained at harvest ages of 12 and 14 months for mid, mid-late and late-maturing cane varieties. However, the optimal sugar yield at the age of 12 months had the highest return rate.

Figure 7 illustrates the Pareto fronts of the objective functions in the present problem using NSGA-II and SPEA2 algorithms. Since these objectives are conflicting, thus improvement of one objective function comes at the expense of another objective function. According to this figure, the dominated solutions space of the problem is located in the down part of the Pareto fronts; and the ideal point of the non-dominated solutions is located in the top-left corner of the Pareto fronts, too. Both algorithms have performed well in terms of conflicting objectives to optimize the harvesting sequence problem. These objectives included decreasing the completion time of harvesting operations due to a reduction in the traveled distance between sugarcane farms for mechanical harvesters and increasing the production of white sugar within a given time horizon. As seen in Figure 7, the NSGA-II algorithm is more efficient than the SPEA2 algorithm in terms of improving convergence towards true Pareto solutions, uniform distribution of solutions on the Pareto front, and producing more non-dominated solutions. The amount of sugar produced in the variety CP69-1062 was obtained 3997.6 tons by both algorithms. This variety possesses area coverage of 45.3% in the Debal Khazaei Agro-Industry Company. Moreover, variety CP48-103 (equal to 3718.6 tons by NSGA-II and 3670.9 tons by SPEA2) with 19.5% area coverage and variety CP57-614 (equal to 3620.9 tons by NSGA-II and 3540.4 tons by SPEA2) with 13.4% area coverage have had the maximum amount of sugar production. The last two varieties have the highest sugar yield compared to their low area coverage. Finally, the sugar production quantity in variety SP70-1143 with 19.1% area coverage obtained 2969.2, and 2415.3 tons using NSGA-II and SPEA2, respectively. Moreover, by NSGA-II, the shortest completion time of harvesting operations lasted 6, 10.41, 64.75, and 7 hours for the above four varieties, which equals 6.3 working days.



(a) CP57-614 (b) SP70-1143

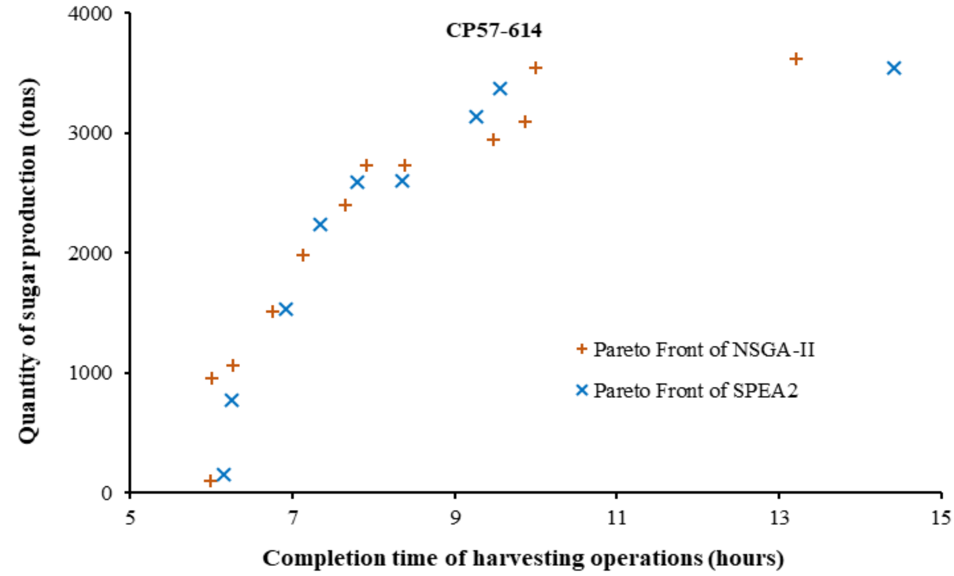
(c) CP69-1062 (d) CP48-103

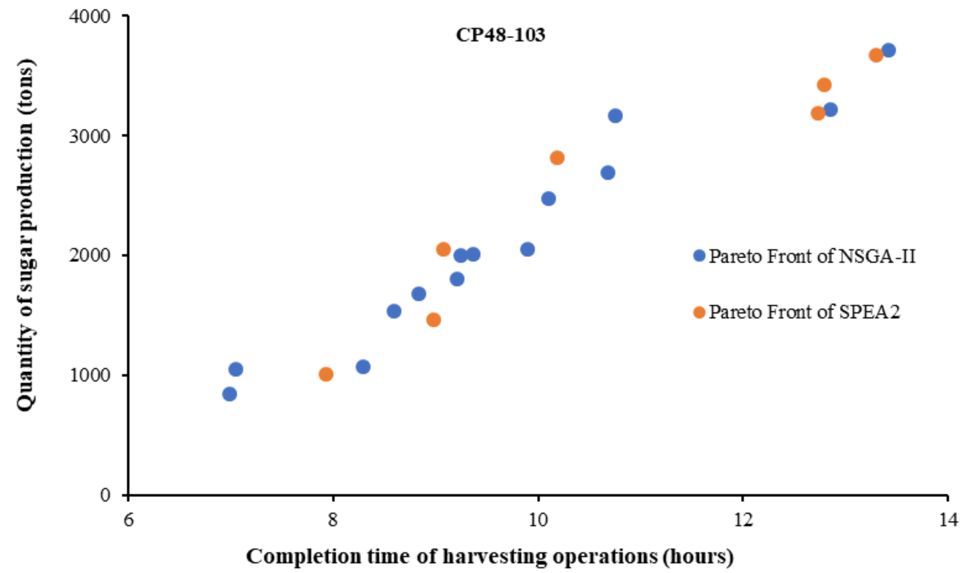
Figure 6 The optimal harvesting routes and obtained distances for the studied varieties

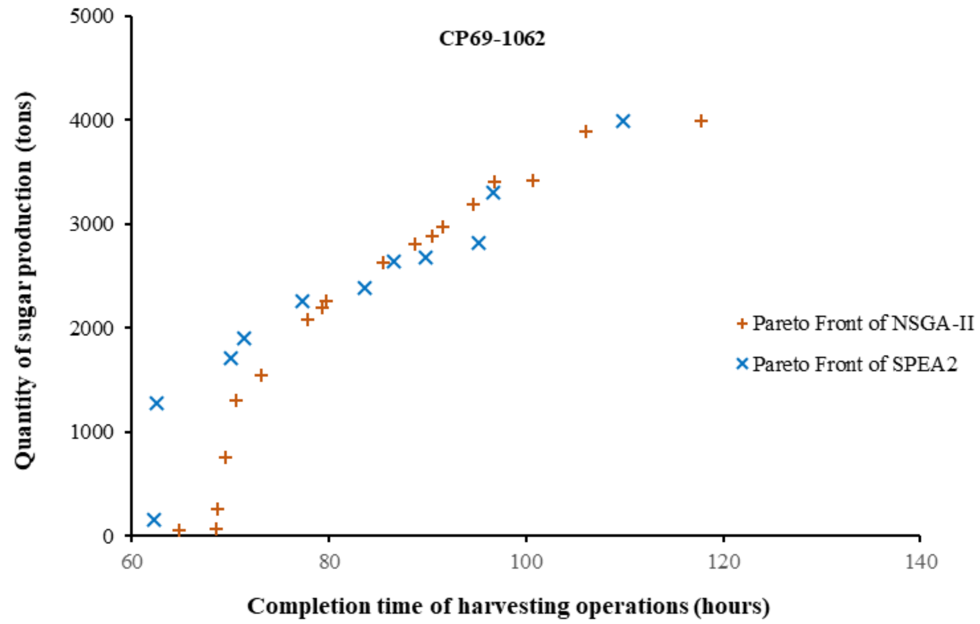
Table 2 The proper harvesting age according to the highest RS% in each sugarcane variety and crop class

	CP57-614					SP70-1143				
Harvesting age (months)	PC	R01	R02	R03	R04	PC	R01	R02	R03	R04
11		10.8	10.3	10.4			10.3	10.3	10.5	10.1
12		10.7	10.3	10	10.2		10.2	10.4	10.8	10.7
13		10.1					10.7	10.7		
14	10.6									
15	10.4					9.9				
16						9.7				
17						9.4				
18										

	CP69-1062					CP48-103				
Harvesting age (months)	PC	R01	R02	R03	R04	PC	R01	R02	R03	R04
11		10.6	10.6	10.6	10.6		10.3	10.3	10.4	
12		10.9	10.8	10.7	11.1		10.7	10.9	10.6	10.3
13							10.5		10.5	10.5
14	10.1									
15	10.5									
16										
17	10.3					10.3				
18	9.7					10				


CP57-614
Quantity of sugar production (tons)
Completion time of harvesting operations (hours)
Pareto Front of NSGA-II
Pareto Front of SPEA2
0
1000
2000
3000
4000
5
7
9
11
13
15


CP48-103
Quantity of sugar production (tons)
Completion time of harvesting operations (hours)
Pareto Front of NSGA-II
Pareto Front of SPEA2
0
1000
2000
3000
4000
6
8
10
12
14


CP69-1062
Quantity of sugar production (tons)
Completion time of harvesting operations (hours)
Pareto Front of NSGA-II
Pareto Front of SPEA2
0
1000
2000
3000
4000
5000
60
80
100
120
140

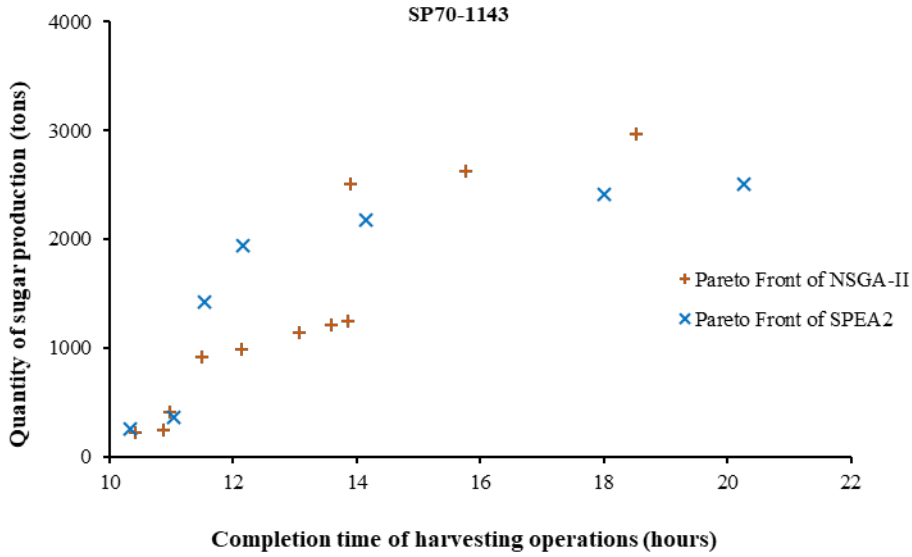
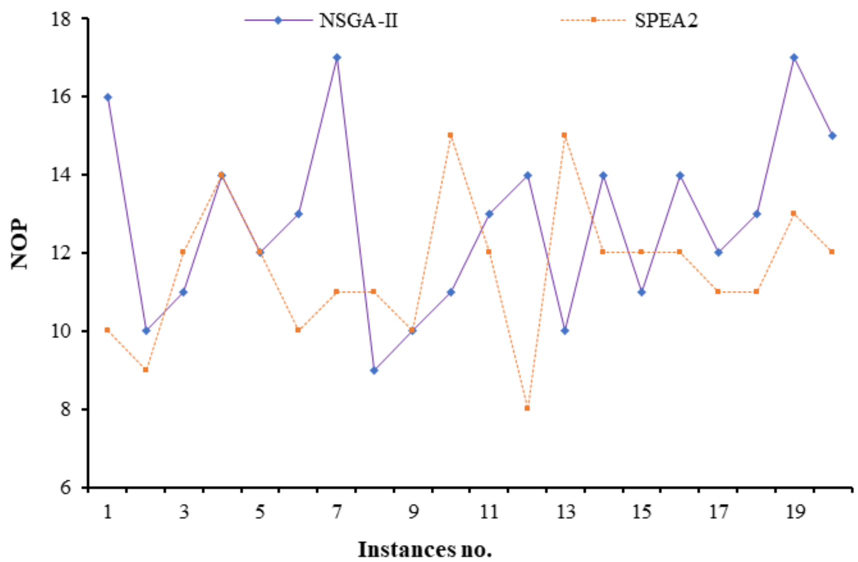
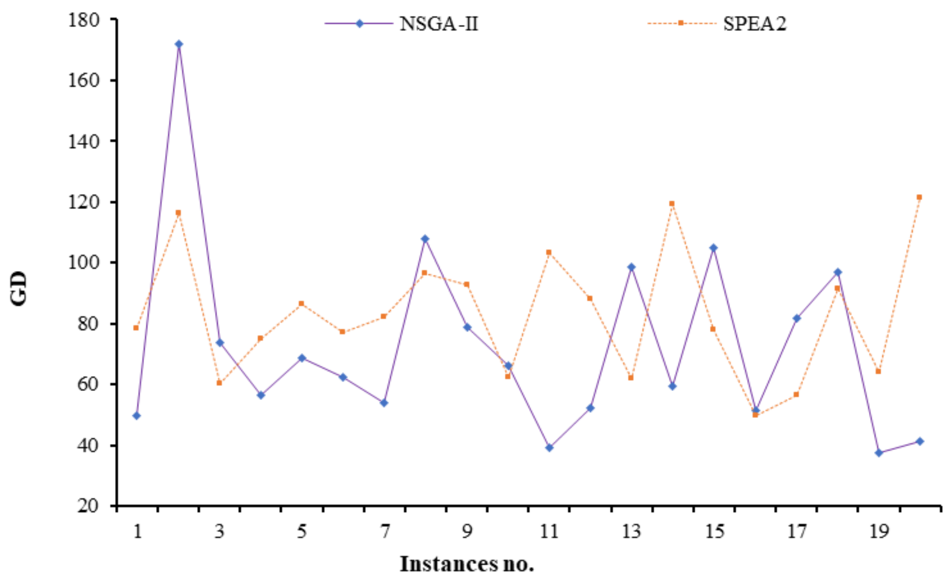


Figure 7 Pareto solutions of NSGA-II and SPEA2 for the studied varieties





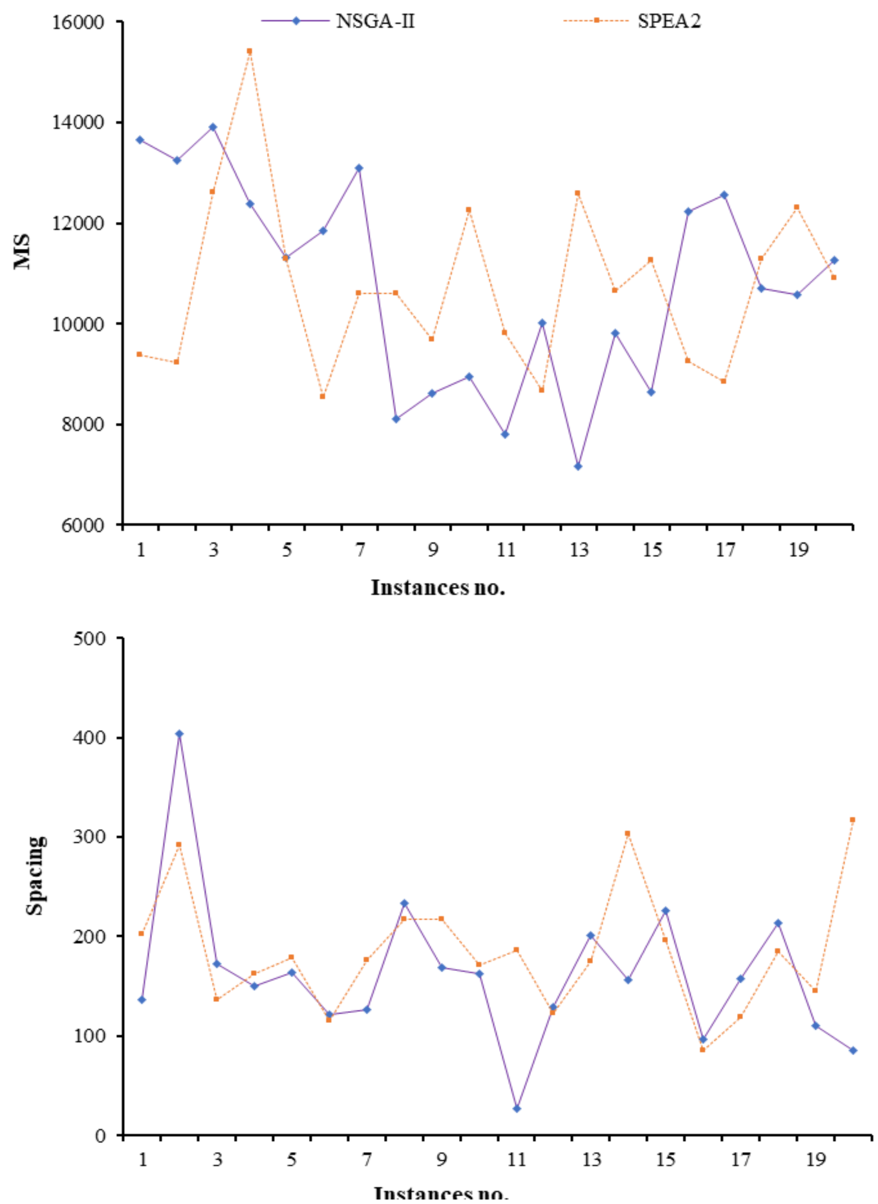


Figure 8 Algorithms comparison in terms of the proposed metrics

Table 3 shows the results of the bi-objective problem for variety CP69-1062 having 11 and 18 solutions by proposed algorithms. This table illustrates non-dominated solutions belonging to the Pareto rank "1". For statistical comparison, both algorithms were tested 20 times and their results can be seen in Figure 8.

Figure 8 shows that the NOP by NSGA-II is higher than SPEA2, so this algorithm is more efficient. Also, this figure demonstrates that the obtained values of the GD by NSGA-II are less than SPEA2; the GD metric is used for convergence measurement towards the true Pareto front. As shown in Figure 8, the obtained values of the MS by NSGA-II are slightly more, but not able to show a significantly better result than SPEA2. Finally, this figure shows that the spacing (S) values by NSGA-II are less than SPEA2; accordingly, this algorithm produces a better result than another. This metric can create a uniform distribution between Pareto solutions. In addition, the running time on average for the NSGA-II was 16.66 seconds and for the SPEA2 was 23.94 seconds. Being less this metric is important for decision-makers.

Table 4 shows the results of sensitivity analysis of the proposed algorithms' performance. It is assumed that the results compared to the T-test have a normal distribution. In Table 4, there are two P-values which the first one implies homogeneity of variances and the second one indicates the difference between variances. According to Table 4, if Levene's test shows that two algorithms have homogeneous variance, we can use the first row of the results. But if Levene's test rejects the homogeneity of variance hypothesis, the second row of results is used. According to the Sig column of Table 4, since the significance values of all metrics except running time metric are more than our considered significant level (α=0.05), the hypothesis about equality of variances of the two algorithms is not rejected. It is assumed that two algorithms have a significant difference in performance. As seen from Table 4, the P-values of the spacing (S), GD, NOP, and running time metrics are less than significant level, so the null hypothesis (i.e., equality of means) can be rejected. Therefore, it can be concluded that the two algorithms are different. Then by testing the average of metrics which is presented in Mean column of Table 4, it can be expressed that NSGA-II performs better than SPEA2 in terms of GD, spacing (S), NOP, and the running time metrics. Regarding the MS metric, the null hypothesis cannot be rejected, so two algorithms have been aligned in this case. However, NSGA-II performs slightly better compared to SPEA2. This result conforms to Bandyopadhyay and Bhattacharya (2013), which used NSGA-II for a parallel machine scheduling problem. Also, it agrees with Zhang et al. (2019), which optimized routing and sequencing synchronously for job shop scheduling problems in dynamic environment.

Table 3 The Pareto solutions obtained for variety CP69-1062 by NSGA-II and SPEA2

NSGA-II algorithm						SPEA2 algorithm		
Solution no.	Sugar production quantity (tons)	Completion time of harvesting operations (hours)	Solution no.	Sugar production quantity (tons)	Completion time of harvesting operations (hours)	Solution no.	Sugar production quantity (tons)	Completion time of harvesting operations (hours)
1	55.15	64.75	12	257.01	68.73	1	150.35	62.24
2	3997.57	117.72	13	3412.60	96.77	2	1273.08	62.50
3	3890.15	106.03	14	2810.03	88.73	3	3997.57	109.78
4	1548.15	73.10	15	2973.33	91.58	4	1715.29	70.01
5	2632.09	85.47	16	63.60	68.57	5	1899.26	71.39
6	752.64	69.50	17	2881.35	90.51	6	2264.02	77.28
7	3413.95	100.65	18	2199.88	79.30	7	3306.53	96.65
8	2076.64	77.81	-	-	-	8	2646.18	86.58
9	1308.85	70.59	-	-	-	9	2680.31	89.73
10	2264.02	79.67	-	-	-	10	2389.70	83.60
11	3186.11	94.63	-	-	-	11	2813.95	95.13

Table 4 Statistical assumption test results for comparing algorithms

		Levene's test for Equality of Variances	T-test for Equality of Means					
Metric	Assumption	Sig.	P-value	95% Confidence Interval of the Difference: Lower	95% Confidence Interval of the Difference: Upper	Algorithm	N	Mean
S	Variances are equal	0.275	0.008	-81.18	-13.19	NSGA-II	20	138.4
	Variances are unequal		0.008	-81.33	-13.04	SPEA2	20	185.6
GD	Variances are equal	0.738	0.001	-35.02	-10.32	NSGA-II	20	61.4
	Variances are unequal		0.001	-35.02	-10.31	SPEA2	20	84.1
MS	Variances are equal	0.353	0.471	-730.32	1550.4	NSGA-II	20	11082
	Variances are unequal		0.471	-731.03	1551.1	SPEA2	20	10672
NOP	Variances are equal	0.109	0.001	0.90	3.39	NSGA-II	20	14.6
	Variances are unequal		0.001	0.90	3.40	SPEA2	20	11.4
Time	Variances are equal	0.048	0.000	-7.86	-6.72	NSGA-II	20	16.7
	Variances are unequal		0.000	-7.86	-6.72	SPEA2	20	23.9

4 Conclusions

To optimize the sugarcane farms harvesting sequence, we needed to determine the proper harvesting age for different sugarcane varieties and find the optimal harvest routes for mechanical harvesters. Subsequently, the model was solved by two multi-objective algorithms, called NSGA-II and SPEA2. The algorithms' performance was tested in terms of the studied metrics. Results showed that the features of NSGA-II were more than SPEA2. These features include the highest NOP and the lowest values of spacing (S), GD, and running time. It can be expressed that NSGA-II efficiency was proved for 62 large-scale sugarcane farms in the present problem. Moreover, results indicated that the maximum amount of sugar production according to area coverage belonged to variety CP57-614, which is the type of early maturing variety. The shortest time lasted approximately 88 hours to perform harvesting operations. The proper age to harvest the ratoon crops in variety CP57-614 was 11 months when recoverable sugar percent reaches its maximum value. Moreover, for ratoon crops of variety SP70-1143, it was 12-13 months. And finally, for ratoon crops in varieties CP69-1062 and CP48-103, it was 12 months.

Acknowledgments

The authors would like to thank Shahid Chamran University of Ahvaz for financial support. Also, they want to express gratitude to the Debal Khazaei Agro-Industry Company that has provided the required information.

References

Abido, M. A., and J. M. Bakhashwain. 2005. Optimal VAR dispatch using a multi-objective evolutionary algorithm. *International Journal of Electrical Power & Energy Systems*, 27(1): 13-20.

Bandyopadhyay, S., and R. Bhattacharya. 2013. Solving multi-objective parallel machine scheduling problem by a modified NSGA-II. *Applied Mathematical Modelling*, 37(10-11): 6718-6729.

Bonyadi, M. R., Z. Michalewicz, and L. Barone. 2013. The travelling thief problem: The first step in the transition from theoretical problems to realistic problems. In *2013 IEEE Congress on Evolutionary Computation*, 1037-1044. Cancun, Mexico, 20-23 June 2013.

Cruse, H. 2006. *Neural Networks as Cybernetic Systems*. 2nd ed. Bielefeld, Germany: Brains, Minds and Media.

Deb, K., Agrawal, S., Pratab, A. and T. Meyarivan. 2000. A Fast Elitist Non- Dominated Sorting Genetic Algorithm NSGA-II. In *International Conference on Parallel Problem Solving from Nature*, eds. M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo and H. P. Schwefel, 849-858. Paris, France: Springer.

Deb, K. 2001. *Multi-objective Optimization Using Evolutionary Algorithms*. New York, NY: Wiley.

Florentino, H. O., and M. V. Pato. 2014. A bi-objective genetic approach for the selection of sugarcane varieties to comply with environmental and economic requirements. *Journal of the Operational Research Society*, 65(6): 842-854.

Gao, J., L. Fang, and J. Wang. 2010. A weight-based multi-objective immune algorithm: WBMOIA. *Engineering Optimization*, 42(8): 719-745.

Glen, J. J. 1987. Mathematical models in farm planning: A survey. *Operational Research*, 35(5): 641-666.

Hagos, H., L. Mengistu, and Y. Mequanint. 2014. Determining optimum harvest age of sugarcane varieties on the newly establishing sugar project in the tropical areas of Tendaho, Ethiopia. *Advances in Crop Science and Technology*, 2(5): 156-159.

Javid, Y. 2021. A bi-objective mathematical model to determine risk-based inspection programs. *Process Safety and Environment Protection*, 146(2): 893-904.

Jena, S., and M. P. de Aragão. 2011. A mixed integer programming approach for sugar cane cultivation and harvest planning. In *Computational Methods for Agricultural Research: Advances and Applications*, eds. H. do Prado, A. Barreto Luiz, and H. Filho. 144-174. New York: IGI Global.

Jena, S. D., and M. Poggi. 2013. Harvest planning in the Brazilian sugarcane industry via mixed integer programming. *European Journal of Operational Research*, 230(2): 374-384.

Junqueira, R. A. R., and R. Morabito. 2017. Optimization approaches for sugarcane harvest front programming and scheduling. *Gestão & Produção*, 24(2): 407-422.

Kaab, A., M. Sharifi, H. Mobli, A. Nabavi-Pelesaraei, and K. W. Chau. 2019. Combined life cycle assessment and artificial intelligence for prediction of output energy and environmental impacts of sugarcane production. *Science of the Total Environment*, 664: 1005-1019.

Khan, A. Q., K. A. Tadesse, and B. L. Robe. 2017. Quality characteristics and their relation with flowering in sugarcane (*Saccharum spp.* hybrid) in Ethiopia. *International Journal of Plant Breeding and Genetics*, 11: 84-91.

Kittilertpaisan, K., and S. Pathumnakul. 2015. Sugarcane harvester planning based on the Vehicle Routing Problem with Time Window (VRPTW) approach. In *Toward Sustainable Operations of Supply Chain and Logistics Systems*, eds. V. Kachitvichyanukul, K. Sethanan, and P. Golinska-Dawson, ch. 23, 335-344. Switzerland: Springer International Publishing Switzerland, Cham.

Knoema. 2020. Iran - Sugar cane production quantity. Available at: https://knoema.com/atlas/Iran/topics/Agriculture/Crops-Production-Quantity-tonnes/Sugar-cane-production/. Accessed 22 March 2022.

Lopez Milan, E., S. M. Fernandez, and L. M. P. Aragones. 2006. Sugar cane transportation in Cuba, a case study. *European Journal of Operational Research*, 174(1): 374-386.

Mei, Y., X. Li, and X. Yao. 2014. Improving efficiency of heuristics for the large scale traveling thief problem. In *Simulated Evolution and Learning 10th International Conference*, eds. G. Dick, W. N. Browne, P. Whigham, M. Zhang, L. T. Bui, H. Ishibuchi, Y. Jin, X. Li, Y. Shi, P. Singh, K. C. Tan and K. Tang, 631-643. Dunedin, New Zealand: Springer.

Mei, Y., X. Li, F. Salim, and X. Yao. 2015. Heuristic evolution with genetic programming for traveling thief problem. In *2015 IEEE Congress on Evolutionary Computation*, 2753-2760. Sendai, Japan, 25-28 May 2015.

Munoz, F., and S. Lee. 2019. Sugarcane harvest scheduling using a distributed control approach. *Procedia Manufacturing*, 39: 1149-1157.

Nagelkerke, N. J. D. 1992. *Maximum Likelihood Estimation of functional Relationships*. New York, NY: Springer.

Psychas, I. D., E. Delimpasi, and Y. Marinakis. 2015. Hybrid evolutionary algorithms for the multi-objective traveling salesman problem. *Expert Systems with Applications*, 42(22): 8956-8970.

Ren, G., Y. Cao, S. Wen, T. Huang, and Z. Zeng. 2018. A Modified Elman Neural Network with a New Learning Rate Scheme. *Neurocomputing*, 286: 11-18.

Salmasnia, A., S. Hasannejad, and H. Mokhtari. 2018. A multi-objective optimization for brush monofilament tufting process design. *Journal of Computational Design and Engineering*, 5(1): 120-136.

Sankararao, B., and S. K. Gupta. 2006. Multi-objective optimization of the dynamic operation of an industrial steam reformer using the jumping gene adaptations of simulated annealing. *Asia-Pacific Journal of Chemical Engineering*, 1(1-2): 21-31.

Schott, J. R. 1995. Fault Tolerant Design Using Single and Multi-criteria Genetic Algorithm Optimization. M.S. thesis. Cambridge Univ., Massachusetts.

Srinivas, N., and K. Deb. 1994. Multi-objective optimization using non-dominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3): 221-248.

Stray, B. J. 2010. Tactical sugarcane harvest scheduling. Ph.D. diss., Stellenbosch Univ., South Africa.

Sungnul, S., W. Pornprakun, S. Prasattong, and C. Baitiang. 2017. Multi-objective mathematical model for the optimal time to harvest sugarcane. *Applied Mathematics*, 8(3): 329-343.

Van Veldhuizen, D. A. 1999. Multi-objective Evolutionary Algorithms: Classifications, Analyses, and New Innovations. Ph.D. diss., Air Force Institute of Technology, Ohio, United States.

Yen, G. G., and Z. He. 2014. Performance Metric Ensemble for Multi-objective Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 18(1): 131-144.

Zhang, F., Y. Mei, and M. Zhang. 2019. Evolving dispatching rules for multi-objective dynamic flexible job shop scheduling via genetic programming hyper-heuristics. In *2019 IEEE Congress on Evolutionary Computation*, 1366-1373. Wellington, New Zealand, 10-13 June 2019.

Zitzler, E. 1999. Evolutionary Algorithms for Multi-objective Optimization: Methods and Applications. Ph.D. diss., Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Zitzler, E., and L. Thiele. 1999. An evolutionary algorithm for multi-objective optimization: the strength Pareto approach. Technical report no. 43. Zurich, Switzerland; Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH).

Zitzler, E., M. Laumanns, and L. Thiele. 2001. SPEA2: improving the strength Pareto evolutionary algorithm. Technical report no. 103. Zurich, Switzerland: Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH).