Journal of Environmental Sciences Studies

Journal home page: www.jess.ir

Estimation of separation of Black-eyed pea with gravity separator table using Random Forest optimized by Genetic algorithm

Saeed AgaAzizi[a,*], Mohamadhosein Kianmehr [b], Amir shyei[c], Mohammad Ali Keyhandoust[d]

[*]a,c,d .Department of Biosystems Engineering, Faculty of Agricultural Technology and Natural Resources, University of Mohaghegh Ardabili, Ardibil, Iran

b. Department of Biosystems Engineering, University of Tehran, College of Abouraihan, Pakdasht, Iran.

[*]Email Address: s.agaazizi@yahoo.com

Received: (18-02-2022)
Accepted: (18-04-2022)

Abstract

Separation (SP) is an undeniable member in the process set after harvesting of bulk products. The gravity table separator machine (GTSM) is one of the devices used to separate modal impurities in grain masses. Due to the continuity in the range of changes in the parameters of the GTSM and the high number of these factors affecting the rate of SP of impurities in the mass of cowpea beans, and considering that it seems almost impossible to examine all the values in this range. The use of machine learning (ML) to predict the process of SP against the changes applied to these factors facilitates the use of the GTSM for black-eyed pea (BeP). The present study is about predicting the performance of a GTSM in separating the BeP. The dependent variables included the cleaned seeds (Y1), weight of the cleaned seeds (Y2), total gross number (Y3), total gross weight (Y4), rotten seeds number (Y5), rotten seeds weight (Y6), broken seeds number (Y7) and broken seeds weight (Y8) and the independent variables included transverse slopes of the table (X1), longitudinal slopes of the table (X2), frequencies of table oscillation (X3) and blower air speeds (X4). The employed methods were single Random forest (RF) and a hybrid Random forest integrated by Genetic algorithm (RF-GA) for optimization of RF parameters. Results were evaluated using correlation coefficient (CC), Scattered Index (SI) and Willmott's Index (WI). According to findings, hybrid method provided higher performance compared with that for the single method and increased the prediction performance, successfully.

Keywords

Black-eyed pea; machine learning; separation; gravity table.

Introduction

Black-eyed pea (BeP), from the variety of Fabaceae and Fabales, is one of the most ancient crops available for humans using, today. This crop originated in Africa near Ethiopia and subsequently developed in the savannah fields of Africa, it is now widely cultivated worldwide as one of the legumes (Alcalde, Wheeler et al. 1999). Nigeria has the largest production of BeP in the world with an annual output of 3.4 million tons (FAO 2017). Sowing, cleaning and separating grains reduces issues raised by the planting process, facilitates operations, reduces storage space requirements, increases crop yield and increases marketability (Ebrahimi and Askari asli ardeh 2015). One of the most important factors affecting different aspects of BeP is the quality of seed. The quality of the product increases or decreases the marketability. Use of a separator is needed to improve the quality of the product and to remove damaged seeds from healthy seeds (Ebrahimi and Askari asli ardeh 2015). Agaazizi et al (Agaazizi, Rasekh et al. 2020) separated and removed the impurities in the wheat grain mass using a gravity table separator. The device had 5 adjustable variables. Their effect on the percentage of separation (SP) of impurities was investigated. The results showed that in the transverse slope of 1.5°, the longitudinal slope of 2.5°, the oscillation frequency of 395 cycles per minute by the amplitude of 5 mm and the air velocity of 6.75 m/s the maximum SP percentage was equal to 87.03%. One of the major problems with BeP production fields is the presence of gross material associated with the crop. The amount of mass contamination in seed silos is an indicator of product quality and plays an important role in its price and marketing. The main impurities in seeds of BeP can be rotten seeds, seeds of monocots, stone and straw noted. Numerous studies have been conducted to increase the purity of harvested legumes. One of the researches carried out by the gravity separator is the comparative research of Falconer (Falconer and engineering 1970) who employed traditional and modern methods of gravity SP to determine the percentage of seed purity with gravity separator and compared both methods in the terms of advantages, disadvantages, principles and variables and it has been shown that analysis and optimization of SP parameters is essential for a continuous production, but the selection of the most suitable separator for specific use needs to be according to an experimental methods. Rasekh et al (Rasekh, Tavakoli et al. 2005) investigated the grain SP of aged wheat from healthy wheat with gravity separator and concluded that gravity separator had the best performance in wheat SP at 2° transverse slope, 3° long slope, oscillation frequency of 450 cycles per minute and air velocity of 8 m/s. Bagheri et al. (BAGHERI, RASEKH et al. 2017) used dimensional analysis to determine effective parameters of gravity separator to minimize impurities in lentil mass and investigated that increasing the transverse slope of the table from 0.5 to 1° and increasing the longitudinal slope of the table from 1 to 2° increase the impurity SP in the lentil mass. The researchers also investigated the effect of the dimensionless number V/aω (velocity of blown air on the product of the amplitude at the oscillation rate of the table) in the SP ratio and showed that when the V/aω is equal to 171, the transverse slope of the table is 1° and the longitudinal slope of the table is 2°, the percentage of impurities in the lentil mass is the highest. Egbali and Borghei (Egbali and Bargaee 2003) investigated the separating of three wheat cultivars in three types of Iranian sweepers and evaluated the effect of nutrition, wind and sieve parameters on the removal of impurities from wheat cultivars. The results showed that, wheat SP was significant in the three mentioned systems and the SP amount depends on wheat variety. Chenari et al., (Chenari, Shahid zadeh et al. 2012) to evaluate and determine the efficiency of wheat seed dispensers in an experiment designed the factorial experiment in a completely randomized factorial design with four factors including type of machine (three levels), feed rate (two levels), wind speed (three levels) and wheat seed variety (three levels) in three replications and the effect of these factors on the SP was obtained for each machine. The results of this study showed that all four factors had significant effect on SP and it was found that Kymbrya winnowing machine had the highest SP rate (86.72%). The researchers also reported that separating in Behrang variety, with 84.31%, was better than other varieties. It was also found that decreasing feed rate and increasing wind speed increased the SP rate. Muhammad et al.,(Muhammad, Abubakar et al. 2013) after designing and manufacturing a cereal winnowing machine, evaluated it with sorghum, soybean, and millet products at three levels of feed rate and three levels of fan speed. The results of their investigations showed the best yield values in SP are of 95, 98 and 91% and the missing seeds are 0.63, 0.81 and 0.75% for sorghum, soybean and wheat, respectively. Rosgar et al. (Rouzegar, Khalifeh et al. 2013) evaluated a laboratory pneumatic separator. Their evaluation in four replications included: moisture content (12 and 18% b/w), feeding rate (13, 25 and 32 kg/min) and suction rate (850, 950 and 1150 rpm). The results showed that the highest purification rate was 99.55% at fan speed of 850 rpm. Also, with increasing the moisture content of the seed and with increasing the feeding rate, the rate of SP increases significantly. Simoniyan & Yiljep (Simonyan and Yiljep 2008) investigated the elimination losses of sorghum seeds in a laboratory sorghum thresher and showed that with increasing sieve oscillation, the rate of purification decreases and mortality increases, and the rate of purification losses increases with increasing feed and air flow.As is described SP plays vital role in determining the value of seed products. BeP is one of the seed products that the presence of impurities in it causes a sharp drop in its selling price. One of the disadvantages of existing devices for cleaning seeds is their high cost and their low efficiency. recently, machine learning (ML) methods provided an effective solution for increasing the efficiency of cleaning devices and reducing their cost (Vlasov and Fadeev 2017, Gievsky, Orobinsky et al. 2018). In the present study, due to the wide range of changes in the parameters of the weight table separator device, which in practice makes it impossible to test each of them, and also due to the fact that the use of different statistical methods due to the high number of input parameters, As it was observed in previous studies in this field, research in the field of statistical research has been done in two stages, in the first stage three parameters and in the next stage 2 other parameters have been studied. Entry seems to be necessary to change the rate of SP of impurities from the mass of cowpea beans. Due to the ability of Machine learning techniques in analyzing the high number of input parameters, it seems that the use of this feature in achieving the best values of the parameters of the GTSM is desirable. On the other hand, due to the fact that studies on the integration of these two issues are very limited, the need for such research is becoming more apparent.The present study wants to model the SP process of BeP by a GTSM using Random forest (RF) and hybrid Genetic Algorithm-Random forest (RF). In our best of knowledge this is the first study which follows this idea and procedure. The employed parameters were transverse slopes and longitudinal slopes of the table, frequencies of table oscillation and blower air speeds as independent variables and cleaned seeds, weight of the cleaned seeds, total gross number, rotten seeds number, rotten seeds weight, broken seeds number and broken seeds weight as the dependent variables. In the case of the BeP, limited research has been conducted on the SP of impurities in its mass, and no study has been conducted on the use of ML in predicting the rate of SP of impurities using GTSM. Therefore, the present study lacks a history of research related to the problem. The innovation of the present study is the use of single and hybrid ML techniques in the category of SP devices and predicting the rate of SP of impurities of BeP affected by 4 factors.

2. Materials and Methods

The employed GTSM was the LA-K model from the Danish company WESTRUP (Figure 1). This machine separates seeds based on the density values.

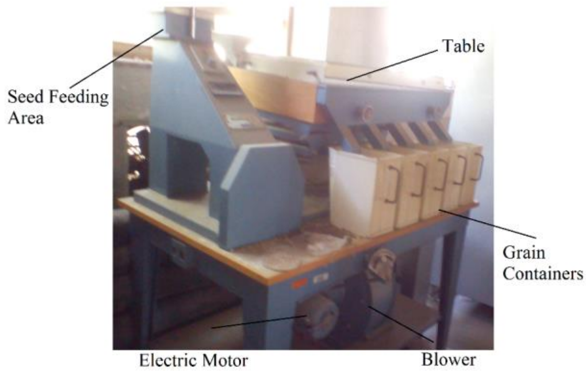


Figure 1. The employed GTSM

The GTSM has a table that is inclined in both longitudinal and transverse directions and its surface is grid and perforated and its lattice is of a special type. The swing table has a reciprocating swing. The upward flow of air hits the seeds from below the table. The amplitude of the machine swing, the longitudinal and transverse angles of the table and the speed of the swing of the table were adjusted by the levers on the machine. Fig. 2 shows the vertical view of the table surface. In this machine, materials are shed from the vibration tank on the machine table. The height of the table increases in the positive direction x (longitudinal slope) and in the positive direction y (transverse slope). So the lower left corner of the table has the lowest height (point O). The material leaves the table along the side of O_x. The table rotates in the x-axis direction and the air flows through the material beneath the table. As a result of desk vibration and air force, lighter materials float on heavier materials and are separated by longitudinal and transverse slopes in the lower left corner of the table. And the heavier parts that remain in contact with the table are moved to the top of the longitudinal slope (lower right corner of the table) as they move out of the table. Digital air-flow meter with accuracy of 0.1 m/s was used to measure air velocity in the Gravity separator.

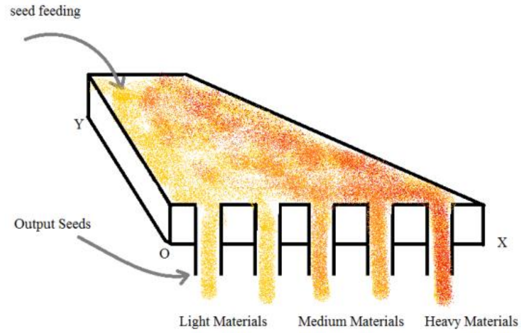


Figure 2. Schematic of GTSM

Factorial experiment based on completely randomized design was used to determine the effect of table longitudinal slope, table slope frequency, table swing frequency, table swing amplitude and blown air velocity on the percentage of impurities in BeP seed mass. In this experiment three transverse slopes of the table (X1)(0.75, 1.5 and 2.25 degrees), three longitudinal slopes of the table (X2)(2.5, 3.5 and 4.5 degrees), four frequencies of table oscillation (X3)(395, 415, 435 and 415 cycle per min.) Four blower air speeds (X4)(5.25, 6.5, 6.75 and 7.5 m/s) in five replications were used to determine the optimum level of each parameter to achieve the highest SP percentage. SP percentage was evaluated using eight factors which have been employed as the independent variables including cleaned seeds (Y1), weight of the cleaned seeds (Y2), total gross number (Y3), total gross weight (Y4), rotten seeds number (Y5), rotten seeds weight (Y6), broken seeds number (Y7) and broken seeds weight (Y8). In each experiment, some BeP seeds along with impurities were removed from the exit of the light material (lower left corner of the table in Figure 2) and counted and selected by electron seeding machine of three 200 samples. Then, by observing individual seeds and separating healthy BeP seeds from 200 masses and counting them, percentage of impurities of BeP seed mass was obtained.

2.1. Modeling procedure

Modeling was performed to predict the dependent variables including cleaned seeds (Y1), weight of the cleaned seeds (Y2), total gross number (Y3), total gross weight (Y4), rotten seeds number (Y5), rotten seeds weight (Y6), broken seeds number (Y7) and broken seeds weight (Y8) using independent variables including transverse slopes of the table (X1), longitudinal slopes of the table (X2), frequencies of table oscillation (X3) and blower air speeds (X4). The employed methods were included RF and a hybrid GA-RF for optimization of RF parameters.

2.2. Random Forest (RF)

RF as a supervised learning algorithm, provides excellent results even without adjusting its meta-parameters. This method is one of the frequently used ML method for Classification and Regression purposes due to its performance and simplicity (Liaw and Wiener 2002, Pal 2005). RF randomly generates a forest which is collection of a group of Decision Trees (DTs). Usually, bagging is employed for forest construction. The main idea of the bagging method is that a hybrid and integrated learning model increases the performance of the model. Simply, a RF is decided by several trees and merged with each other to make more accurate and stable predictions (Svetnik, Liaw et al. 2003, Oshiro, Perez et al. 2012). Figure 3 presents the architecture of a RF model.

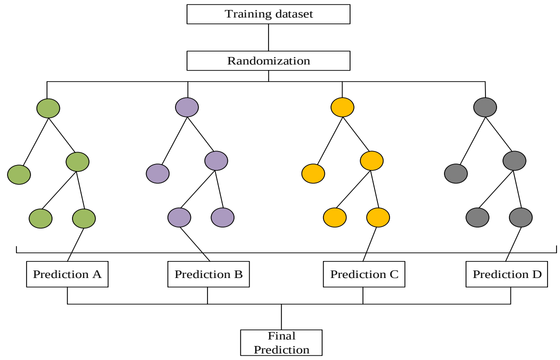


Figure 3. The architecture of RF (reproduced from (Dimitriadis, Liparas et al. 2018)).

According to Figure 3, there is a set of DTs, each of which is injected with a subset of data. Each algorithm performs learning operations. When predicting, that is, when a new set of data is given to an algorithm to predict, each of these learned algorithms predicts a result. The random forest algorithm can finally, using voting, select the class that received the most votes and use it as the final class to perform the classification operation.RF has parameters similar to the DT. On the other hand, there is no need to integrate a DT with a packing bag and classifier-class of the RF can be employed. As mentioned earlier, RF can also be used to solve regression problems. RF adds randomness to the model as trees grow. Instead of searching for the most important features when dividing a "node", this algorithm looks for the best features among a random set of features. This leads to more variety and ultimately a better model. Therefore, in a RF, only one subset of features is considered by the algorithm to divide a node. By adding a random threshold for each feature instead of searching for the best possible threshold, trees can even be made more random (Svetnik, Liaw et al. 2003, Dimitriadis, Liparas et al. 2018).In the present study, RF was employed as the control model to predict Y1 to Y8 by the use of X1 to X4. In the next step, RF was integrated with GA to be optimized.

2.3. RF optimized by Genetic algorithm (GA)

GA is a frequently used and popular evolutionary technique. GA is compound of chromosomes which solve problems by an encoded solution. For some problems, the direct encoding of a solution in a GA's chromosome. The theoretical materials of Genetic algorithm is comprehensively discussed in the study by John h. Holland (Holland 1992). In recent years, GA is successfully employed for optimizing the parameters of different ML methods for making a hybrid ML method which is more robust than the single ML method. Because, a hybrid model benefits the advantages of two ML method and covers the lack of ML methods and improves the performance.

In the present study, it is decided to employ GA for optimizing the parameters of RF method and making a hybrid RF-GA prediction model and comparing the prediction ability of the RF-GA with single RF model. In this process, Y1 to Y8 were considered to be the dependent variables and X1 to X4 were considered to be the independent variables. Fig. 4 indicates the algorithm of the developed RF-GA method.

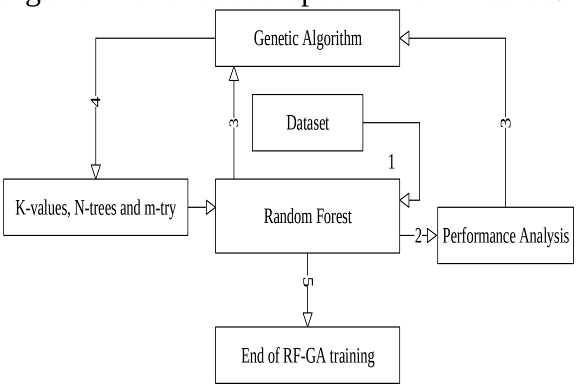


Figure 4. The main algorithm of a hybrid RF-GA method (reproduced from (Elyan and Gaber 2017)).

According to Fig. 4, dataset is imported to RF, then the performance analysis is performed by the output of RF, then the factors of RF is imported to GA along with dataset. In step four, GA produces the factors affecting the performance of RF including K-values, N-trees and m-try. These factors are imported by RF to produce the best output. This cycle is continued to reach an acceptable output. Then, the process of training a hybrid RF-GA method is finished. Table 1 presents the indicators of the developed models.

Table 1. The indicators of the developed models

Dependent variable	RF	GA-RF
Y1	RF-1	GA-RF-1
Y2	RF-2	GA-RF-2
Y3	RF-3	GA-RF-3
Y4	RF-4	GA-RF-4
Y5	RF-5	GA-RF-5
Y6	RF-6	GA-RF-6
Y7	RF-7	GA-RF-7
Y8	RF-8	GA-RF-8

2.4. Evaluation criteria

Evaluation of the performance of the models were performed by the use of correlation coefficient (CC), Scattered Index (SI) and Willmott's Index (WI) of agreement, Eq. 1 to 3 (Ardabili, Mahmoudi et al. 2016, Faizollahzadeh_Ardabili, Najafi et al. 2017).

$$CC = \frac{\left(\sum_{i=1}^{n} O_i P_i - \frac{1}{n}\sum_{i=1}^{n} O_i \sum_{i=1}^{n} P_i\right)}{\left(\sum_{i=1}^{n} O_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} O_i\right)^2\right)\left(\sum_{i=1}^{n} P_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} P_i\right)^2\right)} \quad (1)$$

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\overline{O}} \quad (2)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}\left(\left|P_i - \overline{O}_i\right| + \left|O_i - \overline{O}_i\right|\right)^2}\right] \quad (3)$$

Where, O refers the output value, P refers to the predicted value and n refers to the number of data.

3. Results & Discussion

Statistical analysis of each varable was performed by SPSS® software. Table 2 presents the statistical characteristics of the utilized data.

Table 2. Statistical characteristics of the utilized data

Variable	mean	minimum	maximum	standard deviation	coefficient of variation	skewness
X1	6.375	5.250	7.500	0.841	0.132	0.000
X2	425.0	395.0	455.0	22.439	0.053	0.000
X3	3.406	1.000	4.500	0.897	0.263	-0.243
X4	1.500	0.750	2.250	0.615	0.410	0.000
Y1	72.544	58.00	90.667	6.774	0.093	0.324
Y2	0.812	0.709	0.924	0.045	0.056	0.000
Y3	27.456	9.333	42.000	6.774	0.247	-0.324
Y4	18.807	7.635	29.112	4.542	0.242	0.000
Y5	19.859	6.333	31.667	5.443	0.274	-0.106
Y6	13.005	3.850	21.894	3.670	0.282	0.227
Y7	6.479	1.333	12.333	2.190	0.338	0.165
Y8	5.115	1.206	11.645	1.759	0.344	0.313

1. According to Table 2, the range of X1 variation is from 5.25 to 7.5 with mean value 6.375 and standard deviation 0.841, the range of X2 variation is from 395 to 455 with mean value 425 and standard deviation 22.439, the range of X3 variation is from 1 to 4.5 with mean value 3.406 and standard deviation 0.897, the range of X4 variation is from 0.75 to 2.25 with mean value 1.5 and standard deviation 0.615, the range of Y1 variation is from 58 to 90.667 with mean value 72.544 and standard deviation 6.774, the range of Y2 variation is from 0.709 to 0.924 with mean value 0.812 and standard deviation 0.045, the range of Y3 variation is from 0.933 to 42 with mean value 27.456 and standard deviation 6.774, the range of Y4 variation is from 7.635 to 29.112 with mean value 18.807 and standard deviation 4.542, the range of Y5 variation is from 6.333 to 31.667 with mean value 19.859 and standard deviation 5.443, the range of Y6 variation is from 3.850 to 21.894 with mean value 13.005 and standard deviation 3.670, the range of Y7 variation is from 1.333 to 12.333 with mean value 6.479 and standard deviation 2.190, the range of Y2 variation is from 1.206 to 11.645 with mean value 5.115 and standard deviation 1.759. Modeling process was performed by RF as the control and RF-GA method. Results related to model parameters have been presented in Table 3.

Table 3. parameters of the RF and RF-GA models.

Model	parameter						
	A	B	C	D	E	F	G
RF	100	10	0.1	2	4	3	0.2
GA-RF-1	94	5	0.3463	41	20	3	0.1467
GA-RF-2	94	5	0.3661	41	20	55	0.1515
GA-RF-3	94	5	0.3463	41	20	3	0.1467
GA-RF-4	94	5	0.3661	41	20	55	0.1515
GA-RF-5	94	5	0.3464	41	20	3	0.1497
GA-RF-6	94	5	0.3661	41	20	3	0.1467
GA-RF-7	81	5	0.3463	41	20	3	0.1467
GA-RF-8	81	5	0.3392	41	20	3	0.1653

A: Random Forest.number_of_trees
B: Random Forest.maximal_depth
C: Random Forest.confidence
D: Random Forest.minimal_leaf_size
E: Random Forest.minimal_size_for_split
F: Random Forest.number_of_prepruning_alternatives
G: Random Forest.subset_ratio

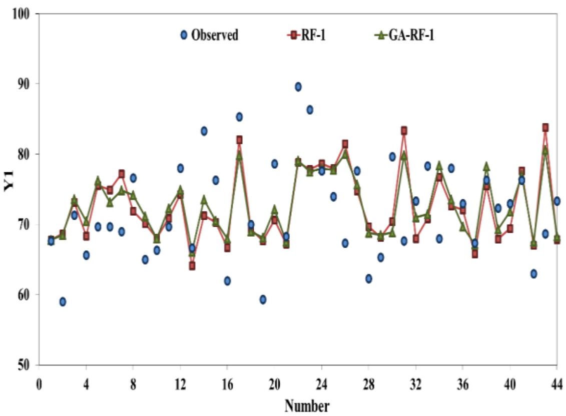
Table 4 presents the results for the performance evaluation of the developed methods. Results have been categorized into three factors including CC, SI and WI for both RF and RF-GA.

Table 4. General results of the computations for the RF and GA-RF models

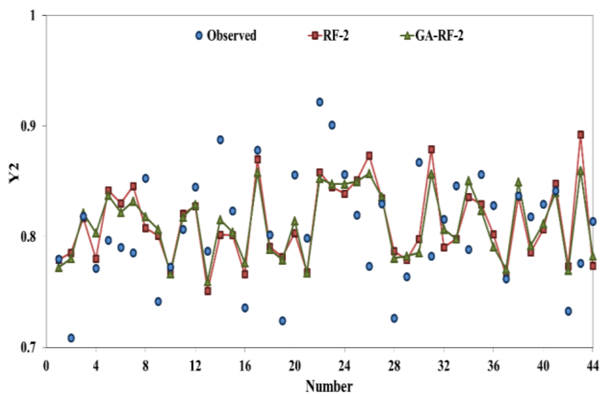
Model	Statistical parameters		
	CC	SI	WI
RF-1	0.416	0.093	0.640
RF-GA-1	0.495	0.086	0.669
RF-2	0.395	0.058	0.623
RF-GA-2	0.485	0.053	0.664
RF-3	0.417	0.239	0.641
RF-GA-3	0.496	0.220	0.670
RF-4	0.396	0.246	0.623
RF-GA-4	0.485	0.224	0.665
RF-5	0.450	0.265	0.630
RF-GA-5	0.473	0.258	0.634
RF-6	0.407	0.293	0.597
RF-GA-6	0.478	0.277	0.622
RF-7	0.320	0.281	0.576
RF-GA-7	0.405	0.259	0.602
RF-8	0.293	0.296	0.575
RF-GA-8	0.344	0.268	0.584

As is clear from Table 4, RF-GA can successfully increase the accuracy by increasing CC by about 19 % in average, reducing SI by about 7% in average and increasing WI by about 5% in average compared with those for the RF. This can be due to the optimizing ability of GA for estimating the optimized factors for the RF method in the hybrid RF-GA technique.

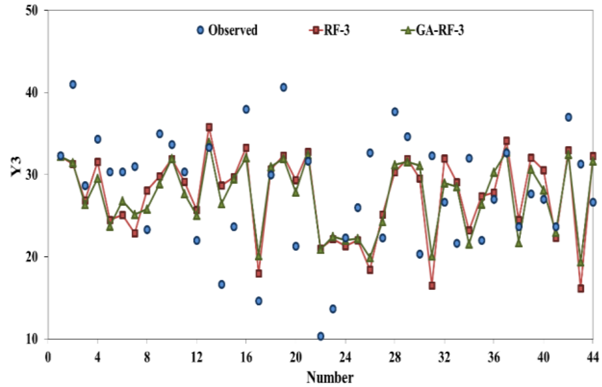
Figure 5 presents the observed and predicted values of the studied dependent variables with RF and GA-RF models, separately. This Figure also indicates the distance of the observed and predicted values for RF and GA-RF for doing a comparative analysis about



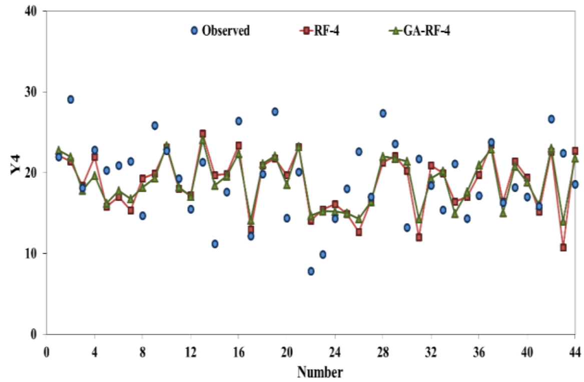
(a)



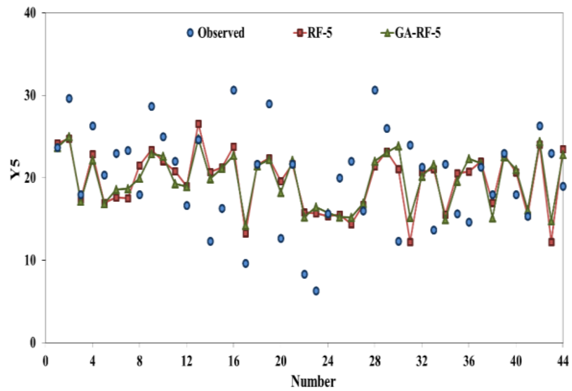
(b)



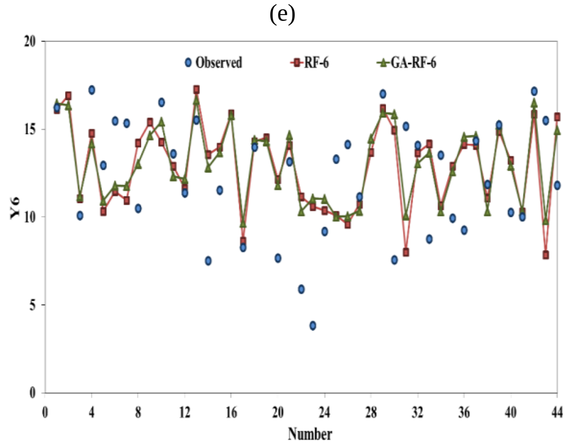
(c)



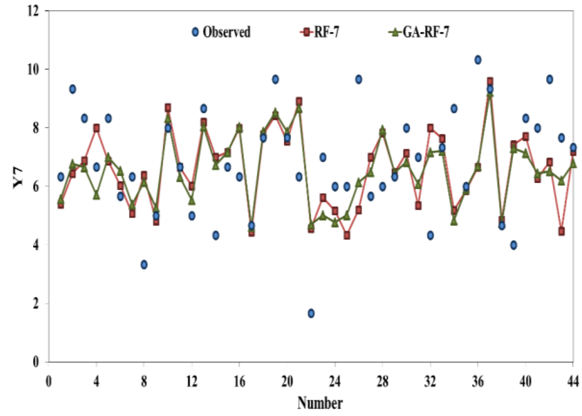
(d)



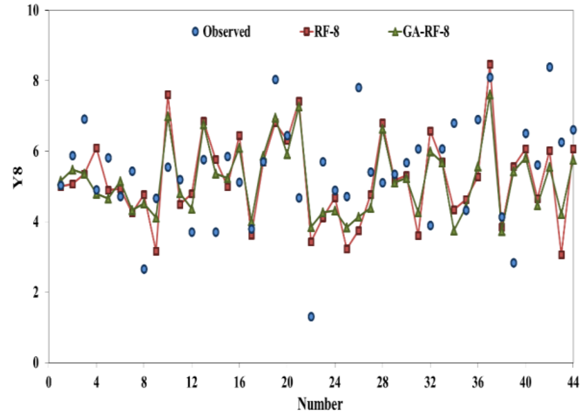
(e)



(f)



(g)

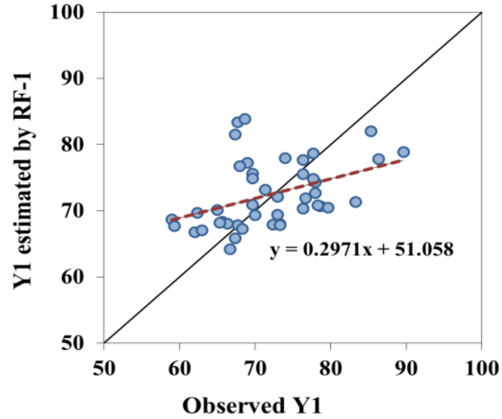


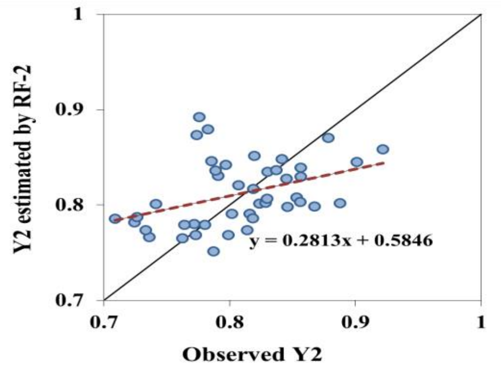
(h)

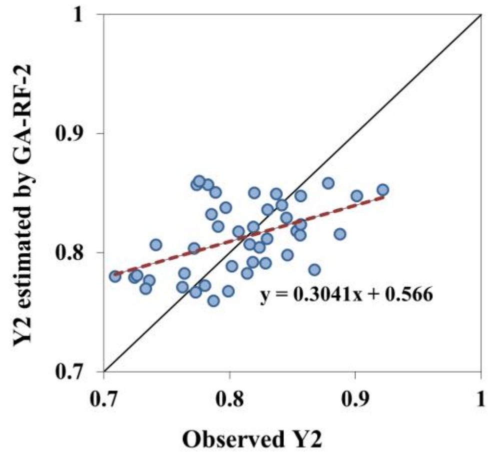
Figure 5 . Observed and estimated values of studied parameters with RF and GA-RF models a) Y1, b) Y2, c) Y3, d) Y4, e) Y5, f) Y6, g) Y7 and h) Y8.

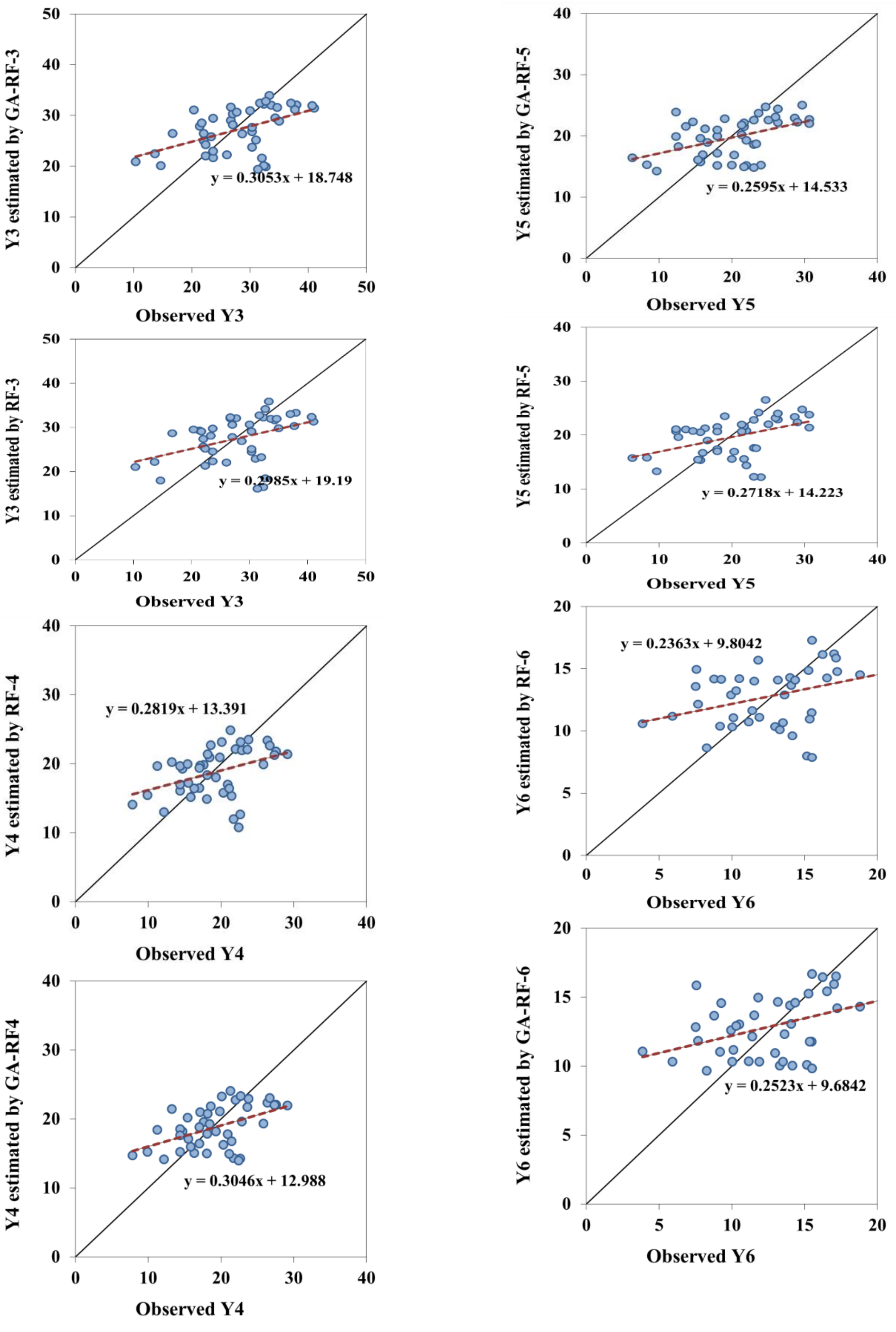
According to Fig. 5, graph a refers to the prediction of Y1, graph b refers to the prediction of Y2, graph c refers to the prediction of Y3, graph d refers to the prediction of Y4, graph e refers to the prediction of Y5, graph f refers to the prediction of Y6, graph g refers to the prediction of Y7 and graph h refers to the prediction of Y8. As is clear from Fig. 5 the predicted values have deviation from the target values. Such that, the deviation of RF form target values is higher than that for the RF-GA, or in other word the vertical distance of points predicted by single RF method from target value is higher than the vertical distance of points predicted by hybrid RF-GA from target value.

Fig. 6 presents the plot diagram for testing the linearity of the predicted and target values. The vertical axis of plots indicates the estimated values by RF or RF-GA and the horizontal axis indicates the target values.








Y3 estimated by GA-RF-3
Observed Y3
y = 0.3053x + 18.748
Y5 estimated by GA-RF-5
Observed Y5
y = 0.2595x + 14.533
Y3 estimated by RF-3
Observed Y3
y = 0.2985x + 19.19
Y5 estimated by RF-5
Observed Y5
y = 0.2718x + 14.223
Y4 estimated by RF-4
Observed Y4
y = 0.2819x + 13.391
Y6 estimated by RF-6
Observed Y6
y = 0.2363x + 9.8042
Y4 estimated by GA-RF4
Observed Y4
y = 0.3046x + 12.988
Y6 estimated by GA-RF-6
Observed Y6
y = 0.2523x + 9.6842

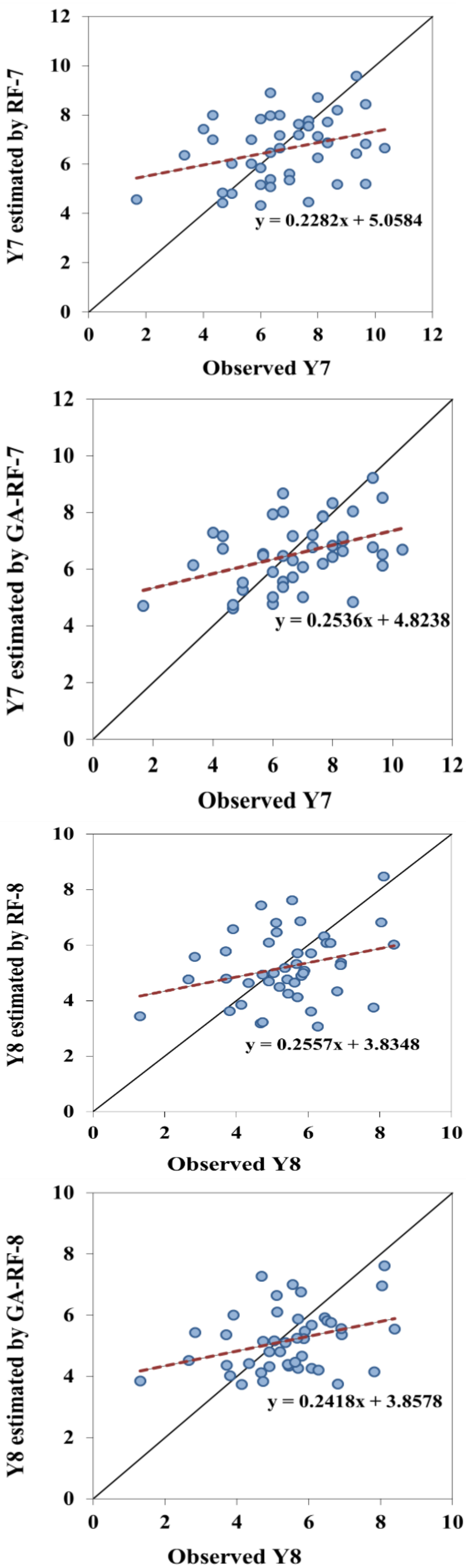
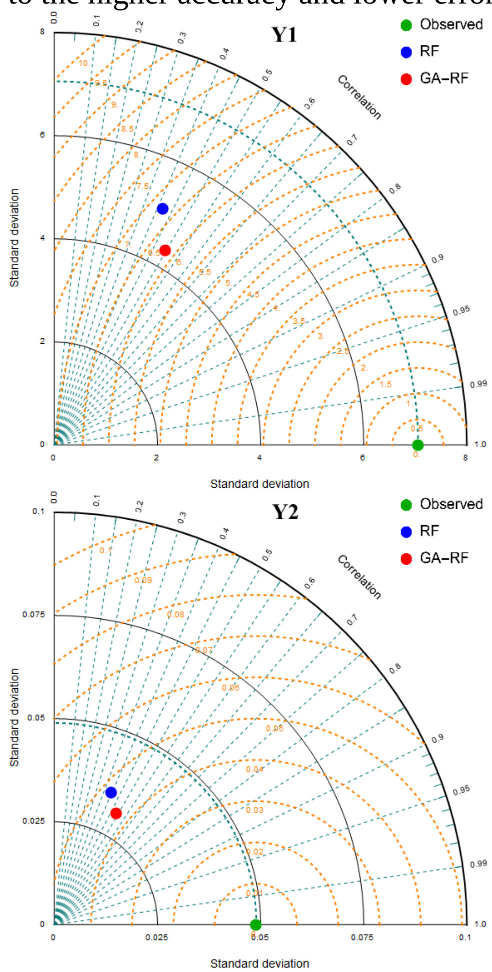


Figure 6. The scatter plots of observed and estimated values RF and RF-GA models.

The graphs in column a refers to the predicted values by single RF and the graphs in column b refers to the predicted values by the hybrid RF-GA model. According to Fig. 6, the distance of points from the 1:1 (or x=y) line shows the linearity degree of the predicted and target values. By reducing the distance from the 1:1 line, linearity increases and the error between predicted and target values reduces. As is clear, the linearity of the results of the hybrid RF-GA is higher than that of the single RF. Accordingly, Fig. 7 presents the Taylor diagrams which is consisted of standard deviation and correlation coefficient between target and predicted values for comparing the performance of the single RF and hybrid RF-GA. Lower standard deviation along with the higher correlation coefficient refer to the higher accuracy and lower error.



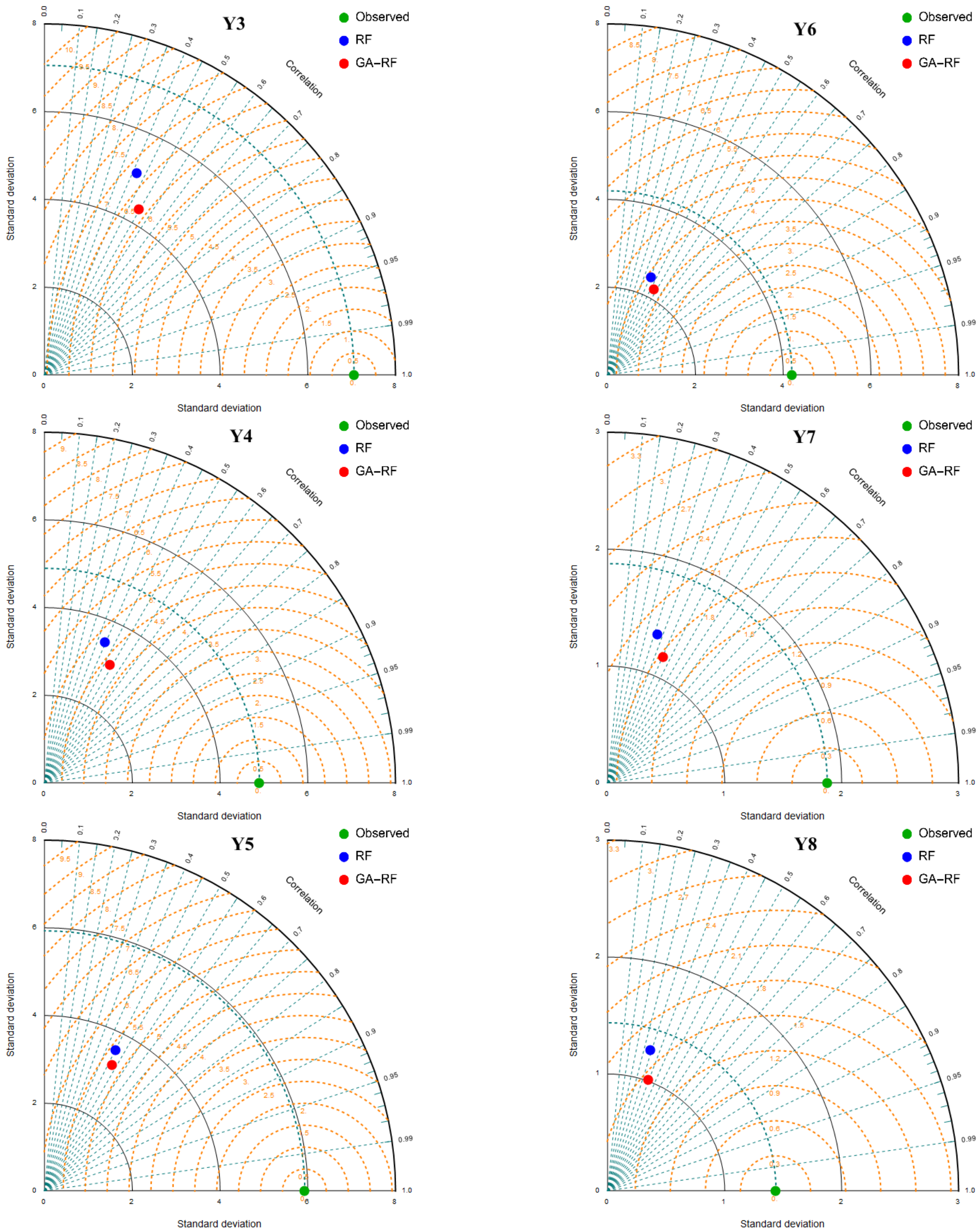


Figure 7. Taylor diagrams of estimated values.

According to Figures 7, the prediction performance of the hybrid RF-GA is higher than that of the single RF method. This can be one of the main achievement of the present study for proposing the use of hybrid methods instead of the single methods oin such problems.

4. Conclusions

The present study is about predicting the performance of a weight table separator in separating the BeP. the dependent variables included the cleaned seeds (Y1), weight of the cleaned seeds (Y2), total gross number (Y3), total gross weight (Y4), rotten seeds number (Y5), rotten seeds weight (Y6), broken seeds number (Y7) and broken seeds weight (Y8) and the independent variables included transverse slopes of the table (X1), longitudinal slopes of the table (X2), frequencies of table oscillation (X3) and blower air speeds (X4). The employed methods were single RF and a hybrid RF-GA for optimization of RF parameters. The reason for using ML methods in predicting the performance of GTSM is its low processing time and costly affectivity which can be important in designing similar devices considering some other parameters. In this regard, limited research work has been done and the topic has more chance for growth and study. For the future perspective, for more accurate predictions, the use of deep learning algorithms can be more effective. Such studies could be among the tasks ahead.

Nomenclatures table	
ANN	Artificial neural network
BeP	Black-eyed pea
RF	Random Forest
GA	Genetic algorithm
GTSM	Gravity Table Separator Machine
RMSE	Root mean square error
DL	Deep learning
DT	Decision tree
SP	Separation

1. Agaazizi, S., et al. (2020). "Investigation on the Separation of Wheat Bulk Impurities with Gravity Separator Table (in farsi)." Iranian Biosystem Engineering: pp. 161-170.
2. Alcalde, J., et al. (1999). "Quantitative effects of the genes Lf, Sn, E, and Hr on time to flowering in pea (Pisum sativum L.)." 50(340): 1691-1700.
3. Ardabili, S. F., et al. (2016). "Modeling and simulation controlling system of HVAC using fuzzy and predictive (radial basis function, RBF) controllers." Journal of Building Engineering 6: 301-308.
4. BAGHERI, H., et al. (2017). "Using of dimensional analysis to determine the parameters of gravity separator table device to minimize impurities in bulk lentils."
5. Chenari, M., et al. (2012). "Evaluate the efficiency of wheat seed cleaning machines (In farsi)." Journal of Agricultural Engineering Research 14(2): 69-80.
6. Dimitriadis, S. I., et al. (2018). "How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database." 13(6): 962.
7. Ebrahimi, R. and E. Askari asli ardeh (2015). "Evaluation of a pneumatic separator wheat (In farsi)." Proceedings of the 3rd International Conference on Modern Ideas in Agriculture, the Environment and Tourism, 31 August. University of Mohaghegh Ardebili, Ardebil, Iran: pp. 1-7.
8. Egbali, A. and A. Bargaee (2003). "Determination of the efficiency and evaluation of wheat seed seedlings in arid and semi-tropical regions (In farsi)." Proceedings of The 3rd student Conference on Agricultural Machinery and Mechanization Engineering, University of Tabriz Tabriz, Iran.: pp. 1-13.

9. Elyan, E. and M. M. J. I. s. Gaber (2017). "A genetic algorithm approach to optimising random forests applied to class engineered data." 384: 220-234.
10. Faizollahzadeh_Ardabili, et al. (2017). "A novel enhanced exergy method in analysing HVAC system using soft computing approaches: a case study on mushroom growing hall." Journal of Building Engineering.
11. Falconer, A. J. P. s. i. s. and engineering (1970). "Gravity separation: old technique/new methods." 12.
12. FAO, E. (2017). Food and Agriculture Organization of the United Nations: Rome, Italy.
13. Gievsky, A., et al. (2018). Substantiation of basic scheme of grain cleaning machine for preparation of agricultural crops seeds. IOP Conference Series: Materials Science and Engineering.
14. Holland, J. H. J. S. a. (1992). "Genetic algorithms." 267(1): 66-73.
15. Liaw, A. and M. J. R. n. Wiener (2002). "Classification and regression by randomForest." 2(3): 18-22.
16. Muhammad, U., et al. (2013). "Design and evaluation of a cleaning machine."
17. Oshiro, T. M., et al. (2012). How many trees in a random forest? International workshop on machine learning and data mining in pattern recognition, Springer.
18. Pal, M. J. I. j. o. r. s. (2005). "Random forest classifier for remote sensing classification." 26(1): 217-222.
19. Rasekh, M., et al. (2005). "Experimental studying of performance of gravity separator apparatus in separation of sunn Pest-Damaged wheat."
20. Rouzegar, M. R., et al. (2013). "Study effects of moisture content, feed rate and fan revolution on separation efficiency in a paddy laboratory winnower." 5(21): 2576-2578.
21. Simonyan, J. and D. J. A. E. I. C. J. Yiljep (2008). "Investigating grain separation and cleaning efficiency distribution of a conventional stationary rasp-bar sorghum thresher."
22. Svetnik, V., et al. (2003). "Random forest: a classification and regression tool for compound classification and QSAR modeling." 43(6): 1947-1958.
23. Vlasov, A. V. and A. S. Fadeev (2017). A machine learning approach for grain crop's seed classification in purifying separation. Journal of Physics: Conference Series, IOP Publishing.