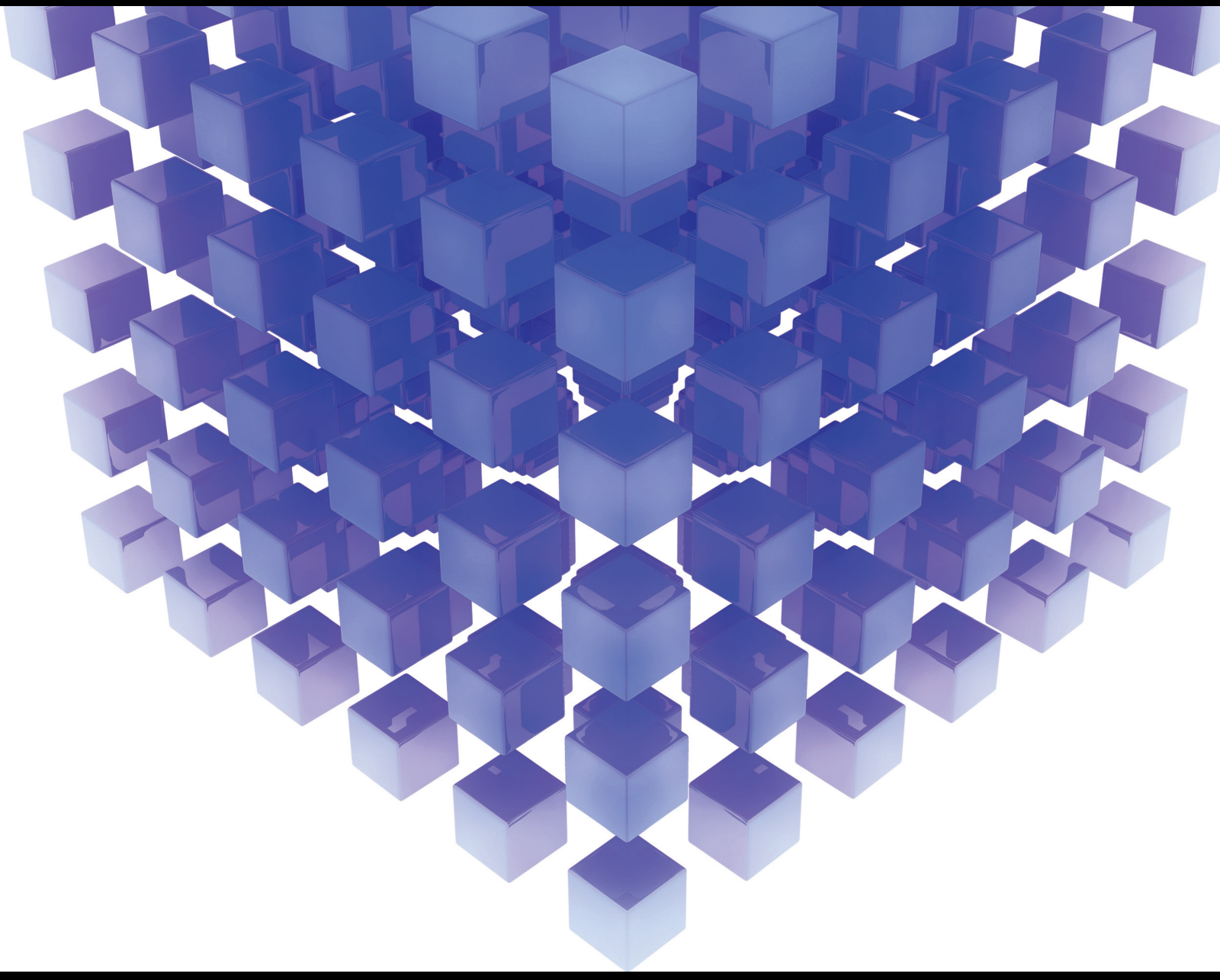
Mathematical Problems in Engineering

Optimization Algorithms and Engineering: Recent Advances and Applications

Lead Guest Editor: Noorbakhsh Amiri Golilarz
Guest Editors: Mohammad Yazdi and Samuel Yousefi



Optimization Algorithms and Engineering: Recent Advances and Applications

Mathematical Problems in Engineering

Optimization Algorithms and Engineering: Recent Advances and Applications

Lead Guest Editor: Noorbakhsh Amiri Golilarz
Guest Editors: Mohammad Yazdi and Samuel Yousefi

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9990906, 14 pages
https://doi.org/10.1155/2021/9990906

Research Article

Metaheuristic Approaches Integrated with ANN in Forecasting Daily Emergency Department Visits

Engin Pekel,[1] **Muhammet Gul**,[2] **Erkan Celik**,[3] **and Samuel Yousefi**[4]

[1]*Department of Industrial Engineering, Hitit University, 19000 Çorum, Turkey*
[2]*Department of Emergency Aid and Disaster Management, Munzur University, 62000 Tunceli, Turkey*
[3]*Department of Transportation and Logistics, Istanbul University, 34320 İstanbul, Turkey*
[4]*Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran*

Correspondence should be addressed to Muhammet Gul; muhammetgul@munzur.edu.tr

Received 28 March 2021; Revised 26 April 2021; Accepted 5 November 2021; Published 27 November 2021

Academic Editor: Laurent Dewasme

The overall service quality level of Emergency Departments (EDs) can be improved by accurate forecasting of patient visits. Accordingly, this study aims to evaluate the use of three metaheuristic approaches integrated with Artificial Neural Network (ANN) in forecasting daily ED visits. To do this, five performance measures are used for evaluating the accuracy of the proposed approaches, including Bayesian ANN, Genetic Algorithm-based ANN (GA-ANN), and Particle Swarm Optimization algorithm-based ANN (PSO-ANN). The outputs of this study show that the PSO-ANN model provides the most dominant performance in both the training and testing process. The lowest error is obtained with a mean absolute percentage error (MAPE) of 6.3%, Mean Absolute Error (MAE) of 42.797, Mean Squared Error (MSE) of 2499.340, Root Mean Square Error (RMSE) of 49.933, and *R*-squared (R^2) of 0.824 on the training dataset. The lowest error with an MAPE of 6.0%, MAE of 40.888, MSE of 2839.998, RMSE of 53.292, and R^2 of 0.791 is also obtained on the testing process.

1. Introduction

Emergency Departments (EDs) are the units that perform very crucial duties within the hospital service system and provide uninterrupted service. Also, these departments are the sole units where the patient traffic and transfer is the most and overcrowding is felt too much [1]. When this is the case, it is vital to improving the provided service quality level by newly adopted methodologies. Improving the service quality means a decreased waiting time, decreased length of stay, and increased ED throughput. These key performance metrics are directly interrelated with the daily patient volume of the EDs. Patient visits at the emergency departments cover 40%–70% of all hospital care [2]. The density of patient visits at the EDs on hourly, daily, weekly, monthly, or yearly basis will assist in arranging and allocating human and material resources (number of doctors, nurses, receptionists, medical devices, ED bed, and medicines). Therefore, accurate forecasting of patient visits gains great importance to ED decision makers.

The main aim of forecasting ED patient visits is to inform about the pattern of changes in the density of visits in the future [3]. The forecasting studies regarding ED visits have several dimensions such as time frame, forecasting methodology, the independent variables used in the modeling, and measurement of the model's accuracy [4, 5].

The remainder of the study is organized as follows: Section 2 presents both an overview on contemporary real-life case studies of ANN and ED patient visit forecast in the light of four dimensions. Section 3 includes ANN-based solution approaches with their pseudocodes. In section 4, the case study is demonstrated. Section 5 provides analysis results and a deep discussion. The final section presents the conclusion, future recommendations, and limitations of the study.

2. Literature Review

2.1. Overview on ED Patient Visit Forecast in the Light of Four Dimensions. ED patient visit forecasting, also called ED

patient volume forecasting or ED patient admission forecasting, is the problem of forecasting the future patient arrival of an ED. For that purpose, the historical data demonstrated as a time series are gathered in a regular time frame of hourly [6, 7], daily [8–13], weekly [14], monthly [15] and yearly [16] basis. Based on the literature, the vast majority of studies focus daily [5, 17]. The time frame also has an impact on the accuracy of the forecasting model. The smaller the interval of the time frame, the lower the relative accuracy of the model as compared to the models with a higher interval of the time frame (for example, annual ED patient visit forecasting) [7, 17, 18]. Furthermore, daily patient visits are the mostly dealt topics for the researchers since the forecasts play an important role in scheduling ED medical personnel, which is one of the most considerable problems faced by hospital management.

Another dimension concerns forecasting methodology. In the work of Nas and Koyuncu [19], it is stated that the studies regarding modeling the ED patient visits generally apply two types of methods. While the first one analyzes the correlations between patient visits and several regression variables, such as calendar or climatic variables, the second one predicts future values from the past values considering patient visits follow a time series [19]. These two groups propose regression-based and time series-based models, respectively. Apart from this, there exist machine learning-based models such as Artificial Neural Network (ANN), Support Vector Machine (SVM), and Long Short-Term Memory Network (LSTM) that are applied to ED patient visit forecasting. A similar grouping is mentioned in Yousefi et al.'s [20] study. They distinguish methods used in ED patient visit forecasting under two groups, namely, linear and nonlinear methods. Linear methods include Holt–Winters, Multiple Linear Regression (MLR), Exponential Smoothing (ES), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), and some other regression-based methods [18, 21–23]. Nonlinear methods include the adaptive Neuro-Fuzzy Inference System (ANFIS), ANN, SVM, and LSTM [20, 24–26]. In addition to these two groups of methods, some hybrid approaches are developed for this problem to benefit from the advantages of the usage of these methods either individually or integrated, improve accuracy, and decrease modeling errors [12, 14, 27]. Regression-based models are incorporated with machine learning-based (e.g., MLR-ANN) and time series-based models (MLR-ARIMA).

One of the most important dimensions of the ED patient visit forecasting studies are regarding the independent variables used in the model. In the literature, the variables include time-related (temporal or calendar), demographic, and climatic variables. In the literature, scholars mostly agree on a result that time-related variables have more impact than weather variables in forecasting ED patient visits [27–29]. Most of the authors deal with temporal variables such as the day of the week, the month of the year, holidays (school or public), the day after the holiday, the day before the holiday, and soccer match day [19]. Climatic variables such as air temperature, humidity, and wind speed are considered secondarily by the scholars [28–30]. Some other variables related to demography, transportation, epidemic, and hospital reputation are also studied [31].

The measurement of the accuracy of the model is also an important dimension in ED patient visit forecasting studies. Different measures, such as Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Mean Absolute Error (MAE), Receiver Operating Characteristic (ROC) curve value, Root Mean Square Error (RMSE), and R-squared, are used in the studies to test the accuracy of the models. The accuracy of the developed models is changed with respect to the chosen variables and some specific parameters (e.g., the number of hidden layers, learning rate, and momentum in ANN modeling). By setting up these, the developed models can be performed better. Most studies regarding ED patient visit forecasting prefer MAPE to measure the accuracy. In the Wargon et al.'s [17] review paper, the investigated studies that focus on daily ED visits result in an MAPE value of between 4.2% and 14.4%. That means an MAPE value which is lower than 10% or around 10% indicates good statistical predictability.

2.2. Research Gaps and Contributions of the Study. This study focuses on the dimension of "forecasting methodology" that is mentioned in the second order. We dealt with the applicability of metaheuristics integrated with ANN in ED patient visit forecasting. In this context, three hybridized ANN-based approaches are applied to the data of daily ED visits for the first time in the literature. These approaches are Bayesian ANN, Genetic Algorithm-based ANN (GA-ANN), and Particle Swarm Optimization algorithm-based ANN (PSO-ANN). Although plenty of hybridized approaches are proposed in the literature, the metaheuristic algorithms are not yet incorporated with regression, time series, and machine learning algorithm-based methods, which are appropriate for the nature of this problem. Therefore, this study will remedy the gap in the literature and contributes a lot by the following aspects:

(i) Approaches utilizing metaheuristics algorithms merged with ANN are applied to the daily ED patient visit forecasting problem (a novelty for the methodological viewpoint)

(ii) A comparative outline is produced by making a benchmark analysis between three approaches in terms of a common forecast accuracy measure "MAPE" (a novelty for the methodological viewpoint)

(iii) A case study in a public hospital in Istanbul (Turkey) is carried out to demonstrate the applicability of the approaches (a novelty for the application viewpoint)

3. ANN-Based Solution Approaches

3.1. Bayesian ANN. ANN provides solutions to problems in many different areas from natural science to engineering [32–34], social science [35, 36], and health science [37]. It has been developed for further improvement of targeted

systems by mimicking the biological nervous systems that occur in the human brain. ANN has process nodes with a simple logic connected to each other. Each node has an activation function that collects an input signal, and the aggregated signal is converted to a different value in the specified transformation function. Thus, a converted output signal is generated. Although each function is implemented very slowly by each neuron, a network can effectively carry out an incredible number of tasks [38, 39]. Initial weights are determined randomly at the first iteration, and the output and error of each latent neuron are calculated. The weight change is calculated according to a specific function, and the change is used to update the weights. The next iteration is then performed with respect to the updated weights.

Let $\overline{w}$ be an optimal weight vector, and it is most likely to catch the set of observed target data $y = (y_1, y_2, \ldots, y_n)$, given the inputs $x = (x_1, x_2, \ldots, x_n)$ in the Bayesian training approach. w is a vector of connection and bias weights that characterizes the data generating relationship. Bayesian training aims to gather the posterior probability distribution of the weights given the observed data $P(w|y, x)$. This process is carried out while updating any knowledge of the weight values before obtaining the data, with the information contained in the data, using Bayes theorem [40].

$$P(w|y, x) = \frac{P(y|w, x) * P(w)}{P(y|x)}. \quad (1)$$

The prior weight distribution and the likelihood function are represented, respectively, by $P(w)$ and $P(y|w, x)$, respectively. $P(y|x)$ shows the prior probability of the training data. The flowchart of Bayesian ANN is shown in Figure 1.

3.2. Genetic Algorithm. GAs have been developed to mimic some of the processes observed in natural evolution [41]. GAs aims to create a competitive set of solutions, and these targeted solutions progress through the natural selection process, in which noneffective solutions emerge and more efficient solutions continue to be reproduced. This process is repeated until the optimal solution set is obtained.

3.2.1. GA-ANN. A hybrid GA-ANN is a backpropagation network, which is the only exception to obtaining the weight matrix from performing genetic processes under optimal convergence conditions [42–44]. The flowchart of GA-ANN is shown in Figure 2. The initial weights are set at random in the first iteration, and the output and error of each latent neuron are calculated. Updated weights are computed according to GA by applying primary selection, reproduction, and mutation. The next iteration is then performed concerning the updated weights. Crossover and mutation operators in GA are used in the selection of weights and result in new offspring weights that offer better fitness value. While the weights are optimized here, the population number, crossover, and mutation values are also tried to be optimized by trial and error.

3.3. Particle Swarm Optimization. The PSO algorithm starts by creating a random population. It refers to a randomly generated population particle, the optimal values of these particles should be determined, and each particle is actually a different decision variable that must represent a vector in the problem-solving area. In the PSO algorithm, the movement of any particle affects the movement of the entire group, and ultimately, each member of the group can benefit from the discoveries and skills of other members [45, 46].

3.3.1. PSO-ANN. The PSO algorithm starts with the generation of starting particles, and the starting rates are assigned to the starting particles. At each iteration, each particle is updated based on the best values obtained. The comparison of the obtained values is performed with regard to the fitness value of the relevant iteration and the fitness value obtained so far during the iteration. One of them is the compliance value obtained so far, and this value is kept as the best solution. The other value is the best value reached by any particle in the population. $p_{\_p\_\text{best}}$ is i particle's best-known position, and $p_{\_g\_\text{best}}$ is the best position known to the swarm. The rand variable generates random values between 0 and 1. C_1 and C_2 equal to 1 and 2, respectively. The position values of the particles are obtained by means of the position equation affected by the velocity. Position values represent the weight values of the network. As the fitness function is optimized, the position values, i.e., the weight values of the net, are also optimized. Different combinations are tested on C_1 and C_2 parameters to obtain better weight values. The flowchart of PSO-ANN is shown in Figure 3.

$$v_i(t+1) = v_i(t) + \left(C_1 * \text{rand} * \left(p_{\_p\_\text{best}} - p_i(t)\right)\right) + \left(C_2 * \text{rand} * \left(p_{\_g\_\text{best}} - p_i(t)\right)\right). \quad (2)$$

Equation (2) shows the calculation of updating the velocity of the particle.

$$p_i(t+1) = p_i(t) + v_i(t). \quad (3)$$

Equation (3) indicates the calculation of updating the position of the particle for the weights of networks.

4. Dataset and Analysis

We used two years' data of a public hospital ED in Istanbul, Turkey. The data belongs to the years 2011 and 2012 (from January 1, 2011, to December 31, 2012). The time series used for modeling is shown in Figure 4. Temporal variables and a climatic variable of maximum temperature were used as dependent variables to forecast the daily ED patient visit. A total of 21 independent variables are used. Binary dummy variables are used in data regarding the month of the year, day of the week, and holiday (weekend holiday). While the value of 0 means that the related date does not belong to that month or that day or weekend holiday, "1" means belonging. The dummy variables are utilized in the dataset instead of the original categorical variable. The data for maximum temperature are obtained from a French meteorological association named Infoclimat

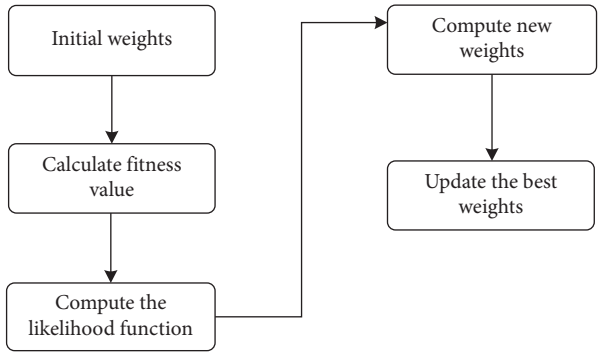


Figure 1: The flowchart of Bayesian ANN.

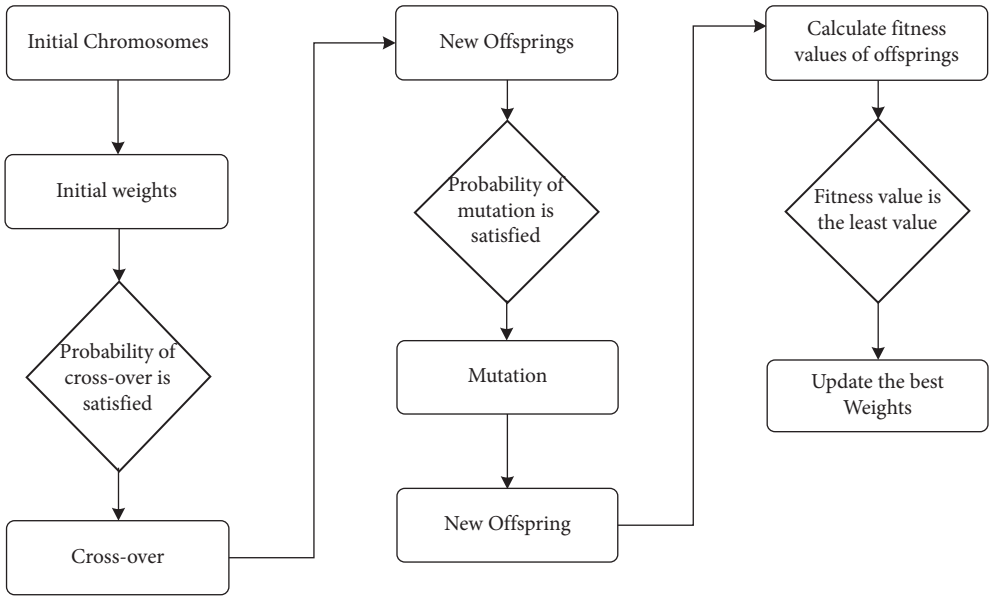


Figure 2: The flowchart of the hybridized GA-ANN.

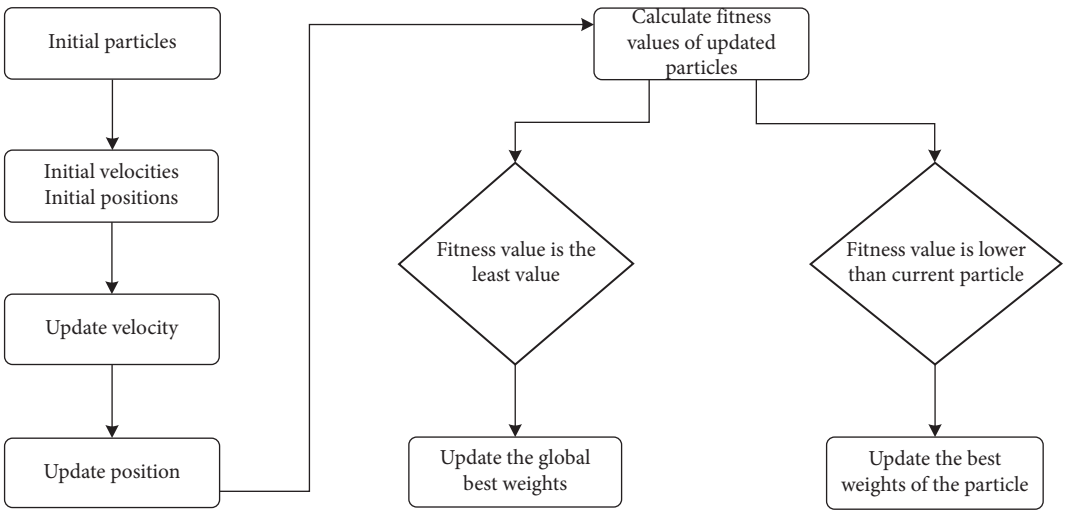


Figure 3: The flowchart of the hybridized PSO-ANN.

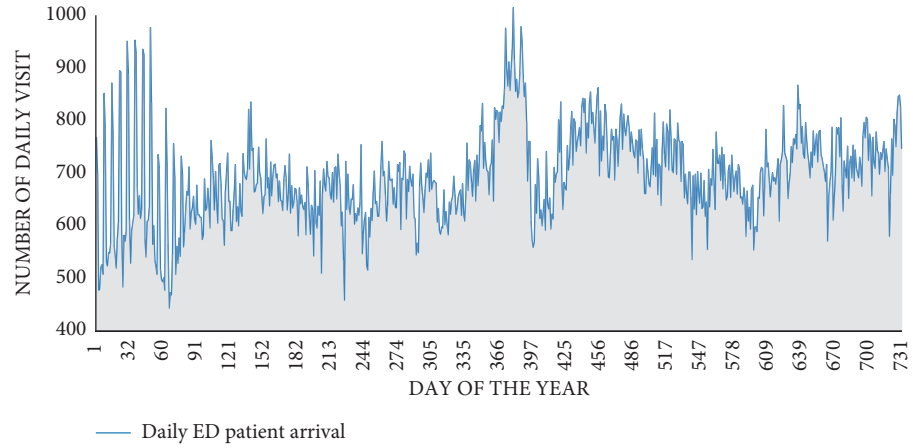


Figure 4: Time-series data of the daily ED patient visits (2011-2012).

Table 1: Descriptive statistics of variables.

Variable	Unit	Input/output	Total number of variables	Mean (standard deviation)
Month of the year	Binary dummy	Input	12	—
Day of the week	Binary dummy	Input	7	—
Holiday (weekend)	Binary dummy	Input	1	—
Maximum temperature (°C)	Numerical	Input	1	19.34 (8.98)
Daily ED patient visit	Numerical	Output	1	692.27 (87.66)

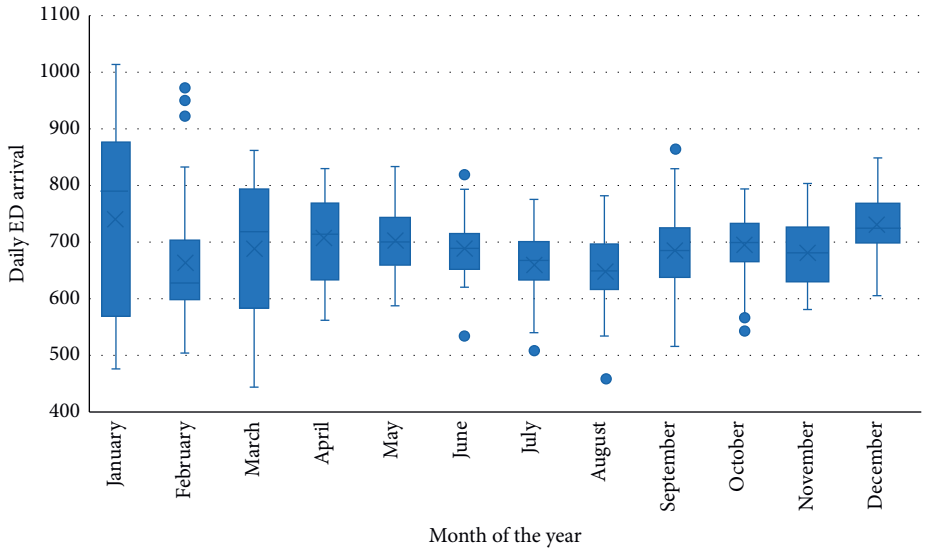


Figure 5: Box plot of the daily ED data on the basis of monthly trend.

(http://www.infoclimat.fr). A detailed description for each variable is provided in Table 1. Also, the time-series data of the dependent variable (daily ED visit) are presented in Figure 4. This figure shows that the highest number of visits at the ED occurred on January 14, 2012, with 1013 visits per day. The plot of daily ED visits shows the monthly fluctuations in the ED: there were more patients in January and December; there were slightly fewer patients in August. The trend of monthly ED visits on average is given as a box plot in Figure 5.

Also, the trend of the maximum temperature variable is demonstrated in Figure 6. According to this figure, the average maximum temperature is obtained as 19.34°C with a standard deviation of 8.98.

5. Results and Discussion

This paper performs the cross validation to the training dataset to prevent overfitting. It divides the training dataset into ten subsets. One of the ten subsets is held, and the rest of

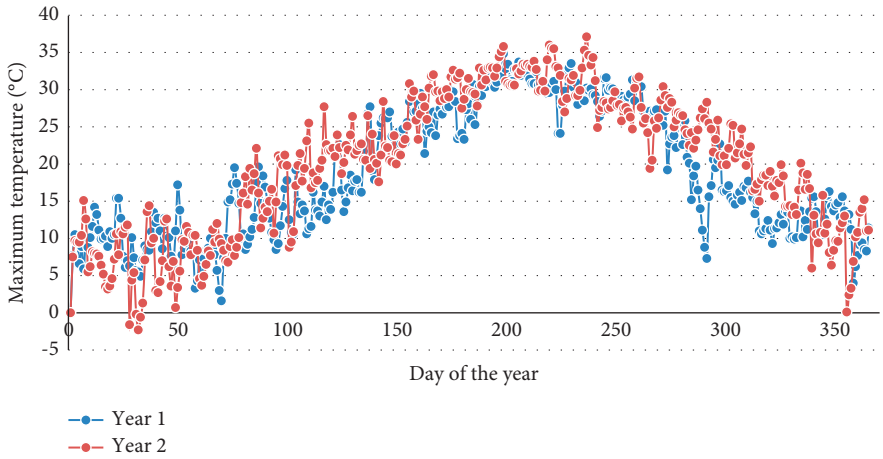


Figure 6: Trend of the maximum temperature (°C) data for the years 2011 and 2012.

Table 2: The trials for Bayesian ANN for testing.

#n	MSE	#n	MSE	#n	MSE	#n	MSE	#n	MSE
2	4911.005	**12**	4841.185	**22**	4806.977	**32**	4851.388	**42**	4813.478
3	4835.644	**13**	4806.369	**23**	4862.533	**33**	4936.914	**43**	6035.613
4	5019.853	**14**	4834.975	**24**	4822.185	**34**	4510.968	**44**	4892.921
5	4837.925	**15**	4885.426	**25**	4908.832	**35**	4504.176	**45**	**4246.407**
6	4928.894	**16**	4939.633	**26**	4941.382	**36**	4981.652	**46**	4884.429
7	4781.13	**17**	4781.687	**27**	4883.294	**37**	4534.402	**47**	5983.782
8	4891.587	**18**	4932.688	**28**	4809.14	**38**	4825.12	**48**	4773.257
9	4818.638	**19**	4882.4	**29**	4873.752	**39**	4884.537	**49**	4799.222
10	4827.339	**20**	4865.005	**30**	4923.553	**40**	4620.716	**50**	4835.748
11	4885.545	**21**	4867.583	**31**	4659.339	**41**	4838.274		

the subsets are trained regarding the fitness function. We applied different number of neurons (**#n**) for obtaining the best solution in Bayesian ANN. The number of neurons is increased one by one, ranging from 2 to 50. In this process, we used MSE for determining the best number of neurons. While logsig is used in the transfer function, purelin is also applied in the activation function. The minimum MSE is obtained with 45 neurons, and all solutions are presented in Table 2. Then, the results of Bayesian ANN are solved using this combination.

After obtaining the best configuration for Bayesian ANN with respect to the MSE, the results of five performance measures are presented in Table 3. The results of the training and testing are also illustrated in Figures 7 and 8, respectively.

The obtained MAPE for training and testing for Bayesian ANN is 7.3% and 8.8%, respectively. The other performance measures are also given in Table 3.

In the GA-ANN approach, we also applied the different combinations of parameters for obtaining the best solution. The number of populations, crossover rate, mutation rate, and the number of neurons are used. The number of populations is considered as 50. Six different crossover rates are used as 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9. Four different mutations' rate is also utilized as 0.1, 0.2, 0.3, and 0.4. The number of neurons is increased one by one, ranging from 2 to 50 as in Bayesian ANN. The logsig and purelin are used, respectively, in the transfer and activation functions in the ANN process. Therefore, 1176 different solutions ($=:4\times6\times49$) are totally obtained for GA-ANN. The best twenty combinations with respect to MSE are presented in Table 4.

Table 3: The results of the performance measures for Bayesian ANN.

	Training	Testing
MAPE	0.073	0.088
MAE	50.016	60.358
MSE	4246.407	7031.078
RMSE	65.164	83.852
R^2	0.672	0.343

After obtaining the best configuration for GA-ANN with respect to the MSE, the results of five performance measures are presented in Table 5. The results of the training and testing are also illustrated in Figures 9 and 10, respectively. The obtained MAPE for training and testing for GA-ANN are 7.8% and 6.9% respectively. The other performance measures are also given in Table 5.

In the PSO-ANN approach, we also applied different combinations of parameters for obtaining the best solution. The number of populations, the weighting coefficient for the

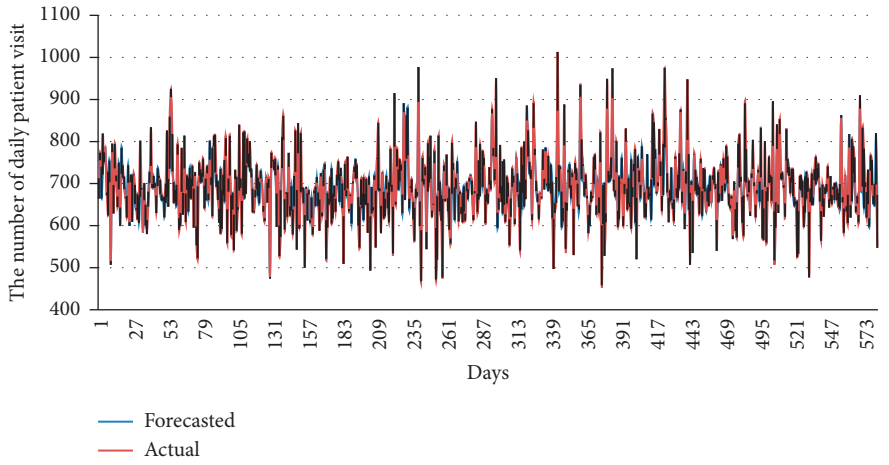


Figure 7: The actual and forecasted daily patient visits of Bayesian ANN for training.

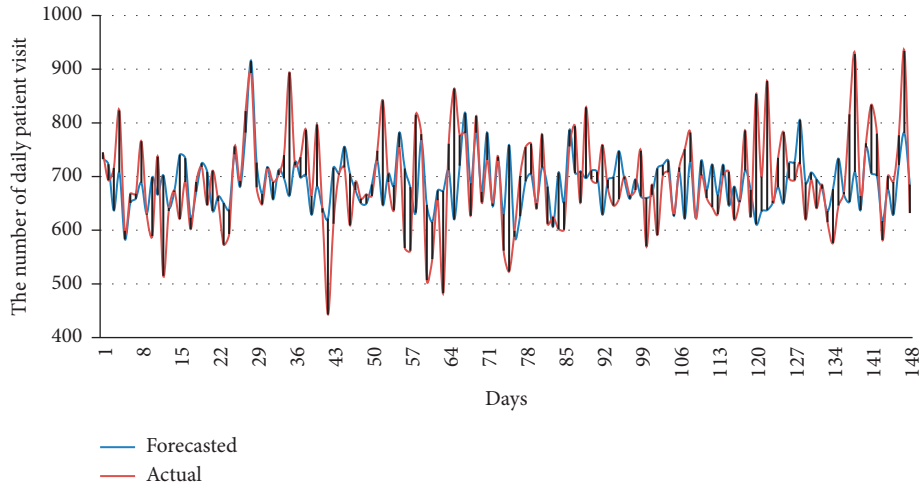


Figure 8: The actual and forecasted daily patient visits of Bayesian ANN for testing.

Table 4: The trials for GA-ANN for testing.

Population	Crossover rate	Mutation rate	#n	MSE	Population	Crossover rate	Mutation rate	#n	MSE
50	**0.8**	**0.1**	**14**	**3765.061**	50	0.6	0.2	31	3809.621
50	0.9	0.1	41	3766.242	50	0.4	0.4	11	3813.64
50	0.8	0.4	35	3778.22	50	0.9	0.3	50	3814.467
50	0.7	0.3	6	3778.834	50	0.8	0.2	2	3814.507
50	0.4	0.2	4	3780.327	50	0.6	0.3	42	3818.202
50	0.7	0.3	20	3797.909	50	0.8	0.3	10	3819.057
50	0.7	0.4	3	3800.592	50	0.7	0.4	45	3820.331
50	0.9	0.4	37	3801.734	50	0.9	0.1	11	3820.392
50	0.9	0.2	36	3804.538	50	0.7	0.1	46	3820.961
50	0.9	0.3	9	3808.179	50	0.8	0.1	21	3822.238

Table 5: The results of the performance measures for GA-ANN.

	Training	Testing
MAPE	0.078	0.069
MAE	53.183	47.643
MSE	3765.061	3321.312
RMSE	61.360	57.631
R^2	0.716	0.736

local best solution ($C1$), the weighting coefficient for the global best solution ($C2$), and the number of neurons are used. The number of populations is considered as 50. Three different weighting coefficients for the local best solution are used as 1.0, 1.5, and 2.0. Nine different weighting coefficients for the global best solution are also utilized that is between 1 and 5 as 0.5 increasing. The number of neurons is increased

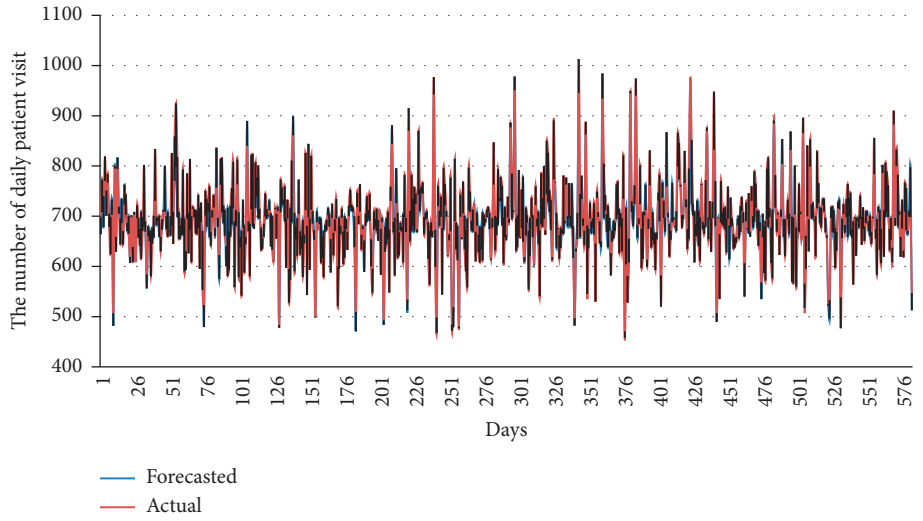


Figure 9: The actual and forecasted daily patient visits of GA-ANN for training.

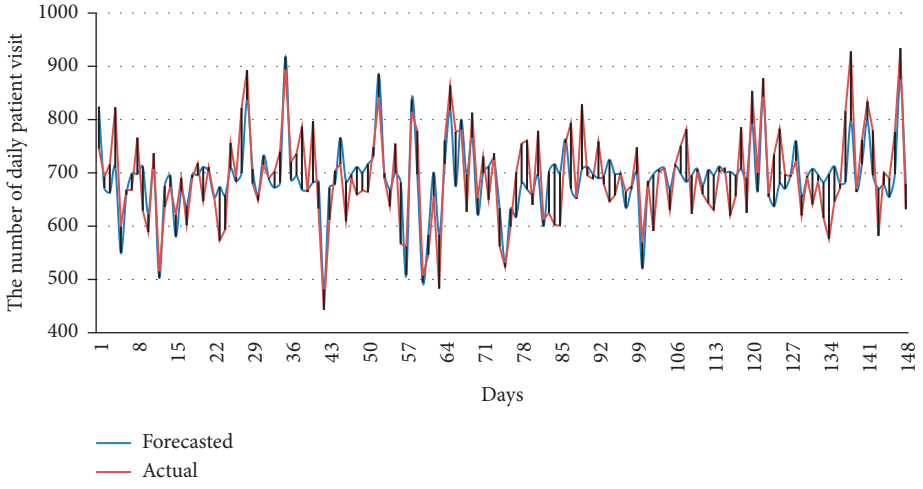


Figure 10: The actual and forecasted daily patient visits of GA-ANN for testing.

Table 6: The trials for PSO-ANN for the best twenty combinations for testing.

Population	$C1$	$C2$	$\#n$	MSE	Population	$C1$	$C2$	$\#n$	MSE
50	**1**	**1.5**	**25**	**2499.34**	50	1	1.5	15	2653.19
50	1	1.5	9	2554.78	50	1.5	1	16	2656.31
50	1.5	1.5	24	2569.01	50	1	1.5	39	2658.15
50	1.5	1	8	2572.10	50	1	1	20	2661.10
50	1.5	1.5	10	2582.99	50	1.5	1.5	5	2661.90
50	1	1.5	11	2587.58	50	1	1	19	2666.78
50	1	1.5	23	2593.40	50	1.5	1	31	2670.81
50	1	1.5	19	2597.32	50	1.5	1	4	2672.37
50	1	1.5	27	2614.94	50	1.5	1	29	2673.48
50	1	2	10	2620.29	50	1	1	17	2676.55
50	1.5	1.5	11	2626.79	50	1.5	1.5	8	2679.84
50	1	1.5	20	2642.21	50	1	1.5	16	2680.05
50	1.5	1.5	12	2647.97					

Table 7: The results of the performance measures for PSO-ANN.

	Training	Testing
MAPE	0.063	0.069
MAE	42.797	47.643
MSE	2499.340	3321.312
RMSE	49.993	57.631
R^2	0.824	0.736

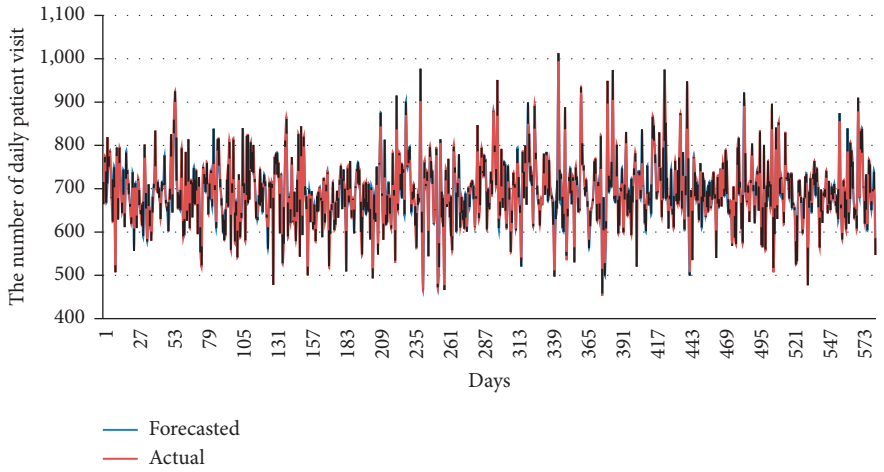


Figure 11: The actual and forecasted daily patient visits of PSO-ANN for training.

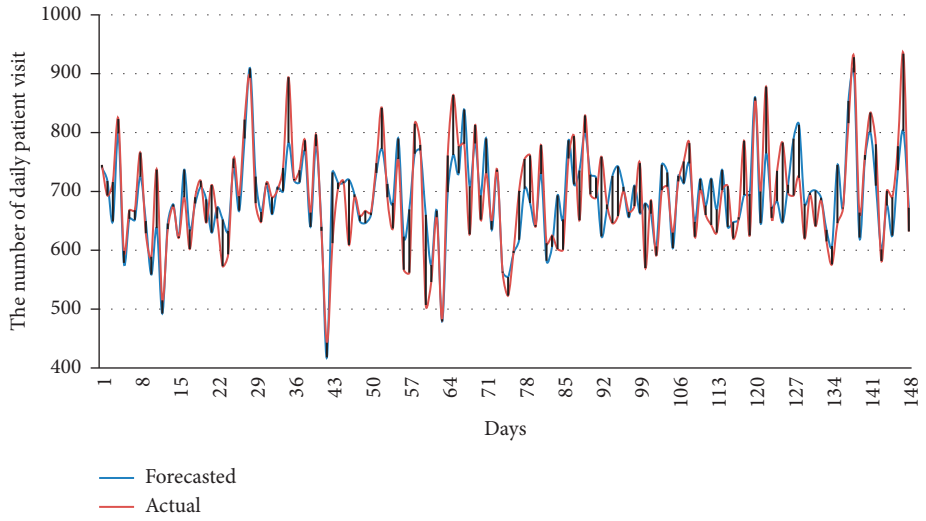


Figure 12: The actual and forecasted daily patient visits of PSO-ANN for testing.

one by one, ranging from 2 to 50 as in PSO-ANN. In the ANN process, the tansig are used for both transfer and activation functions. Therefore, 1182 different solutions are obtained for PSO-ANN. The best twenty combinations with respect to MSE are presented in Table 6.

After obtaining the best configuration for PSO-ANN with respect to the MSE, the results of five performance measures are presented in Table 7. The results of the training and testing are also illustrated in Figures 11 and 12, respectively. The obtained MAPE for training and testing for PSO-ANN is 6.3% and 6.9%, respectively. The other performance measures are also given in Table 7.

After the models are implemented for the forecasting of daily patient visits in emergency departments, their results are compared and evaluated together through the performance measures. A detailed evaluation of the proposed models based on both training and the testing dataset is implemented in this section. Table 8 presents the forecasting

Table 8: The results of three approaches in training and testing processes.

	Training									
	MAPE	Rank	MAE	Rank	MSE	Rank	RMSE	Rank	R^2	Rank
Bayesian ANN	0.073	2	50.016	2	4246.407	3	65.164	3	0.672	3
GA-ANN	0.078	3	53.183	3	3765.061	2	61.360	2	0.716	2
PSO-ANN	**0.063**	**1**	**42.797**	**1**	**2499.340**	**1**	**49.993**	**1**	**0.824**	**1**
	Testing									
	MAPE	Rank	MAE	Rank	MSE	Rank	RMSE	Rank	R^2	Rank
Bayesian ANN	0.088	3	60.358	3	7031.078	3	83.852	3	0.343	3
GA-ANN	0.069	2	47.643	2	3321.312	2	57.631	2	0.736	2
PSO-ANN	**0.060**	**1**	**40.888**	**1**	**2839.998**	**1**	**53.292**	**1**	**0.791**	**1**

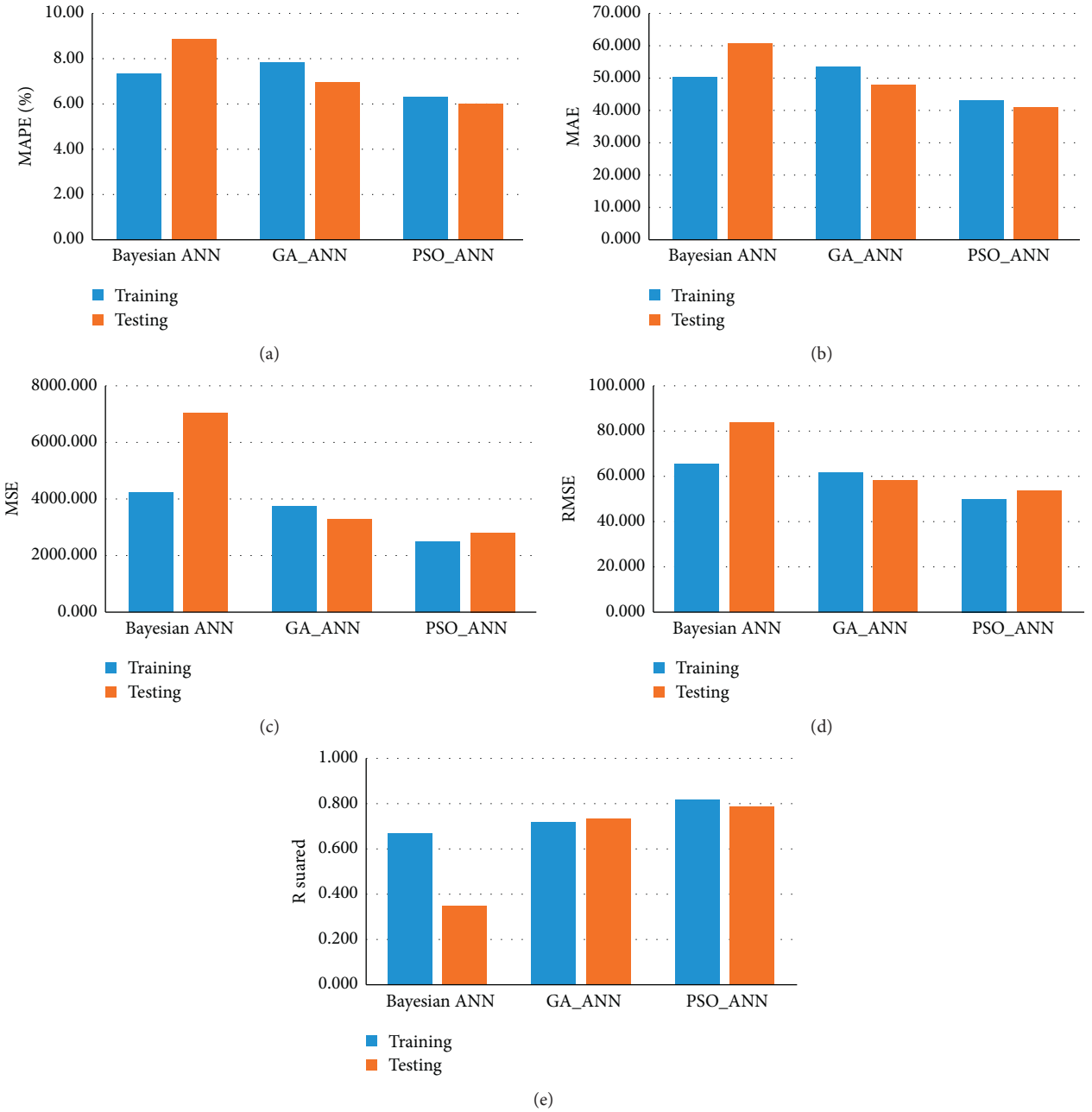


Figure 13: The comparison of approaches with respect to five performance measures: (a) MAPE, (b) MAE, (c) MSE, (d) RMSE, and (e) R-squared.

Table 9: Comparison of previous studies from the literature.

Study	Data	Method(s)	Compared methods	Time frame	Variables	Performance measure	Results
Choudhury and Urena [6]	Data from January 2014 to August 2017	TBATS, Holt–Winters, neural net, and ARIMA	—	Hourly	Only temporal variables	ME and RMSE	1.75, 1.19, 1.40, and 1.00 (ME for each model, respectively); 2.28, 27.86, 3.26, and 1.55 (RMSE for each model, respectively)
Yousefi et al. [20]	Data from January 2014 to November 2016	Long short-term memory (LSTM)	MLR, ARIMA, SVR, GLM, GEE, SARIMA, and ARIMA-LR	Daily	Weekend, holiday, soccer match day, the day after holiday, and the day before holiday	MAPE and R-squared	5.55% (MAPE) and 0.940 (R-squared)
Zhang et al. [26]	Data from January 1, 2013, to December 31, 2016	ARIMA-SVR hybrid approach	ARIMA and SVR	Daily	Only temporal variables	MAPE, RMSE, and MAE	7.02% (MAPE), 19.20 (RMSE), and 14.97 (MAE)
Jilani et al. [23]	Data between Jan 2011 and December 2015 from four hospitals	A modified heuristics based on a fuzzy time-series model	ARIMA and ANN	Weekly and monthly	Only temporal variables	MAPE and RMSE	2.5% to 7% (MAPE-daily) and 2.09% to 2.81% (MAPE-monthly)
Khaldi et al. [14]	Seven years of aggregated weekly demand from 2010 to 2016	ANN with ensemble empirical mode decomposition (EEMD)	ANN with discrete wavelet transform (DWT) decomposition, ANN, and ARIMA	Weekly	Only temporal variables	RMSE, MAE, and R (correlation coefficient)	52.86 (RMSE), 39.88 (MAE), and 0.96 (R)
Tideman et al. [31]	Seven years of historical daily ED arrivals	Least absolute shrinkage and selection operator (LASSO) regression	—	Daily	Climatic variables, Google trends, and calendar variables	MAPE, RMSE, R-squared, and percent absolute percent error (PAPE)	7.58% to 10.99% (MAPE), 12.08 to 16.73 (RMSE), 0.13 to 0.57 (R-squared), and 3.29% to 11.23% (PAPE)
Carvalho-Silva et al. [22]	Data for ED arrivals in 2 years (2012-2013)	ARIMA	Moving average, multiplicative winters, Holt–Winters, and exponential smoothing	Daily	Only temporal variables	MAPE	5.92% to 10.63% (MAPE)
Sarıyer [11]	Data between 01/12/2016 and 28/02/2017	ARIMA	—	Daily	Only temporal variables	MAPE	5.01% to 8.16% (MAPE)
Juang et al. [15]	Monthly ED visits from January 2009 to December 2016	ARIMA	—	Monthly	Only temporal variables	MAPE	8.91% (MAPE)
Xu et al. [27]	Daily ED visits from January 1, 2012, to December 31, 2013	ARIMA-LR hybrid approach	GLM, ARIMA, ARIMAX, and ARIMA-ANN	Daily	Calendar, holiday and temperature variables	MAPE and RMSE	6.5% to 9.3% (MAPE of ED-1), 12.3% to 13.1% (MAPE of ED-2), 67.1 to 98.2 (RMSE of ED-1), and 5.49 to 5.73 (RMSE of ED-2)

Table 9: Continued.

Study	Data	Method(s)	Compared methods	Time frame	Variables	Performance measure	Results
Calegari et al. [28]	Period from January 1, 2013, to May 31, 2015	SS, SMHW, SARIMA, and MSARIMA	—	Daily	Calendar and climatic variables	MAPE	2.91% to 11.16% (MAPE)
Current study	Two years' data from January 1, 2011, to December 31, 2012	Bayesian ANN, GA-based ANN, and PSO-based ANN	—	Daily	Calendar, holiday, and temperature variables	MAPE, MAE, MSE, RMSE, and R-squared	6% to 8.8% (MAPE), 40.888 to 60.358 (MAE), 2499.340 to 7031.078 (MSE), 49.993 to 83.852 (RMSE), and 0.343 to 0.824 (R-squared)

results of daily patient visits by the metaheuristic approaches integrated with ANN. The results of the performance measures are also illustrated in Figure 13. From Table 8, it can be seen that the PSO-ANN model provided the most dominant performance in the both the training and testing process. It obtained the lowest error with an MAPE of 6.3%, MAE of 42.797, MSE of 2499.340, RMSE of 49.933, and R^2 of 0.824 on the training dataset. It also obtained the lowest error with an MAPE of 6.0%, MAE of 40.888, MSE of 2839.998, RMSE of 53.292, and R^2 of 0.791 on the testing dataset. The Bayesian ANN and GA-ANN metaheuristics algorithms yielded lower performance in the ANN model optimization in the training process. The weakest model in this optimization process is the Bayesian ANN for both the training and testing process.

As a creative contribution to the literature, a comparison with some previously published papers is performed to highlight the studies in terms of data, applied method(s), compared methods, time frame, independent variables, performance measures, and analysis results. A total of twelve studied including the current study are investigated under these dimensions. Most of the studies focus on the daily basis of ED visits. Some studies tackle the hourly [6], weekly [14, 23], and monthly trend of ED visits [15, 23]. The data-gathering period varied from one study to another. The current study used two years' data of daily arrivals. In light of the eleven studies from the literature summarized in Table 9, it is understood that the data collection period of the current study is sufficient. When analyzing the methodology used, the current study bridges the gap of the literature. Since it does not show any attempts that apply metaheuristics incorporated with ANN in forecasting daily ED visits, this study has novelty for this application domain.

In the work of Wargon et al. [17], variability in MAPE and RMSE ranging from 4.2% to 14.4% is considered acceptable in ED daily visit forecasting studies. In this context, the approaches used in the current study meet the criteria. The MAPE values obtained from the three approaches are between 6% and 8.8% that means acceptable.

6. Conclusions

In this study, we apply three approaches named Bayesian ANN, GA-based ANN, and PSO-based ANN to the daily ED visit forecasting problem. Two years of daily ED visit data are gathered to use in these models. Temporal, climatic, and holiday variables are used in the models as independent variables. Results of each model are analyzed under five different performance measures called MAPE, MAE, MSE, RMSE, and R-squared. Results of the approaches show that PSO-based ANN is superior according to all five performance measures. GA-based ANN yields more successful results compared to the Bayesian ANN model. We conclude that the use of metaheuristics integrated with ANN in ED visit forecasting improves the accuracy of the model considerably. The proposed approaches eliminate the problem of getting stuck in local extremums and crossing plateaus of the error function in classical ANN. The initial weights of the ANN are computed using GA and ANN instead of the trial and error process. There are some limitations. Firstly, the data come from only one institution and consist of two-year data. Second, the input variables for ED data are all generally utilized for model parameters in the ED literature. However, the number of parameters can be enriched to improve the accuracy of proposed approaches. Thirdly, the number of analyzed parameters of proposed approaches (Bayesian ANN, GA-ANN, and PSO-ANN) are limited. We have obtained a total of 2497 solutions with respect to the proposed approaches. A parameter optimization for the proposed approaches can be implemented. For future work, we plan to include more variables in our models and compare the performance of current models with some time-series and machine learning algorithms. This study is novel in the literature from the aspect of applying hybrid metaheuristics-based approaches for the first time. Findings of the current research will also contribute to ED decision makers in the practice to plan and schedule medical staff to reach an efficient resource planning and service quality.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] S.-Y. Lee, R. B. Chinnam, E. Dalkiran, S. Krupp, and M. Nauss, "Prediction of emergency department patient disposition decision for proactive resource allocation for admission," *Health Care Management Science*, vol. 23, no. 3, pp. 339–359, 2020.

[2] D. Bouzon Nagem Assad and T. Spiegel, "Improving emergency department resource planning: a multiple case study," *Health Systems*, vol. 9, no. 1, pp. 2–30, 2020.

[3] H. T. Karsanti, I. Ardiyanto, and L. E. Nugroho, "Deep learning-based patient visits forecasting using long short term memory," in *Proceedings of the 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT)*, pp. 344–349, IEEE, Yogyakarta, Indonesia, March 2019.

[4] J. Berglind, "Temporal convolutional networks for forecasting of patient volumes in a digital healthcare setting," *Degree Project in Computer Science and Engineering*, Second Cycle, Stockholm, Sweden, 2019.

[5] M. Gul and E. Celik, "An exhaustive review and analysis on applications of statistical forecasting in hospital emergency departments," *Health Systems*, vol. 9, no. 4, pp. 263–284, 2020.

[6] A. Choudhury and E. Urena, "Forecasting hourly emergency department arrival using time series analysis," *British Journal of Healthcare Management*, vol. 26, no. 1, pp. 34–43, 2020.

[7] M. Hertzum, "Forecasting hourly patient visits in the emergency department to counteract crowding," *The Ergonomics Open Journal*, vol. 10, no. 1, pp. 1–13, 2017.

[8] M. Ordu, E. Demir, and C. Tofallis, "A decision support system for demand and capacity modelling of an accident and emergency department," *Health Systems*, vol. 9, no. 1, pp. 31–56, 2020.

[9] M. Ordu, E. Demir, and C. Tofallis, "A comprehensive modelling framework to forecast the demand for all hospital services," *The International Journal of Health Planning and Management*, vol. 34, no. 2, pp. e1257–e1271, 2019.

[10] E. Pekel, M. Gul, and E. Celik, "Forecasting daily patient visits in an emergency department by GA-ANN hybrid approach," in *Proceedings of the 14th International Symposium on Operational Research*, pp. 473–478, SOR, Slovenia, Bled, September 2017.

[11] G. Sarıyer, "Acil servislerde talebin zaman serileri modelleri ile tahmin edilmesi," *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, vol. 10, no. 1, pp. 66–77, 2018.

[12] M. Yucesan, M. Gul, and E. Celik, "A multi-method patient arrival forecasting outline for hospital emergency departments," *International Journal of Healthcare Management*, vol. 13, no. 1.

[13] M. Yucesan, M. Gul, S. Mete, and E. Celik, "A forecasting model for patient arrivals of an emergency department in healthcare management systems," in *Intelligent Systems for Healthcare Management and Delivery*, pp. 266–284, IGI Global, Hershey, PA, USA, 2019.

[14] R. Khaldi, A. E. Afia, and R. Chiheb, "Forecasting of weekly patient visits to emergency department: real case study," *Procedia Computer Science*, vol. 148, pp. 532–541, 2019.

[15] W.-C. Juang, S.-J. Huang, F.-D. Huang, P.-W. Cheng, and S.-R. Wann, "Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in southern Taiwan," *BMJ open*, vol. 7, no. 11, Article ID e018628, 2017.

[16] M. Gul and A. F. Guneri, "Planning the future of emergency departments: forecasting ED patient arrivals by using regression and neural network models," *International Journal of Industrial Engineering: Theory, Applications and Practice*, vol. 23, no. 2, pp. 137–154, 2016.

[17] M. Wargon, B. Guidet, T. D. Hoang, and G. Hejblum, "A systematic review of models for forecasting the number of emergency department visits," *Emergency Medicine Journal*, vol. 26, no. 6, pp. 395–399, 2009.

[18] F. Kadri, F. Harrou, S. Chaabane, and C. Tahon, "Time series modelling and forecasting of emergency department overcrowding," *Journal of Medical Systems*, vol. 38, no. 9, p. 107, 2014.

[19] S. Nas and M. Koyuncu, "Emergency department capacity planning: a recurrent neural network and simulation approach," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 4359719, 13 pages, 2019.

[20] M. Yousefi, M. Yousefi, M. Fathi, and F. S. Fogliatto, "Patient visit forecasting in an emergency department using a deep neural network approach," *Kybernetes*, vol. 49, no. 9, pp. 2335–2348, 2019.

[21] P. Aboagye-Sarfo, Q. Mai, F. M. Sanfilippo, D. B. Preen, L. M. Stewart, and D. M. Fatovich, "A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in western Australia," *Journal of Biomedical Informatics*, vol. 57, pp. 62–73, 2015.

[22] M. Carvalho-Silva, M. T. T. Monteiro, F. d. Sá-Soares, and S. Dória-Nóbrega, "Assessment of forecasting models for patients arrival at emergency department," *Operations Research for Health Care*, vol. 18, pp. 112–118, 2018.

[23] T. Jilani, G. Housley, G. Figueredo, P.-S. Tang, J. Hatton, and D. Shaw, "Short and long term predictions of hospital emergency department attendances," *International Journal of Medical Informatics*, vol. 129, pp. 167–174, 2019.

[24] A. Ekström, M. Nordberg, and O. Eriksson, "Shorter waiting time, better emergency healthcare: forecasting Stockholm's emergency department visits," *Model Assisted Statistics and Applications*, vol. 13, no. 4, pp. 377–385, 2018.

[25] D. Golmohammadi, "Predicting hospital admissions to reduce emergency department boarding," *International Journal of Production Economics*, vol. 182, pp. 535–544, 2016.

[26] Y. Zhang, L. Luo, J. Yang, D. Liu, R. Kong, and Y. Feng, "A hybrid ARIMA-SVR approach for forecasting emergency patient flow," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 8, pp. 3315–3323, 2019.

[27] Q. Xu, K.-L. Tsui, W. Jiang, and H. Guo, "A hybrid approach for forecasting patient visits in emergency department," *Quality and Reliability Engineering International*, vol. 32, no. 8, pp. 2751–2759, 2016.

[28] R. Calegari, F. S. Fogliatto, F. R. Lucini, J. Neyeloff, R. S. Kuchenbecker, and B. D. Schaan, "Forecasting daily volume and acuity of patients in the emergency department," *Computational and Mathematical Methods in Medicine*, vol. 2016, Article ID 3863268, 8 pages, 2016.

[29] I. Marcilio, S. Hajat, and N. Gouveia, "Forecasting daily emergency department visits using calendar variables and ambient temperature readings," *Academic Emergency Medicine*, vol. 20, no. 8, pp. 769–777, 2013.

[30] S. A. Jones, M. P. Joy, and J. Pearson, "Forecasting demand of emergency care," *Health Care Management Science*, vol. 5, no. 4, pp. 297–305, 2002.

[31] S. Tideman, M. Santillana, J. Bickel, and B. Reis, "Internet search query data improve forecasts of daily emergency department volume," *Journal of the American Medical Informatics Association*, vol. 26, no. 12, pp. 1574–1583, 2019.

[32] D. M. Himmelblau, "Accounts of experiences in the application of artificial neural networks in chemical engineering," *Industrial & Engineering Chemistry Research*, vol. 47, no. 16, pp. 5782–5796, 2008.

[33] O. Onat and M. Gul, "Application of artificial neural networks to the prediction of out-of-plane response of infill walls subjected to shake table," *Smart Structures and Systems*, vol. 21, no. 4, pp. 521–535, 2018.

[34] B. S. Waziri, K. Bala, and S. A. Bustani, "Artificial neural networks in construction engineering and management," *International Journal of Architecture, Engineering and Construction*, vol. 6, no. 1, pp. 50–60, 2017.

[35] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: a survey," *Heliyon*, vol. 4, no. 11, Article ID e00938, 2018.

[36] U. J. Frey and H. Rusch, "Using artificial neural networks for the analysis of social-ecological systems," *Ecology and Society*, vol. 18, no. 2, 2013.

[37] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: a scoping review," *PLoS One*, vol. 14, no. 2, Article ID e0212356, 2019.

[38] D. L. Reilly and L. N. Cooper, "An overview of neural networks: early models to real world systems," in *How We Learn; How We Remember: Toward an Understanding of Brain and Neural Systems: Selected Papers of Leon N Cooper*, pp. 300–321, World Scientific, Singapore, 1995.

[39] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: the state of the art," *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[40] G. B. Kingston, M. F. Lambert, and H. R. Maier, "Bayesian training of artificial neural networks used for water resources modeling," *Water Resources Research*, vol. 41, no. 12, pp. 1–11, 2005.

[41] L. T. Le, H. Nguyen, J. Dou, and J. Zhou, "A comparative study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings' energy efficiency for smart city planning," *Applied Sciences*, vol. 9, no. 13, p. 2630, 2019.

[42] A. Kadiyala, D. Kaur, and A. Kumar, "Development of hybrid genetic-algorithm-based neural networks using regression trees for modeling air quality inside a public transportation bus," *Journal of the Air & Waste Management Association*, vol. 63, no. 2, pp. 205–218, 2013.

[43] E. Pekel and S. S. Kara, "Passenger flow prediction based on newly adopted algorithms," *Applied Artificial Intelligence*, vol. 31, no. 1, pp. 64–79, 2017.

[44] E. Pekel and S. Soner Kara, "Solving capacitated location routing problem by variable neighborhood descent and GA-artificial neural network hybrid method," *Promet—Traffic & Transportation*, vol. 30, no. 5, pp. 563–578, 2018.

[45] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39–43, IEEE, Nagoya, Japan, October 1995.

[46] M. Abbaspour Onari, S. Yousefi, and M. Jahangoshai Rezaee, "Risk assessment in discrete production processes considering uncertainty and reliability: *Z*-number multi-stage fuzzy cognitive map with fuzzy learning algorithm," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 1349–1383, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5021868, 11 pages
https://doi.org/10.1155/2021/5021868

Research Article

Experimental Microemulsion Flooding Study to Increase Low Viscosity Oil Recovery Using Glass Micromodel

Hamed Hematpur,[1] **Reza Abdollahi**,[1] **Mohsen Safari-Beidokhti**,[1] **and Hamid Esfandyari**[2]

[1]*EOR Research Department, Research Institute of Petroleum Industry, Tehran, Iran*
[2]*Abadan Faculty of Petroleum Engineering, Petroleum University of Technology, Abadan, Iran*

Correspondence should be addressed to Hamid Esfandyari; esfandyari_shirazu@yahoo.com

Received 19 May 2021; Accepted 4 September 2021; Published 13 September 2021

Academic Editor: Mohammad Yazdi

The growing demand for clean energy can be met by improving the recovery of current resources. One of the effective methods in recovering the unswept reserves is chemical flooding. Microemulsion flooding is an alternative for surfactant flooding in a chemical-enhanced oil recovery method and can entirely sweep the remaining oil in porous media. The efficiency of microemulsion flooding is guaranteed through phase behavior analysis and customization regarding the actual field conditions. Reviewing the literature, there is a lack of experience that compared the macroscopic and microscopic efficiency of microemulsion flooding, especially in low viscous oil reservoirs. In the current study, one-quarter five-spot glass micromodel was implemented for investigating the effect of different parameters on microemulsion efficiency, including surfactant types, injection rate, and micromodel pattern. Image analysis techniques were applied to represent the phase saturations throughout the microemulsion flooding tests. The results confirm the appropriate efficiency of microemulsion flooding in improving the ultimate recovery. LABS microemulsion has the highest efficiency, and the increment of the injection rate has an adverse effect on oil recovery. According to the pore structure's tests, it seems that permeability has little impact on recovery. The results of this study can be used in enhanced oil recovery designs in low-viscosity oil fields. It shows the impact of crucial parameters in microemulsion flooding.

1. Introduction

The decreasing trend of fossil resources discovery in parallel with the expanding energy demand raises the eminence of techniques and approaches (reservoir-based and well-based) applied to improve the recovery of current resources [1–10]. Approximately, 70% of the reserves are not producible naturally [11–15]. This significant trapped reserve is an attractive target for applying enhanced oil recovery (EOR) methods [16–23].

Generally, EOR techniques are majorly categorized into thermal and nonthermal. The main subdivisions of nonthermal EOR are miscible and chemical techniques. Chemical EOR techniques refer to the application of standalone or combination of different materials (caustic, surfactant, polymer, and micellar/emulsion). These techniques substantially enhance the recovery by improving the mobility ratio or by decreasing the saturation of residual oil. Many parameters, including mineralogy, permeability, viscosity ranges, temperature, and salinity, affect the efficiency of chemical flooding [4, 24–28].

One of the effective chemical EOR methods is microemulsion flooding [29, 30]. This method was first suggested as an alternative for surfactant flooding. It is more efficient in comparison to surfactants [30]. The microemulsion solution consists of three main components, including water, hydrocarbon, and surfactant. Two other components (cosurfactant and electrolyte) may be added to modify the solution according to the reservoir conditions. This multiphase microemulsion can be water external, oil external, or a three-phase solution consisting of water, hydrocarbon, and surfactant [4, 24, 25]. Theoretically, this method is capable of increasing the recovery up to 100%. Microemulsions can remarkably induce the interfacial tension (IFT) to values of

10^{-2}–10^{-3} mN m^{-1}. The IFT reduction arouses spontaneous emulsification and consequently unswept oil displacement [29]. The main detrimental phenomenon that affects the efficiency of microemulsion flooding is the retention of surfactant molecules on the reservoir rock. Other factors that affect microemulsion flooding include salinity and surfactant concentration [31–33].

Microemulsion can be flooded in two general approaches. First, the solution is prepared on the surface and then injected [30, 34]. In the second method, the microemulsion is generated in the reservoir by continuous injection of surfactants [35]. Microemulsion flooding needs a high concentration of surfactants, and it is the main restricting factor for the field application of this technique [29, 36–38].

So far, many attempts have been made to investigate all prospects of microemulsion flooding. The equilibrium condition of microemulsion along the excess oil, water, or both was studied by Winsor [39]. Winsor categorized the microemulsion system into four types (Figure 1):

Type I systems are made of an oil-water microemulsion that coexists with an additional oil phase

Type II systems, an oil-water microemulsion, exist with an extra water phase

Type III systems, three discrete phases of microemulsion, water, and oil coexist together

Type IV systems, a homogeneous and single phase of microemulsion exists

Healy and Reed investigated some microemulsion fundamentals (IFT, salinity, and viscosity). They related the phase behavior results to Winsor's concepts [40]. Holm used sodium sulfonates to produce microemulsion for the EOR test and reported the effect of temperature on microemulsion efficiency [41]. Healy et al. conducted the microemulsion flooding (with the addition of monoethanolamine salt of alkylorthoxylene sulfonic acid) and stated the impact of surfactant retention on the reservoir rock [42]. The influence of water salinity on surfactant retention and the consequent phase behavior was studied by Glover et al. [31]. Verkruyse and Salter showed that although the microemulsion (with the addition of ethoxylated alcohols) could considerably reduce the IFT, the ultimate recovery was not changed remarkably [43]. Bouabboune et al. compared surfactant with microemulsion flooding (alkali surfactant NM as a surfactant) and concluded that the microemulsion had a better performance [44]. Mandal et al. utilized the nanoemulsion in the EOR process (with the addition of Tergitol), and they found that this solution can improve the ultimate recovery by more than 30% [45]. Karambeigi et al. studied the phase behavior of different solutions (including surfactant brine, biodiesel, and cosolvent) and the consequent efficiencies in carbonated rock [46]. Hu et al. used nanoparticles (iron oxide) for microemulsion and evaluated their application in EOR. They reported that this system significantly increased the oil recovery reaching 28.9% and presented a more stable pressure profile [47]. Ferreira et al. prepared microemulsion with glycerin, and the prepared solution was injected in a sandstone rock. The maximum attained recovery in this system was 49% [48]. Castro Dantas et al. examined the effect of acid microemulsion on EOR in carbonate reservoirs, and the maximum reported recovery was 30% [49].

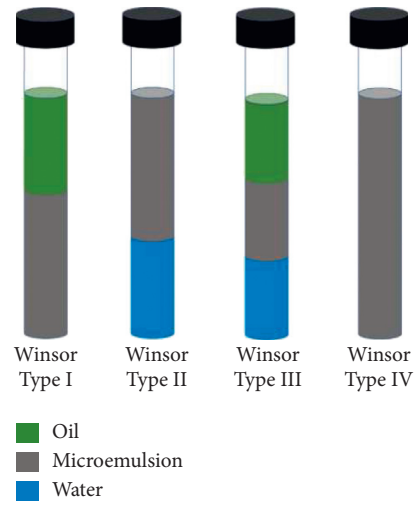


Figure 1: Classified microemulsion systems by Winsor [39].

One of the main methods that have been widely used for visualizing chemical flooding is the micromodel. This setup consists of an etched flow pattern that is visible by a microscope. The pattern can be synthetic or extracted from a real pore stricter. The visualization of fluid flow is very constructive in understanding the mechanism of the EOR process at the pore level. However, the visual micromodel encounters many limitations. The most important one is the difference in the flow pattern in the micromodel and field porous media [50–56]. It can be mentioned that few studies have been appropriately applied to quantify microemulsion flooding in micromodel experiments.

In this study, the one-quarter five-spot glass micromodel was used to investigate the effect of different factors including, solution injection rate, types of solution, and micromodel structure, on the performance of microemulsion flooding. In the end, they compared with each other to suggest the best conditions for one of the low viscous Iranian oil reservoirs.

2. Experimental Setup and Procedure

2.1. Low-Pressure Micromodel Apparatus. Figure 2 shows the micromodel setup used for microemulsion flooding in this study. All experiments were conducted at atmospheric pressure.

Six microemulsion flooding tests were performed to study the impact of the following parameters:

Microemulsion type

Injection rate

Pore structure

For each microemulsion flooding experiment, the following steps were repeated.

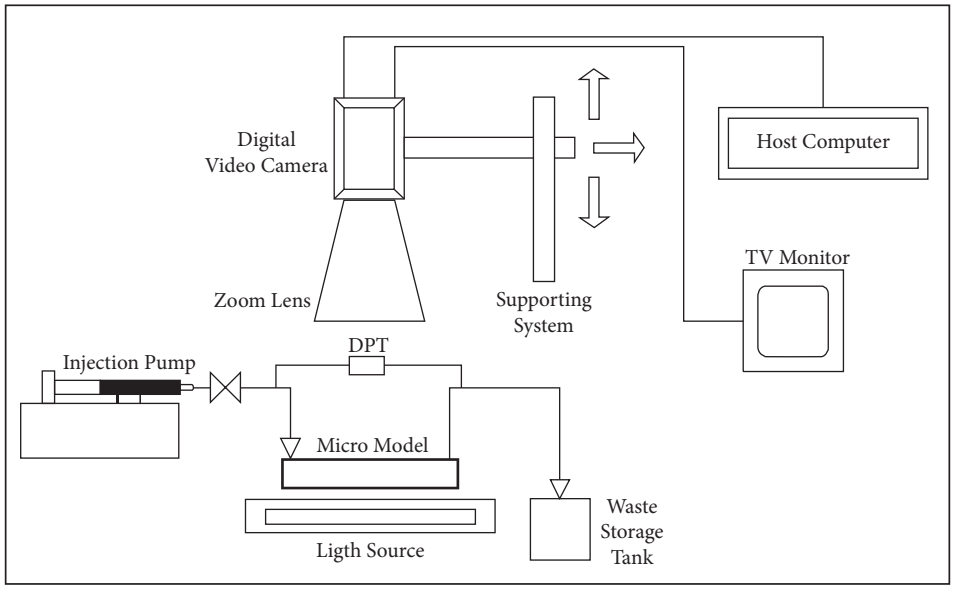


Figure 2: Schematic of micromodel assembly.

The first step was the saturation of the micromodel with brine. The saturated media was then flooded by oil to reach the initial water saturation. The saturated micromodel was kept for two weeks to restore wettability. Finally, the microemulsion was injected into the micromodel at a constant rate. The recoveries were recorded throughout the flooding process. The microscope of the zooming range of 200X to 500X was used to take pictures of high resolution to observe oil, connate water, solvent saturation, and asphaltene deposition after each experiment.

2.2. Crude Oil Properties. Experiments were conducted on the low viscous oil samples prepared from an Iranian oil field. The oil properties are represented in Table 1.

2.3. Synthetic Brine. Regarding the composition and properties of real field water samples, the synthetic brine (136000 ppm) with the following composition was prepared (Table 2).

2.4. Cleaning Fluid. Before conducting each experiment, the micromodel was sequentially washed with the injection of toluene, alcohol, and distilled water.

2.5. Microemulsion. Two solutions were prepared for investigating the efficiency of microemulsion in EOR improvement. The coconut fatty acid diethanolamide (cocamide DEA) [57] and linear alkylbenzenes (LABS) were used as surfactant and n-hexane as a hydrocarbon. The critical micelle concentration (CMC) of these two surfactants is 1.4% w/w and 0.65% w/w, respectively. Figure 3 represents the structure of these surfactants.

To construct the ternary phase diagram, different concentrations (volume percent) of brine, surfactant, and n-hexane are mixed. The volume measurements were done primarily on a mass scale, considering the density of the components at the desired temperature. Different phases (Winsor microemulsion types) are recorded, and consequently, the ternary diagram is plotted considering the phases and volumes. Since the salinity of brine is fixed for injection (produced formation water), the sensitivity on salinity is not in the scope of this study.

Table 1: Oil properties.

Temperature: 119°C		
Bubble point pressure (psia)	API	Viscosity (cp)
1959.7	43.6	0.554

Table 2: Synthetic brine composition (136000 ppm).

Component	Molecular weight (g/mol)	Concentration (g/lit)
$MgCl_2.6H_2O$	203.3	1
$CaCl_2.2H_2O$	147.03	3
Na_2SO_4	142.04	30
NaCl	58.44	102

The phase diagram of microemulsion containing cocamide DEA is given in Figure 4.

This solution has different behaviors in comparison to the ideal behavior presented by Green and Willhite [58]. It included several multiphase and single-phase regions. The volume percent of each component was estimated regarding the single-phase behavior of microemulsion.

Hadi et al. applied the cocamide DEA emulsion for heavy oil [57]. For light oil with high water saturation, it is better to apply a new microemulsion. In the second microemulsion, similar to Hadi's study [57], n-hexane was used as hydrocarbon, but LABS as surfactant.

Several titrations were done to obtain the phase behavior of LABS. As shown in Figure 5, the result was closer to the ideal behavior in comparison to the previous microemulsion. Also, laboratory pictures from results of one, two, and three phases are given in this figure.

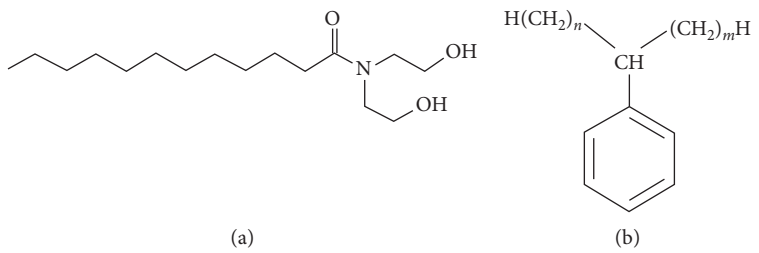


Figure 3: The chemical structure of cocamide DEA and LABS. (a) Cocamide DEA. (b) LABS.

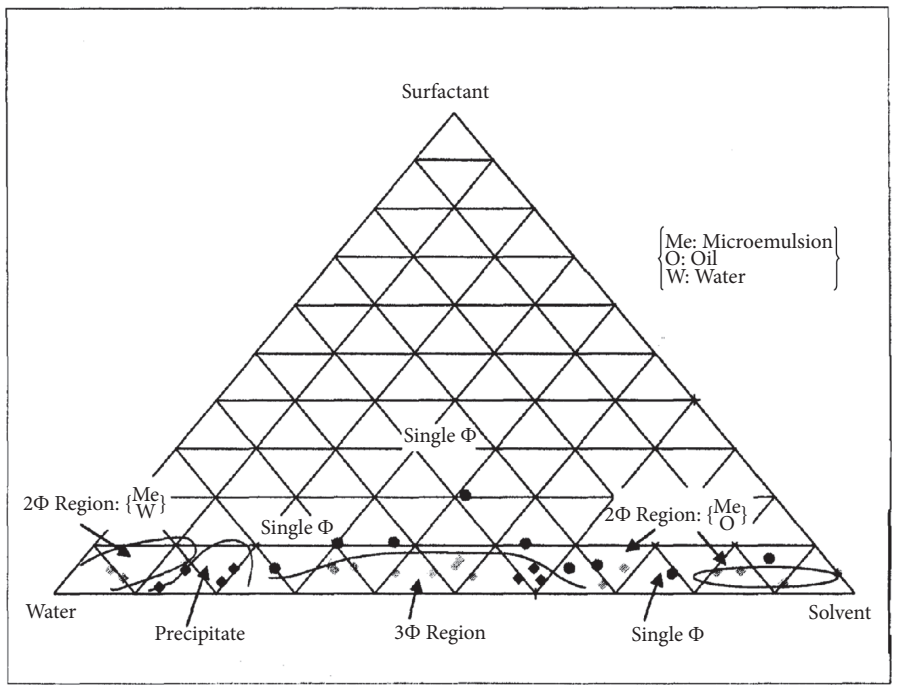


Figure 4: Ternary diagram for cocamide DEA [57].

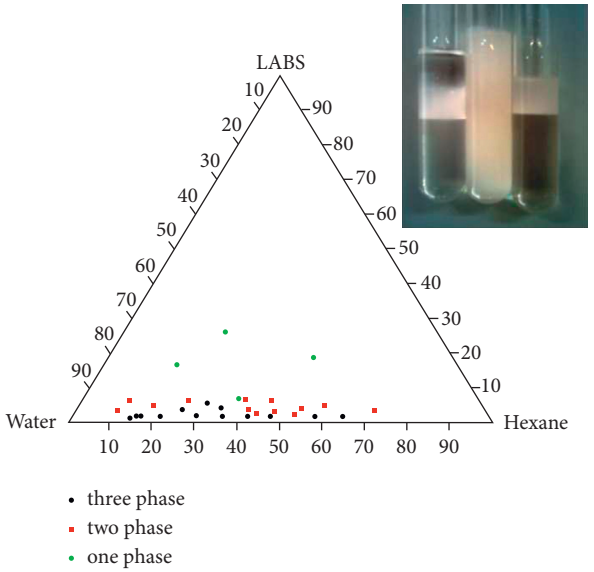


Figure 5: The ternary diagram for LABS.

The final compositions chosen for injection are given in Table 3. The composition and the volume percentage of constitute materials were selected regarding the following criteria:

Single-phase behavior

Minimum possible surfactant concentration (economic criterion)

Desirable viscosity

2.6. Pattern Design. In this study, two synthetic hexagonal patterns with different coordination numbers and pore throat ratios were used for microemilsion tests. The patterns are shown in Figures 6 and 7 (the permeability of A and B patterns are 1500 and 2000 md, respectively).

Each pattern (with 6.5 ∗ 6.5 cm area) was etched on a thick float plate (dimensions: 4 ∗ 7 ∗ 0.183 inches) and covered with the same size plate. The cover had two ports for fluid injection and production (with 3 mm diameter). The etched and cover plates were connected and sealed to form an enclosed pore space.

3. Result and Discussion

3.1. Effect of Microemulsion Type. Pattern B was used to evaluate the efficiencies of suggested microemulsions. The injection rate of microemulsion into the saturated micromodel was 0.0008 cc/min. Microemulsion A consists of hexane (hydrocarbon), LABS (surfactant), and brine. Microemulsion B consists of hexane (hydrocarbon), coconut fatty acid diethanolamide (surfactant), and brine. The main mechanism for the recovery improvement is the reduction of IFT. Once the microemulsion front reaches the remained oil, it lowers the IFT, and consequently, the capillary force decreases. Capillary force plays an important role in phase trapping through the porous media, and reduction in capillary force leads to higher recovery. Viscous force is another force that plays a role in hydrocarbon recovery. Increasing the viscosity of injection fluid (lower mobility ratio) improves the recovery factor. Results show that type A has better efficiency than type B due to more IFT reduction and more elevated viscosity. However, this difference is not significant after 1 PV of injection. The main difference between these two microemulsions is breakthrough time. Type B shows a higher breakthrough time (0.93 PV) in comparison to type A (0.31 PV) because it has a higher amount of hydrocarbon than type A. It seems that the increase of hydrocarbon concentration in microemulsion ascends the breakthrough time. However, the increase of breakthrough time does not affect the ultimate recovery significantly.

Figure 8 illustrates recoveries vs. the number of pore volume injections. According to this figure, it can be concluded that microemulsion A has a higher recovery than type B. Furthermore, type A has lower hydrocarbon content, and it can be the advantage of this type.

3.2. Effect of Injection Rate. In order to investigate the effect of injection rate on ultimate recovery, two different rates (0.0008 and 0.001 cc/min) were tested. Pattern B and microemulsion type A were selected for these experiments, and other conditions were the same as before. The results show that breakthrough is highly dependent on the rate of injection. In the case of changing the injection rate, the viscous force plays the main role. Increasing the injections' rate leads to a higher probability of viscous fingering throughout the flooding period. The viscous fingering brings the earlier breakthrough and lower recovery. This analysis is based on the pore volume of the injected microemulsion, not on the time of injection. Comparing the breakthrough according to the time of injection might be inaccurate. The breakthrough time for 0.0008 and 0.001 cc/min is 0.31 and 0.26 PV, respectively.

Figure 9 shows that, by increasing the injection rate, the recovery goes down. In addition, the graph shows that changing the injection rate does not have a sensible effect on the trend of the chart, and both scenarios have the same trend.

3.3. Effect of Pore Structure. Two patterns A and B, with different permeabilities, were used for evaluating the impact of flow networks on ultimate recovery of microemulsion flooding. Microemulsion type A with 0.0008 cc/min rate was injected in these tests. Figure 10 illustrates the effect of the pore structure and permeability on the consequent recovery. The permeability variation affects the viscous forces, so the IFT reduction mechanism for both cases is the same. However, the results showed that there was no obvious change in recovery by changing the permeability. Roughly two graphs in Figure 10 are the same. Since the IFT reduction is the main mechanism for microemulsion flooding and leads to a high recovery factor (about 80% in our study), the impact of permeability on the ultimate recovery can be negligible in high recovery factor values.

3.4. Qualitative Observation of Connate Water's Effect. Qualitative microscopic observations were conducted throughout the flooding period of each test. Figure 11 illustrates six stages of microemulsion (A) flooding in pattern A. The figure clearly shows the movement of connate water in the frontal path; also, it shows that the front moves very stable and sweeps all water and oil through its way. The high-viscosity and low-mobility ratio of microemulsion causes the front stability in this case. Figure 12 shows six stages of flooding with microemulsion (B). It shows that the front is not very stable, and viscous fingering occurs to some extent. The lower viscosity of this microemulsion is the reason for this behavior.

Figures 13 and 14 show the microscopic images to illustrate the mechanism of the connate water (trapped or made the film on grain) discharging during the microemulsion flooding.

Figure 13 shows when the microemulsion faces the connate water, the microemulsion causes IFT reduction, and consequently, connate water becomes movable. The water which is not in touch with the microemulsion cannot move easily because of high interfacial tension. Figure 14 depicts

Table 3: Composition of microemulsions.

Composition	Surfactant	Surfactant vol. (%)	n-Hexane vol. (%)	Water vol. (%)	Viscosity (cp)
Microemulsion (A)	**LABS**	6.7	56.1	37	8
Microemulsion (B)	**Cocamide DEA**	5.86	64.89	29.25	7.5

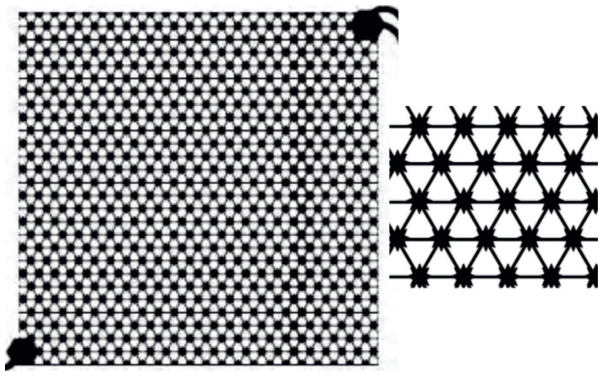


Figure 6: Triple hexagonal pattern (A) with the coordination number of 6 and pore-throat ratio of 4.

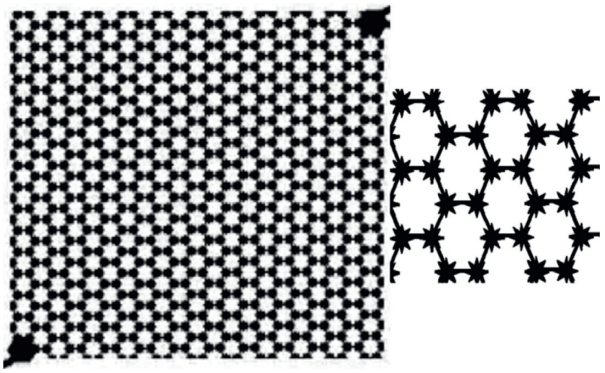


Figure 7: Single hexagonal pattern (B) with the coordination number of 3 and pore-throat ratio of 6.

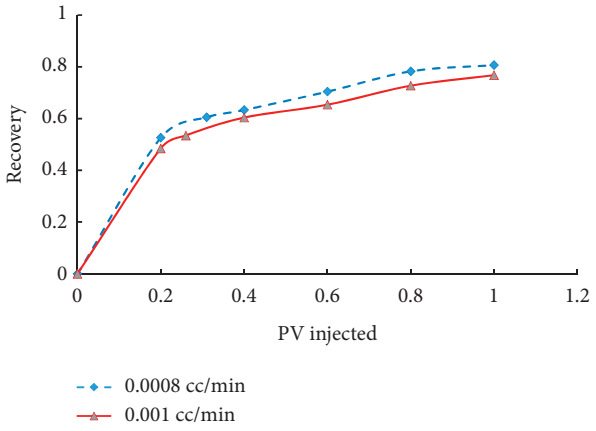


Figure 9: Effect of the injection rate on recovery.

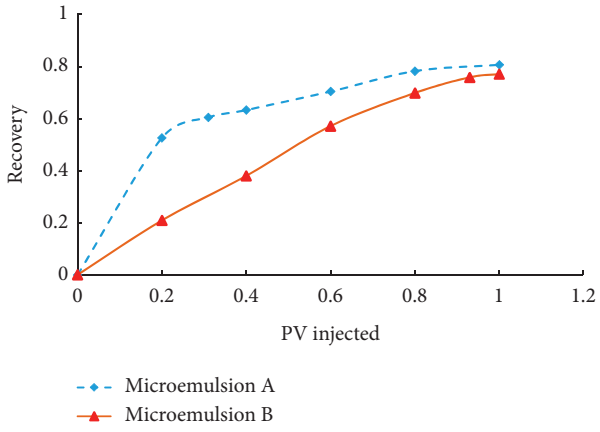


Figure 8: Effect of microemulsion types on recovery.

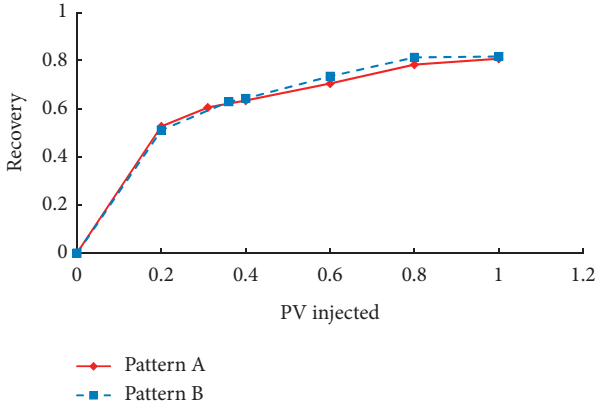


Figure 10: Effect of pore structure on oil recovery in the microemulsion flooding.

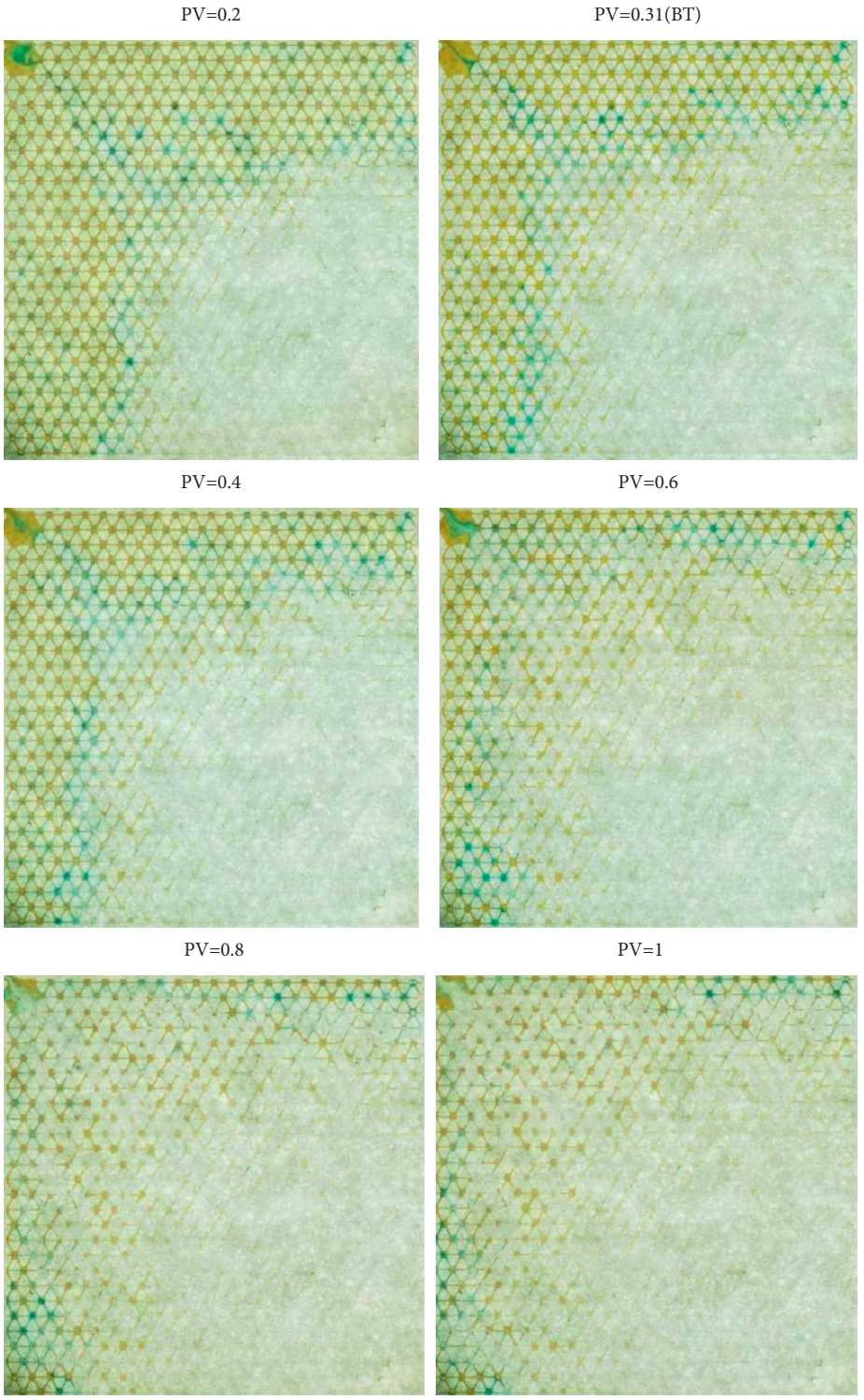


Figure 11: 6-stages of microemulsion flooding (A) (green: water, light brown: oil, and white: emulsion).

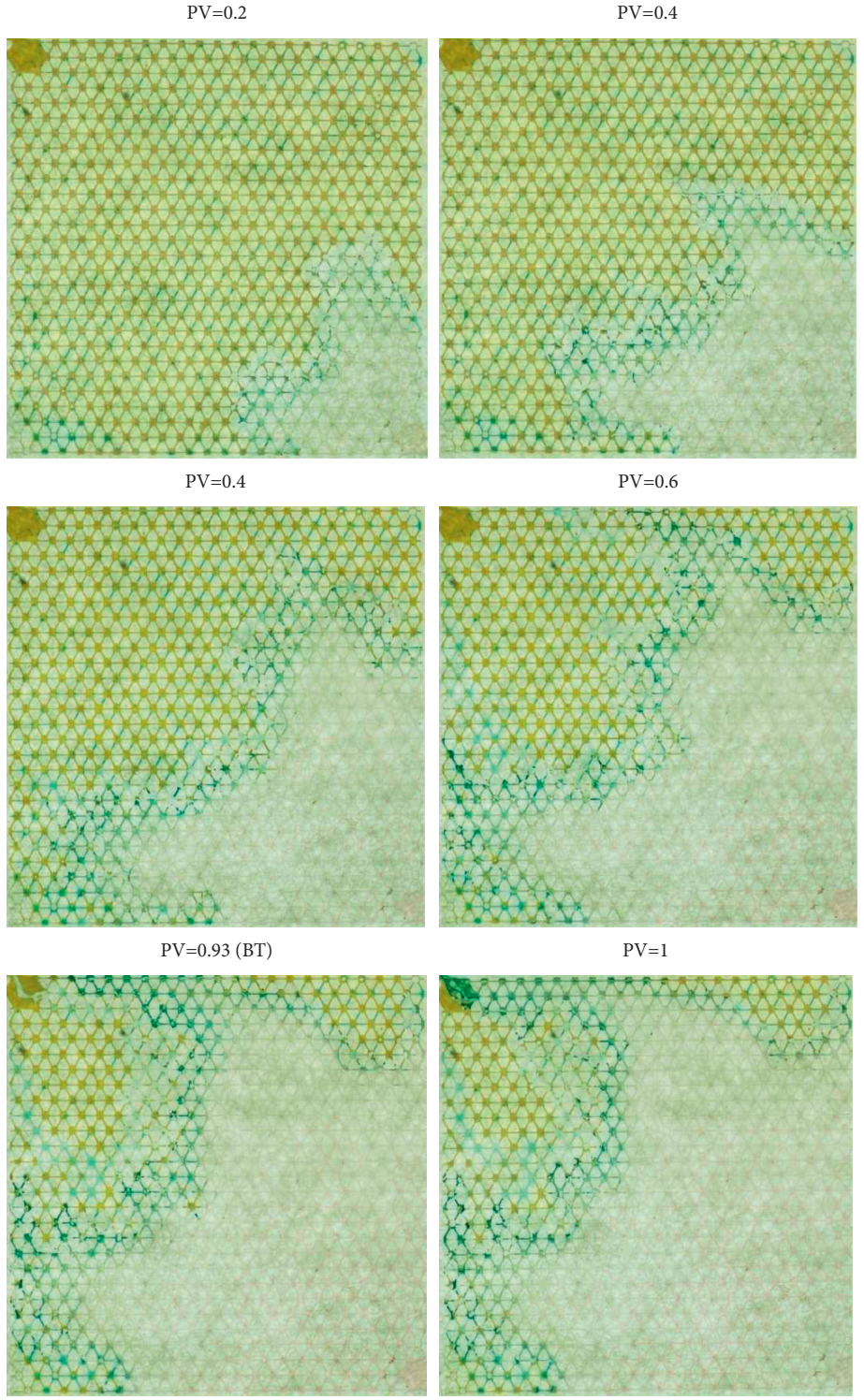


Figure 12: 6-stages of microemulsion flooding (B) (green: water, light brown: oil, and white: emulsion).

this state. In this pore, microemulsion does not reach the connate water.

These experiments show how microemulsion affects the water saturation in two sides of microemulsion front. Behind the front, the microemulsion changes the end-point of the residual water saturation by reducing the IFT value. Consequently, the swept water is moved to the head of the front.

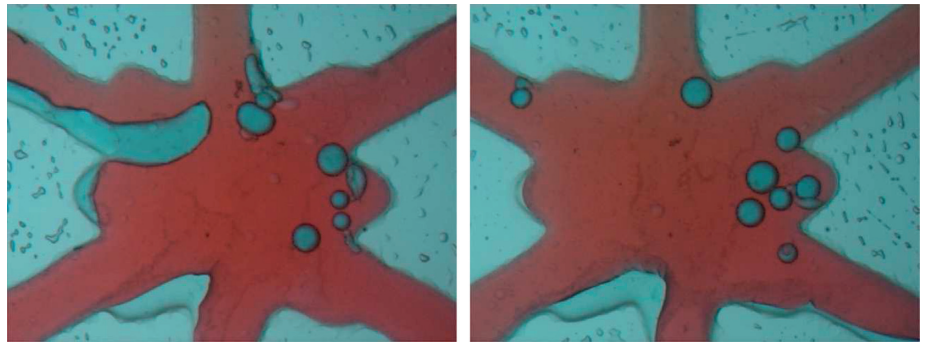


Figure 13: Demonstration of connate water discharging by microemulsion.

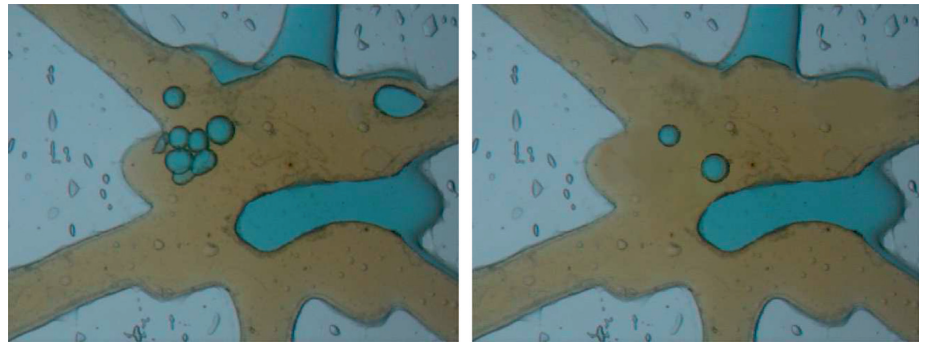


Figure 14: Origin zone, which was not in touch with microemulsion.

4. Conclusion

Regarding the objectives of this study, the following conclusions were obtained:

This study investigates the application of microemulsion flooding for low viscosity oil. The microemulsion flooding leads to interfacial tension reduction and consequently increasing oil recovery.

Two types of surfactants were used in this study. The results show that the type of surfactant has a significant effect on the recovery factor. The LABS surfactant increases the recovery by two times after 0.2 PV injection.

In the case of injection rates, increasing the rate from 0.0008 to 0.001 cc/min leads to a slight increase in oil recovery.

The permeability is another parameter that was studied using different pore structures. The pore structure does not show a sensible effect on recovery.

Behind the front of the microemulsion, the end-point saturations are changed due to changing of IFT.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Y. Chen, H. Wang, Y. Wang, and G. Ma, "Numerical evaluation of a fracture acidizing treatment in a three-dimensional fractured carbonate reservoir," *Journal of Natural Gas Science and Engineering*, vol. 81, 2020.

[2] X. Zhuang, S. Zhou, M. Sheng, and G. Li, "On the hydraulic fracturing in naturally-layered porous media using the phase field method," *The Engineering Geologist*, vol. 266, 2020.

[3] O. Massarweh and A. S. Abushaikha, "The use of surfactants in enhanced oil recovery: a review of recent advances," *Energy Report*, vol. 6, 2020.

[4] H. K. van Poollen, *Fundamentals of Enhanced Oil Recovery*, PenWell Books, Tulsa, OK, USA, 1980.

[5] R. Abdollahi and S. R. Shadizadeh, "The effect of spent acid on carbonate rock wettability during a matrix acidizing treatment," *Petroleum Science and Technology*, vol. 32, 2014.

[6] R. Abdollahi, H. Esfandyari, M. Nadri Pari, and A. Davarpanah, "Conventional diverting techniques and novel fibr-assisted self-diverting system in carbonate reservoir acidizing with successful case studies," *Petroleum Research*, 2021, In press.

[7] R. Abdollahi, M. Nadri, H. Gholghanddashti, M. Safari, and M. Zare Reisabadi, "A stepwise approach for identification of water production mechanisms in gas reservoirs," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2021, In press.

[8] H. Esfandyari, S. R. Shadizadeh, F. Esmaeilzadeh, and A. Davarpanah, "Implications of anionic and natural surfactants to measure wettability alteration in EOR processes," *Fuel*, vol. 278, 2020.

[9] H. Hematpour, R. Arabjamloei, M. Nematzadeh, H. Esmaili, and M. Mardi, "An experimental investigation of surfactant

flooding efficiency in low viscosity oil using a glass micromodel," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 34, 2012.

[10] M. Yazdi, N. A. Golilarz, A. Nedjati, and K. A. Adesina, "An improved lasso regression model for evaluating the efficiency of intervention actions in a system reliability analysis," *Neural Computing & Applications*, vol. 33, pp. 7913–7928, 2021.

[11] X. Hu, M. Li, C. Peng, and A. Davarpanah, "Hybrid thermal-chemical enhanced oil recovery methods; an experimental study for tight reservoirs," *Symmetry (Basel)*, vol. 12, 2020.

[12] S. Nesic, A. Zolotukhin, V. Mitrovic, D. Govedarica, and A. Davarpanah, "An analytical model to predict the effects of suspended solids in injected water on the oil displacement efficiency during waterflooding," *Processes*, vol. 8, 2020.

[13] J. C. S. Rosestolato, A. Pérez-Gramatges, E. R. Lachter, and R. S. V. Nascimento, "Lipid nanostructures as surfactant carriers for enhanced oil recovery," *Fuel*, vol. 239, 2019.

[14] N. Saxena, A. Saxena, and A. Mandal, "Synthesis, characterization and enhanced oil recovery potential analysis through simulation of a natural anionic surfactant," *Journal of Molecular Liquids*, vol. 282, 2019.

[15] H. Esfandyari, A. Haghighat Hoseini, S. R. Shadizadeh, and A. Davarpanah, "Simultaneous evaluation of capillary pressure and wettability alteration based on the USBM and imbibition tests on carbonate minerals," *Journal of Petroleum Science and Engineering*, vol. 200, no. 12, 2020.

[16] Y. A. Alzahid, P. Mostaghimi, S. D. C. Walsh, and R. T. Armstrong, "Flow regimes during surfactant flooding: the influence of phase behaviour," *Fuel*, vol. 236, 2019.

[17] V. Alvarado and E. Manrique, "Enhanced oil recovery: an update review," *Energies*, vol. 3, 2010.

[18] A. Barati-Harooni, A. Najafi-Marghmaleki, A. Tatar, and A. H. Mohammadi, "Experimental and modeling studies on adsorption of a nonionic surfactant on sandstone minerals in enhanced oil recovery process with surfactant flooding," *Journal of Molecular Liquids*, vol. 220, 2016.

[19] H. Esfandyari, A. Moghani, F. Esmaeilzadeh, and A. Davarpanah, "A laboratory approach to measure carbonate rocks' adsorption density by surfactant and polymer," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5539245, 7 pages, 2021.

[20] R. Eghbali, P. Khanmohammadi Hazaveh, F. Rashchi, and A. Ataie, "Recovery of manganese from a low-grade waste and valorization via the synthesis of a nanostructured magnetic manganese ferrite," *Materials Science and Engineering: B*, vol. 269, Article ID 115177, 2021.

[21] F. Pan, Z. Zhang, X. Zhang, and A. Davarpanah, "Impact of anionic and cationic surfactants interfacial tension on the oil recovery enhancement," *Powder Technology*, vol. 373, 2020.

[22] H. Hematpour, M. Mardi, S. Edalatkhah, and R. Arabjamaloei, "Experimental study of polymer flooding in low-viscosity oil using one-quarter five-spot glass micromodel," *Petroleum Science and Technology*, vol. 29, 2011.

[23] M. R. SaharKafashi, E. Eshraghi, L. Kuhar, A. Bona, and N. Aleksandar, "Visual study of TiO2 nanofluid stabilization methods on inhibition of asphaltene precipitation in porous media," *Minerals Engineering*, vol. 169, Article ID 106953, 2021.

[24] S. Kafashi, M.-R. Rasaei, and G. R. Karimi, "Experimental study of nanoclay absorbents and additives' effects on modification of rheological properties of drilling fluids in porous media using glass micromodel," *Journal of Porous Media*, vol. 23, no. 6, pp. 627–639, 2020.

[25] S. Thomas, "Enhanced oil recovery-an overview," *Oil and Gas Science and Technology*, vol. 63, no. 1, pp. 9–19, 2008.

[26] N. Hadian Nasr, S. M. Mahmood, and H. Hematpur, "A rigorous approach to analyze bulk and coreflood foam screening tests," *Journal of Petroleum Exploration and Production Technology*, vol. 9, 2019.

[27] H. Hematpour, S. M. Mahmood, N. H. Nasr, M. Karimi, and M. Bataee, "Experimental investigation on the mobility reduction factor of surfactant-alternating-gas foam flooding," *Kuwait Journal of Science*, vol. 46, 2019.

[28] H. Hematpur, S. M. Mahmood, N. H. Nasr, and K. A. Elraies, "Foam flow in porous media: concepts, models and challenges," *Journal of Natural Gas Science and Engineering*, vol. 53, 2018.

[29] A. Bera, T. Kumar, K. Ojha, and A. Mandal, "Screening of microemulsion properties for application in enhanced oil recovery," *Fuel*, vol. 121, 2014.

[30] V. C. Santanna, F. D. S. Curbelo, T. N. Castro Dantas, A. A. Dantas Neto, H. S. Albuquerque, and A. I. C. Garnica, "Microemulsion flooding for enhanced oil recovery," *Journal of Petroleum Science and Engineering*, vol. 66, 2009.

[31] C. J. Glover, M. C. Puerto, J. M. Maerker, and E. L. Sandvik, "Surfactant phase behavior and retention in porous media," *Society of Petroleum Engineers of AIME Journal*, vol. 19, 1979.

[32] F. D. S. Curbelo, V. C. Santanna, E. L. B. Neto et al., "Adsorption of nonionic surfactants in sandstones," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 293, 2007.

[33] A. S. Zelenev, L. M. Champagne, and M. Hamilton, "Investigation of interactions of diluted microemulsions with shale rock and sand by adsorption and wettability measurements," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 391, 2011.

[34] M. Safari-Beidokhti, A. Hashemi, R. Abdollahi, H. Hematpur, and E. Hamid, "Numerical well test analysis of condensate dropout effects in dual-permeability model of naturally fractured gas condensate reservoirs: case studies in the south of Iran," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9916914, 10 pages, 2021.

[35] A. A. Roozshenas, H. Hematpur, R. Abdollahi, and E. Hamid, "Water production problem in gas reservoirs: concepts, challenges, and practical solutions," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9075560, 20 pages, 2021.

[36] V. C. Santanna, A. C. M. Silva, H. M. Lopes, and F. A. Sampaio Neto, "Microemulsion flow in porous medium for enhanced oil recovery," *Journal of Petroleum Science and Engineering*, vol. 105, 2013.

[37] M. F. Nazar, S. S. Shah, and M. A. Khosa, "Microemulsions in enhanced oil recovery: a review," *Petroleum Science and Technology*, vol. 29, 2011.

[38] P. Kamranfar and M. Jamialahmadi, "Effect of surfactant micelle shape transition on the microemulsion viscosity and its application in enhanced oil recovery processes," *Journal of Molecular Liquids*, vol. 198, 2014.

[39] C. Stubenrauch, *Microemulsions: Background, New Concepts, Applications, Perspectives*, Wiley, Hoboken, NJ, USA, 2009.

[40] R. L. Reed and R. N. Healy, *Some Physicochemical Aspects of Microemulsion Flooding: A Review in: Improved Oil Recover by Surfactant Polymer Flooding*, Academic Press, Cambridge, MA, USA, 1977.

[41] L. W. Holm, "Use of soluble oils for oil recovery," *Journal of Petroleum Technology*, vol. 23, 1971.

[42] R. N. Healy, R. L. Reed, and C. W. Carpenter, "Laboratory study of microemulsion flooding," *Society of Petroleum Engineers of AIME Journal*, vol. 15, 1975.

[43] L. A. Verkruyse and S. J. Salter, "Potential use of nonionic surfactants in micellar flooding," in *Proceedings of the SPE Oilfield and Geothermal Chemistry Symposium*, Phoenix, Arizona, April 1985.

[44] M. Bouabboune, N. Hammouch, and S. Benhadid, "Comparision between micro-emulsion and surfactant solution flooding effeciency for enhanced oil recovery in tinfouyé oil field," in *Proceedings of the Canadian International Petroleum Conference*, Calgary, Alberta, June 2006.

[45] A. Mandal, A. Bera, K. Ojha, and T. Kumar, "Characterization of surfactant stabilized nanoemulsion and its use in enhanced oil recovery," in *Proceedings of the SPE International Oilfield Nanotechnology Conference and Exhibition*, Noordwijk, The Netherlands, June 2012.

[46] M. S. Karambeigi, M. Nasiri, A. Haghighi Asl, and M. A. Emadi, "Enhanced oil recovery in high temperature carbonates using microemulsions formulated with a new hydrophobic component," *Journal of Industrial and Engineering Chemistry*, vol. 39, 2016.

[47] Z. Hu, E. Nourafkan, H. Gao, and D. Wen, "Microemulsions stabilized by in-situ synthesized nanoparticles for enhanced oil recovery," *Fuel*, vol. 210, 2017.

[48] G. F. D. Ferreira, D. R. Q. Souza, R. Lima, A. K. C. L. Lobato, A. C. M. Silva, and L. C. L. Santos, "Novel glycerin-based microemulsion formulation for enhanced oil recovery," *Journal of Petroleum Science and Engineering*, vol. 167, 2018.

[49] T. N. de Castro Dantas, A. C. de Oliveira, T. T. C. de Souza, C. R. dos Santos Lucas, E. de Andrade Araújo, and P. T. P. Aum, "Experimental study of the effects of acid microemulsion flooding to enhancement of oil recovery in carbonate reservoirs," *Journal of Petroleum Exploration and Production Technology*, vol. 10, 2020.

[50] W. Yang, J. Lu, B. Wei, H. Yu, and T. Liang, "Micromodel studies of surfactant flooding for enhanced oil recovery: a review," *ACS Omega*, vol. 6, no. 9, pp. 6064–6069, 2021.

[51] H. Amani, "Study of enhanced oil recovery by rhamnolipids in a homogeneous 2D micromodel," *Journal of Petroleum Science and Engineering*, vol. 128, 2015.

[52] M. Dong, Q. Liu, and A. Li, "Displacement mechanisms of enhanced heavy oil recovery by alkaline flooding in a micromodel," *Particuology*, vol. 10, 2012.

[53] G. Cheraghian, "An experimental study of surfactant polymer for enhanced heavy oil recovery using a glass micromodel by adding nanoclay," *Petroleum Science and Technology*, vol. 33, 2015.

[54] M. Mohajeri, M. Hemmati, and A. S. Shekarabi, "An experimental study on using a nanosurfactant in an EOR process of heavy oil in a fractured micromodel," *Journal of Petroleum Science and Engineering*, vol. 126, 2015.

[55] A. Mehranfar, M. H. Ghazanfari, M. Masihi, and D. Rashtchian, "Macroscopic and microscopic investigation of alkaline-surfactant-polymer flooding in heavy oil recovery using five-spot micromodels: the effect of shale geometry and connatewater saturation," *Journal of Porous Media*, vol. 18, 2015.

[56] A. Kianinejad, M. H. Ghazanfari, R. Kharrat, and D. Rashtchian, "An experimental investigation of surfactant flooding as a good candidate for enhancing oil recovery from fractured reservoirs using one-quarter five spot micromodels: the role of fracture geometrical properties," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 35, 2013.

[57] E. Hadi, *Experimental Investigation of Solvent Injection into Heavy Oil Layered Reservoirs Using Micromodel*, Master's thesis, IFP School, Rueil-Malmaison, France, Petroleum University of Technology, Abadan, Iran, 2009.

[58] D. Green and G. Willhite, "Enhanced Oil Recovery," *Functional Polymers*, SPE, London, UK, 2nd edition, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9984934, 21 pages
https://doi.org/10.1155/2021/9984934

Research Article

Optimum Design of Straight Circular Channels Incorporating Constant and Variable Roughness Scenarios: Assessment of Machine Learning Models

Majid Niazkar

Department of Civil and Environmental Engineering, Shiraz University, Shiraz, Iran

Correspondence should be addressed to Majid Niazkar; mniazkar@shirazu.ac.ir

Received 4 April 2021; Accepted 12 August 2021; Published 25 August 2021

Academic Editor: Samuel Yousefi

In this study, two machine learning (ML) models named as artificial neural network (ANN) and genetic programming (GP) were applied to design optimum canals with circular shapes. In this application, the earthwork and lining costs were considered as the objective function, while Manning's equation was utilized as the hydraulic constraint. In this design problem, two different scenarios were considered for Manning's coefficient: (1) constant Manning's coefficient and (2) the experimentally proved variation of Manning's coefficient with water depth. The defined design problem was solved for a wide range of different dimensionless variables involved to produce a large enough database. The first part of these data was used to train the ML models, while the second part was utilized to compare the performances of ANN and GP in optimum design of circular channels with those of explicit design relations available in the literature. The comparison obviously indicated that the ML models improved the accuracy of the circular channel design from 55% to 91% based on two performance evaluation criteria. Finally, application of the ML models to optimum design of circular channels demonstrates a considerable improvement over the explicit design equations available in the literature.

1. Introduction

The need for conveying water using manmade lined canals is inevitable. Basically, channel design is the determination of channel properties so that not only can it convey required amount of water properly but also it is designed based on some controlling criteria. The former demands taking into account hydraulic conditions of flow passing through the channel under consideration, while the latter delineates how to define the problem statement. For instance, when the budget of channel construction is the most important key factor, a cost-minimization problem is particularly sought. As a result, the quest for optimum design of channels has brought about an active field of research in hydraulic engineering [1, 2]. These studies can help hydraulic engineers to design channels with different section shapes including (1) linear sections [3], (2) curved sections [4], and (3) linear-curved sections [5, 6]. Since the focus of this study is devoted to optimum design of circular channels, the literature is reviewed considering studies conducted exclusively on design of circular canals.

Among various studies on optimum channel design, design of circular channels, as one of conventional sections, was considered among the earliest attempts. Swamee [7] developed design equations for the best hydraulic circular section by minimizing flow area. Swamee [8] derived section variables of an optimal circular channel for viscous flow. Swamee et al. [8] defined a general construction cost incorporating a lining cost and suggested explicit equations for optimum design of lined circular channels. These equations were derived by minimizing the general construction cost. Swamee et al. [9] presented explicit relations for the optimum design of canals with circular shapes by minimizing earthwork cost of channel construction. In two previous

studies, Swamee's resistance equation was considered as the hydraulic constraint. Swamee and Kashyap [10] solved the differential equation that governs seepage flow using a finite difference scheme for a large number of independent dimensionless variables of a circular section. The outcome of this analysis yielded the proposal of an explicit equation for seepage from circular canals, which was used for developing explicit equations for optimum design of circular channels with minimum seepage loss. These equations were improved to take into account the impact of a drainage layer occurred at a shallow depth [11]. Aksoy and Altan-Sakarya [3] proposed two models for calculating the optimal section variables of circular channels by minimizing earthwork and lining costs, while Manning's equation was the hydraulic constraint. Niazkar and Afzali [12] utilized the Modified Honey Bee Mating Optimization algorithm to minimize the generalized form of the construction cost. They developed new explicit equations for optimum design of lined circular channels and compared the performance of their equations with that of Swamee et al.'s [8] equations and Aksoy and Altan-Sakarya's [3] equations. Their comparison indicated that Niazkar and Afzali's [12] explicit relations outperformed the other ones. Swamee and Chahar [13] recommended explicit equations for the optimum design of circular sections when they should transport a requisite sediment discharge. Since the sediment velocity has a direct relationship with the channel hydraulic radius, the optimum design problem reduced to the maximization of the hydraulic radius. Since no reliable resistance equation is available for computing flow in rigid boundary channels that carry sediment, a sediment discharge relation was selected as the constraint. This relation was obtained by substituting a limit deposit velocity into the head loss formula of sediment transport through pipes in heterogeneous suspension [13]. More recently, Niazkar et al. [1] minimized the earthwork and lining construction costs by taking into account the variation of Manning's coefficient with water depth. They proposed explicit relations for the optimum design of circular channels. According to the conducted literature review, the optimum design of circular channels was basically introduced as an optimization problem, while a hydraulic constraint was imposed based on the design priority. Additionally, most of previous studies recommended equations to design circular channels explicitly.

Despite previous efforts on estimating optimum channel properties with circular shapes, there is still a need to explore further in favor of improving the optimum results. In this regard, this study aims not only to revisit the optimum design problem of open channels with circular cross sections but also to possibly improve estimations of channel properties. In this regard, artificial intelligence (AI) and machine learning (ML) models, which have been proven to be powerful estimation tools for different water resources problems in literatures [14–17], have not been applied for optimum design of canals with circular sections. Since ML models have been successfully used in estimation of the optimum geometric variables of lined channels with triangular, rectangular, and trapezoidal sections [18] and predicting channel geometries with trapezoidal and rectangular sections considering water loss [19], applying AI models to the optimum design of lined circular channels may improve the accuracy of this design.

Since determining the optimum values of channel properties in a water-conveyance project can significantly lower the required budget for excavating and lining, exploring for estimating the optimum values of channel properties is of great importance. In this study, the optimum design of lined canals with circular shapes is tackled by applying two ML methods (artificial neural network (ANN) and genetic programming (GP)) for estimating channel properties in for two scenarios: (1) constant and (2) variable Manning's coefficients. To the best of author's knowledge, it is the first time that ML methods has been exploited to estimate channel properties of lined circular channels. The performances of these ML models were compared with those of the explicit design equations present in the literature. In this regard, the problem statement of the circular channel design is introduced in the next section. In the same section, different models used for estimating circular channel properties are reviewed. Afterwards, the results of applying ML to the channel design with circular shapes are presented and discussed for constant and variable roughness scenarios.

2. Methods and Materials

2.1. Problem Statement of Optimum Channel Design. Channels are one of the most widely used hydraulic structures to convey water through either short or long distances. Generally, the construction cost of canals is counted as one of the main parameters, which plays the key role in real-life water-conveyance projects. The quest for a reality-based definition for channel design problem has been addressed in the literature by various studies. In this regard, one of the approaches for channel design is to treat it as an optimization problem, while the construction cost is assumed as an objective function. Although various factors may play roles in the cost of constructing a channel, considering all of them in a problem statement is not possible mostly because some of them are not predictable [12]. Nevertheless, considering dominant factors is required to address a close-to-reality solution for canal construction projects.

The total construction cost per unit length of a typical lined channel (C) is one the most generalized form of construction costs available in literatures [1, 3, 8, 12]. It assumed three construction components: (1) the earthwork cost per unit volume (β_E), (2) the additional earthwork cost associated with different earthwork costs in different depths (β_A), and (3) the lining cost per unit area (β_L). Therefore, the objective function consists of three costs [8]:

$$C = \beta_L P + \beta_E A + \beta_A \int_0^{y_n} a\,d\eta, \tag{1}$$

where P is the wetted perimeter, A is the channel cross section area, y_n is the water normal depth, a is the flow area at height η, and $d\eta$ is the unit length of earthwork at height η, where η represents the vertical axis of channel geometry.

In order to preserve a hydraulically valid condition for flow throughout the channel, considering a resistance equation is inevitably required in a channel design problem. In this regard, Manning's equation, which is the most common resistance equation in open channels [20], is utilized as the problem constraint:

$$Q - \frac{1}{n} AR^{2/3} \sqrt{S} = 0, \tag{2}$$

where Q is the discharge, n is Manning's coefficient, R is the hydraulic radius, and S is the channel slope.

Using a length-scale parameter $\lambda = (Qn/\sqrt{S})^{3/8}$, the involved dimensional parameters are converted to new dimensionless parameters shown in Table 1. In this table, y is the water depth, r is the channel radius, and $*$ subscript denotes the dimensionless form of a variable. In the optimum design of channels, S in Manning's equation is substituted by the bottom channel slope. Therefore, water depth resulted from Manning's equation is the normal water depth, while the index associated with the normal depth is omitted from y and r for simplification.

Generally, Manning's coefficient may be flow-dependent or flow-independent in open-channel hydraulics [20, 21]. As a result, λ can either be constant or vary with flow. Although

Table 1: Dimensionless parameters involved in the design problem statement.

$C_* = C/\beta_E\lambda^2$	$\beta_{A*} = \beta_A\lambda/\beta_E$	$\beta_{L*} = \beta_L/\beta_E\lambda$	$A_* = A/\lambda^2$
$P_* = P/\lambda$	$y_* = y/\lambda$	$r_* = r/\lambda$	

the former assumption may seem to be more streamlining for practical purposes, the latter one is based on experiments conducted in partially filled circular channels. Based on these experiments [22–25], Manning's coefficient varies with water depth angle (θ) [1]:

$$\frac{n}{n_f} = 1 + 0.18(2\pi - \theta)\left[0.1 + \exp(-0.3\theta)\sin^2(0.38\theta)\right], \tag{3}$$

where n and n_f are Manning's coefficients associated with a partially and completely full cross section, respectively.

Using the dimensionless parameters shown in Table 1, the objective function and the constraint of the design problem of lined circular channels considering variable Manning's coefficients are shown in equation (4) and equation (5), respectively [1]:

$$\text{minimize } C_* = \beta_{L*}\theta r_* + 0.5 r_*(\theta - \sin\theta) + 0.5\beta_{A*} r_* \left[(y_* - r_*)\theta + \frac{4r_*^2 + 2(y_* - r_*)^2}{3r_*^2}\sqrt{r_*^2 - (y_* - r_*)^2}\right], \tag{4}$$

$$\text{subjected to } n_f \times \left\{1 + 0.18(2\pi - \theta)\left[0.1 + \exp(-0.3\theta)\sin^2(0.38\theta)\right]\right\} - \left[0.5 r_*(\theta - \sin\theta)\right]^{5/3}(\theta r_*)^{-2/3} = 0. \tag{5}$$

For better clarification, Figure 1 presents a schematic view of geometric variables involved in the design problem of partially filled circular channels. Particularly, Figure 1 introduces θ for two scenarios: (a) $y > r$ and (b) $y < r$. Based on Figure 1, $\theta/2$ is less than 90 degree for the former and more than 90 degree for the latter, respectively.

When n is assumed to be flow-independent, the objective function of the problem statement is still equation (4) because n does not appear in the cost of channel construction. However, the constraint, which plays the role of a hydraulically feasible condition for flow moving in channels, inevitably depends on Manning's coefficient. Consequently, the constraint for flow-independent n is presented in the following equation [3]:

$$\left\{r_*\left[\pi - 2\sin^{-1}\left(1 - \frac{y_*}{r_*}\right)\right]\right\}^{2/3} - \left\{0.5 r_*^2\left[\pi - 2\sin^{-1}\left(1 - \frac{y_*}{r_*}\right) - 2\left(1 - \frac{y_*}{r_*}\right)\frac{\sqrt{r_*^2 - (r_* - y_*)^2}}{r_*}\right]\right\}^{5/3} = 0. \tag{6}$$

The design problem of lined circular channels for both constant and variable n has been solved in the literature, and various explicit formulas have been suggested for computing the optimum circular channel geometries. These relations are summarized in Table 2. In this table, L in Swamee et al.'s [8] model is equal to $\lambda_1((\varepsilon/\lambda_1) + (8\upsilon\lambda_1/Q))^{0.04}$, where $\lambda_1 = (Q/\sqrt{gS})^{0.4} = \lambda^{32/30}/n^{0.4}g^{0.2}$, g is the gravitational acceleration, ε is the average roughness height of canal surface, and υ is the kinematic viscosity. Among the equations reviewed in Table 2, equations (ix)–(xii) consider variable n, while the rest assume constant n with respect to flow.

2.2. Optimization Algorithm. In this study, the described design problem was solved for two scenarios (constant and variable n) by a well-known optimization algorithm, named Modified Honey Bee Mating Optimization (MHBMO) algorithm. This optimization algorithm has been successfully applied to this specific problem [1, 12]. The MHBMO algorithm

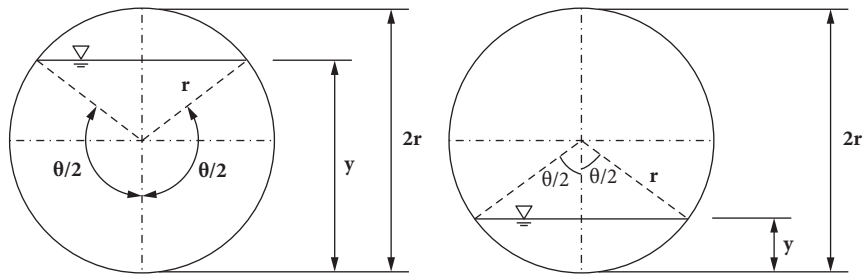


Figure 1: Schematic view of geometric variables of a partially filled circular channel: (a) $y > r$ and (b) $y < r$.

Table 2: Chronological review of explicit equations available for computing circular channel geometries.

Researchers	Equation number	Proposed relation
(a) For flow-independent Manning's coefficient		
Swamee et al. [8]	(i)	$2r = 0.78065L + (0.19375\beta_A L^3/(\beta_E\ L + 13.6232\beta_L))$
	(ii)	$y = 0.39032L(1 + (0.12631\beta_A L^2/(\beta_E L + 12.9379\beta_L)))^{-1}$
Aksoy and Altan-Sakarya [3]-first model	(iii)	$r_* = 1.004 + (0.113\beta_{A*}/\beta_{L*})$
	(iv)	$y_* = 1.004(1 + (0.055\beta_{A*}/\beta_{L*}))^{-1}$
Aksoy and Altan-Sakarya [3]-second model	(v)	$r_* = 1.004 + (0.58\beta_{A*}/(1 + 5.008\beta_{L*}))$
	(vi)	$y_* = 1.004(1 + (0.277\beta_{A*}/1 + 4.937\beta_{L*}))^{-1}$
Niazkar and Afzali [12]	(vii)	$r_* = 1.004 + 0.2358\beta_{A*}^{0.9978}\beta_{L*}^{-0.7749}$
	(viii)	$y_* = 1.004(1.0091 + 3.3182\beta_{A*}^{1.3175}\beta_{L*}^{-1.0878})^{-0.0708}$
(b) For flow-dependent Manning's coefficient		
Niazkar et al. [26]-first model	(ix)	$r_* = 0.9493\beta_{L*}^{0.0202} + 0.3542\beta_{A*}^{1.0351}\beta_{L*}^{-0.702}$
	(x)	$y_* = 1.1663\beta_{L*}^{-0.01} - 0.1557\beta_{A*}^{0.9322}\beta_{L*}^{-0.5489}$
Niazkar et al. [26]-second model	(xi)	$r_* = 0.9539\beta_{L*}^{0.0175} + 0.3501\beta_{A*}^{1.0352}\beta_{L*}^{-0.705}$
	(xii)	$y_* = 1.1598\beta_{L*}^{-0.0114} - 0.1618\beta_{A*}^{1.0695}\beta_{L*}^{-0.5676}$

can be classified as a zero-order optimization algorithm, which was originally inspired from the mating process of honey bees. It has five main controlling parameters, which include the size of the initial population, the number of workers, the queen's speed at the start and end of the mating flight, and the speed reduction factor. The MHBMO controlling parameters were set as recommended in previous studies [12]. Finally, the MHBMO algorithm was merely utilized in this study to compute the optimum values of channel properties with circular shapes, while two ML methods were applied to the optimum properties computed for the circular channels.

2.3. Machine Learning Models. In this study, two ML models are used for estimating the optimum properties of lined circular channels: (1) ANN and (2) GP. Although the applications of ANN and GP have been presented well in the literature of hydraulic engineering [15, 18], this is the first time that these AI models have been utilized for the optimum design of lined circular channels. The feature characteristics of ANN such as training capability, parallel operation, and distributed memory result in their fault resiliency and prediction accuracy. In essence, it compromises a random mapping between two vectors of input and output values. The architecture of ANNs, as shown in Figure 2(a), commonly consists of three layers of neurons, which are invariantly called input, hidden, and output layers, while each layer includes several neurons, and the layers are interrelated by sets of correlation weights. This structure, which provides a high degree of freedom, may yield a robust prediction capability. In this study, a three-layer network with feed-forward back-propagation characteristic was used to train ANN, while the controlling parameters of ANN were selected as the same as the ones used in literature [18].

The second ML model used in this study, i.e., GP, is basically a suitable extension to the genetic algorithm. In essence, it adopts principles of crossover Figure 2(c), mutation Figure 2(d), and survival of the fittest individuals not only to find an optimum solution but also to predict an output vector based on a known input vector. The tree-like structure of GP, as shown in Figure 2(b), consists of functions and terminals, while branches of each tree are connected to a root point. The function set can contain arithmetic operations, Boolean logical operators, and logical or deductive conditions or mathematical functions or any other functions. Furthermore, the terminal set is usually composed of variables, numerical constants, and so on. This characteristic inherently makes GP a powerful estimation tool. The main advantage of GP over regression-based equations is that it can seek for both structures and parameters of the relationship under investigation without the need to assume a shape limit in advance [15]. For applying GP, Discipulus [27] software was used to predict the optimal circular channels, while it has been employed for various

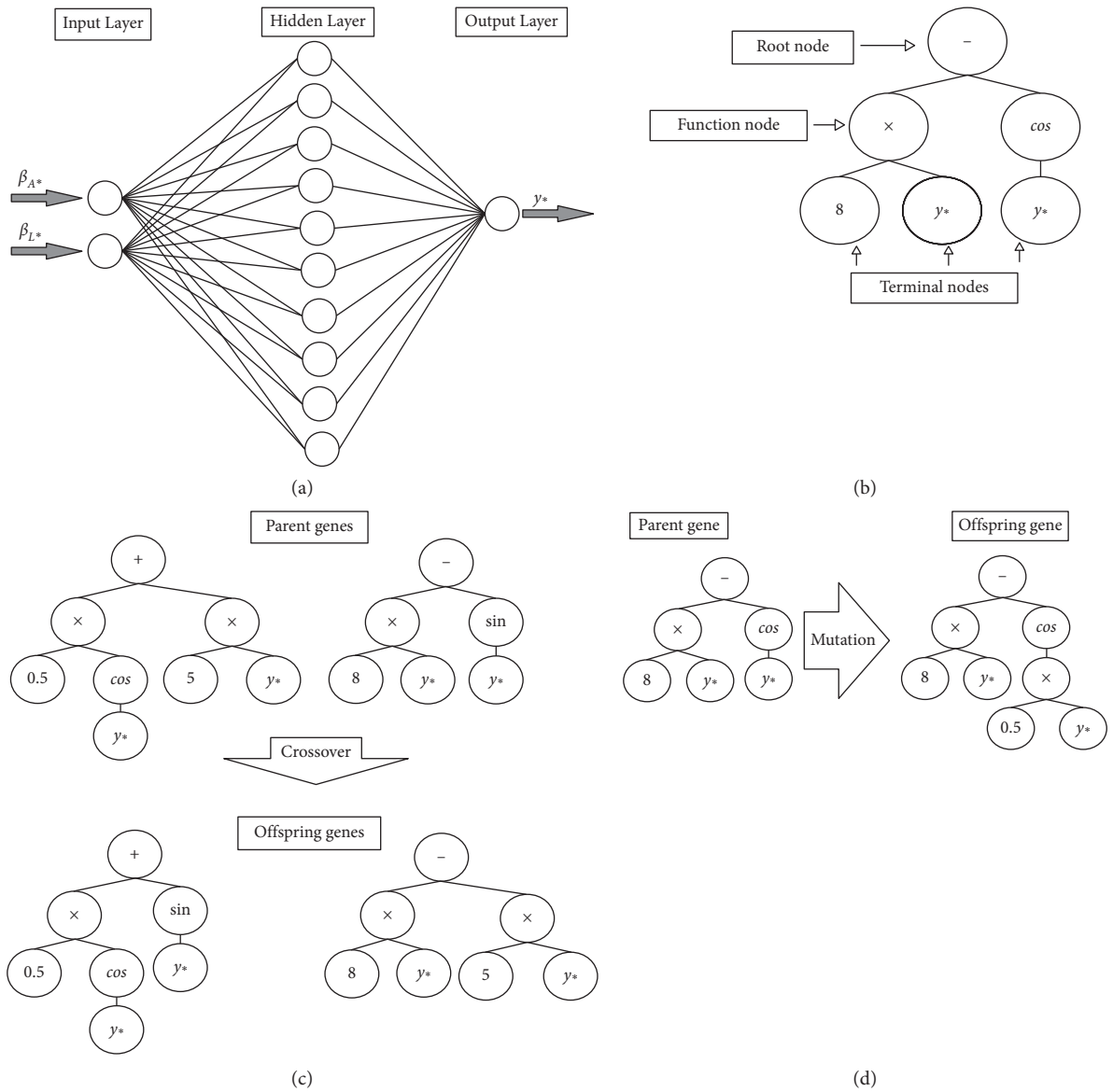


Figure 2: Schematic view of ML models: (a) ANN structure; (b) GP structure; (c) crossover process in GP; (d) mutation process in GP.

applications in water-related fields of study [1]. Finally, the default values of GP parameters presumed in Discipulus software were considered.

2.4. Application of ML Method to Solve the Design Problem. The described design problem of lined circular channels was treated as optimization problem, which was solved by the MHBMO algorithm for two constant and variable roughness scenarios. Each scenario is solved for 210 pairs of β_{A*} and β_{L*}, while the ratio of β_{A*}/β_{L*} is within (0, 2) for all pairs of data, similar to previous studies [3, 12]. By solving this design problem, the optimum channel properties (r_* and y_*) were obtained for various values of β_{A*} and β_{L*}. The optimum values of channel properties and dimensionless cost factors build up a dataset, which was utilized for estimating optimum values of channel properties using different models including ML methods (ANN and GP). The developed data were normalized within the range of (0.0, 1.0) before being used by the ML models. These 210 data for each scenario are depicted in Figure 3. As shown, variations of channel roughness have an inevitable impact on optimum values of r_* and y_*. The data illustrated in Figure 3 are randomly divided into two parts: (1) train data (150 data for each scenario) and (2) test data (60 data for each scenario). The former was utilized for training the ML models, while the latter was employed for testing stage. The detailed description of the data division is presented in literature [26].

2.5. Performance Evaluation Criteria. In order to compare performances of explicit equations with those of the ML models for calculating the optimum water depth and channel radius of circular channels, five performance evaluation criteria are used in this study [26]. These criteria

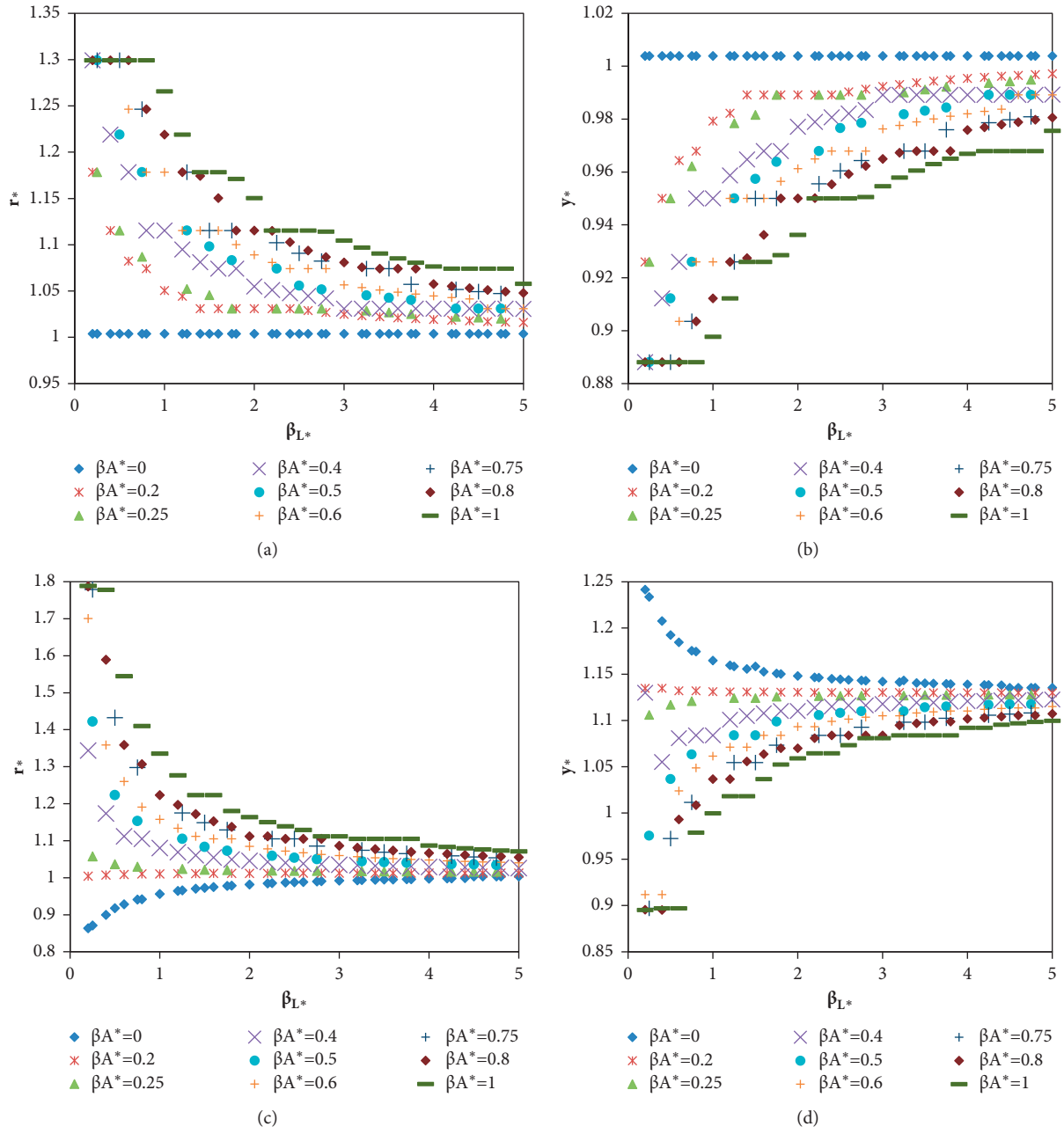


FIGURE 3: Variations of r_* and y_* for different β_{A*} and β_{L*}: (a) constant roughness; (b) constant roughness; (c) variable roughness; (d) variable roughness.

include (1) root mean square error (RMSE), (2) mean absolute error (MAE), (3) mean absolute relative error (MARE), (4) relative error (RE), and (5) coefficient of determination (R^2). Among these criteria, RE can be computed for each set of y_* and r_*, whereas other four criteria are reported for either train or test data. The five indices are written in terms of y_* in the following equations, respectively, while they can be written for r_*, too [26]:

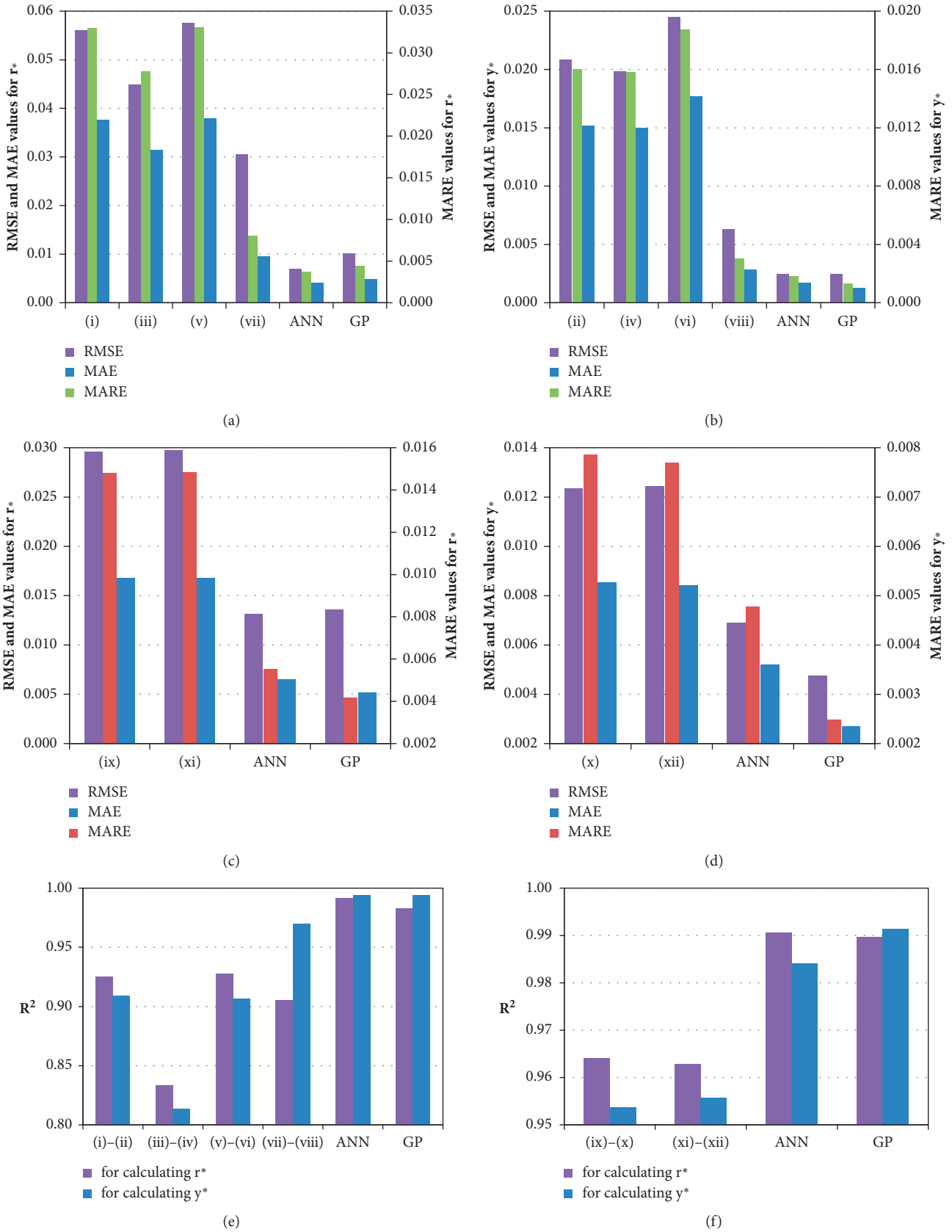


Figure 4: Comparison of different models for calculating the optimum r_* and y_*: (a) constant roughness; (b) constant roughness; (c) variable roughness; (d) variable roughness; (e) constant roughness; (f) variable roughness.

$$\begin{aligned}
\mathrm{RMSE} &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{*,\mathrm{database}} - y_{*,\mathrm{estimated}}\right)^2},\\
\mathrm{MAE} &= \frac{1}{N}\sum_{i=1}^{N}\left|y_{*,\mathrm{database}} - y_{*,\mathrm{estimated}}\right|,\\
\mathrm{MARE} &= \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_{*,\mathrm{database}} - y_{*,\mathrm{estimated}}}{y_{*,\mathrm{database}}}\right| \times 100,\\
\mathrm{RE} &= \frac{y_{*,\mathrm{estimated}} - y_{*,\mathrm{database}}}{y_{*,\mathrm{database}}},\\
R^2 &= \left(\frac{\sum_{i=1}^{N}\left[\left(y_{*,\mathrm{database}} - \left(\sum_{i=1}^{N} y_{*,\mathrm{database}}/N\right)\right)\left(y_{*,\mathrm{estimated}} - \left(\sum_{i=1}^{N} y_{*,\mathrm{estimated}}/N\right)\right)\right]}{\sqrt{\sum_{i=1}^{N}\left[\left(y_{*,\mathrm{database}} - \left(\sum_{i=1}^{N} y_{*,\mathrm{database}}/N\right)\right)^2\left(y_{*,\mathrm{estimated}} - \left(\sum_{i=1}^{N} y_{*,\mathrm{estimated}}/N\right)\right)^2\right]}}\right)^2,
\end{aligned} \tag{7}$$

where $y_{*,\mathrm{database}}$ and $y_{*,\mathrm{estimated}}$ are the database and estimated dimensionless water depth, respectively.

3. Results and Discussion

The results obtained by applying ML methods to the design problems are plotted in Appendix A. Furthermore, the output equations achieved by ANN for each roughness scenario (constant or variable n) and each channel property (r_* or y_*) are presented in Appendix B. Finally, the results of different methods are compared separately for each roughness scenario in the following.

3.1. Results of the Constant Roughness Scenario. As shown in Table 2, four models are available in the literature for computing circular channel geometries when n is assumed constant in respect with flow. The performances of these explicit equations are compared with those of the ML models for computing r_* and y_* in Figures 4(a) and 4(b) for the test data, respectively. According to Figure 4, ML models suggested in this study perform much better than all explicit equations based on all four criteria for flow-independent n. Among explicit equations for calculating optimum r_* for constant n, equation (vii) (Table 2) achieved the best RMSE, MAE, and MARE, while equation (v) (Table 2) yielded the best R2, as shown in Figure 4(e). Also, Figure 4 shows that equation (viii) (Table 2) reaches the best values for the four criteria considered for calculating optimum y_* for constant n. Additionally, Figure 4 demonstrates that ANN calculates closer r_* to the optimum solutions for constant n, while ANN and GP have quite the same performances for calculating optimum y_* for constant n. Based on RMSE values shown in Figure 4, the improvement made for predicting r_* and y_* varies between 61% and 88%, while the ML models improved the MARE values of r_* and y_* estimated by the available explicit equations between 54% and 91%. Consequently, Figure 4 obviously indicates that the ML models result in much closer values of r_* and y_* to the optimum solutions for design of lined circular channels in comparison with other models available in the literature for constant roughness.

Figures 5(a) and 5(b) depict relative errors of r_* and y_* predicted by the ML models for the test data by assuming flow-independent n. As shown, RE values computed by ANN for predicting r_* and y_* are placed within [−0.0179, 0.0293] and [−0.0106, 0.0040], respectively. Moreover, the corresponding bounds of RE values for calculating r_* and y_* by GP are [−0.0455, 0.0198] and [−0.0151, 0.0067], respectively. Comparing these four ranges of RE values shows that the AI models achieved a lower bound of RE values for calculating r_* when n is assumed to be invariant of flow. Furthermore, the average of absolute RE achieved by ANN is 0.037 and 0.0018 for computing r_* and y_*, while the corresponding values obtained by GP are 0.0044 and 0.0013, respectively. These values also indicate that the ML models reached a lower average of absolute RE when they calculated y_* rather than r_*. Among these two ML models, GP yielded a lower average of absolute RE for predicting y_*, while ANN resulted a lower average of absolute RE for calculating r_*.

The confidence limits of r_* and y_* estimated by different models are compared with the benchmark solutions for constant roughness in Figures 6(a) and 6(b), respectively. According to Figure 6(a) (equation (vii) in (Table 2)), ANN and GP achieved much closer confidence limits to that of the benchmark solutions in comparison to other models. Furthermore, Figure 6(b) depicts that equation (viii) (Table 2) and GP obtained close confidence limits to the benchmark confidence limit of y_*. However, Figure 6(b) shows that ANN failed to predict y_* close to the benchmark solutions under the constant roughness scenario. Thus, investigating the confidence limits of r_* and y_* for constant roughness reveals that GP performed very well, whereas ANN merely provided accurate estimations for r_*.

Bland–Altman plots for r_* and y_* estimated by ANN and GP are presented in Figures 7 and 8, respectively. As

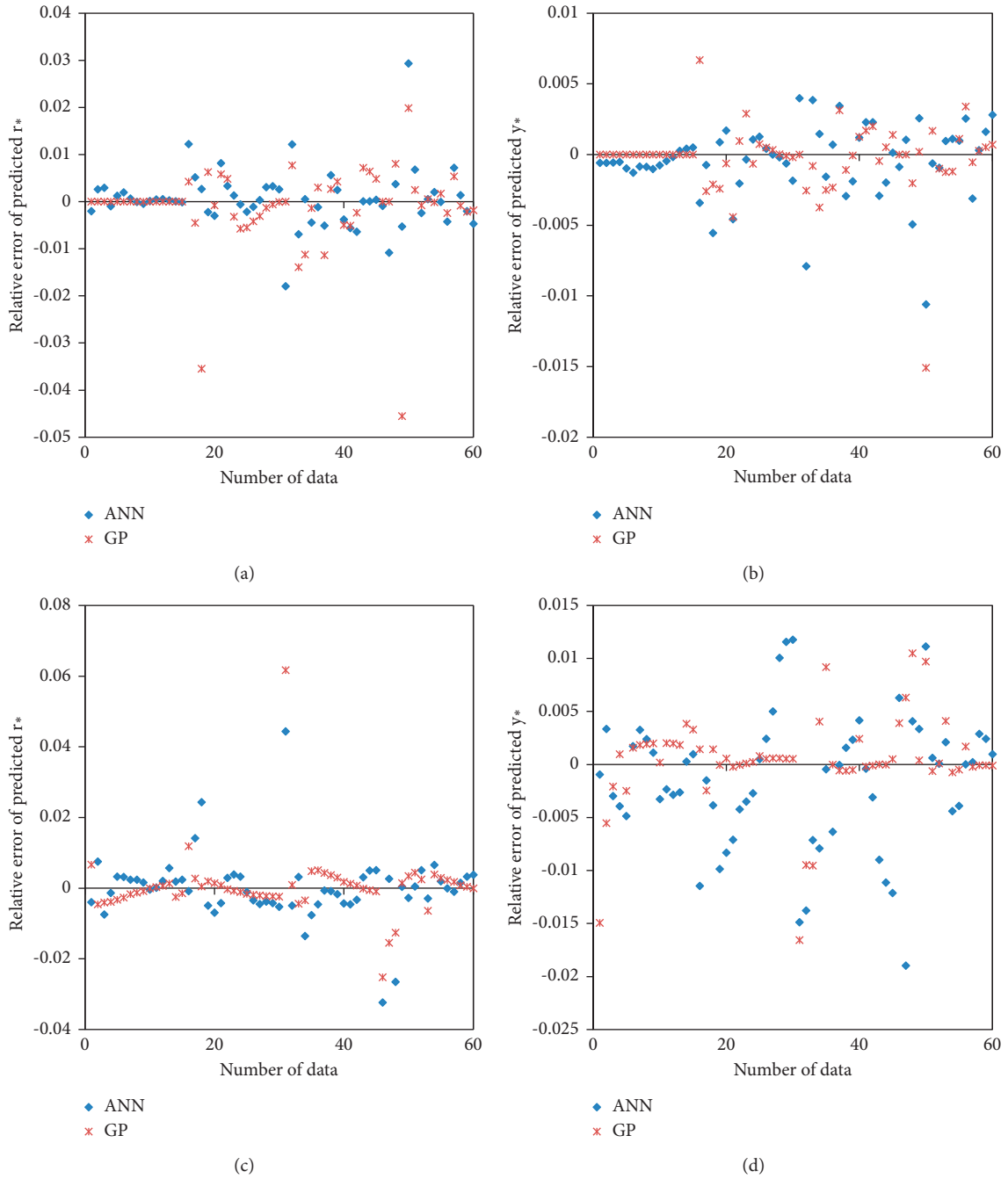


FIGURE 5: Relative error for the optimum r_* and y_* predicted by ANN and GP for the test data: (a) constant roughness; (b) constant roughness; (c) variable roughness; (d) variable roughness.

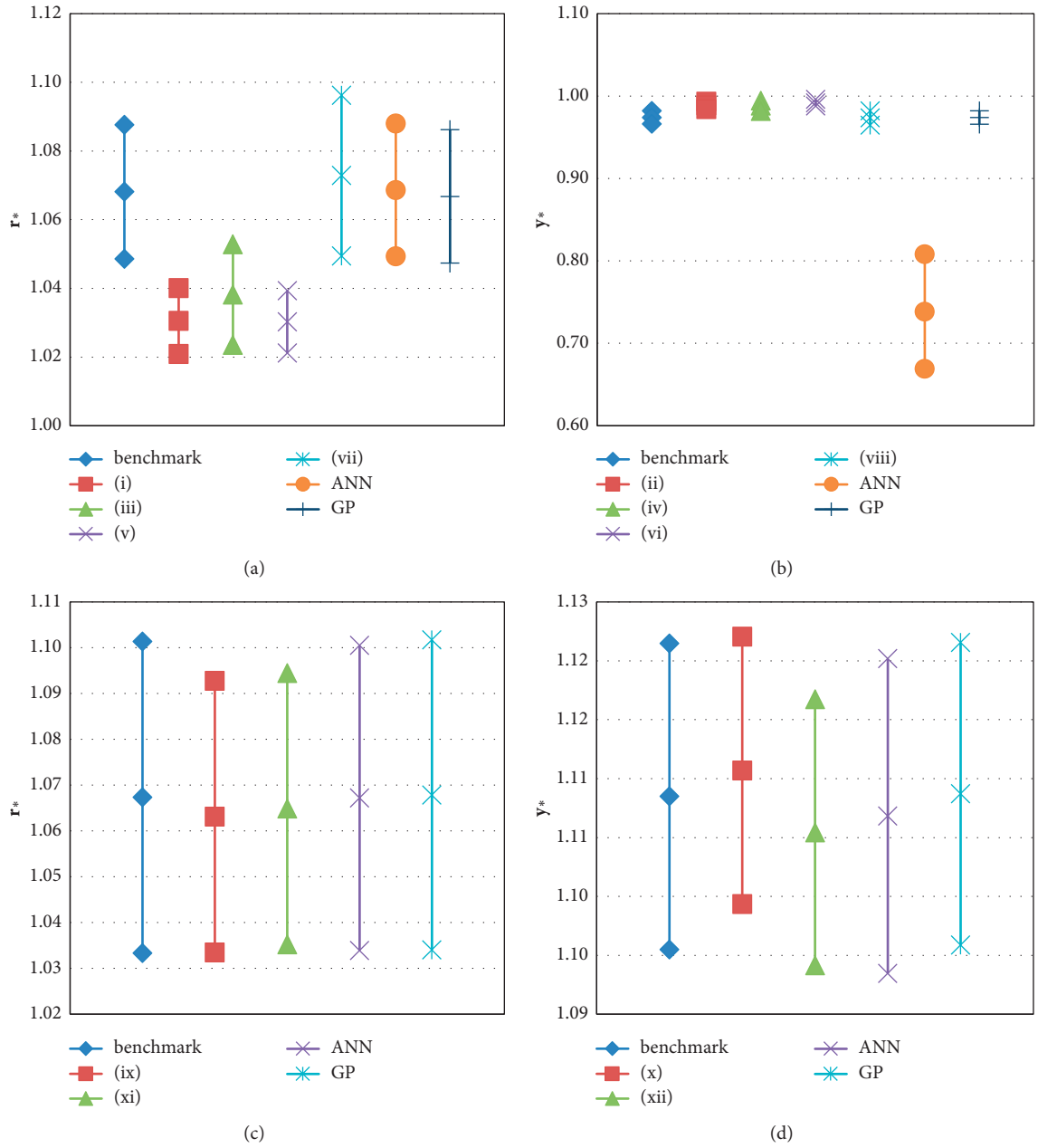


FIGURE 6: Confidence limits for the optimum r_* and y_* predicted by different models for the test data: (a) constant roughness; (b) constant roughness; (c) variable roughness; (d) variable roughness.

shown, the horizontal axis of a Bland–Altman plot is average of measured and estimated values for each data point, while the difference of measured and estimated values is illustrated on the vertical axis. In addition, the mean difference of all data points is a horizontal solid line, while two dashed lines denote ±1.96 standard deviation from the mean difference. In this regard, when 95% of data points of the Bland–Altman plot are placed between the two mentioned dashed lines, the estimated values are in agreement with the benchmark solutions. In other words, this scatter plot can help determine whether r_* and y_* predicted by the ML methods are consistent with the benchmark solutions. According to Figure 7(a), the r_* values estimated by ANN are in agreement with the benchmark solutions as most of points on the corresponding Bland–Altman plot fall between dashed lines associated with ±1.96 standard deviation from the mean difference. However, Figure 7(b) shows that there is a difference between the ANN estimations of y_* and the benchmark solutions for the constant roughness scenario. Furthermore, Figures 8(a) and 8(b) imply that r_* and y_* predicted by GP are in good agreement with the benchmark solutions, respectively. Therefore, GP performed acceptably consistent with benchmark solutions for predicting channel properties under the constant roughness scenario, while ANN only estimated consistent r_* values in comparison with the benchmark solutions.

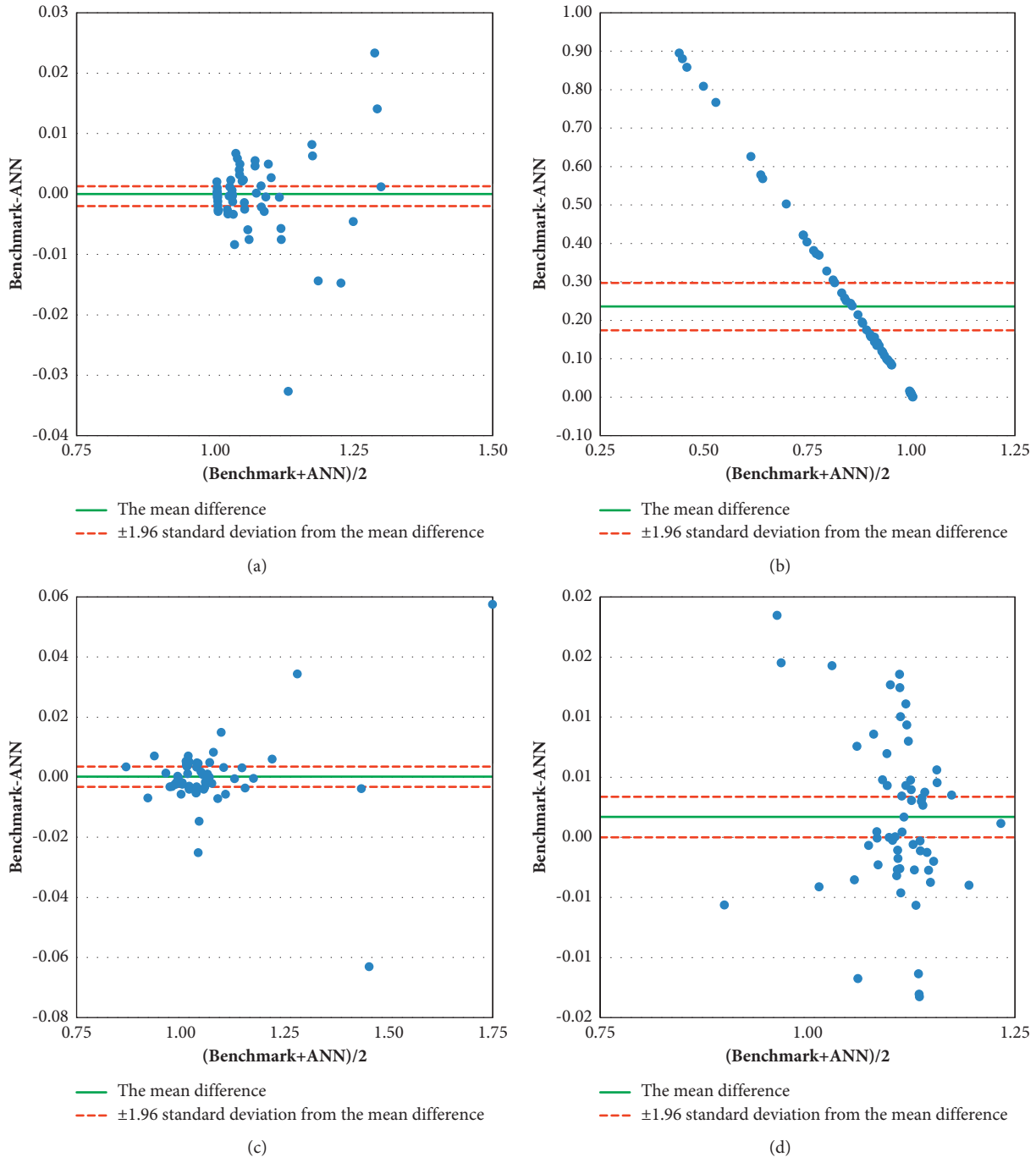


Figure 7: Bland–Altman plots for the optimum r_* and y_* predicted by ANN for the test data: (a) r_* for constant roughness; (b) y_* for constant roughness; (c) r_* for variable roughness; (d) y_* for variable roughness.

3.2. Results of the Variable Roughness Scenario. Two models presented in Table 2 have been proposed to design circular lined canals for flow-dependent n. Figure 4(c) compares the performances of the ML models with those of equations (ix) and (xi) (Table 2) for computing r_* for the test data. As shown, the ML models outperformed the explicit equations available in the literature for predicting optimum r_* values based on all four criteria considered. Additionally, ANN and GP have quite similar performances in calculating r_* for the variable roughness scenario. The performances of the ML models are compared with those of explicit equations in Figure 4(d) for estimating y_* for the test data. The results indicate that both ML models achieved much closer values to the optimum solutions, while GP obtained slightly better results than ANN for calculating optimum y_* for variable n. According to Figure 4, the ML models suggested in this study improved the calculation of r_* up to 55% and 62% based on RMSE and MARE, respectively, while more than 61% and 68% improvements were made to

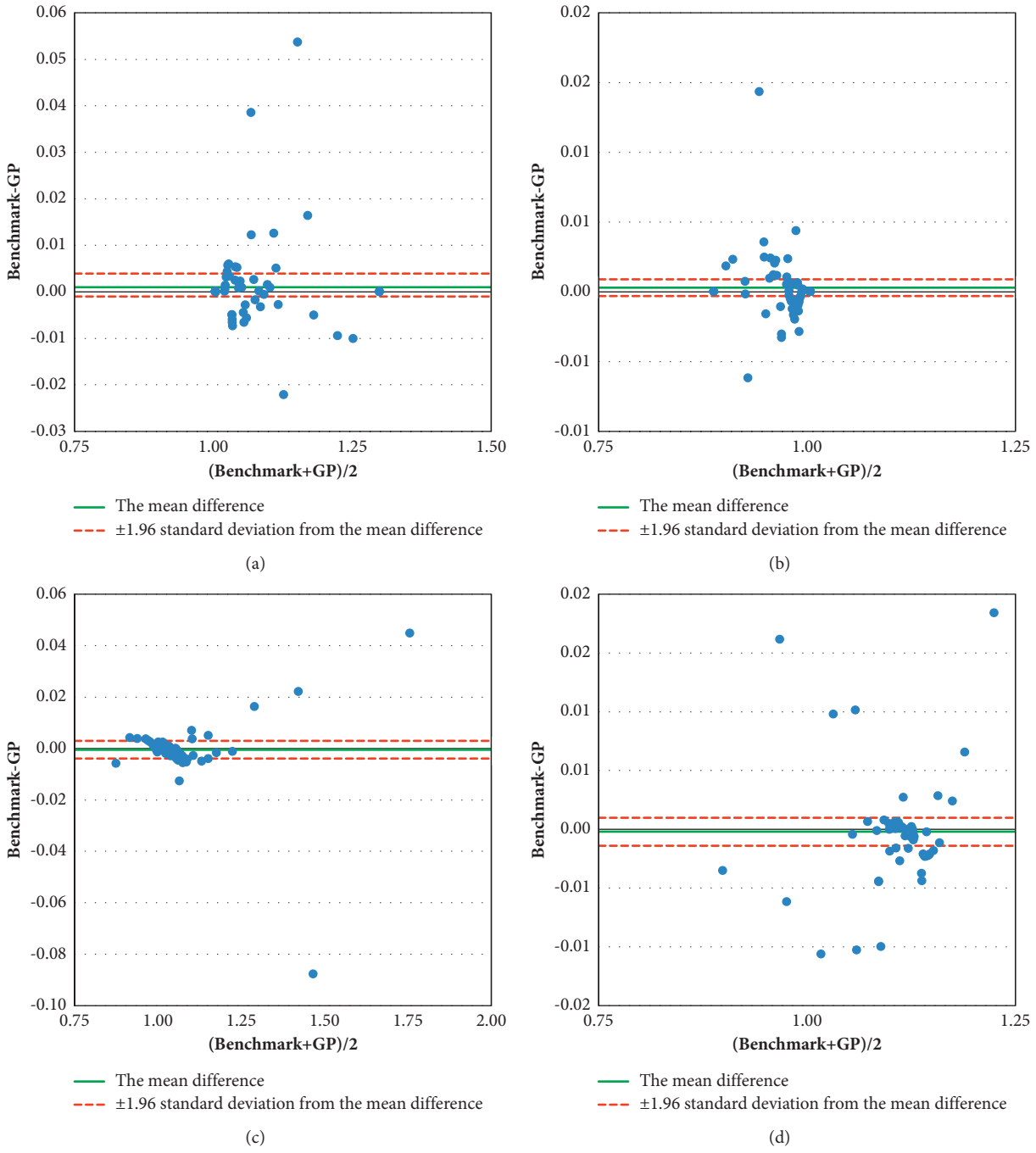


Figure 8: Bland–Altman plots for the optimum r_* and y_* predicted by GP for the test data: (a) r_* for constant roughness; (b) y_* for constant roughness; (c) r_* for variable roughness; (d) y_* for variable roughness.

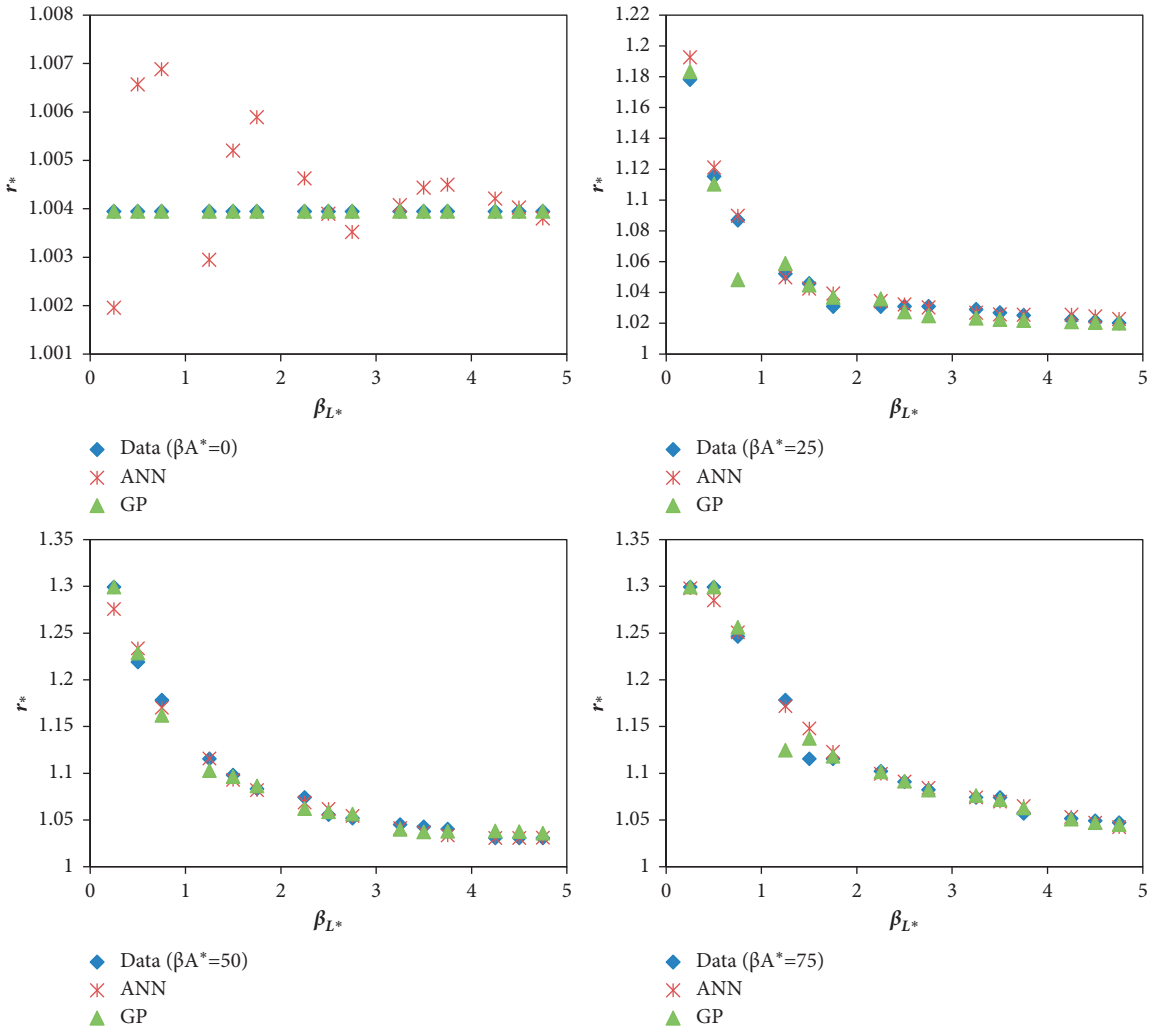


FIGURE 9: Comparison of the benchmark values of r^* with the estimations made by ANN and GP for the constant n scenario for the test data.

RMSE and MARE by the ML models in estimating y_* in comparison with the explicit equations. Therefore, Figure 4 clearly demonstrates that the ML models reached much closer results to the optimum solutions in design of a lined circular channel for flow-dependent n.

For flow-independent n, the relative errors of r_* and y_* calculated by the ML models are presented in Figures 5(c) and 5(d), respectively. As shown, the horizontal axis denotes the number of the test data, while the vertical axis is RE. According to Figure 5(c), the ranges of RE for r_* predicted by ANN and GP are [−0.0324, 0.0444] and [−0.0252, 0.0617], respectively. These bounds demonstrate that the minimum RE of r_* was achieved by ANN, while the maximum RE of r_* was obtained by GP. Moreover, the average values of absolute RE of r_* estimated by ANN and GP are 0.0055 and 0.0042, respectively. These results indicate that ANN performed slightly better than GP for computing r_* in terms of the average of absolute RE for variable roughness scenario. Based on Figure 5(d), RE values calculated by ANN and GP for predicting y_* vary within [−0.0190, 0.0118] and [−0.0166, 0.0105], respectively. Thus, both maximum and minimum RE values were achieved by ANN, which implies that GP has a lower bound of RE values. Furthermore, the average values of absolute RE obtained by ANN and GP for predicting y_* are 0.0048 and 0.0025, respectively. This also indicates that GP performs marginally better than ANN based on the average of absolute RE. In summary, GP yielded a lower average of absolute RE values for predicting y_*, while ANN reached a lower average of absolute RE values for calculating r_*. Regarding each scenario and metrics considered, GP and ANN had slightly different performances, while they both outperformed the explicit equations

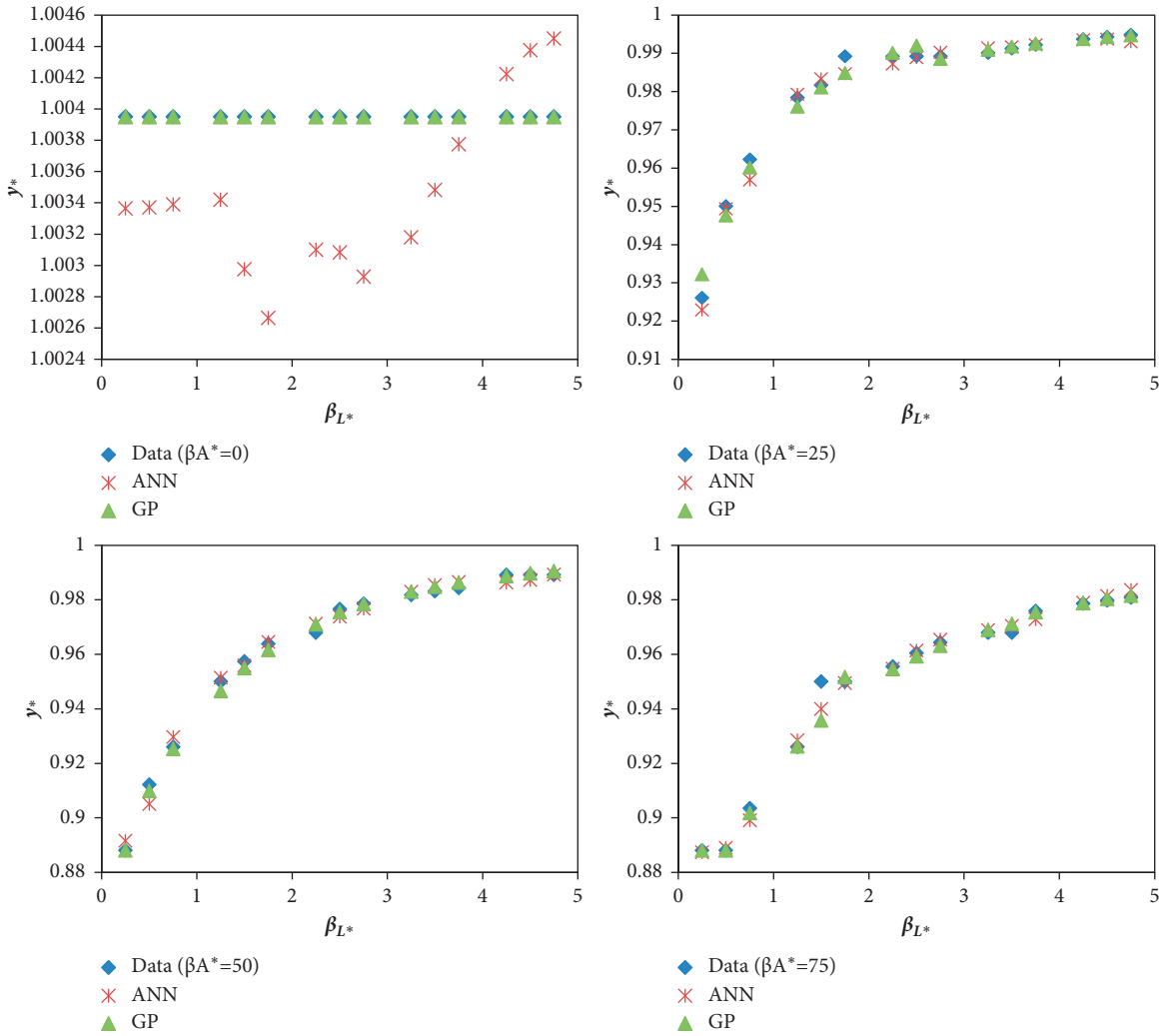


FIGURE 10: Comparison of the benchmark values of y^* with the estimations made by ANN and GP for the constant n scenario for the test data.

available in the literature for predicting r_* and y_* for both constant and variable roughness scenarios.

Figures 6(c) and 6(d) depict the confidence limits of different models under the variable roughness scenario. As shown, both ANN and GP estimated r_* and y_* within the benchmark solutions. Therefore, the ML methods predicted channel properties with high accuracy when a variable n is assumed. Moreover, Figure 6(c) indicates that equations (ix) and (xi) (Table 2) underestimated the maximum r_* in comparison with the benchmark solutions, whereas the ML methods performed adequately for estimating both minimum and maximum r_*. Additionally, Figure 6(d) shows that the maximum and minimum y_* of the benchmark solutions were not estimated acceptably by equations (x) and (xii) (Table 2), respectively, whereas the ML methods, GP in particular, predicted the maximum and minimum y_* close to those of benchmark solutions. Thus, the comparison of confidence limits of channel properties under the variable roughness scenario demonstrates high performance of the ML methods over the existing methods in the literature.

Investigating the Bland–Altman plots for the optimum r_* under the variable roughness scenario Figures 7(c) and 8(c) reveals that the ML estimations are placed with the 95% confidence limit of the corresponding plots. Consequently, GP and ANN provided r_* estimations consistent with the benchmark solution. Likewise, Figure 8(d) obviously indicates the agreement between the y_* values estimated by GP and the benchmarks solutions under the variable roughness scenario. On the contrary, Figure 7(d) shows that most of ANN estimations for y_* fall out of the accepted range, which indicates a difference between ANN predictions and the benchmark solutions. In summary, Figures 7 and 8 demonstrate that GP predicted channel properties are consistent with the benchmark solutions under constant and variable roughness scenarios. These figures also indicate that only the r_* values predicted by ANN are in agreement with the

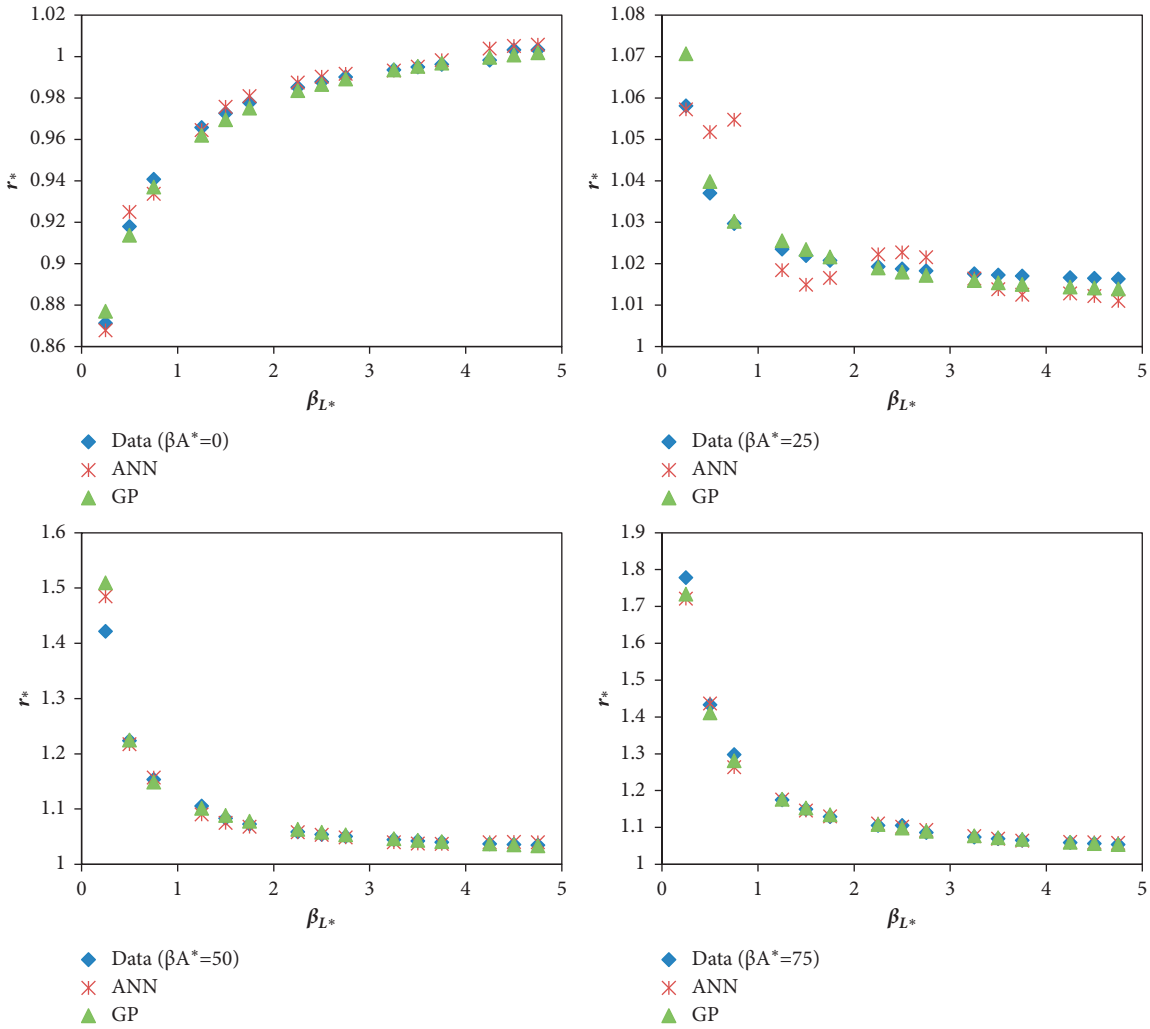


Figure 11: Comparison of the benchmark values of r^* with the estimations made by ANN and GP for the variable n scenario for the test data.

benchmark solutions. In conclusion, the accuracy achieved by applying the ML models over the explicit equations implies that these models can be considered as a suitable alternative for design of open channels with circular sections.

3.3. Advantages, Limitations, and Recommendation for Future Works. Since variables involved in the design problem were normalized, the solutions can embrace a variety combination of values of variables with dimensions, which shows the generality of the solutions presented in this study. According to the results of the comparative analysis, the ML methods perform better than equations available in the literature for estimating the optimum values of circular channel properties. This is an advantage of ML methods over the existing models for the same purpose. Furthermore, the ML methods (ANN and GP) have more flexibility to capture the variations of r_* and y_* for different β_{A*} and β_{L*} values in comparison to available formulas. This makes the ML methods more suitable for predicting optimum values of r_* and y_*. Additionally, estimating using the ML methods considered in this study require running the learned machine for any set of β_{A*} and β_{L*} values. This characteristic indicates that the ML methods performed as a black box method and did not present an ML-based simple equation. This feature is likely to be interpreted as a disadvantage for an engineer who prefers to work with mathematical equations when it comes to design optimum circular channels. In a bid to overcome this shortcoming, it is suggested to employ improved version of the ML methods, e.g., Multi-Gene GP, which provide mathematical equations. Finally, the uncertainty in β_{A*} and β_{L*} values is suggested to be investigated in future studies.

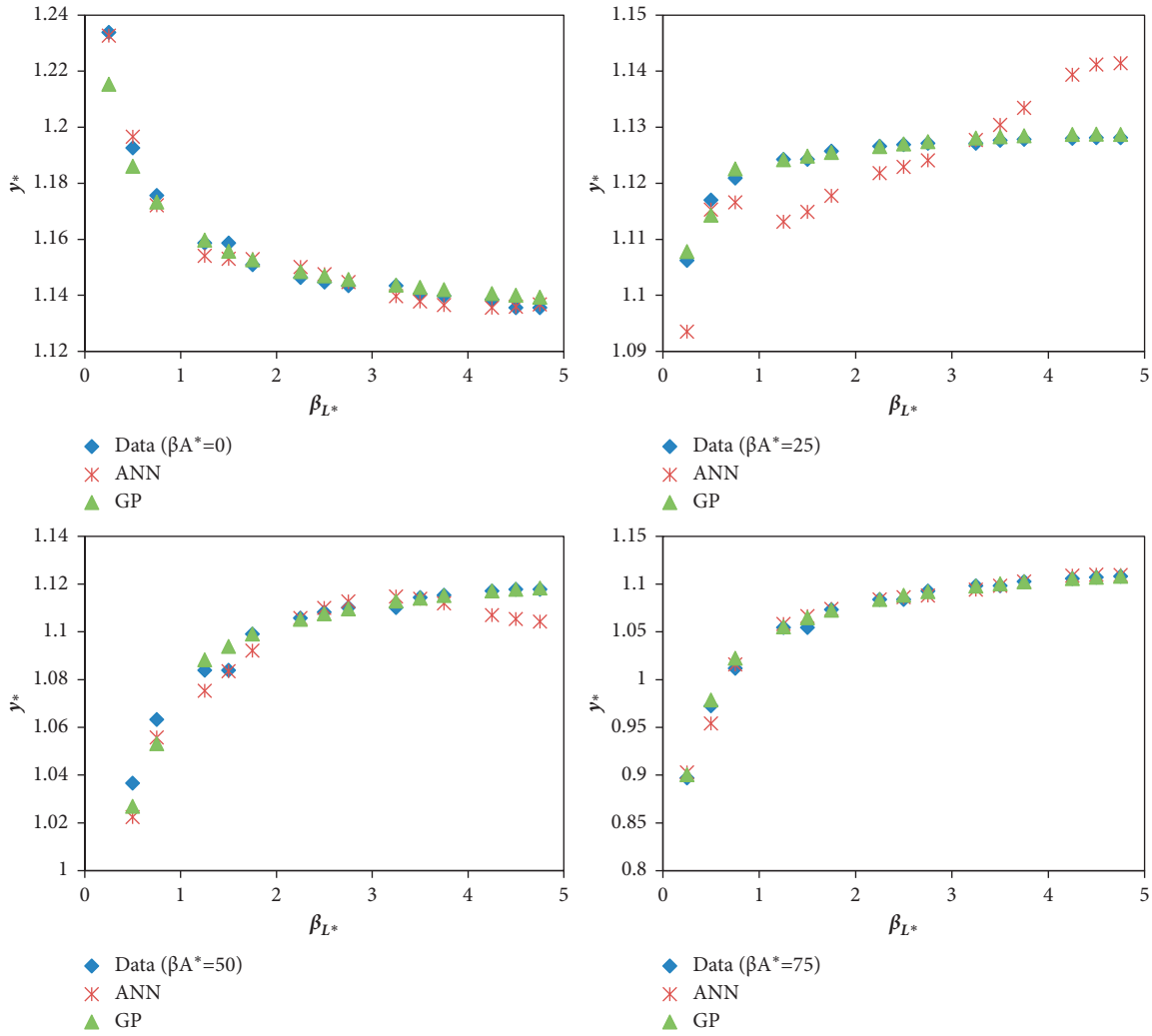


Figure 12: Comparison of the benchmark values of y^* with the estimations made by ANN and GP for the variable n scenario for the test data.

4. Conclusions

Conveying water from natural and artificial sources is substantially inevitable because of the distance between where water is in need and where water is available. In this study, the application of two ML models has been assessed for the optimum design of lined circular channels for the first time. The performances of these models were compared with those of explicit equations available in the literature for two scenarios: (1) flow-independent Manning's coefficient and (2) flow-dependent Manning's coefficient. Although the former may seem to be more suitable and streamlined to be used in practice, the latter has been approved by experimental studies. Since the variables involved in the design problem were normalized using a length-scale parameter, the solutions are not limited to a specific case study, whereas they embrace a variety combination of values for variables involved. According to the comparison, the ML models improved the prediction of geometries of circular channels estimated by the explicit equations between 55% and 88% based on RMSE, while the corresponding improvement range was between 62% and 91% based on MARE for both scenarios. Furthermore, the investigation of relative errors achieved by the ML models indicates that ANN and GP had quite similar performances. However, they had slightly different performances based on the scenario and channel geometry. To be more specific, for the constant roughness scenario, ANN performed slightly better than GP for calculating optimum values of dimensionless channel radius, while GP gave better results for optimum values of dimensionless water depth. On the other hand, GP results were a bit closer to the optimum solutions for calculating optimum values of circular geometries for the variable roughness scenario. Finally, the obtained results obviously demonstrate that not only applying AI models to the optimum design of circular channels yields closer results to the optimum solutions than using the available explicit equation but also considering a flow-dependent channel roughness in optimum design of lined circular channels may lead to solutions different from when a constant bed roughness is assumed.

Appendix

A. Comparison of the Benchmark Solutions with the ML Estimations

The benchmark solutions and outputs estimated by ANN and GP versus input variables are plotted for different scenarios. These plots are depicted in Figures 9–12. These figures provide the opportunity to investigate the effect of boundaries on the performance of the ML models. As shown, GP and ANN did perform acceptably in estimating the minimum and maximum values of output variables in most cases, while the former schematically performs even better than the latter in this regard. Since these figures do not provide qualitative information about the performances of the ML methods, they are presented here in Appendix A.

B. The ANN Output Equations

The output equation obtained by ANN is presented in the following equation [14]:

$$\text{output} = \text{linear}\{[\tanh(\text{input} \times IW + b_1)] \times LW + b_2\}, \tag{B.1}$$

where output is the channel property (r_* or y_*), $\text{linear}(x) = x$, $\tanh(x) = (2/1 + e^{-2x}) - 1$, input = $[\beta_{A*} \;\; \beta_{L*}]$ is the input matrix, and IW, LW, b_1, and b_2 are constant matrices (input weights, output weights, and biases). The constant matrices appeared in equation (B.1) depend on the roughness scenario (constant or variable n) and the channel property (r_* or y_*):

(a) Normalized radius (r_*) for constant roughness is as follows:

$$IW = \begin{bmatrix} -1.2747 & 4.7714 \\ 3.8194 & -1.9625 \\ -3.5514 & -1.5797 \\ 2.2211 & -3.1057 \\ -4.3712 & -1.3558 \\ -2.6416 & 1.9691 \\ 4.4975 & -0.7146 \\ 4.4073 & 2.2163 \\ -6.3307 & 1.4429 \\ 7.9585 & -1.7410 \end{bmatrix},$$

$$LW = [0.1089 \;\; 0.0313 \;\; 0.0138 \;\; -0.1659 \;\; 0.0145 \;\; 0.2045 \;\; 0.1898 \;\; 2.1255 \;\; 2.7497 \;\; 2.0036],$$

$$b_1 = \begin{bmatrix} 3.8295 \\ -3.0611 \\ 3.0218 \\ -0.1294 \\ 0.3099 \\ -1.5848 \\ 2.3836 \\ 7.7226 \\ -4.5416 \\ 5.6053 \end{bmatrix},$$

$$b_2 = [-2.1230]. \tag{B.2}$$

(b) Normalized water depth (y_*) for constant roughness is as follows:

$$IW = \begin{bmatrix} 1.8972 & -6.8594 \\ 4.5393 & -2.4272 \\ -4.1690 & -3.1827 \\ -3.0383 & -2.7364 \\ -1.3725 & 1.5407 \\ 8.3576 & -4.3048 \\ -4.4968 & 0.6641 \\ 10.8951 & -2.4065 \\ 6.4566 & -3.2644 \\ -4.0052 & -5.8913 \end{bmatrix},$$

$$LW = \begin{bmatrix} 0.0908 & -0.0453 & 0.2740 & -0.3087 & -0.3872 & -1.4087 & 0.1038 & 0.3004 & 1.7986 & 2.0548 \end{bmatrix},$$

$$b_1 = \begin{bmatrix} -5.9635 \\ -2.2562 \\ 2.6413 \\ 2.0764 \\ 0.1281 \\ 4.8110 \\ -1.4938 \\ 9.2226 \\ 3.7423 \\ -10.6689 \end{bmatrix}, \tag{B.3}$$

$$b_2 = [2.0829].$$

(c) Normalized radius (r_*) for variable roughness is as follows:

$$IW = \begin{bmatrix} 2.1051 & -3.4872 \\ 4.4230 & -0.4737 \\ -4.6699 & -1.4348 \\ 0.6999 & -2.8961 \\ -0.8608 & 1.9382 \\ 7.1420 & -1.1776 \\ -4.1166 & -1.1951 \\ 5.3566 & 1.3649 \\ 6.1985 & -1.1128 \\ -5.2261 & -1.5273 \end{bmatrix},$$

$$LW = \begin{bmatrix} -0.0366 & 0.0226 & 0.0043 & 0.2108 & 0.3665 & 2.6476 & -0.0298 & -2.0896 & -3.4226 & -2.1832 \end{bmatrix},$$

$$b_1 = \begin{bmatrix} -2.7516 \\ -2.6639 \\ 0.9685 \\ 1.6542 \\ -1.0562 \\ 6.2280 \\ -1.9247 \\ 5.5766 \\ 5.5902 \\ -5.6061 \end{bmatrix}, \tag{B.4}$$

$$b_2 = [0.1331].$$

(d) Normalized water depth (y_*) for variable roughness is as follows:

$$IW = \begin{bmatrix} 1.4323 & -5.3379 \\ 2.6917 & 2.9538 \\ -1.3738 & 5.4332 \\ 1.2035 & 0.4187 \\ 2.6227 & 3.3157 \\ -1.0381 & -3.5604 \\ -3.3328 & 0.2628 \\ 0.1158 & 4.9020 \\ 7.5539 & -1.3193 \\ 3.5985 & 1.9346 \end{bmatrix},$$

$$LW = \begin{bmatrix} -0.1947 & -0.1309 & 0.1458 & 0.3537 & -0.1333 & 0.3330 & -0.1289 & -0.1725 & 0.4865 & -1.6517 \end{bmatrix},$$

$$b_1 = \begin{bmatrix} -4.5318 \\ -3.8305 \\ 0.9470 \\ -0.7408 \\ -1.4023 \\ -2.7596 \\ -1.4157 \\ 0.6905 \\ 7.1377 \\ 6.2423 \end{bmatrix},$$

$$
\begin{aligned}
b_2 &= [1.4027],\\
C_* &= \frac{C}{\beta_E \lambda^2},\\
\beta_{A*} &= \frac{\beta_A \lambda}{\beta_E},\\
\beta_{L*} &= \frac{\beta_L}{\beta_E \lambda},\\
A_* &= \frac{A}{\lambda^2},\\
P_* &= \frac{P}{\lambda},\\
y_* &= \frac{y}{\lambda},\\
r_* &= \frac{r}{\lambda}.
\end{aligned}
\tag{B.5}
$$

Data Availability

The data used to support this study are obtained from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] M. Niazkar, G. R. Rakhshandehroo, and S. H. Afzali, "Deriving explicit equations for optimum design of a circular channel incorporating a variable roughness," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 42, no. 2, pp. 133–142, 2018.

[2] S. M. Easa, "Open Channel cross section design: review of recent developments," in *Advances in Measurements and Instrumentation: Review (Chapter 6)*, S. Y. Yurish, Ed., vol. 1, International Frequency Sensor Association Publishing, Barcelona, Spain, 2018.

[3] B. Aksoy and A. B. Altan-Sakarya, "Optimal lined channel design," *Canadian Journal of Civil Engineering*, vol. 33, no. 5, pp. 535–545, 2006.

[4] Y.-C. Han and S. M. Easa, "Exact solution of optimum hydraulic power-law section with general exponent parameter," *Journal of Irrigation and Drainage Engineering*, vol. 144, no. 12, Article ID 04018035, 2018.

[5] B. R. Chahar and S. Basu, "Optimal design of curved bed trapezoidal canal sections," *Proceedings of the Institution of Civil Engineers-Water Management*, vol. 162, no. 3, pp. 233–240, 2009.

[6] S. M. Easa and A. R. Vatankhah, "New open channel with elliptic sides and horizontal bottom," *KSCE Journal of Civil Engineering*, vol. 18, no. 4, pp. 1197–1204, 2014.

[7] P. K. Swamee, "Optimal irrigation canal sections," *Journal of Irrigation and Drainage Engineering*, vol. 121, no. 6, pp. 467–469, 1995.

[8] P. K. Swamee, G. C. Mishra, and B. R. Chahar, "Minimum cost design of lined canal sections," *Water Resources Management*, vol. 14, no. 1, pp. 1–12, 2000.

[9] P. K. Swamee, G. C. Mishra, and B. R. Chahar, "Design of minimum earthwork cost canal sections," *Water Resources Management*, vol. 15, no. 1, pp. 17–30, 2001.

[10] P. K. Swamee and D. Kashyap, "Design of minimum seepage-loss nonpolygonal canal sections," *Journal of Irrigation and Drainage Engineering*, vol. 127, no. 2, pp. 113–117, 2001.

[11] P. K. Swamee and D. Kashyap, "Design of minimum seepage-loss nonpolygonal canal sections with drainage layer at shallow depth," *Journal of Irrigation and Drainage Engineering*, vol. 130, no. 2, pp. 166–170, 2004.

[12] M. Niazkar and S. H. Afzali, "Optimum design of lined channel sections," *Water Resources Management*, vol. 29, no. 6, pp. 1921–1932, 2015.

[13] P. K. Swamee and B. R. Chahar, *Design of Canals*, Springer, New Delhi, India, 2015.

[14] I. Ebtehaj, H. Bonakdari, and A. H. Zaji, "A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes," *Alexandria Engineering Journal*, vol. 57, no. 3, pp. 1783–1795, 2018.

[15] M. Niazkar and M. Zakwan, "Assessment of artificial intelligence models for developing single-value and loop rating curves," *Complexity*, vol. 2021, Article ID 6627011, 21 pages, 2021.

[16] M. Zakwan and M. Niazkar, "A comparative analysis of data-driven empirical and artificial intelligence models for estimating infiltration rates," *Complexity*, vol. 2021, Article ID 9945218, 13 pages, 2021.

[17] M. Niazkar, F. Hajizadeh mishi, and G. Eryılmaz Türkkan, "Assessment of artificial intelligence models for estimating lengths of gradually varied flow profiles," *Complexity*, vol. 2021, Article ID 5547889, 11 pages, 2021.

[18] M. Niazkar, "Assessment of artificial intelligence models for calculating optimum properties of lined channels," *Journal of Hydroinformatics*, vol. 22, no. 5, pp. 1410–1423, 2020.

[19] A. M. Tawfik, "Design of channel section for minimum water loss using Lagrange optimization and artificial neural networks," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 415–422, 2021.

[20] M. Niazkar, N. Talebbeydokhti, and S. H. Afzali, "One dimensional hydraulic flow routing incorporating a variable grain roughness coefficient," *Water Resources Management*, vol. 33, no. 13, pp. 4599–4620, 2019.

[21] S. K. McKay and J. C. Fischenich, *Robust Prediction of Hydraulic Roughness, ERDC/CHL CHETN-VII-11*, U.S. Army Engineer Research and Development Center, Vicksburg, MS, USA, 2011.

[22] D. L. Yarnell and S. M. Woodward, *The Flow of Water in Drain Tile*, US Department of Agriculture, Washington, DC, USA, 1920.

[23] E. R. Wilcox, "A comparative test of the flow of water in 8-inch concrete and vitrified clay sewer pipe," *Engineering Experiment Station. Engineering Experiment Station Series. Bulletin*, Seattle, Wash., The University Washington (State) University, Seattle, WA, USA, 1924.

[24] N. A. Zaghloul, "Gradually varied flow in circular channels with variable roughness," *Advances in Engineering Software*, vol. 15, no. 1, pp. 33–42, 1992.

[25] Ö. Akgiray, "Simple formulae for velocity, depth of flow, and slope calculations in partially filled circular pipes," *Environmental Engineering Science*, vol. 21, no. 3, pp. 371–385, 2004.

[26] M. Niazkar and S. H. Afzali, "Developing a new accuracy-improved model for estimating scour depth around piers using a hybrid method," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 43, no. 2, pp. 179–189, 2018.

[27] F. D. Francone, *Discipulus Owner's Manual*, Machine Learning Technologies, Inc, Littleton, CO, USA, 1998.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9048808, 18 pages
https://doi.org/10.1155/2021/9048808

Research Article

Optimized Vegetation Density to Dissipate Energy of Flood Flow in Open Canals

Mahdi Feizbahr,[1] **Navid Tonekaboni**,[2] **Guang-Jun Jiang**,[3,4] **and Hong-Xia Chen**[3,4]

[1]*School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia*
[2]*Mechanical-Energy Conversion, Islamic Azad University, Nour Branch, Nour, Iran*
[3]*School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010051, China*
[4]*Inner Mongolia Key Laboratory of Advanced Manufacturing Technology, Hohhot, Inner Mongolia 010051, China*

Correspondence should be addressed to Guang-Jun Jiang; jianggj_2003@163.com

Received 8 April 2021; Revised 18 June 2021; Accepted 17 July 2021; Published 11 August 2021

Academic Editor: Mohammad Yazdi

Vegetation along the river increases the roughness and reduces the average flow velocity, reduces flow energy, and changes the flow velocity profile in the cross section of the river. Many canals and rivers in nature are covered with vegetation during the floods. Canal's roughness is strongly affected by plants and therefore it has a great effect on flow resistance during flood. Roughness resistance against the flow due to the plants depends on the flow conditions and plant, so the model should simulate the current velocity by considering the effects of velocity, depth of flow, and type of vegetation along the canal. Total of 48 models have been simulated to investigate the effect of roughness in the canal. The results indicated that, by enhancing the velocity, the effect of vegetation in decreasing the bed velocity is negligible, while when the current has lower speed, the effect of vegetation on decreasing the bed velocity is obviously considerable.

1. Introduction

Considering the impact of each variable is a very popular field within the analytical and statistical methods and intelligent systems [1–14]. This can help research for better modeling considering the relation of variables or interaction of them toward reaching a better condition for the objective function in control and engineering [15–27]. Consequently, it is necessary to study the effects of the passive factors on the active domain [28–36]. Because of the effect of vegetation on reducing the discharge capacity of rivers [37], pruning plants was necessary to improve the condition of rivers. One of the important effects of vegetation in river protection is the action of roots, which cause soil consolidation and soil structure improvement and, by enhancing the shear strength of soil, increase the resistance of canal walls against the erosive force of water. The outer limbs of the plant increase the roughness of the canal walls and reduce the flow velocity and deplete the flow energy in vicinity of the walls. Vegetation by reducing the shear stress of the canal bed reduces flood discharge and sedimentation in the intervals between vegetation and increases the stability of the walls [38–41].

One of the main factors influencing the speed, depth, and extent of flood in this method is Manning's roughness coefficient. On the other hand, soil cover [42], especially vegetation, is one of the most determining factors in Manning's roughness coefficient. Therefore, it is expected that those seasonal changes in the vegetation of the region will play an important role in the calculated value of Manning's roughness coefficient and ultimately in predicting the flood wave behavior [43–45]. The roughness caused by plants' resistance to flood current depends on the flow and plant conditions. Flow conditions include depth and velocity of the plant, and plant conditions include plant type, hardness or flexibility, dimensions, density, and shape of the plant [46]. In general, the issue discussed in this research is the optimization of flood-induced flow in canals

by considering the effect of vegetation-induced roughness. Therefore, the effect of plants on the roughness coefficient and canal transmission coefficient and in consequence the flow depth should be evaluated [47, 48].

Current resistance is generally known by its roughness coefficient. The equation that is mainly used in this field is Manning equation. The ratio of shear velocity to average current velocity $(V*/V)$ is another form of current resistance. The reason for using the $(V*/V)$ ratio is that it is dimensionless and has a strong theoretical basis. The reason for using Manning roughness coefficient is its pervasiveness. According to Freeman et al. [49], the Manning roughness coefficient for plants was calculated according to the Kouwen and Unny [50] method for incremental resistance. This method involves increasing the roughness for various surface and plant irregularities. Manning's roughness coefficient has all the factors affecting the resistance of the canal. Therefore, the appropriate way to more accurately estimate this coefficient is to know the factors affecting this coefficient [51].

To calculate the flow rate, velocity, and depth of flow in canals as well as flood and sediment estimation, it is important to evaluate the flow resistance. To determine the flow resistance in open ducts, Manning, Chézy, and Darcy–Weisbach relations are used [52]. In these relations, there are parameters such as Manning's roughness coefficient (n), Chézy roughness coefficient (C), and Darcy–Weisbach coefficient (f). All three of these coefficients are a kind of flow resistance coefficient that is widely used in the equations governing flow in rivers [53].

The three relations that express the relationship between the average flow velocity (V) and the resistance and geometric and hydraulic coefficients of the canal are as follows:

$$V = \frac{K_n}{n} R^{(2/3)} S_f^{(1/2)}, \quad \text{Manning}, \tag{1}$$

$$V = \sqrt{\frac{8g}{f}} \sqrt{RS_f}, \quad \text{Darcy} - \text{Weisbach}, \tag{2}$$

$$V = C\sqrt{RS_f}, \quad \text{Chezy}, \tag{3}$$

where n, f, and c are Manning, Darcy–Weisbach, and Chézy coefficients, respectively. V = average flow velocity, R = hydraulic radius, S_f = slope of energy line, which in uniform flow is equal to the slope of the canal bed, g = gravitational acceleration, and K_n is a coefficient whose value is equal to 1 in the SI system and 1.486 in the English system. The coefficients of resistance in equations (1) to (3) are related as follows:

$$\sqrt{\frac{f}{8}} = \frac{n}{R^{(1/6)}} \frac{\sqrt{g}}{K_n} = \frac{\sqrt{g}}{C} = \frac{\sqrt{gRS_f}}{V}. \tag{4}$$

Based on the boundary layer theory, the flow resistance for rough substrates is determined from the following general relation:

$$\sqrt{\frac{8}{f}} = 2.5Ln\frac{y}{K_s} + A, \tag{5}$$

where f = Darcy–Weisbach coefficient of friction, y = flow depth, K_s = bed roughness size, and A = constant coefficient.

On the other hand, the relationship between the Darcy–Weisbach coefficient of friction and the shear velocity of the flow is as follows:

$$\sqrt{\frac{8}{f}} = \frac{V}{U^*}. \tag{6}$$

By using equation (6), equation (5) is converted as follows:

$$\sqrt{\frac{8}{f}} = \frac{V}{U^*} 2.5Ln\frac{y}{K_s} + A. \tag{7}$$

Investigation on the effect of vegetation arrangement on shear velocity of flow in laboratory conditions showed that, with increasing the shear Reynolds number (Re^*), the numerical value of the V/U^* ratio also increases; in other words the amount of roughness coefficient increases with a slight difference in the cases without vegetation, checkered arrangement, and cross arrangement, respectively [54].

Roughness in river vegetation is simulated in mathematical models with a variable floor slope flume by different densities and discharges. The vegetation considered submerged in the bed of the flume. Results showed that, with increasing vegetation density, canal roughness and flow shear speed increase and with increasing flow rate and depth, Manning's roughness coefficient decreases. Factors affecting the roughness caused by vegetation include the effect of plant density and arrangement on flow resistance, the effect of flow velocity on flow resistance, and the effect of depth [45, 55].

One of the works that has been done on the effect of vegetation on the roughness coefficient is Darby [56] study, which investigates a flood wave model that considers all the effects of vegetation on the roughness coefficient. There are currently two methods for estimating vegetation roughness. One method is to add the thrust force effect to Manning's equation [47, 57, 58] and the other method is to increase the canal bed roughness (Manning-Strickler coefficient) [45, 59–61]. These two methods provide acceptable results in models designed to simulate floodplain flow. Wang et al. [62] simulate the floodplain with submerged vegetation using these two methods and to increase the accuracy of the results, they suggested using the effective height of the plant under running water instead of using the actual height of the plant. Freeman et al. [49] provided equations for determining the coefficient of vegetation roughness under different conditions. Lee et al. [63] proposed a method for calculating the Manning coefficient using the flow velocity ratio at different depths. Much research has been done on the Manning roughness coefficient in rivers, and researchers [49, 63–66] sought to obtain a specific number for n to use in river engineering. However, since the depth and geometric conditions of rivers are completely variable in different

places, the values of Manning roughness coefficient have changed subsequently, and it has not been possible to choose a fixed number. In river engineering software, the Manning roughness coefficient is determined only for specific and constant conditions or normal flow. Lee et al. [63] stated that seasonal conditions, density, and type of vegetation should also be considered. Hydraulic roughness and Manning roughness coefficient n of the plant were obtained by estimating the total Manning roughness coefficient from the matching of the measured water surface curve and water surface height. The following equation is used for the flow surface curve:

$$\frac{\partial y}{\partial x} = \frac{S_0 - S_f}{1 - \mathrm{Fr}^2}, \tag{8}$$

where $\partial y/\partial x$ is the depth of water change, S_0 is the slope of the canal floor, Sf is the slope of the energy line, and Fr is the Froude number which is obtained from the following equation:

$$Fr = \frac{V}{\sqrt{g\ D}}, \tag{9}$$

where D is the characteristic length of the canal. Flood flow velocity is one of the important parameters of flood waves, which is very important in calculating the water level profile and energy consumption. In the cases where there are many limitations for researchers due to the wide range of experimental dimensions and the variety of design parameters, the use of numerical methods that are able to estimate the rest of the unknown results with acceptable accuracy is economically justified.

FLOW-3D software uses Finite Difference Method (FDM) for numerical solution of two-dimensional and three-dimensional flow. This software is dedicated to computational fluid dynamics (CFD) and is provided by Flow Science [67]. The flow is divided into networks with tubular cells. For each cell there are values of dependent variables and all variables are calculated in the center of the cell, except for the velocity, which is calculated at the center of the cell. In this software, two numerical techniques have been used for geometric simulation, FAVOR™ (Fractional-Area-Volume-Obstacle-Representation) and the VOF (Volume-of-Fluid) method. The equations used at this model for this research include the principle of mass survival and the magnitude of motion as follows. The fluid motion equations in three dimensions, including the Navier–Stokes equations with some additional terms, are as follows:

$$\begin{aligned}
&\frac{\partial u}{\partial t} + \frac{1}{V_F}\left[uA_x\frac{\partial u}{\partial x} + vA_yR\frac{\partial u}{\partial y} + wA_z\frac{\partial u}{\partial z}\right] - \xi\frac{A_y v^2}{xV_F} = -\frac{1}{\rho}\frac{\partial P}{\partial x} + G_x + f_x - \frac{\mathrm{RSOR}}{\rho V_F}u,\\
&\frac{\partial v}{\partial t} + \frac{1}{V_F}\left[uA_x\frac{\partial v}{\partial x} + vA_yR\frac{\partial v}{\partial y} + wA_z\frac{\partial v}{\partial z}\right] - \xi\frac{A_y vu}{xV_F} = -\frac{1}{\rho}\left(R\frac{\partial P}{\partial y}\right) + G_y + f_y - \frac{\mathrm{RSOR}}{\rho V_F}v,\\
&\frac{\partial w}{\partial t} + \frac{1}{V_F}\left[uA_x\frac{\partial w}{\partial x} + vA_yR\frac{\partial w}{\partial y} + wA_z\frac{\partial w}{\partial z}\right] = -\frac{1}{\rho}\frac{\partial P}{\partial z} + G_z + f_z - \frac{\mathrm{RSOR}}{\rho V_F}w,
\end{aligned} \tag{10}$$

where G_x, G_y, G_Z are mass accelerations in the directions x, y, z and f_x, f_y, f_z are viscosity accelerations in the directions x, y, z and are obtained from the following equations:

$$\begin{aligned}
\rho V_F f_x &= wsx - \left[\frac{\partial}{\partial x}\left(A_x\tau_{xx}\right) + R\frac{\partial}{\partial y}\left(A_y\tau_{xy}\right) + \frac{\partial}{\partial z}\left(A_z\tau_{xz}\right) + \frac{\xi}{x}\left(A_x\tau_{xx} - A_y\tau_{yy}\right)\right],\\
\rho V_F f_y &= wsy - \left[\frac{\partial}{\partial x}\left(A_x\tau_{xy}\right) + R\frac{\partial}{\partial y}\left(A_y\tau_{yy}\right) + \frac{\partial}{\partial z}\left(A_z\tau_{yz}\right) + \frac{\xi}{x}\left(A_x - A_y\tau_{xy}\right)\right],\\
\rho V_F f_z &= wsz - \left[\frac{\partial}{\partial x}\left(A_x\tau_{xz}\right) + R\frac{\partial}{\partial y}\left(A_y\tau_{yz}\right) + \frac{\partial}{\partial z}\left(A_z\tau_{zz}\right) + \frac{\xi}{x}\left(A_x\tau_{xz}\right)\right].
\end{aligned} \tag{11}$$

Shear stresses $\tau_{xx}, \tau_{yy}, \tau_{zz}, \tau_{xy}, \tau_{xz}, \tau_{yz}$ in equation (11) are obtained from the following equations:

$$\tau_{xx} = -2\mu\left[\frac{\partial u}{\partial x} - \frac{1}{3}\left(\frac{\partial u}{\partial x} + R\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} + \frac{\xi u}{x}\right)\right],$$

$$\tau_{yy} = -2\mu\left[\frac{\partial v}{\partial y} + \xi\frac{u}{x} - \frac{1}{3}\left(\frac{\partial u}{\partial x} + R\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} + \frac{\xi u}{x}\right)\right],$$

$$\tau_{zz} = -2\mu\left[\frac{\partial w}{\partial z} - \frac{1}{3}\left(\frac{\partial u}{\partial x} + R\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} + \frac{\xi u}{x}\right)\right],$$

$$\tau_{xy} = -\mu\left[\frac{\partial v}{\partial x} + R\frac{\partial u}{\partial y} - \frac{\xi v}{x}\right], \tag{12}$$

$$\tau_{xz} = -\mu\left[\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right],$$

$$\tau_{yz} = -\mu\left[\frac{\partial v}{\partial z} + R\frac{\partial w}{\partial y}\right].$$

The standard model is used for high Reynolds currents, but in this model, RNG theory allows the analytical differential formula to be used for the effective viscosity that occurs at low Reynolds numbers. Therefore, the RNG model can be used for low and high Reynolds currents.

Weather changes are high and this affects many factors continuously. The presence of vegetation in any area reduces the velocity of surface flows and prevents soil erosion, so vegetation will have a significant impact on reducing destructive floods. One of the methods of erosion protection in floodplain watersheds is the use of biological methods. The presence of vegetation in watersheds reduces the flow rate during floods and prevents soil erosion. The external organs of plants increase the roughness and decrease the velocity of water flow and thus reduce its shear stress energy. One of the important factors with which the hydraulic resistance of plants is expressed is the roughness coefficient. Measuring the roughness coefficient of plants and investigating their effect on reducing velocity and shear stress of flow is of special importance.

Roughness coefficients in canals are affected by two main factors, namely, flow conditions and vegetation characteristics [68]. So far, much research has been done on the effect of the roughness factor created by vegetation, but the issue of plant density has received less attention. For this purpose, this study was conducted to investigate the effect of vegetation density on flow velocity changes.

In a study conducted using a software model on three density modes in the submerged state effect on flow velocity changes in 48 different modes was investigated (Table 1).

The number of cells used in this simulation is equal to 1955888 cells. The boundary conditions were introduced to the model as a constant speed and depth (Figure 1). At the output boundary, due to the presence of supercritical current, no parameter for the current is considered. Absolute roughness for floors and walls was introduced to the model (Figure 1). In this case, the flow was assumed to be nonviscous and air entry into the flow was not considered. After 7.7×10^{-4} seconds, this model reached a convergence accuracy of 2.39×10^{-5}.

Due to the fact that it is not possible to model the vegetation in FLOW-3D software, in this research, the vegetation of small soft plants was studied so that Manning's coefficients can be entered into the canal bed in the form of roughness coefficients obtained from the studies of Chow [69] in similar conditions. In practice, in such modeling, the effect of plant height is eliminated due to the small height of herbaceous plants, and modeling can provide relatively acceptable results in these conditions.

48 models with input velocities proportional to the height of the regular semihexagonal canal were considered to create supercritical conditions. Manning coefficients were applied based on Chow [69] studies in order to control the canal bed. Speed profiles were drawn and discussed.

Any control and simulation system has some inputs that we should determine to test any technology [70–77]. Determination and true implementation of such parameters is one of the key steps of any simulation [23, 78–81] and computing procedure [82–86]. The input current is created by applying the flow rate through the VFR (Volume Flow Rate) option and the output flow is considered Output and for other borders the Symmetry option is considered.

Simulation of the models and checking their action and responses and observing how a process behaves is one of the accepted methods in engineering and science [87, 88]. For verification of FLOW-3D software, the results of computer simulations are compared with laboratory measurements and according to the values of computational error, convergence error, and the time required for convergence, the most appropriate option for real-time simulation is selected (Figures 2 and 3).

The canal is 7 meters long, 0.5 meters wide, and 0.8 meters deep. This test was used to validate the application of the software to predict the flow rate parameters. In this experiment, instead of using the plant, cylindrical pipes were used in the bottom of the canal.

The conditions of this modeling are similar to the laboratory conditions and the boundary conditions used in the laboratory were used for numerical modeling. The critical flow enters the simulation model from the upstream boundary, so in the upstream boundary conditions, critical velocity and depth are considered. The flow at the downstream boundary is supercritical, so no parameters are applied to the downstream boundary.

The software well predicts the process of changing the speed profile in the open canal along with the considered obstacles. The error in the calculated speed values can be due to the complexity of the flow and the interaction of the turbulence caused by the roughness of the floor with the

Table 1: The studied models.

Model index	Study models	Z	G	Hydraulic depth	Input speed	Froude number	Plant density	Manning coefficient based on Chow's experiences
Model1-V3-Ch 0.03	Model 1	1	9.81	0.88	3	1.016409	0.25	0.03
Model2-V3-Ch 0.035	Model 2	1	9.81	0.88	3	1.016409	0.5	0.035
Model3-V3-Ch 0.04	Model 3	1	9.81	0.88	3	1.016409	0.75	0.04
Model4-V3-Ch 0.02	Model 4	1	9.81	0.88	3	1.016409	0	0.02
Model5-V3.1-Ch 0.03	Model 5	1	9.81	0.88	3.1	1.0502893	0.25	0.03
Model6-V3.1-Ch 0.035	Model 6	1	9.81	0.88	3.1	1.0502893	0.5	0.035
Model7-V3.1-Ch 0.04	Model 7	1	9.81	0.88	3.1	1.0502893	0.75	0.04
Model8-V3.1-Ch 0.02	Model 8	1	9.81	0.88	3.1	1.0502893	0	0.02
Model9-V3.2-Ch 0.03	Model 9	1	9.81	0.88	3.2	1.0841696	0.25	0.03
Model10-V3.2-Ch 0.035	Model 10	1	9.81	0.88	3.2	1.0841696	0.5	0.035
Model11-V3.2-Ch 0.04	Model 11	1	9.81	0.88	3.2	1.0841696	0.75	0.04
Model12-V3.2-Ch 0.02	Model 12	1	9.81	0.88	3.2	1.0841696	0	0.02
Model13-V3.3-Ch 0.03	Model 13	1	9.81	0.88	3.3	1.1180499	0.25	0.03
Model14-V3.3-Ch 0.035	Model 14	1	9.81	0.88	3.3	1.1180499	0.5	0.035
Model15-V3.3-Ch 0.04	Model 15	1	9.81	0.88	3.3	1.1180499	0.75	0.04
Model16-V3.3-Ch 0.02	Model 16	1	9.81	0.88	3.3	1.1180499	0	0.02
Model17-V4-Ch 0.03	Model 17	2	9.81	1.46	4	1.055455338	0.25	0.03
Model18-V4-Ch 0.035	Model 18	2	9.81	1.46	4	1.055455338	0.5	0.035
Model19-V4-Ch 0.04	Model 19	2	9.81	1.46	4	1.055455338	0.75	0.04
Model20-V4-Ch 0.02	Model 20	2	9.81	1.46	4	1.055455338	0	0.02
Model21-V4.1-Ch 0.03	Model 21	2	9.81	1.46	4.1	1.081841721	0.25	0.03
Model22-V4.1-Ch 0.035	Model 22	2	9.81	1.46	4.1	1.081841721	0.5	0.035
Model23-V4.1-Ch 0.04	Model 23	2	9.81	1.46	4.1	1.081841721	0.75	0.04
Model24-V4.1-Ch 0.02	Model 24	2	9.81	1.46	4.1	1.081841721	0	0.02
Model25-V4.2-Ch 0.03	Model 25	2	9.81	1.46	4.2	1.108228105	0.25	0.03
Model26-V4.2-Ch 0.035	Model 26	2	9.81	1.46	4.2	1.108228105	0.5	0.035
Model27-V4.2-Ch 0.04	Model 27	2	9.81	1.46	4.2	1.108228105	0.75	0.04
Model28-V4.2-Ch 0.02	Model 28	2	9.81	1.46	4.2	1.108228105	0	0.02
Model29-V4.3-Ch 0.03	Model 29	2	9.81	1.46	4.3	1.134614488	0.25	0.03
Model30-V4.3-Ch 0.035	Model 30	2	9.81	1.46	4.3	1.134614488	0.5	0.035
Model31-V4.3-Ch 0.04	Model 31	2	9.81	1.46	4.3	1.134614488	0.75	0.04
Model32-V4.3-Ch 0.02	Model 32	2	9.81	1.46	4.3	1.134614488	0	0.02
Model33-V5-Ch 0.03	Model 33	3	9.81	1.91	5	1.153022908	0.25	0.03
Model34-V5-Ch 0.035	Model 34	3	9.81	1.91	5	1.153022908	0.5	0.035
Model35-V5-Ch 0.04	Model 35	3	9.81	1.91	5	1.153022908	0.75	0.04
Model36-V5-Ch 0.02	Model 36	3	9.81	1.91	5	1.153022908	0	0.02
Model37-V5.1-Ch 0.03	Model 37	3	9.81	1.91	5.1	1.176083366	0.25	0.03
Model38-V5.1-Ch 0.035	Model 38	3	9.81	1.91	5.1	1.176083366	0.5	0.035
Model39-V5.1-Ch 0.04	Model 39	3	9.81	1.91	5.1	1.176083366	0.75	0.04
Model40-V5.1-Ch 0.02	Model 40	3	9.81	1.91	5.1	1.176083366	0	0.02
Model41-V5.2-Ch 0.03	Model 41	3	9.81	1.91	5.2	1.199143825	0.25	0.03
Model42-V5.2-Ch 0.035	Model 42	3	9.81	1.91	5.2	1.199143825	0.5	0.035
Model43-V5.2-Ch 0.04	Model 43	3	9.81	1.91	5.2	1.199143825	0.75	0.04
Model44-V5.2-Ch 0.02	Model 44	3	9.81	1.91	5.2	1.199143825	0	0.02
Model45-V5.3-Ch 0.03	Model 45	3	9.81	1.91	5.3	1.222204283	0.25	0.03
Model46-V5.3-Ch 0.035	Model 46	3	9.81	1.91	5.3	1.222204283	0.5	0.035
Model47-V5.3-Ch 0.04	Model 47	3	9.81	1.91	5.3	1.222204283	0.75	0.04
Model48-V5.3-Ch 0.02	Model 48	3	9.81	1.91	5.3	1.222204283	0	0.02

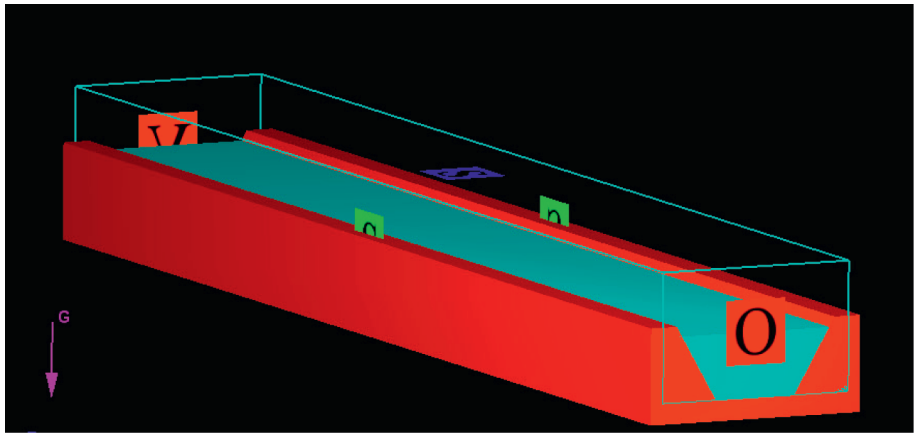


Figure 1: The simulated model and its boundary conditions.

turbulence caused by the three-dimensional cycles in the hydraulic jump. As a result, the software is able to predict the speed distribution in open canals.

2. Modeling Results

After analyzing the models, the results were shown in graphs (Figures 4–14). The total number of experiments in this study was 48 due to the limitations of modeling.

To investigate the effects of roughness with flow velocity, the trend of flow velocity changes at different depths and with supercritical flow to a Froude number proportional to the depth of the section has been obtained.

According to the velocity profiles of Figure 5, it can be seen that, with the increasing of Manning's coefficient, the canal bed speed decreases.

According to Figures 5 to 8, it can be found that, with increasing the Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of the models 1 to 12, which can be justified by increasing the speed and of course increasing the Froude number.

According to Figure 10, we see that, with increasing Manning's coefficient, the canal bed speed decreases.

According to Figure 11, we see that, with increasing Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of Figures 5–10, which can be justified by increasing the speed and, of course, increasing the Froude number.

With increasing Manning's coefficient, the canal bed speed decreases (Figure 12). But this deceleration is more noticeable than the deceleration of the higher models (Figures 5–8 and 10, 11), which can be justified by increasing the speed and, of course, increasing the Froude number.

According to Figure 13, with increasing Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of Figures 5 to 12, which can be justified by increasing the speed and, of course, increasing the Froude number.

According to Figure 15, with increasing Manning's coefficient, the canal bed speed decreases.

According to Figure 16, with increasing Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of the higher model, which can be justified by increasing the speed and, of course, increasing the Froude number.

According to Figure 17, it is clear that, with increasing Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of the higher models, which can be justified by increasing the speed and, of course, increasing the Froude number.

According to Figure 18, with increasing Manning's coefficient, the canal bed speed decreases. But this deceleration is more noticeable than the deceleration of the higher models, which can be justified by increasing the speed and, of course, increasing the Froude number.

According to Figure 19, it can be seen that the vegetation placed in front of the flow input velocity has negligible effect on the reduction of velocity, which of course can be justified due to the flexibility of the vegetation. The only unusual thing is the unexpected decrease in floor speed of 3 m/s compared to higher speeds.

According to Figure 20, by increasing the speed of vegetation, the effect of vegetation on reducing the flow rate becomes more noticeable. And the role of input current does not have much effect in reducing speed.

According to Figure 21, it can be seen that, with increasing speed, the effect of vegetation on reducing the bed flow rate becomes more noticeable and the role of the input current does not have much effect. In general, it can be seen that, by increasing the speed of the input current, the slope of the profiles increases from the bed to the water surface and due to the fact that, in software, the roughness coefficient applies to the channel floor only in the boundary conditions, this can be perfectly justified. Of course, it can be noted that, due to the flexible conditions of the vegetation of the bed, this modeling can show acceptable results for such grasses in the canal floor. In the next directions, we may try application of swarm-based optimization methods for modeling and finding the most effective factors in this research [2, 7, 8, 15, 18, 89–94]. In future, we can also apply the

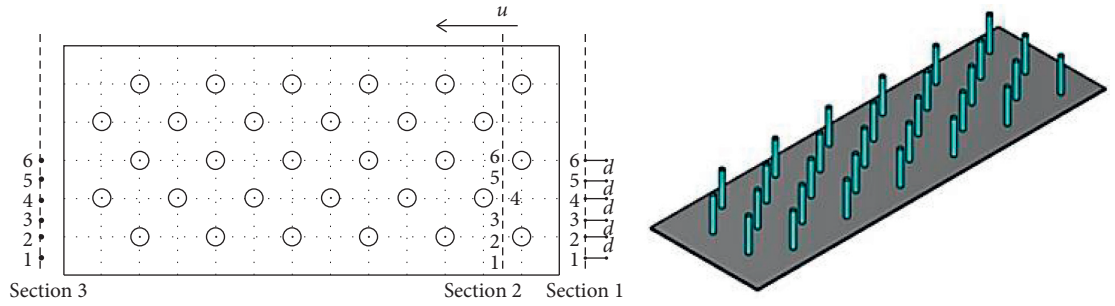


Figure 2: Modeling the plant with cylindrical tubes at the bottom of the canal.

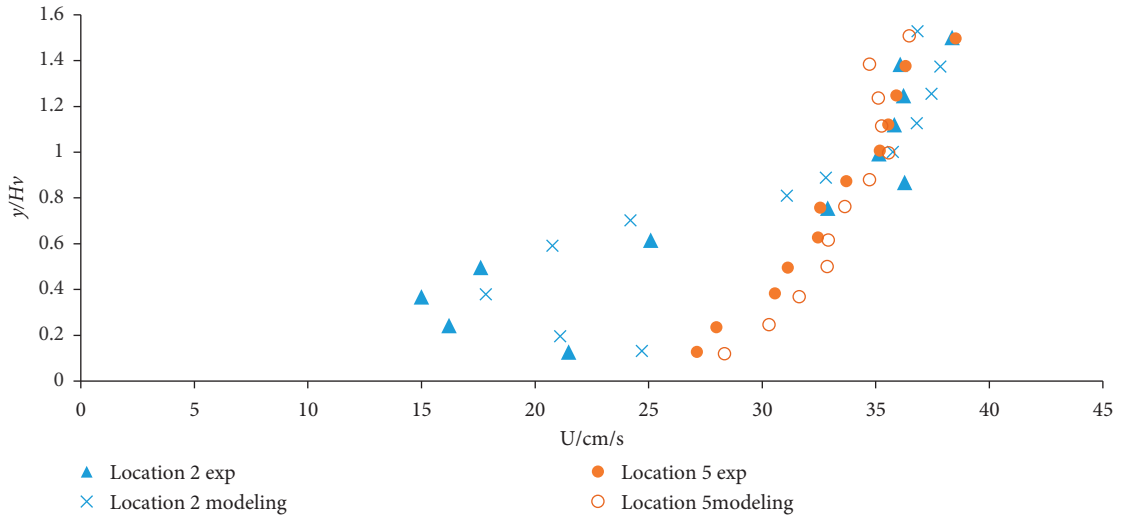


Figure 3: Velocity profiles in positions 2 and 5.

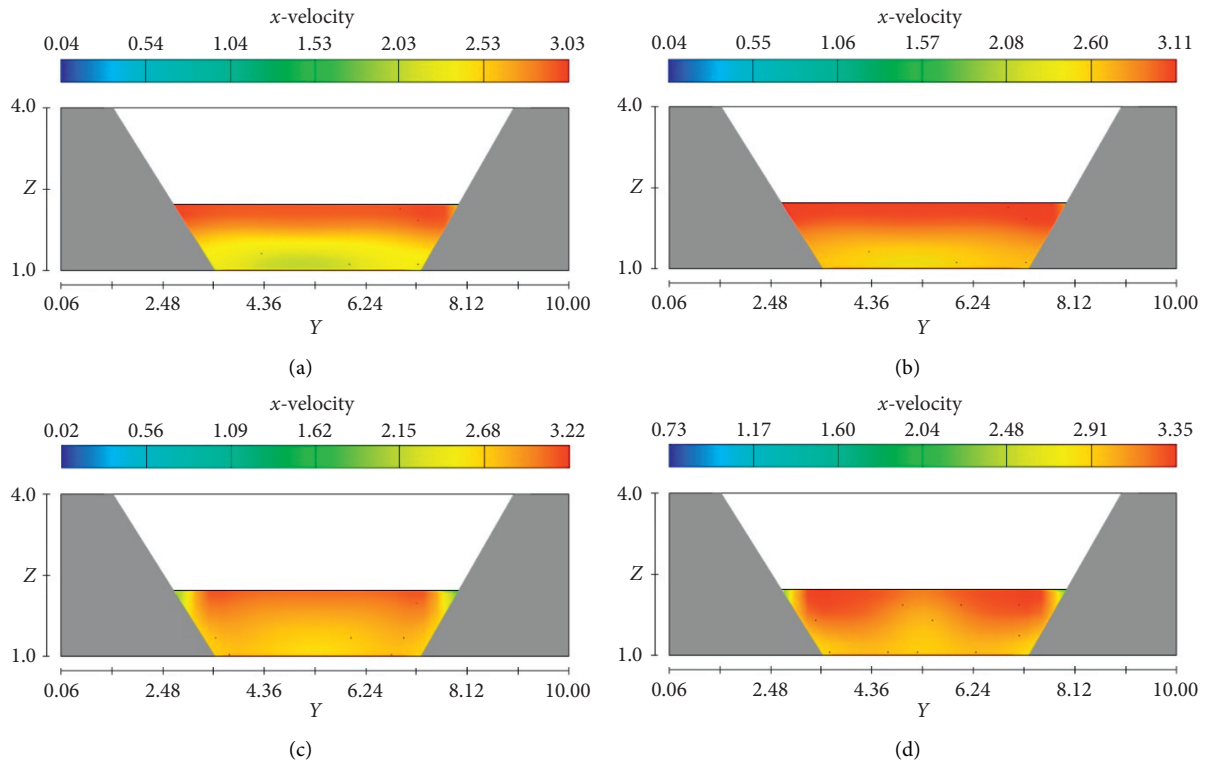


Figure 4: Flow velocity profiles for canals with a depth of 1 m and flow velocities of 3–3.3 m/s. Canal with a depth of 1 meter and a flow velocity of (a) 3 meters per second, (b) 3.1 meters per second, (c) 3.2 meters per second, and (d) 3.3 meters per second.

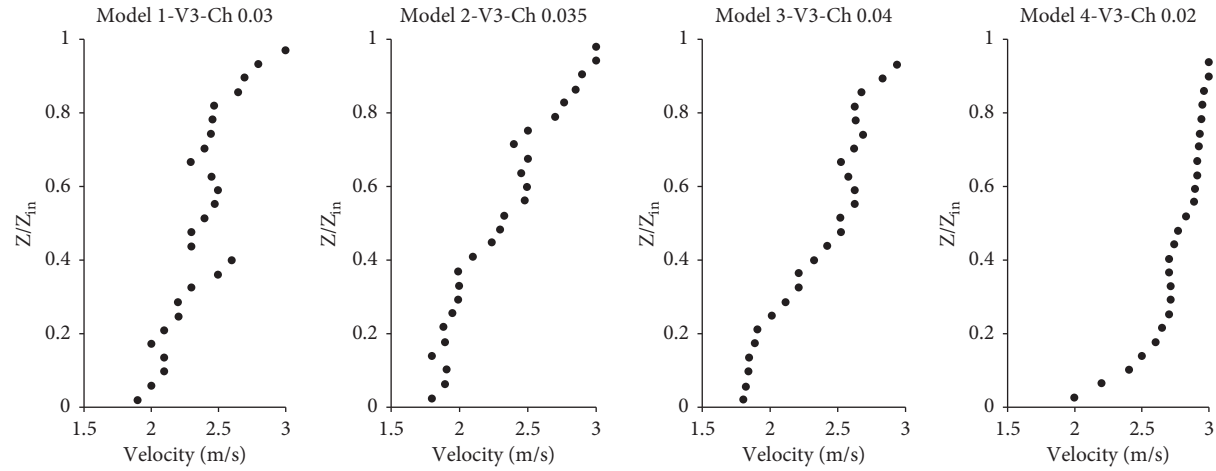


Figure 5: Canal diagram with a depth of 1 meter and a flow rate of 3 meters per second.

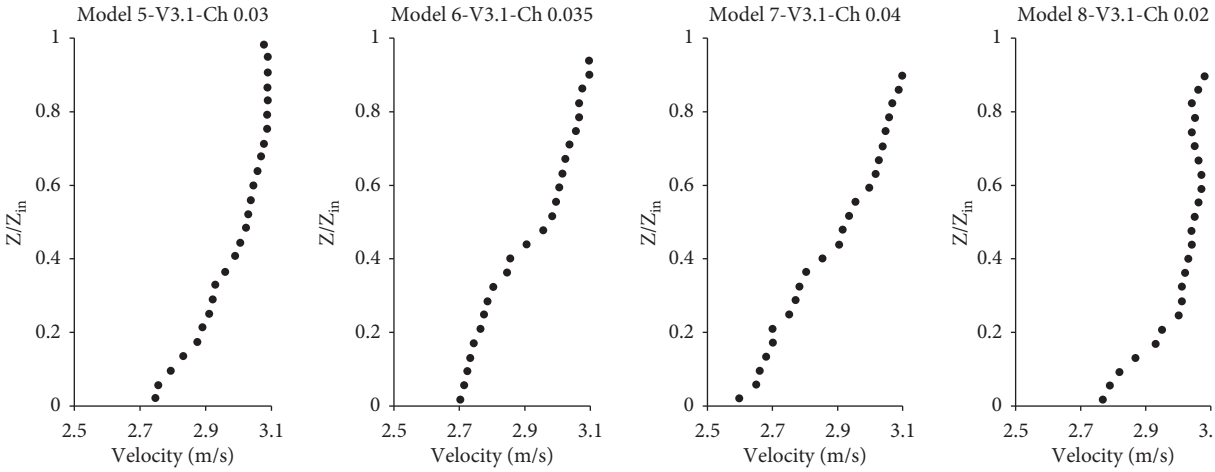


Figure 6: Canal diagram with a depth of 1 meter and a flow rate of 3.1 meters per second.

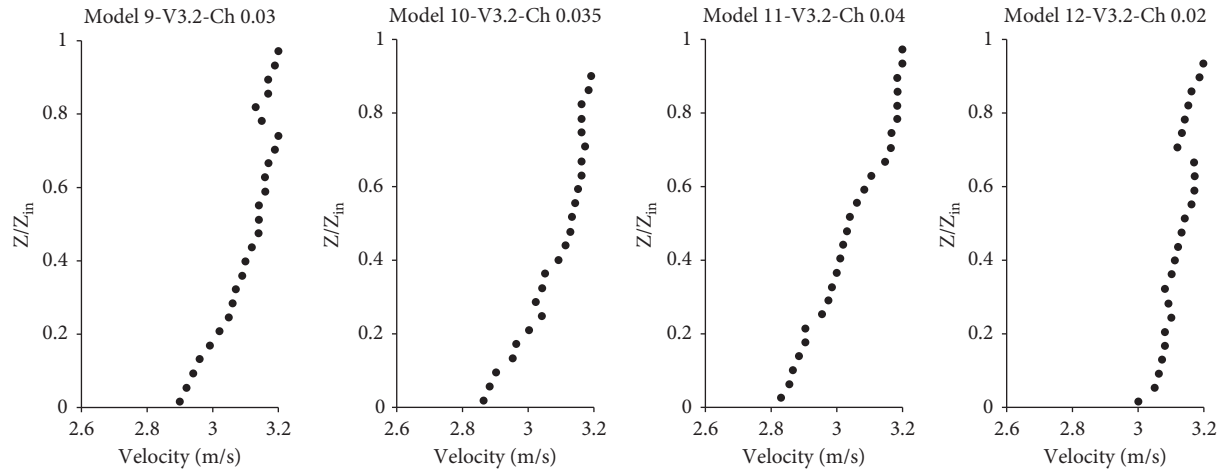


Figure 7: Canal diagram with a depth of 1 meter and a flow rate of 3.2 meters per second.

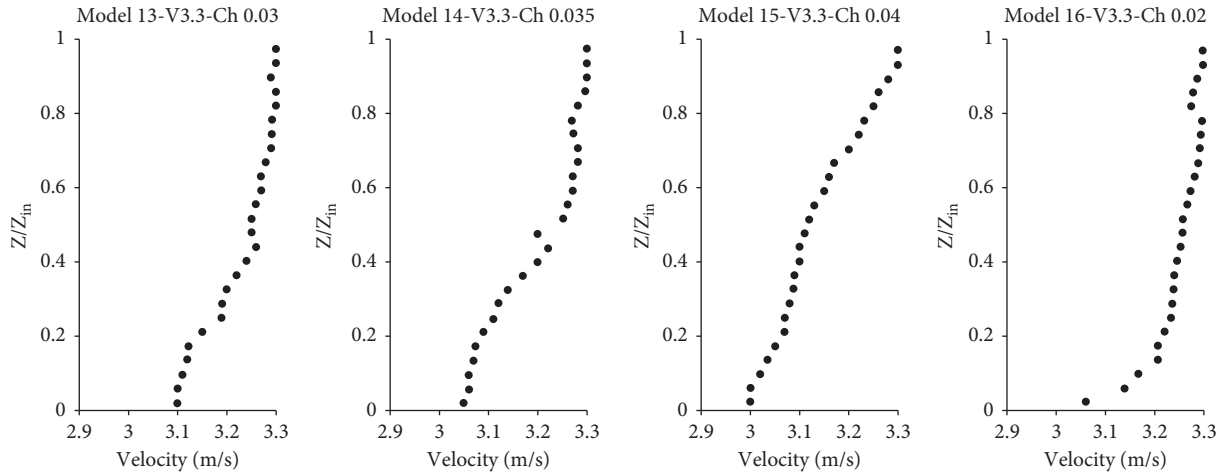


Figure 8: Canal diagram with a depth of 1 meter and a flow rate of 3.3 meters per second.

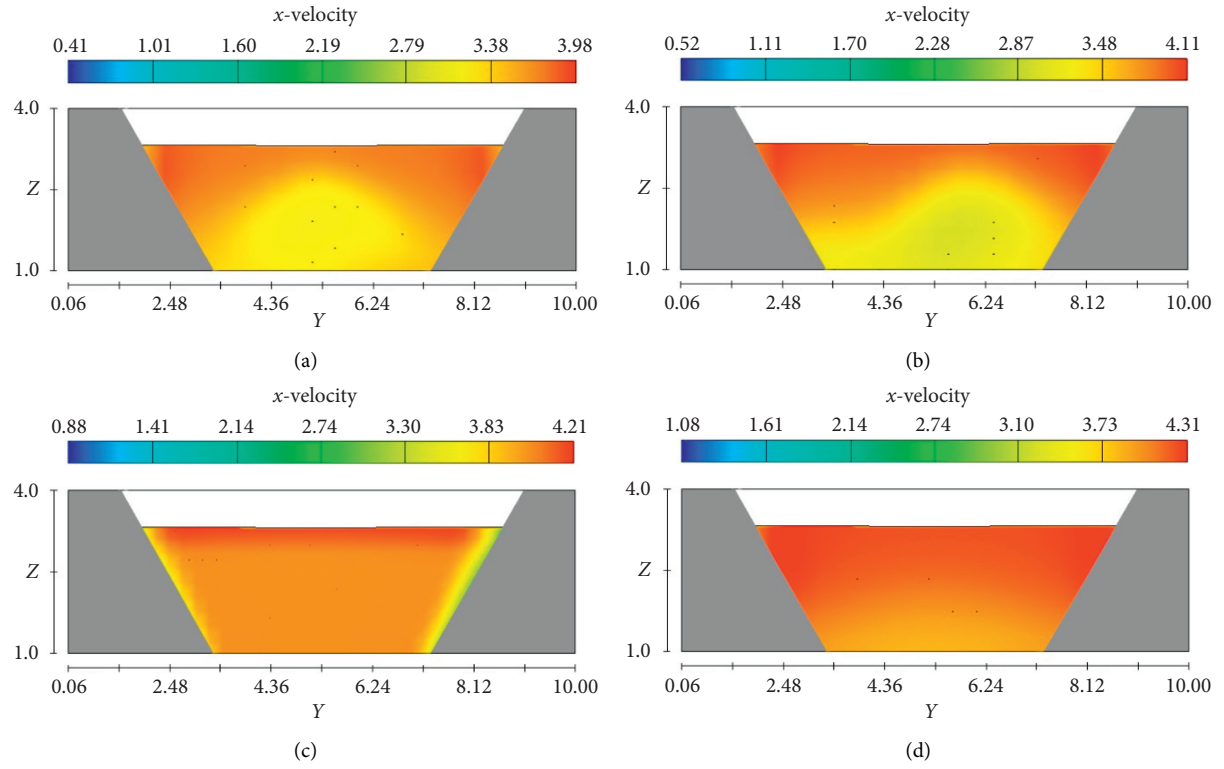


Figure 9: Flow velocity profiles for canals with a depth of 2 m and flow velocities of 4–4.3 m/s. Canal with a depth of 2 meters and a flow rate of (a) 4 meters per second, (b) 4.1 meters per second, (c) 4.2 meters per second, and (d) 4.3 meters per second.

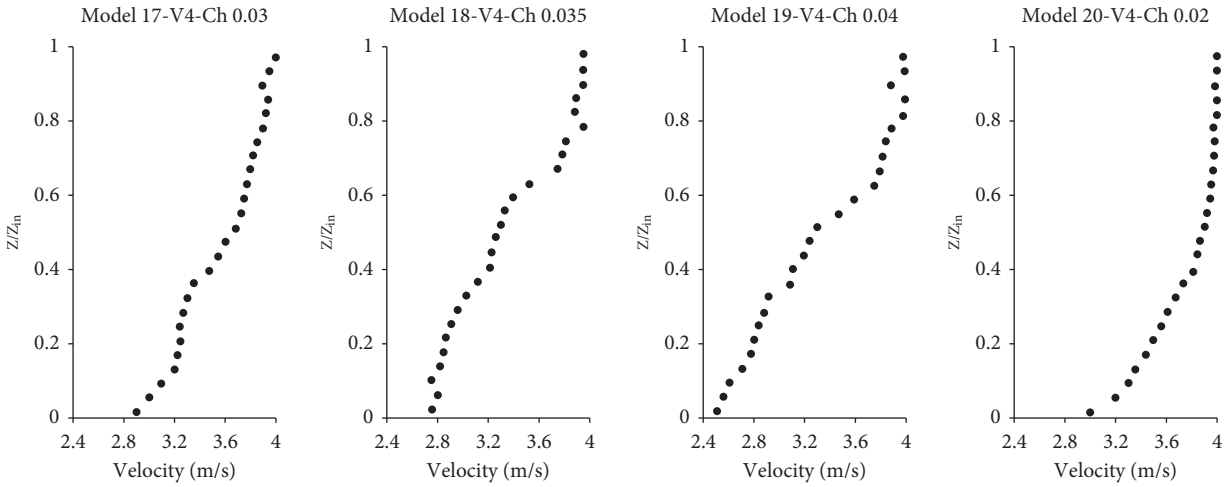


Figure 10: Canal diagram with a depth of 2 meters and a flow rate of 4 meters per second.

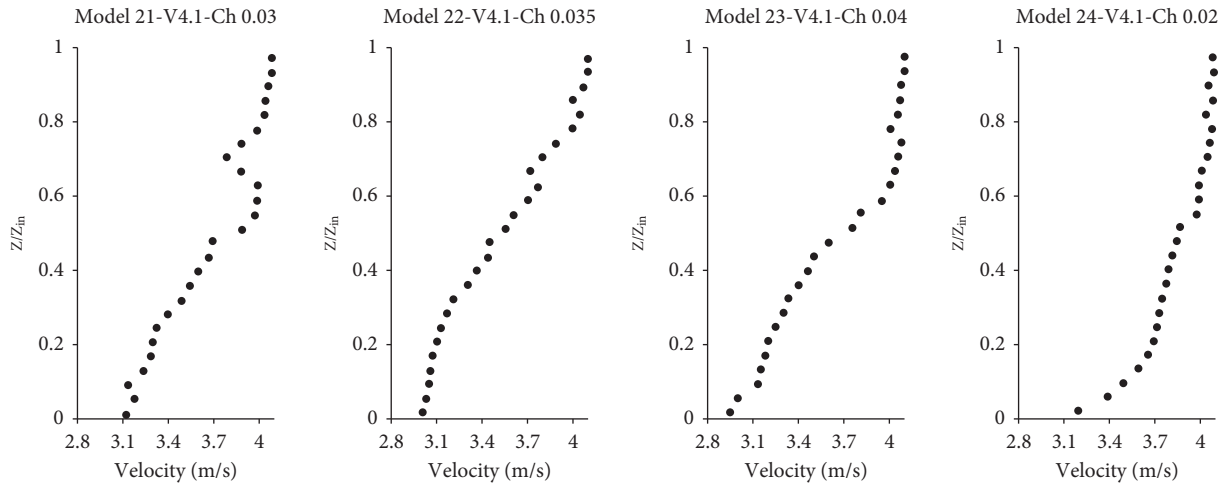


Figure 11: Canal diagram with a depth of 2 meters and a flow rate of 4.1 meters per second.

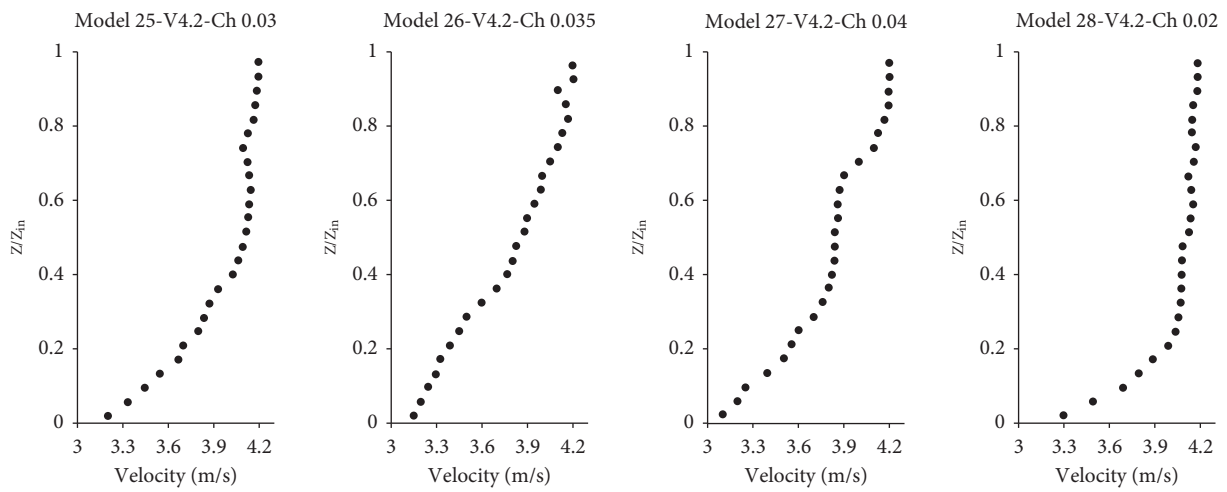


Figure 12: Canal diagram with a depth of 2 meters and a flow rate of 4.2 meters per second.

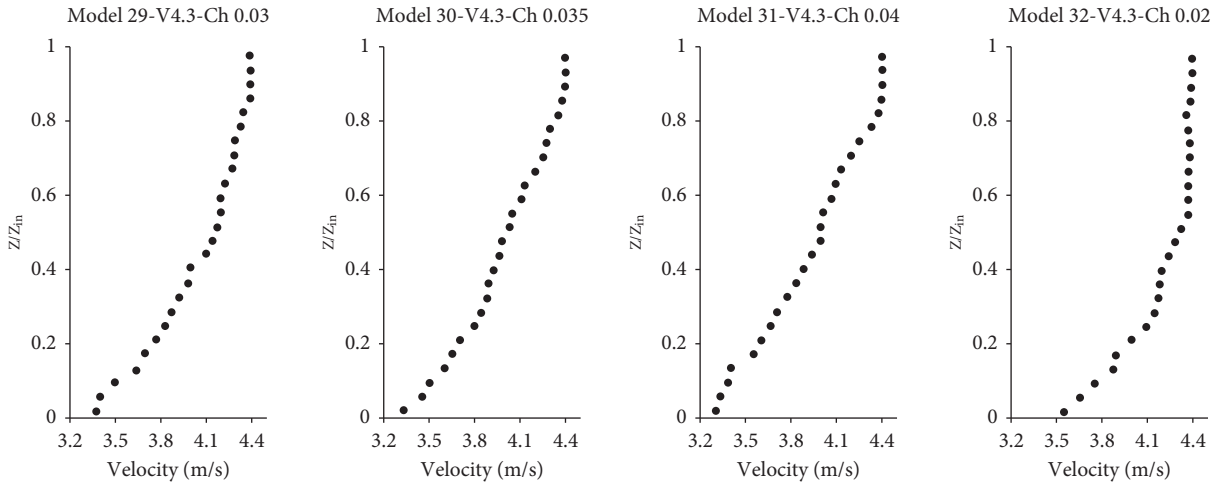


Figure 13: Canal diagram with a depth of 2 meters and a flow rate of 4.3 meters per second.

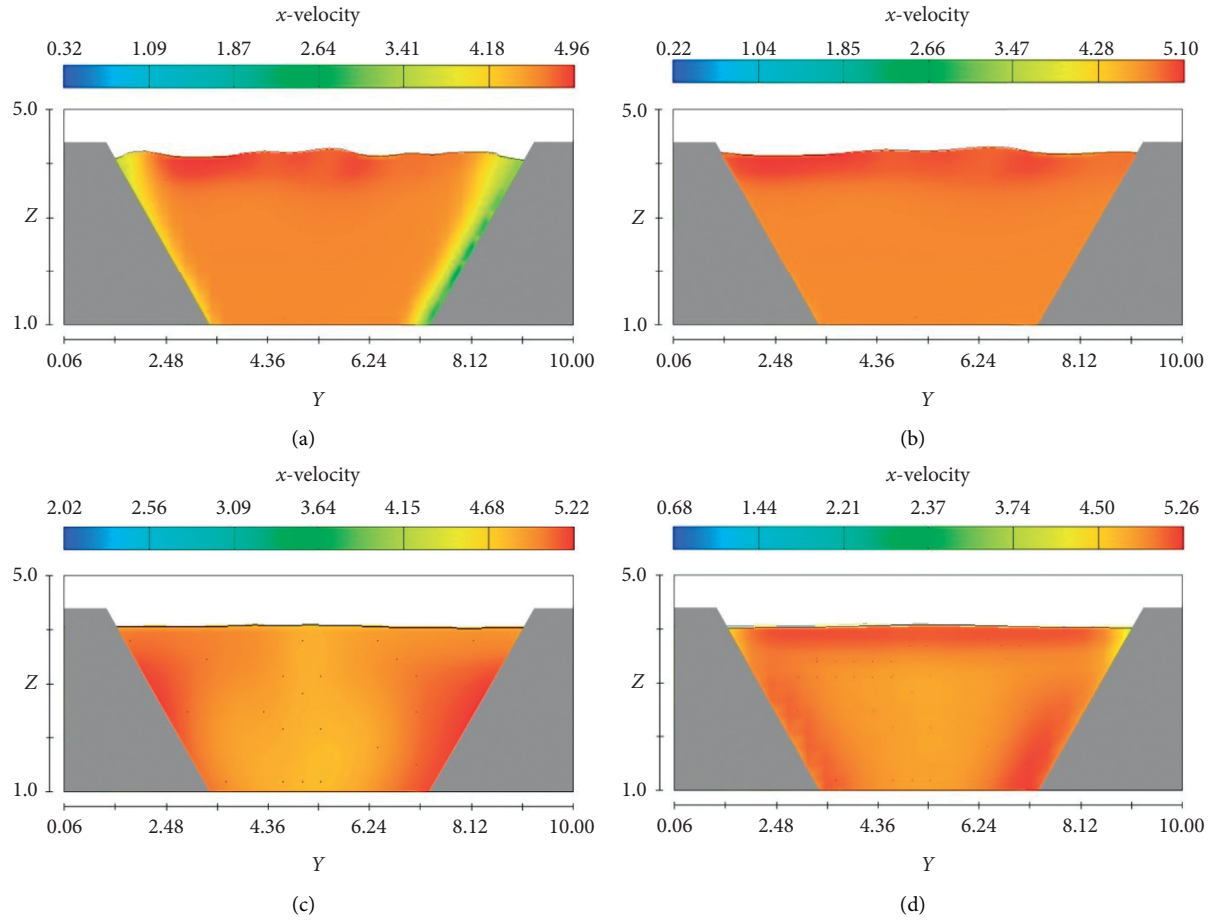


Figure 14: Flow velocity profiles for canals with a depth of 3 m and flow velocities of 5–5.3 m/s. Canal with a depth of 2 meters and a flow rate of (a) 4 meters per second, (b) 4.1 meters per second, (c) 4.2 meters per second, and (d) 4.3 meters per second.

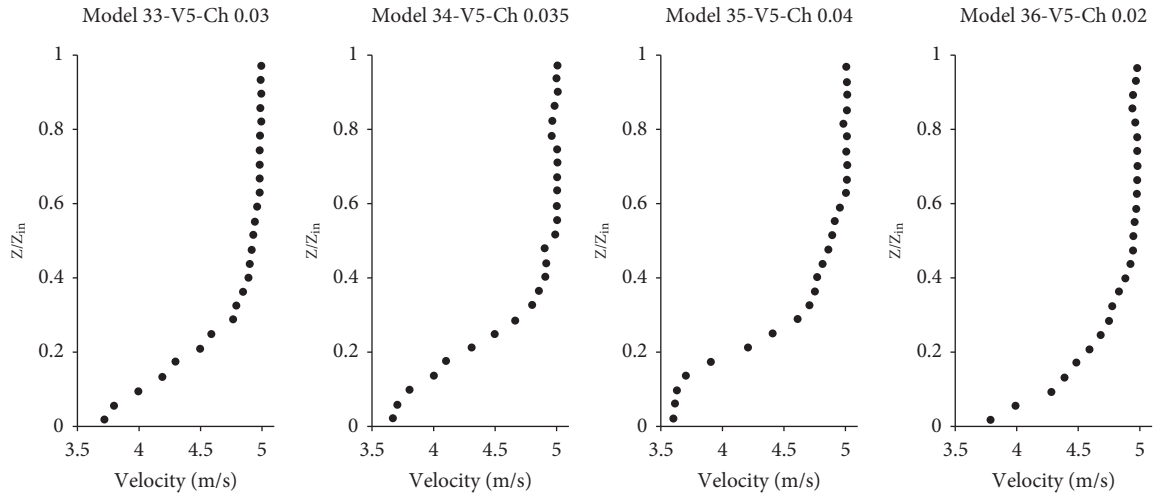


Figure 15: Canal diagram with a depth of 3 meters and a flow rate of 5 meters per second.

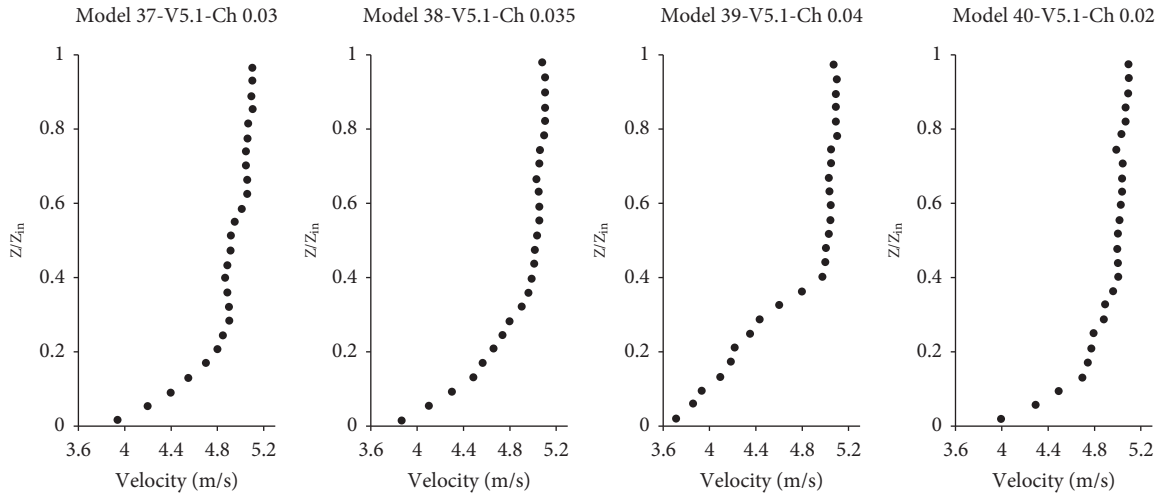


Figure 16: Canal diagram with a depth of 3 meters and a flow rate of 5.1 meters per second.

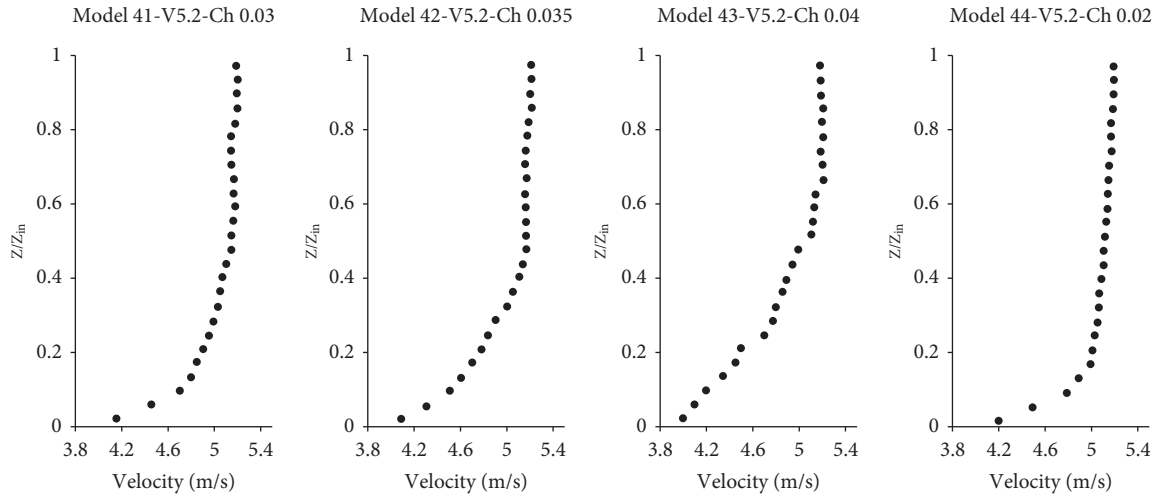


Figure 17: Canal diagram with a depth of 3 meters and a flow rate of 5.2 meters per second.

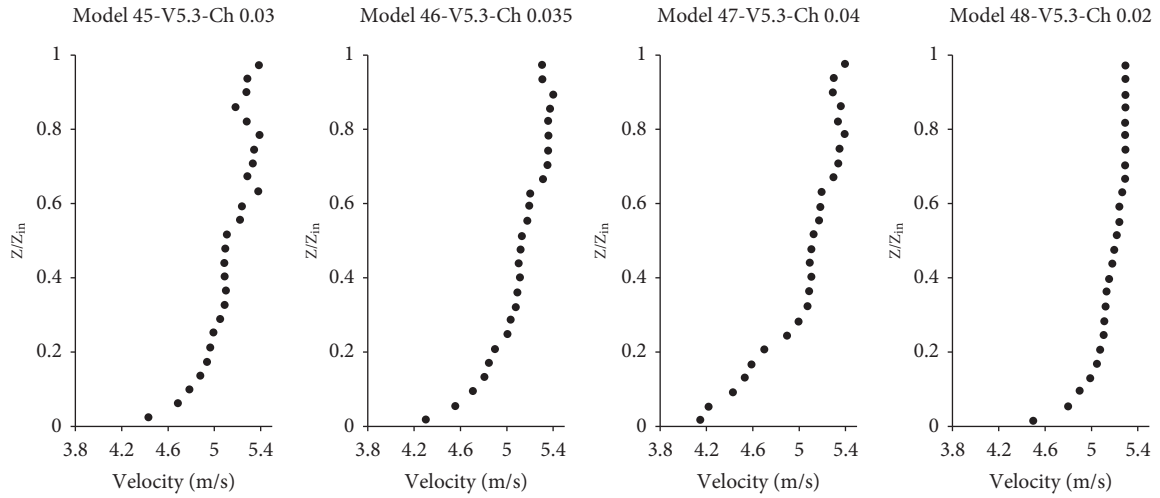


Figure 18: Canal diagram with a depth of 3 meters and a flow rate of 5.3 meters per second.

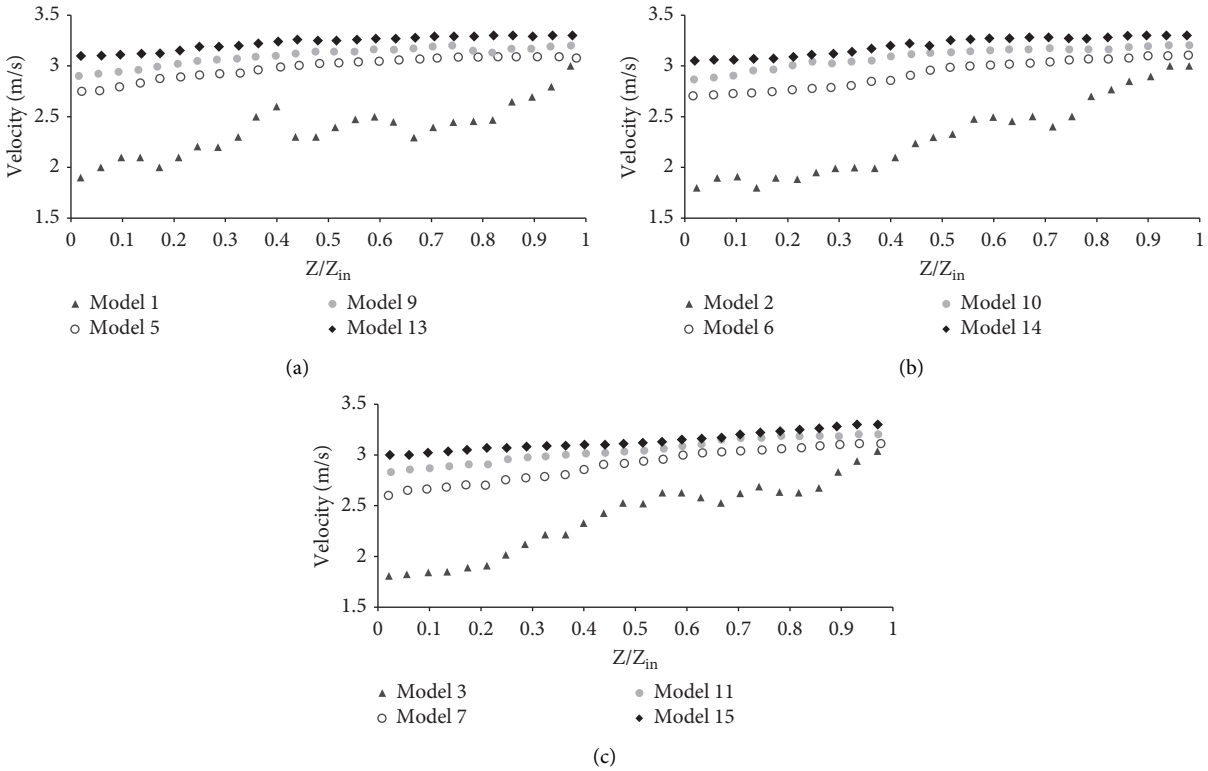


Figure 19: Comparison of velocity profiles with the same plant densities (depth 1 m). Comparison of velocity profiles with (a) plant densities of 25%, depth 1 m; (b) plant densities of 50%, depth 1 m; and (c) plant densities of 75%, depth 1 m.

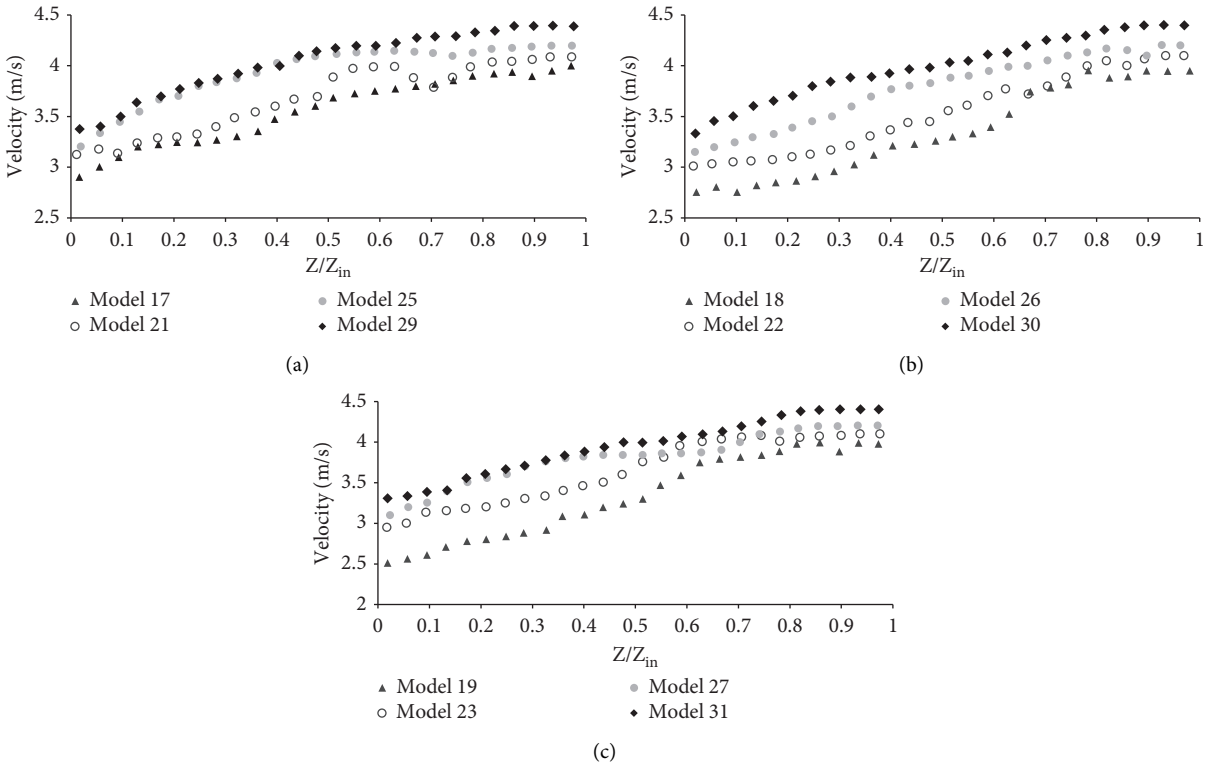


Figure 20: Comparison of velocity profiles with the same plant densities (depth 2 m). Comparison of velocity profiles with (a) plant densities of 25%, depth 2 m; (b) plant densities of 50%, depth 2 m; and (c) plant densities of 75%, depth 2 m.

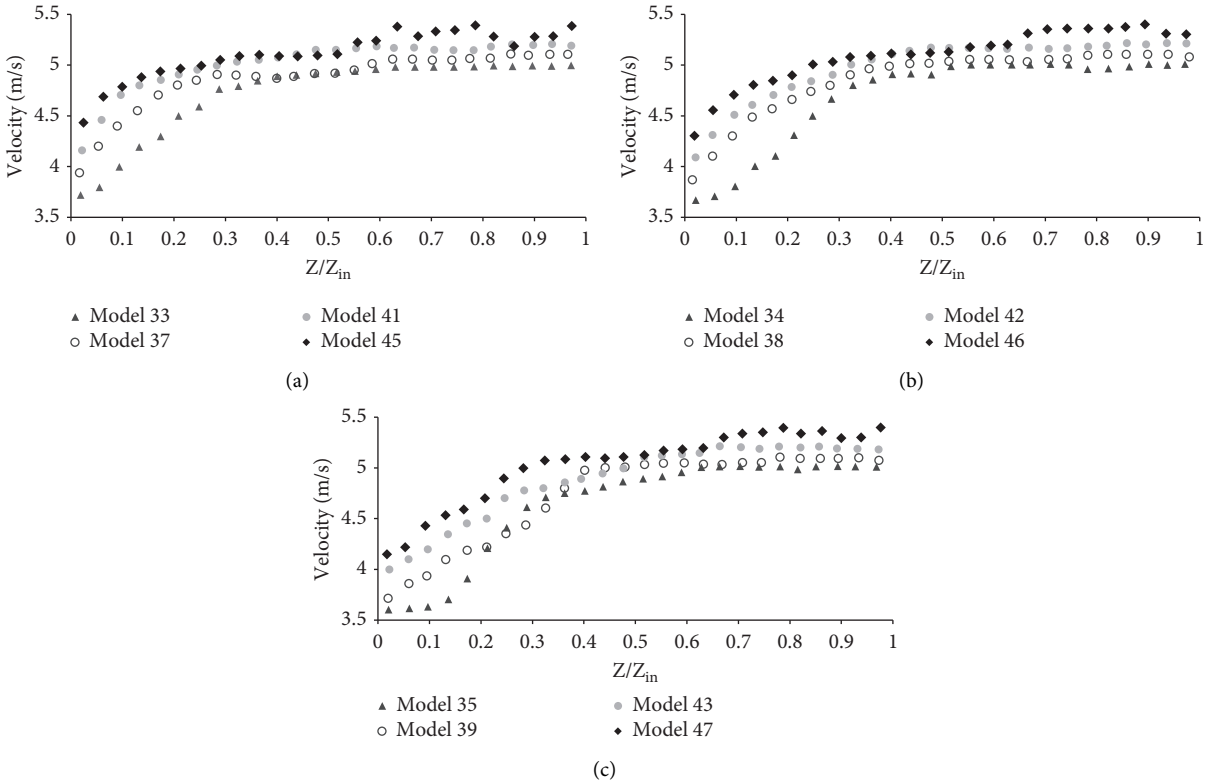


Figure 21: Comparison of velocity profiles with the same plant densities (depth 3 m). Comparison of velocity profiles with (a) plant densities of 25%, depth 3 m; (b) plant densities of 50%, depth 3 m; and (c) plant densities of 75%, depth 3 m.

simulation logic and software of this research for other domains such as power engineering [95–99].

3. Conclusion

The effects of vegetation on the flood canal were investigated by numerical modeling with FLOW-3D software. After analyzing the results, the following conclusions were reached:

(i) Increasing the density of vegetation reduces the velocity of the canal floor but has no effect on the velocity of the canal surface.

(ii) Increasing the Froude number is directly related to increasing the speed of the canal floor.

(iii) In the canal with a depth of one meter, a sudden increase in speed can be observed from the lowest speed and higher speed, which is justified by the sudden increase in Froude number.

(iv) As the inlet flow rate increases, the slope of the profiles from the bed to the water surface increases.

(v) By reducing the Froude number, the effect of vegetation on reducing the flow bed rate becomes more noticeable. And the input velocity in reducing the velocity of the canal floor does not have much effect.

(vi) At a flow rate between 3 and 3.3 meters per second due to the shallow depth of the canal and the higher landing number a more critical area is observed in which the flow bed velocity in this area is between 2.86 and 3.1 m/s.

(vii) Due to the critical flow velocity and the slight effect of the roughness of the horseshoe vortex floor, it is not visible and is only partially observed in models 1-2-3 and 21.

(viii) As the flow rate increases, the effect of vegetation on the rate of bed reduction decreases.

(ix) In conditions where less current intensity is passing, vegetation has a greater effect on reducing current intensity and energy consumption increases.

(x) In the case of using the flow rate of 0.8 cubic meters per second, the velocity distribution and flow regime show about 20% more energy consumption than in the case of using the flow rate of 1.3 cubic meters per second.

Nomenclature

n: Manning's roughness coefficient
C: Chézy roughness coefficient
f: Darcy–Weisbach coefficient
V: Flow velocity
R: Hydraulic radius
g: Gravitational acceleration
y: Flow depth
Ks: Bed roughness
A: Constant coefficient
Re^*: Reynolds number
$\partial y/\partial x$: Depth of water change
S_0: Slope of the canal floor
Sf: Slope of energy line
Fr: Froude number
D: Characteristic length of the canal
G: Mass acceleration
τ: Shear stresses.

Data Availability

All data are included within the paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Contract no. 71761030 and Natural Science Foundation of Inner Mongolia under Contract no. 2019LH07003.

References

[1] H. Yu, L. Jie, W. Gui et al., "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: method and analysis," *Engineering with Computers*, vol. 20, pp. 1–29, 2020.

[2] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing*, vol. 24, pp. 585–596, 2014.

[3] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, Article ID 106684, 2021.

[4] C. Yu, M. Chen, K. Chen et al., "SGOA: annealing-behaved grasshopper optimizer for global tasks," *Engineering with Computers*, vol. 4, pp. 1–28, 2021.

[5] W. Shan, Z. Qiao, A. A. Heidari, H. Chen, H. Turabieh, and Y. Teng, "Double adaptive weights for stabilization of moth flame optimizer: balance analysis, engineering cases, and medical diagnosis," *Knowledge-Based Systems*, vol. 8, Article ID 106728, 2020.

[6] J. Tu, H. Chen, J. Liu et al., "Evolutionary biogeography-based whale optimization methods with communication structure: towards measuring the balance," *Knowledge-Based Systems*, vol. 212, Article ID 106642, 2021.

[7] Y. Zhang, R. Liu, X. Wang et al., "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, 2020.

[8] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[9] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme

learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[10] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[11] X. Xu and H.-L. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Computing*, vol. 18, no. 4, pp. 797–807, 2014.

[12] M. Wang, H. Chen, B. Yang et al., "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, 2017.

[13] L. Chao, K. Zhang, Z. Li, Y. Zhu, J. Wang, and Z. Yu, "Geographically weighted regression based methods for merging satellite and gauge precipitation," *Journal of Hydrology*, vol. 558, pp. 275–289, 2018.

[14] F. J. Golrokh, G. Azeem, and A. Hasan, "Eco-efficiency evaluation in cement industries: DEA malmquist productivity index using optimization models," *ENG Transactions*, vol. 1, 2020.

[15] D. Zhao, L. Lei, F. Yu et al., "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowledge-Based Systems*, vol. 8, Article ID 106510, 2020.

[16] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection," *Engineering with Computers*, vol. 517, pp. 1–30, 2020.

[17] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[18] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

[19] X. Zhao, X. Zhang, Z. Cai et al., "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Computational Biology and Chemistry*, vol. 78, pp. 481–490, 2019.

[20] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced Moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.

[21] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing Journal*, vol. 88, Article ID 105946, 2020.

[22] Y. Chen, J. Li, H. Lu, and P. Yan, "Coupling system dynamics analysis and risk aversion programming for optimizing the mixed noise-driven shale gas-water supply chains," *Journal of Cleaner Production*, vol. 278, Article ID 123209, 2020.

[23] H. Tang, Y. Xu, A. Lin et al., "Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers," *IEEE Access*, vol. 8, pp. 35546–35562, 2020.

[24] H.-J. Ma and G.-H. Yang, "Adaptive fault tolerant control of cooperative heterogeneous systems with actuator faults and unreliable interconnections," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, pp. 3240–3255, 2015.

[25] H.-J. Ma and L.-X. Xu, "Decentralized adaptive fault-tolerant control for a class of strong interconnected nonlinear systems via graph theory," *IEEE Transactions on Automatic Control*, vol. 66, 2020.

[26] H. J. Ma, L. X. Xu, and G. H. Yang, "Multiple environment integral reinforcement learning-based fault-tolerant control for affine nonlinear systems," *IEEE Transactions on Cybernetics*, vol. 51, pp. 1–16, 2019.

[27] J. Hu, M. Wang, C. Zhao, Q. Pan, and C. Du, "Formation control and collision avoidance for multi-UAV systems based on Voronoi partition," *Science China Technological Sciences*, vol. 63, no. 1, pp. 65–72, 2020.

[28] C. Zhang, H. Li, Y. Qian, C. Chen, and X. Zhou, "Locality-constrained discriminative matrix regression for robust face identification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1–15, 2020.

[29] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 238–255, 2019.

[30] X. Zhang, J. Wang, T. Wang, R. Jiang, J. Xu, and L. Zhao, "Robust feature learning for adversarial defense via hierarchical feature alignment," *Information Sciences*, vol. 560, 2020.

[31] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 03, p. 1, 2020.

[32] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Computer Vision and Image Understanding*, vol. 197-198, Article ID 103003, 2020.

[33] X. Zhang, T. Wang, W. Luo, and P. Huang, "Multi-level fusion and attention-guided CNN for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 3, p. 1, 2020.

[34] L. He, J. Shen, and Y. Zhang, "Ecological vulnerability assessment for ecological conservation and environmental management," *Journal of Environmental Management*, vol. 206, pp. 1115–1125, 2018.

[35] Y. Chen, W. Zheng, W. Li, and Y. Huang, "Large group Activity security risk assessment and risk early warning based on random forest algorithm," *Pattern Recognition Letters*, vol. 144, pp. 1–5, 2021.

[36] J. Hu, H. Zhang, Z. Li, C. Zhao, Z. Xu, and Q. Pan, "Object traversing by monocular UAV in outdoor environment," *Asian Journal of Control*, vol. 25, 2020.

[37] P. Tian, H. Lu, W. Feng, Y. Guan, and Y. Xue, "Large decrease in streamflow and sediment load of Qinghai-Tibetan Plateau driven by future climate change: a case study in Lhasa River Basin," *Catena*, vol. 187, Article ID 104340, 2020.

[38] A. Stokes, C. Atger, A. G. Bengough, T. Fourcaud, and R. C. Sidle, "Desirable plant root traits for protecting natural and engineered slopes against landslides," *Plant and Soil*, vol. 324, no. 1, pp. 1–30, 2009.

[39] T. B. Devi, A. Sharma, and B. Kumar, "Studies on emergent flow over vegetative channel bed with downward seepage," *Hydrological Sciences Journal*, vol. 62, no. 3, pp. 408–420, 2017.

[40] G. Ireland, M. Volpi, and G. Petropoulos, "Examining the capability of supervised machine learning classifiers in extracting flooded areas from Landsat TM imagery: a case study from a Mediterranean flood," *Remote Sensing*, vol. 7, no. 3, pp. 3372–3399, 2015.

[41] L. Goodarzi and S. Javadi, "Assessment of aquifer vulnerability using the DRASTIC model; a case study of the Dezful-Andimeshk Aquifer," *Computational Research Progress in Applied Science & Engineering*, vol. 2, no. 1, pp. 17–22, 2016.

[42] K. Zhang, Q. Wang, L. Chao et al., "Ground observation-based analysis of soil moisture spatiotemporal variability

across a humid to semi-humid transitional zone in China," *Journal of Hydrology*, vol. 574, pp. 903–914, 2019.

[43] L. De Doncker, P. Troch, R. Verhoeven, K. Bal, P. Meire, and J. Quintelier, "Determination of the Manning roughness coefficient influenced by vegetation in the river Aa and Biebrza river," *Environmental Fluid Mechanics*, vol. 9, no. 5, pp. 549–567, 2009.

[44] M. Fathi-Moghadam and K. Drikvandi, "Manning roughness coefficient for rivers and flood plains with non-submerged vegetation," *International Journal of Hydraulic Engineering*, vol. 1, no. 1, pp. 1–4, 2012.

[45] F.-C. Wu, H. W. Shen, and Y.-J. Chou, "Variation of roughness coefficients for unsubmerged and submerged vegetation," *Journal of Hydraulic Engineering*, vol. 125, no. 9, pp. 934–942, 1999.

[46] M. K. Wood, "Rangeland vegetation-hydrologic interactions," in *Vegetation Science Applications for Rangeland Analysis and Management*vol. 3, , pp. 469–491, Springer, 1988.

[47] C. Wilson, O. Yagci, H.-P. Rauch, and N. Olsen, "3D numerical modelling of a willow vegetated river/floodplain system," *Journal of Hydrology*, vol. 327, no. 1-2, pp. 13–21, 2006.

[48] R. Yazarloo, M. Khamehchian, and M. R. Nikoodel, "Observational-computational 3d engineering geological model and geotechnical characteristics of young sediments of golestan province," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 03, 2017.

[49] G. E. Freeman, W. H. Rahmeyer, and R. R. Copeland, "Determination of resistance due to shrubs and woody vegetation," *International Journal of River Basin Management*, vol. 19, 2000.

[50] N. Kouwen and T. E. Unny, "Flexible roughness in open channels," *Journal of the Hydraulics Division*, vol. 99, no. 5, pp. 713–728, 1973.

[51] S. Hosseini and J. Abrishami, *Open Channel Hydraulics*, Elsevier, Amsterdam, Netherlands, 2007.

[52] C. S. James, A. L. Birkhead, A. A. Jordanova, and J. J. O'Sullivan, "Flow resistance of emergent vegetation," *Journal of Hydraulic Research*, vol. 42, no. 4, pp. 390–398, 2004.

[53] F. Huthoff and D. Augustijn, "Channel roughness in 1D steady uniform flow: Manning or Chézy?," *NCR-days*, vol. 102, 2004.

[54] M. S. Sabegh, M. Saneie, M. Habibi, A. A. Abbasi, and M. Ghadimkhani, "Experimental investigation on the effect of river bank tree planting array, on shear velocity," *Journal of Watershed Engineering and Management*, vol. 2, no. 4, 2011.

[55] A. Errico, V. Pasquino, M. Maxwald, G. B. Chirico, L. Solari, and F. Preti, "The effect of flexible vegetation on flow in drainage channels: estimation of roughness coefficients at the real scale," *Ecological Engineering*, vol. 120, pp. 411–421, 2018.

[56] S. E. Darby, "Effect of riparian vegetation on flow resistance and flood potential," *Journal of Hydraulic Engineering*, vol. 125, no. 5, pp. 443–454, 1999.

[57] V. Kutija and H. Thi Minh Hong, "A numerical model for assessing the additional resistance to flow introduced by flexible vegetation," *Journal of Hydraulic Research*, vol. 34, no. 1, pp. 99–114, 1996.

[58] T. Fischer-Antze, T. Stoesser, P. Bates, and N. R. B. Olsen, "3D numerical modelling of open-channel flow with submerged vegetation," *Journal of Hydraulic Research*, vol. 39, no. 3, pp. 303–310, 2001.

[59] U. Stephan and D. Gutknecht, "Hydraulic resistance of submerged flexible vegetation," *Journal of Hydrology*, vol. 269, no. 1-2, pp. 27–43, 2002.

[60] F. G. Carollo, V. Ferro, and D. Termini, "Flow resistance law in channels with flexible submerged vegetation," *Journal of Hydraulic Engineering*, vol. 131, no. 7, pp. 554–564, 2005.

[61] W. Fu-sheng, "Flow resistance of flexible vegetation in open channel," *Journal of Hydraulic Engineering*, vol. S1, 2007.

[62] P.-f. Wang, C. Wang, and D. Z. Zhu, "Hydraulic resistance of submerged vegetation related to effective height," *Journal of Hydrodynamics*, vol. 22, no. 2, pp. 265–273, 2010.

[63] J. K. Lee, L. C. Roig, H. L. Jenter, and H. M. Visser, "Drag coefficients for modeling flow through emergent vegetation in the Florida Everglades," *Ecological Engineering*, vol. 22, no. 4-5, pp. 237–248, 2004.

[64] G. J. Arcement and V. R. Schneider, *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*, US Government Printing Office, Washington, DC, USA, 1989.

[65] Y. Ding and S. S. Y. Wang, "Identification of Manning's roughness coefficients in channel network using adjoint analysis," *International Journal of Computational Fluid Dynamics*, vol. 19, no. 1, pp. 3–13, 2005.

[66] E. T. Engman, "Roughness coefficients for routing surface runoff," *Journal of Irrigation and Drainage Engineering*, vol. 112, no. 1, pp. 39–53, 1986.

[67] M. Feizbahr, C. Kok Keong, F. Rostami, and M. Shahrokhi, "Wave energy dissipation using perforated and non perforated piles," *International Journal of Engineering*, vol. 31, no. 2, pp. 212–219, 2018.

[68] M. Farzadkhoo, A. Keshavarzi, H. Hamidifar, and M. Javan, "Sudden pollutant discharge in vegetated compound meandering rivers," *Catena*, vol. 182, Article ID 104155, 2019.

[69] V. T. Chow, *Open-channel Hydraulics*, Mcgraw-Hill Civil Engineering Series, Chennai, TN, India, 1959.

[70] X. Zhang, R. Jing, Z. Li, Z. Li, X. Chen, and C.-Y. Su, "Adaptive pseudo inverse control for a class of nonlinear asymmetric and saturated nonlinear hysteretic systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 4, pp. 916–928, 2020.

[71] C. Zuo, Q. Chen, L. Tian, L. Waller, and A. Asundi, "Transport of intensity phase retrieval and computational imaging for partially coherent fields: the phase space perspective," *Optics and Lasers in Engineering*, vol. 71, pp. 20–32, 2015.

[72] C. Zuo, J. Sun, J. Li, J. Zhang, A. Asundi, and Q. Chen, "High-resolution transport-of-intensity quantitative phase microscopy with annular illumination," *Scientific Reports*, vol. 7, no. 1, pp. 7654–7722, 2017.

[73] B.-H. Li, Y. Liu, A.-M. Zhang, W.-H. Wang, and S. Wan, "A survey on blocking technology of entity resolution," *Journal of Computer Science and Technology*, vol. 35, no. 4, pp. 769–793, 2020.

[74] Y. Liu, B. Zhang, Y. Feng et al., "Development of 340-GHz transceiver front end based on GaAs monolithic integration technology for THz active imaging array," *Applied Sciences*, vol. 10, no. 21, p. 7924, 2020.

[75] J. Hu, H. Zhang, L. Liu, X. Zhu, C. Zhao, and Q. Pan, "Convergent multiagent formation control with collision avoidance," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1805–1818, 2020.

[76] M. B. Movahhed, J. Ayoubinejad, F. N. Asl, and M. Feizbahr, "The effect of rain on pedestrians crossing speed," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 3, 2020.

[77] A. Li, D. Spano, J. Krivochiza et al., "A tutorial on interference exploitation via symbol-level precoding: overview, state-of-the-art and future directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 796–839, 2020.

[78] W. Zhu, C. Ma, X. Zhao et al., "Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine," *IEEE Access*, vol. 8, pp. 61107–61123, 2020.

[79] G. Liu, W. Jia, M. Wang et al., "Predicting cervical hyperextension injury: a covariance guided sine cosine support vector machine," *IEEE Access*, vol. 8, pp. 46895–46908, 2020.

[80] Y. Wei, H. Lv, M. Chen et al., "Predicting entrepreneurial intention of students: an extreme learning machine with Gaussian barebone harris hawks optimizer," *IEEE Access*, vol. 8, pp. 76841–76855, 2020.

[81] A. Lin, Q. Wu, A. A. Heidari et al., "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-Nearest neighbor classifier," *Ieee Access*, vol. 7, pp. 67235–67248, 2019.

[82] Y. Fan, P. Wang, A. A. Heidari et al., "Rationalized fruit fly optimization with sine cosine algorithm: a comprehensive analysis," *Expert Systems with Applications*, vol. 157, Article ID 113486, 2020.

[83] E. Rodríguez-Esparza, L. A. Zanella-Calzada, D. Oliva et al., "An efficient Harris hawks-inspired image segmentation method," *Expert Systems with Applications*, vol. 155, Article ID 113428, 2020.

[84] S. Jiao, G. Chong, C. Huang et al., "Orthogonally adapted Harris hawks optimization for parameter estimation of photovoltaic models," *Energy*, vol. 203, Article ID 117804, 2020.

[85] Z. Xu, Z. Hu, A. A. Heidari et al., "Orthogonally-designed adapted grasshopper optimization: a comprehensive analysis," *Expert Systems with Applications*, vol. 150, Article ID 113282, 2020.

[86] A. Abbassi, R. Abbassi, A. A. Heidari et al., "Parameters identification of photovoltaic cell models using enhanced exploratory salp chains-based approach," *Energy*, vol. 198, Article ID 117333, 2020.

[87] M. Mahmoodi and K. K. Aminjan, "Numerical simulation of flow through sukhoi 24 air inlet," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 03, 2017.

[88] F. J. Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *ENG Transactions*, vol. 1, 2020.

[89] H. Chen, A. A. Heidari, H. Chen, M. Wang, Z. Pan, and A. H. Gandomi, "Multi-population differential evolution-assisted Harris hawks optimization: framework and case studies," *Future Generation Computer Systems*, vol. 111, pp. 175–198, 2020.

[90] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "Bayesian hierarchical model-based information fusion for degradation analysis considering non-competing relationship," *IEEE Access*, vol. 7, pp. 175222–175227, 2019.

[91] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "A Bayesian approach for degradation analysis with individual differences," *IEEE Access*, vol. 7, pp. 175033–175040, 2019.

[92] M. M. A. Malakoutian, Y. Malakoutian, P. Mostafapour, and S. Z. D. Abed, "Prediction for monthly rainfall of six meteorological regions and TRNC (case study: north Cyprus)," *ENG Transactions*, vol. 2, no. 2, 2021.

[93] H. Arslan, M. Ranjbar, and Z. Mutlum, "Maximum sound transmission loss in multi-chamber reactive silencers: are two chambers enough?," *ENG Transactions*, vol. 2, no. 1, 2021.

[94] N. Tonekaboni, M. Feizbahr, N. Tonekaboni, G.-J. Jiang, and H.-X. Chen, "Optimization of solar CCHP systems with collector enhanced by porous media and nanofluid," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9984840, 12 pages, 2021.

[95] Z. Niu, B. Zhang, J. Wang et al., "The research on 220GHz multicarrier high-speed communication system," *China Communications*, vol. 17, no. 3, pp. 131–139, 2020.

[96] B. Zhang, Z. Niu, J. Wang et al., "Four-hundred gigahertz broadband multi-branch waveguide coupler," *IET Microwaves, Antennas & Propagation*, vol. 14, no. 11, pp. 1175–1179, 2020.

[97] Z.-Q. Niu, L. Yang, B. Zhang et al., "A mechanical reliability study of 3dB waveguide hybrid couplers in the submillimeter and terahertz band," *Journal of Zhejiang University Science*, vol. 1, no. 1, 1998.

[98] B. Zhang, D. Ji, D. Fang, S. Liang, Y. Fan, and X. Chen, "A novel 220-GHz GaN diode on-chip tripler with high driven power," *IEEE Electron Device Letters*, vol. 40, no. 5, pp. 780–783, 2019.

[99] M. Taleghani and A. Taleghani, "Identification and ranking of factors affecting the implementation of knowledge management engineering based on TOPSIS technique," *ENG Transactions*, vol. 1, no. 1, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5572200, 10 pages
https://doi.org/10.1155/2021/5572200

Research Article

Optimization Algorithm Design for the Taxi-Sharing Problem and Application

Yongjie Wang and Maolin Li

Yuncheng Vocational and Technical University, Yuncheng, Shanxi 044000, China

Correspondence should be addressed to Yongjie Wang; yjiewang67@outlook.com

Received 27 February 2021; Revised 17 June 2021; Accepted 28 July 2021; Published 3 August 2021

Academic Editor: Samuel Yousefi

With the development of mobility techniques, the transportation systems become smarter, pursuing higher goals, such as convenience for passengers and low cost. In this work, we investigate the taxi-sharing system, which is a promising system recently. The passengers can share the same taxis to different destinations to save cost. Considering the property of taxis' routes, the corresponding model is established and our aim is to design the trip for each taxi to reduce the total number of taxi trips in the whole system if one taxi can be shared by several passengers. Compared with the previous work, we do not have any constraint about the taxi stations. The taxi trips have more flexibility in reality. We analyze this problem and prove it is NP-Complete. There are two proposed algorithms to solve this problem, one is a heuristic algorithm and the other is an approximate algorithm. In the experiment, two real-world taxi data sets are tested, and our algorithm shows the superiority of our taxi-sharing system. Using the taxi-sharing system, the number of trips can be reduced by about 30%.

1. Introduction

In recent years, with the increment of traveling and commuting, the numbers of private cars and public transportation are increased. According to the Federal Highway Administration (FHWA)'s report, the total rural and urban vehicles travel 3,262 billion miles [1], which is a huge number. The road congestion problem [2, 3] and the limited parking spaces [4, 5] are becoming new problems for people.

With the increased price of gas line and limited parking space, an efficient car-sharing system has been introduced to a large number of people. This system is a method to share one car with more than one passenger who can follow a common route to similar or close destinations [6]. Although there are some issues about this system, such as reservation strategy, passengers having to walk to the nearest parking lot, it can also bring a lot of benefits. It makes the private transportation flexible, produce less pollution, and traffic congestion. Therefore, it is becoming more and more popular, and people are willing to engage in this new mode of transportation. In practice, the conventional car-sharing systems mainly combine the passengers who have close pick-up locations and destinations because the private cars have predetermined route and cannot stop at any location on the road. In addition, people using this system mainly focus on their commutation to the work location, which means that the routes are stable. Thus, there are a lot of restrictions for such private car-sharing systems. These restrictions make this system not practical in real life.

In this work, we mainly focus on the taxi-sharing system. For taxi drivers, their goal is to carry more passengers to destinations, obtaining more benefits. Given that taxi drivers travel all the time continuously, passengers do not need to have the close pick-up locations and destinations because they can be picked up along the way. Once passengers' trips have some continuous common routes and do not make a large detour, they can share taxis without the loss of much time. Thus, for the taxi-sharing system, it has fewer restrictions than the car-sharing system. In most of the time, passengers can provide their source and destination to the platform and according to the current trip of the taxi, the sharing trips can be dispatched to the corresponding taxi

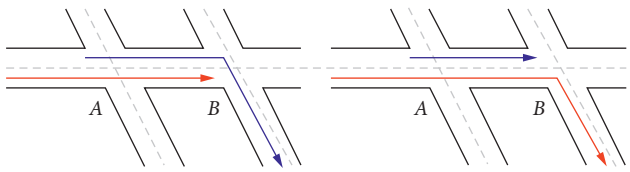


Figure 1: The red line and the blue line represent the trips of two passengers. Currently, the red passenger is picked up by the taxi driver, and the blue passenger is waiting for taxis at crossing A. In the left plot, two passengers can share one taxi because the red passenger can be dropped off at the crossing B and the taxi driver can continue to deliver the blue passenger. In the right plot, the taxi driver can drop off the blue passenger at the crossing B and continue to deliver the red passenger.

driver. In this way, the empty-load rate of a taxi can be reduced. Meanwhile, there can be fewer taxis on the road, which improves the environment.

In real-world application, this taxi-sharing system can still meet some challenges. First, passengers' requirements for taxi-sharing are different. Some passengers are sensitive to time. In the sharing process, drivers should deliver them to destinations on time. Thus, there is no detour in the process. Some passengers do not want to share with too many people, or frequently picking up other passengers. Thus, there should be some limitations on the number of passengers on taxis. How to guarantee the sharing process is very important in this system, meeting the different demands of passengers.

Second, if all the information about passengers is known, including pick-up locations, destinations, and pick-up time, it is important to find a schedule to dispatch taxi drivers to pick-up them to destinations. It is not easy to make such a schedule. In the conventional car-sharing system, passengers are connected according to their information, similar in time and space. In the taxi-sharing system, some passengers can be picked up in the middle, and do not need to be dropped off at the same destinations. Thus, for that, we wish to design the sharing modes for passengers, i.e., what kind of similarity between their trips can make them share the same taxi. There is an illustration figure shown in Figure 1. In these two cases, two passengers can be shared with one taxi if their time is matched.

Last, taxi demand is huge in city-scale. Given the pick-up location, time, and destinations of passengers, it is still hard to schedule taxi drivers and dispatch the delivery tasks. There are many traditional ways to do such schedules. At the same time, for online tasks, it becomes more difficult. We can use the greedy idea to pick-up the waiting passengers as more as possible. However, there is no guarantee for the results, which brings difficult for us to estimate the number of taxis needed. Thus, some better algorithms for the taxi-sharing scheduling problem can help us in city planning and traffic management.

Our Contribution. First, we design the taxi-sharing model based on real-world application. Given a large number of passengers in a region within a time duration, their pick-up locations, time, and destinations are known in advance or given sequentially. We assume that between any pair of pick-up location and destination, the shortest path is unique. All the passengers are sensitive to time, so the taxi drivers cannot make any detour to deliver passengers to destinations, as well as pick-up other passengers. At the same time, the sharing mode requirements are determined. For the trips of the current passengers, only when the other passengers can be satisfied to arrive at the destination on time, no matter they are dropped off earlier or later than the current passenger. This model is suitable for real-world applications. The taxi drivers just need to consider the current passengers and try to pick-up more passengers along the way. Of course, the number of passengers on the taxi is constrained, and the waiting time of passengers is also limited. If a passenger waits for a sharing taxi too long, an empty taxi will be dispatched to her.

Based on this model, our taxi-sharing problem is formally defined, which is to minimize the number of taxi-sharing trips. For the passengers who can share one taxi, they need only one taxi-sharing trip instead of the number of passengers. This problem is proven to be NP-complete, which does not have an optimal solution. Hence, we design two algorithms to solve it, one is a greedy algorithm and the other is an approximation algorithm. For the greedy algorithm, the taxi drivers are dispatched to each passenger if this passenger cannot share taxis with another passenger. Then, in the delivery process, taxi drivers pick-up all the satisfied passengers without exceeding the limitation of passengers on a taxi. In addition, this algorithm can also be applied into the online setting. For the approximation algorithm, we first try to find all the possible taxi-sharing trips, which is a time-consuming task in the raw data. To speed up this process, we convert these passengers into a graph network. Each directed edge represents whether these two passengers can share one taxi. Based on this graph, it becomes easier to obtain all possible taxi-sharing trips from the graph. Later the corresponding combinatorial algorithm is proposed, and the approximation ratio $O(\log m)$ is proved, where m is the number of passengers. However, the result of this algorithm performs well in the experiments.

In the end, we implement our model and algorithms into two real-world data set, which is taxi trajectories in San Francisco, America, and Porto, Portugal. There are more than 536 taxis in the city of San Francisco, over one month. Similarly, there are 442 taxis in the city of Porto, over one complete year. There are more than 170,000 taxi trajectories in the data set. All the trips of passengers are stored with their pick-up locations, time, destinations, and trajectories. Their trajectories are regarded as the shortest path from the pick-up location to the destinations. Thus, we check the number of trips that can be reduced if we count the taxi-sharing trip as one trip. When there are about 30,000 trips that happened in one day, in the original system, the same number of taxi trips needs to be dispatched. Now, with a taxi-sharing system, more than 10% trips can be shared with other passengers. With more trips given in the system, the probability of sharing is getting larger and the reduction is becoming more obvious, more than 20%. Similarly, if the requirement of taxi-sharing becomes loose, such as waiting time, waiting locations, and the maximum number of

passengers on a taxi, the reduction can also be enlarger, especially when there are a very large number of trips within a short period.

In the following, we start by reviewing the related work in Section 2. In Section 3, we introduce our taxi-sharing model and propose the problem definition. Then, two algorithms are provided in Section 4. The experiments are presented in Section 5, and Section 6 concludes this paper.

2. Related Work

Nowadays, there are many enterprises providing car-sharing service, e.g., ZipCar, EVCARD, Turo, Sixt, and so on. According to the rule of vehicle returning, the car-sharing system can be divided into two types [7], one is one-way and the other is round-trip [8, 9]. For the round-trip system, the car needs to be returned to the station where it is initially rented, while this limitation is removed in the one-way system. In our work, our system can be regarded as one-way system. At the same time, we also do not constrain the initial rental location. We consider the taxi that can pick up the passengers anywhere.

In recent years, many researchers began to investigate this area. The most important problem that they focus on is the vehicle dispatching problem. It is necessary for the car-sharing system. We need to consider how to deploy these cars to provide the service as soon as possible. In 2015, Nourinejad et al. [10] build an optimization model to investigate the trade-off between the vehicle fleet size and the human demand, to minimize the investigation cost. The vehicle allocation problem is also related to the capacity of the stop station [11]. For the real-time requests or online settings, some dynamic models are proposed to optimize the car dispatch problem with heuristic idea [12, 13]. With the development of the neural networks, many works [14, 15] are trying to predict the demand in real time, which can help the dispatch tasks get prepared in advance. There are some works to investigate the trajectories similarity, which can help saving the vehicles number [16]. For our work, although we do not have the limitation for the number of cars, our goal is to reduce the number of cars. According to the demand, we can dispatch the car to pick up the passengers within a comfortable waiting time.

With the development of electric cars, more and more car-sharing system began to consider the charging problems. The reason is that each electric car can only execute about ten hours, then it needs to get charged for about one hour. Thus, this charging time is a long time, which needs to be scheduled better [17]. Zhao et al. [18] present a mathematical model for the integrated electric car rebalancing and staff relocation for one-way station-based systems.

In our work, the taxi-sharing problem is different from the previous car-sharing problem. The taxi drivers do not have a fixed route and time restrictions. Thus, they can satisfy more passengers without the requirement for pick-up locations and destinations. There are many researchers studying the constraints of the taxi-sharing system. For example, the sharing agreement is based on social networks [19]. Passengers have permission to select other passengers, according to some features such as sex and age [20]. Some software systems were proposed to manage the schedules for sharing taxi and monitor whether the drivers make a detour [21].

3. Problem Definition

In this section, we introduce some preliminary concepts of our taxi-sharing system. The taxi drivers are allowed to carry more than one passenger to improve their benefit.

Given a region area Ω, many passengers are waiting for taxis to pick-up them and deliver them to their destinations. Suppose that there are m passengers in the region Ω within a long time duration, denoted $P = \{p_1, p_2, \ldots, p_m\}$. For each passenger p_i, she has her own pick-up location, pick-up time, and destination, represented by s_i, st_i, and d_i. In most of the time, it is hard to control the arrival time by taxi, but taxi drivers guarantee the passengers will be carried to destinations as soon as possible without any detour. Given a waiting time threshold δ, which is the maximum waiting for a sharing taxi, it means that it is possible for a passenger to wait for a taxi driver δ time with available spaces; otherwise, this passenger is not satisfied. For each passenger, she is willing to take a sharing taxi instead of an empty taxi because she can save some cost by sharing a trip with others.

Then let us focus on the taxi driver side. For each taxi-sharing trip, it should start from a pick-up location of one passenger and goes to the destination of another passenger. If there is only one passenger on this taxi-sharing trip, the pick-up location and the destination belong to the same passenger. If there are more passengers, the pick-up location and destination might belong to different passengers. In the taxi-sharing system, drivers are allowed to pick-up more passengers along the way, only if they can deliver passengers to their corresponding destinations on time. Thus, taxi drivers cannot make a detour for any passenger, which means that the taxis should be one of the shortest path to the destinations of all the passengers on the whole trip. In this work, we regard a continuous trip with at least one passenger as one taxi-sharing trip. If all the passengers are dropped off, this taxi-sharing trip is finished and the driver should wait for the next dispatch task. Now, we define the trip:

Definition 1 (trip). Each trip T, as a trajectory of the taxi driver within a short time, refers to a sequence of positional points that chronologically sampled during a time period, denoted as $T = \{(l_1, t_1), (l_2, t_2), \ldots, (l_r, t_r)\}$, where l_i is the location of the taxi driver at timestamp t_i and r is the length of this trip. This location l_i can be GPS location or other format of position.

Now, let us take a look at taxi-sharing trips. Given a parameter k, as the maximum number of passengers in one taxi, it is 3 or 4 for the general taxi cars. For a taxi-sharing trip, there is at least one passenger and at most k passengers on the taxi along all the way. Such trips are called taxi-sharing trips.

Fora taxi driver carrying the passenger p_i with trip T, there is $s_i = l_1$ and $st_i = t_1$, which means that the pick-up location and time of p_i are the start point of the trip T. Then,

along the way, the drive can meet another passenger p_j. If this driver wants to pick up the passenger p_j to her destination, there are two conditions, pick-up condition and destination condition.

The pick-up condition is easy to understand. When the taxi driver arrives at the pick-up location of the passenger p_j, the passenger p_j waits less than the maximum waiting time δ, i.e., $s_j = l_a$ and $t_a - st_j \le \delta$, where $1 \le a \le r$. If the passenger p_j is shared this taxi with p_i, their following trip should be same until one of them arrive the destination, which is the destination condition. There are two cases are shown in Figure 1:

(1)The passenger p_i arrives at the destination first, as shown in the left plot of Figure 1. Thus, the destination of p_i should be on the shortest path of passenger p_j.

(2) The passenger p_j arrives at the destination first, as shown in the right plot of Figure 1. Thus, the destination of p_j should be one the shortest path of passenger p_i.

When there is more than one passenger on the taxi, it also needs to satisfy the pick-up condition and destination condition to pick-up more passengers. The following trips of all the current passengers are the same at any timestamp. We hope to minimize the number of taxi-sharing trips satisfying all the passengers in the region. We define our problem under the offline setting, as shown below:

Definition 2 (taxi-sharing problem). Given a region Ω, for any two locations in this region, the shortest path between them are unique and known. With a set of m passengers $P = \{p_1, p_2, \ldots, p_m\}$ in the region, for any passenger p_i, her pick-up locations, pick-up time, and destinations are represented as s_i, st_i, and d_i. We need to design a set of taxi-sharing trips $\mathcal{T} = \{T_1, \ldots, T_n\}$, such that all the passengers are satisfied with the pick-up condition and destination condition. Our goal is to minimize the number of taxi-sharing trips, i.e., n, in the set.

$$\begin{aligned} &\min n = |\mathcal{T}| \\ &\text{s.t. } \forall p_i, \exists T_j, s_i = l_a, (l_a, t_a) \in T_j \\ &t_a - st_j \le \delta, (l_a, t_a) \in T_j. \end{aligned} \tag{1}$$

For the offline taxi-sharing problem, with the finite number of passengers, we can find all possible taxi-sharing trips through the enumerate or other better methods. Then, the optimal solution is one of the combinations of these taxi-sharing trips. We can show the difficulty of this problem below:

Theorem 1. *The offline taxi-sharing trip problem is NP-complete.*

Proof We can reduce this problem from a known NP-complete problem, named the set cover problem [22]. Given a set of element E and a collection of sets $\mathcal{S}$, the goal of set cover problem is to select the minimum number of sets from $\mathcal{S}$ such that each element can be covered by these sets.

For offline taxi-sharing problem, when we obtain all the possible taxi-sharing trips, we know which passengers can be satisfied by the special taxi-sharing trip. For any instance $\{\mathcal{S}, E\}$, each element can be regarded as a passenger, and the set can be a taxi-sharing trip. Suppose we have an algorithm $\mathcal{A}$ to compute offline taxi-sharing problem, the output can be seen as the solution to the set cover problem as well. Therefore, according to Cook's reduction, offline taxi-sharing problem is NP-complete.

4. Algorithm Design

In this section, we propose a heuristic algorithm for online taxi-sharing problem. Another approximation is proposed for offline taxi-sharing problem.

First, in the problem definition, it requires that each passenger needs to be delivered via the shortest path. It means that each taxi-sharing trip can be composed of multiple concatenated shortest path. To simplify this problem, we suppose that each taxi-sharing trip need to travel from a source point to the destination via the shortest paths. In this way, the taxi-sharing trip is a shortest path between two locations. The taxi drivers try to pick up as many passengers as possible along the way.

Each passenger has their own pick-up location and destination. We need to design the shortest path for taxi drivers to pick all the passengers up and take them to their destinations. We will introduce the concept of VC-dimension [23], used in our following analysis.

Definition 3 (VC-dimension). Given a set system $(X, \mathcal{R})$, let A be a subset of X. We say that A is shattered by $\mathcal{R}$ if $\forall Y \subseteq A, \exists R \in \mathcal{R}$ such that $R \cup A = Y$. The VC-dimension of $(X, \mathcal{R})$ is the cardinality of the largest set that can be shattered by $\mathcal{R}$.

Now, we consider the VC-dimension of our case.

Theorem 2. *For the passenger set system $(P, \mathcal{T})$, if each taxi-sharing trip $T_i \in \mathcal{T}$ is the shortest path, then the VC-dimension of this set system $(P, \mathcal{T})$ is at most 2.*

Proof We regard each passenger as an element in the set. Any set of passengers is tried to be shattered by the taxi-sharing trip. Now, each taxi-sharing trip is a shortest path between two locations. For any set of three passengers, a taxi-sharing trip cannot shatter it. In this case, the taxi-sharing trip is the shortest path between two locations. If these three passengers are all along this shortest path, we cannot shatter any two of them from the set. Thus, there is no taxi-sharing trip with the shortest path can shatter the set of three passengers.

Actually, our taxi-sharing trips are composed of multiple concatenated shortest paths. Suppose each taxi-sharing trips are at most k concatenated shortest paths. Then we can check its VC-dimension.

Theorem 3. *For the passenger system $(P, \mathcal{T})$, if each taxi-sharing trip $T_i \in T$ is k concatenated shortest paths, then the VC-dimension of this set system $(P, \mathcal{T})$ is at most $2k^2$.*

Proof Similarly, for any set of $2k^2 + 1$ passengers, there are at least $2k + 1$ passengers on the same shortest path, according to the Pigeonhole principle. Then we select the alternative passengers or passengers of the odd index. There are at least $k + 1$ passengers and they cannot be shattered by k shortest paths. The reason is that any two passengers cannot share one shortest path, so each one passenger needs one shortest path. Thus, the VC-dimension of this set system is at most $2k^2$.

4.1. Heuristic Algorithm. Basically, for each taxi driver, when she picks up the first passenger, she wants to find the next passenger who satisfies the pick-up condition and destination condition. Similarly, when one passenger is waiting for a taxi, the first option is a sharing taxi. Thus, our heuristic algorithm is based on this greedy idea: when a driver picks up one passenger, she will try to find another passenger along the way.

With the request arriving sequentially, there are many taxi-sharing trips on the road, if this request can be satisfied by one of the trips, the passenger can be picked up by one sharing taxi, otherwise, an empty taxi will go to the pick-up location. The pseudo-code is shown in Algorithm 1.

Here, the set of passengers are still given in advance. In this way, it can show that this algorithm can also solve the offline taxi-sharing problem. In Line 1, all the passengers are sorted by their pick-up time, which is suitable for the online setting. The set $f[i]$ represents the current sharing passengers on the taxi with the passenger p_i. Thus, we check the passengers according to their pick-up location in Line 4. First, all the existing taxi-sharing trips need to be checked in Line 5 and the current timestamp is st_i, i.e., the pick-up time of the passenger p_i. For each taxi-sharing trip, some passengers arrive at their destination before the time st_i, so we remove these passengers from the corresponding set $f[j]$. If there are still k passengers on the taxi, i.e., $\|f[j]\| = k$, it is impossible to accommodate the passenger p_i on this taxi. If there are still some available seats, we can check whether the pick-up condition and destination condition are satisfied in Line 9. If it is satisfied, the passenger p_j is included, and all the passengers are stored in the set $f[j]$. If it cannot be satisfied with all the taxi-sharing trips, an empty taxi will be dispatched to pick her up in Line 13–15. The number of trips is added in Line 15 and return the result in Line 16. For the offline taxi-sharing problem, with the finite number of passengers, we can find all possible taxi-sharing trips through the enumerate or other better methods. Then, the optimal solution is one of the combinations of these taxi-sharing trips. We can show the difficulty of this problem below:

We can see that this algorithm also works for the offline taxi-sharing problem. This idea is suitable for our daily life. Passengers are willing to wait for a taxi to share with someone, saving the cost. The taxi drivers do not make any detours and deliver every passenger to the destination on time.

4.2. Approximation Algorithm. In Section 3, we mention that we can enumerate all possible taxi-sharing trips for the given m passengers under the offline setting. It costs too much time to enumerate them. Thus, we first propose a better way to find all possible taxi-sharing trips, which is easy to understand.

Given that there are m passengers, all the information of them is known, including pick-up locations, pick-up time, destinations, and the shortest paths. It is not difficult to check whether any pair of passengers can share one taxi, i.e., checking the pick-up condition and destination condition. Thus, we can build a graph $G(P)$ for all the passengers to represent this relationship, which helps us to find all possible taxi-sharing trips.

For the passenger graph $G(P) = \{V, E\}$, where V is the set of vertices, represented by each passenger, and E is the set of directed edges. Owing to that the pick-up order of passengers needs to be considered, these edges have direction. If the passenger p_j can be satisfied by the taxi-sharing trip of the passenger p_i, there is an edge from v_i to v_j. Hence, our passenger graph is established. Each vertex has two attributes, pick-up time, and arrival time. This arrival time is computed by the speed of the taxi and the shortest path. In this way, we know when the passenger is dropped off. An illustration figure is shown in Figure 2.

Then, we can use depth-first search and topology sort to find all possible trips in the graph $G(P)$. As shown in Figure 2, we take a look at vertices v_1, v_2, and v_3. It shows that the trip with the passenger p_1 can pick up the passenger p_2, but cannot pick up the passenger p_3. Here, we need to check the arrival time of p_1 and the pick-up time of p_3. The below plot is much complex. There are edges from v_4 to v_5 and v_7, but there is no edge between v_5 and v_7. It means that the passengers v_5 and v_7 cannot share one taxi. Thus, the taxi-sharing trip of p_4 can only pick-up one of passengers p_5 and p_7. The case of the passenger p_6 is the same with the passenger p_3. Now, let uss look at the vertex v_8, which shows that the passengers p_4 and p_7 can share a taxi-sharing trip with p_8. Thus, we do not need to check the time of p_8.

In this search process, we need to use the technique of topology sort. If we want to include this vertex, we need to check the arrival time and pick-up time to remove some vertices in the set. Then, if all the vertices in the set have a directed edge to this vertex, this vertex can be included. At the same time, we also need to consider the maximum number of passengers on a taxi. Using this, we can find all possible taxi-sharing trips with a different combination of passengers, speeding up this process. Of course, some memory techniques can be applied in this search process to improve the time complexity.

Based on the set of all possible taxi-sharing trips Γ, we need to select the minimum number of taxi-sharing trips such that all the passengers are satisfied. For each taxi-sharing trip T, we denote $w(T)$ as the number of passengers can be satisfied by it. Then we design the approximation algorithm to select these taxi-sharing trips one by one. In each iteration, we select the taxi-sharing trip that can satisfy the most waiting passengers. Or we can imagine each taxi-sharing trip is a set. We need to select the minimum number

Input: The set of pick-up location s_i and destinations d_i from m passengers P
Output: Number of trips needed
(1) Sorted the passenger set P according to s_i
(2) $f[i] = \varnothing$, $1 \le i \le m$
(3) res = 0
(4) **for** $i = 1, \ldots, m-1$ **do**
(5) **for** $j = 1, \ldots, i-1$ **do**
(6) Remove the passengers who arrive destinations from $f[j]$;
(7) **if** $\|f[j]\| == k$ **then**
(8) Continue
(9) **if** p_j can share with passengers in $f[i]$ **then**
(10) $f[j] = f[i] \cup \{p_j\}$;
(11) $f[i] = \varnothing$;
(12) Continue;
(13) **if** $f[j] == \varnothing$ **then**
(14) $f[j] = \{p_j\}$
(15) res = res + 1
(16) **return** res

Algorithm 1: Greedy algorithm

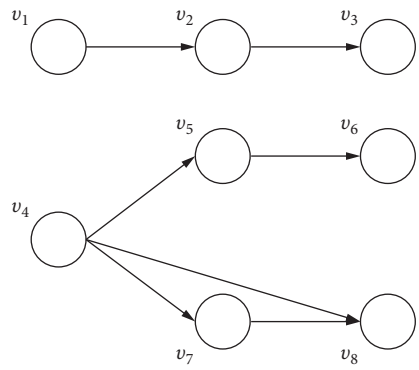


Figure 2: The passenger graph $P(G)$: here we omit the attributes on vertices. The vertices between one edge can share one taxi. If the passenger p_1 is dropped off before picking up the passenger p_3, then passengers p_1, p_2, and p_3 can share one taxi-sharing trip.

of sets to cover all the vertices in the passenger graph $G(P)$. In each iteration, we select a set that can cover the maximum number of uncovered vertices. We now show that this algorithm is an $O(\log m)$ approximation, where m is the number of passengers, which turns out to be the best approximation one can achieve (unless P = NP).

Theorem 4. *This approximation algorithm can achieve* $O(\log m)$ *approximation.*

Proof Consider a step of the algorithm. Let c be the number of waiting passengers before selecting a taxi-sharing trip on this step. Among the taxi-sharing trips that are not selected by the algorithm, the optimal solution uses some set of these taxi-sharing trips to satisfy the remaining waiting passengers. Let R_{OPT} be the optimal set of taxi-sharing trips based on the previous selection.

Let OPT be the number of taxi-sharing trips in the optimal solution. Clearly, we have that $|R_{\mathrm{OPT}}| \le \mathrm{OPT}$ since these taxi-sharing trips are a subset of the optimal solution, where $|R_{\mathrm{OPT}}|$ is the size of set R_{OPT}. In addition, we observe that there is $\sum_{T \in R_{\mathrm{OPT}}} w(T) \ge c$ since by definition these trips satisfy the remaining waiting passengers.

Therefore, we have:

$$\max_{T \in R_{\mathrm{OPT}}} w(T) \ge \frac{\sum T \in R_{\mathrm{OPT}} w(T)}{|R_{\mathrm{OPT}}|} \ge \frac{c}{\mathrm{OPT}}, \tag{2}$$

where the first inequality follows from an average argument, and the second inequality follows from our above observations. Thus, on this step where there are c remaining waiting passengers, there must be a taxi-sharing trip that can satisfy the average number which is at least c/OPT.

We try to divide up the cost of taxi-sharing trip to each passenger. The cost of each taxi-sharing trip is all the unit cost. Thus, each passengers gets a charge as $1/w(T)$. Now, we can claim that the jth passengers to be satisfied can receive a charge at most $\mathrm{OPT}/(m - j + 1)$. It follows equation (2) and the fact that when the jth waiting passenger is satisfied, there must be at least $(m - j + 1)$ waiting passengers on this step.

Thus, we can sum up the total charge of all the passengers. The total charge is

$$\sum_{j=1}^{m} \frac{\mathrm{OPT}}{m-j+1} = \mathrm{OPT} \cdot \sum_{j=1}^{m} \frac{1}{j} = \mathrm{OPT} \cdot O(\log m), \quad (3)$$

where the last equality follows the fact that the mth harmonic number is $O(\log m)$.

This approximation seems not good enough because the number of trips are very large. However, for the independent vertices, i.e., some passengers who cannot share with others are not counted into the approximation ratio. The most sharing cases are easy to find because there is no too many combinations. Thus, the performance of this algorithm in the real application is good enough to satisfy the requirement.

5. Numerical Experiments

In this section, we test our model and algorithms in two real-world data sets. First, we introduce our experimental setup, including data sets, hardware, and baseline algorithms as references.

Hardware. We implemented our algorithm in Python with version 3.8. We ran the experiments on the machine equipped with Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz-, and 32 GB of RAM.

Data set. We test our model and algorithms on two real-world data set. The first one is a set of taxi trajectories from San Francisco [24]. There are 536 taxis delivering passengers over one month in an area of $6,327 \times 6,827\mathrm{km}^2$. In each day, there are about 21,843 trips of these taxis, which is a huge number. The visualization map of San Francisco is shown in Figure 3. Each line represents a trip of taxi, and the color shows the frequency of traveling. We can see that the central part of the city is very dense and there is some high way around the city.

The second one is also a set of taxi trajectories in the center of Porto, Portugal [25], which is an area of size $8,116 \times 8,068\mathrm{km}^2$. There are 442 taxis running in the city over a complete year (from 01/07/2013 to 30/06/2014). More than 170,000 trajectories are included in this data set, which is visualized in Figure 4. We can see that these trajectories are radial shaped around the city. In the central part of the city, the trajectories are denser, which means that more passengers are traveling in a similar route. It has a large probability to take a sharing taxi to destinations.

For these two data sets, the passengers' demands are given, including their pick-up locations, pick-up time, and destinations. At the same time, their trips based on GPS locations and times are also provided. In our implementation, we can regard their trips from pick-up locations to destinations as the shortest paths. We also know the exact time of taxis arriving at each location, which helps us to determine whether another passenger can be picked up. The taxi drivers cannot make any detour in this process. The sharing modes should be based on their trips in the real-world data sets.

Tool chain. We implement our proposed algorithms based on Python, and the source code can be shared after publication. We provide a tool chain of our framework, including data processing, graph building, algorithm implementation, and comparison. The main processes are introduced below:

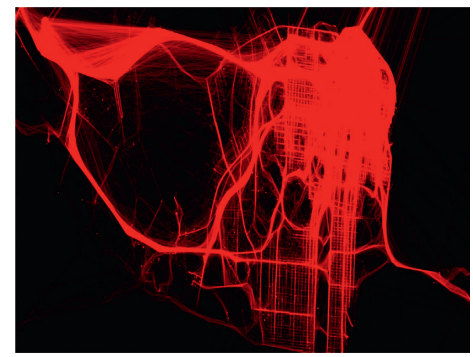


Figure 3: The visualization map of San Francisco, America.

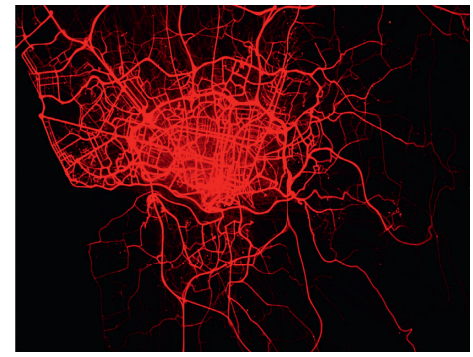


Figure 4: The visualization map of Porto, Portugal.

(1) Data set Processing. Given the initial real-world data sets, we first need to do some preprocessing jobs for data sets, including removing some outlier trips (with unrealistic speed or at an impossible location). For the San Francisco data set, the whole trajectories of taxis are given including empty states. Thus, we also need to extract the trips from the data sets. Then we have to sort all the trajectories according to the pick-up time of each passenger for these two data sets.

(2) Graph Building. For our approximation algorithms, we hope to embed all the passengers with their trip information into a graph $G(P)$. Each vertex represents passengers, and edges between vertices show whether two passengers can share one taxi. Similarly, our greedy algorithm can also do this step first and then try to include more passengers in one trip. Thus, in this step, we need to check whether one passenger can share the trip with the other passenger. Then, using this information to build our passenger graph. To speed up this process, we can use the hashing technique to store passengers, with the keys as the pick-up locations and destination separately. The lookup can be improved to constant time.

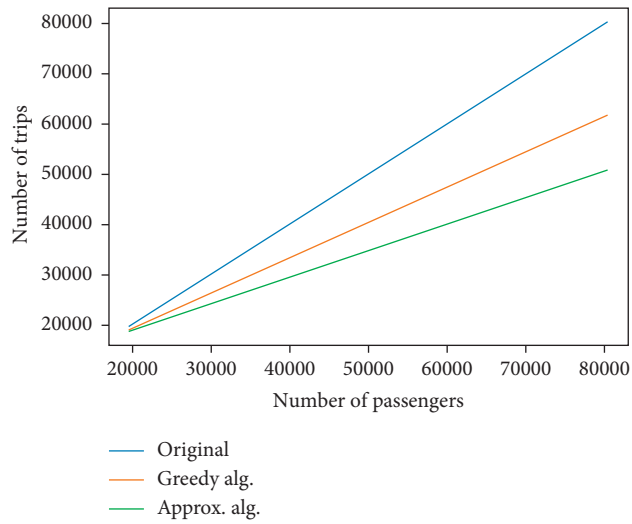


Figure 5: Number of taxi-sharing trips in San Francisco: the number of taxi-sharing trips is increased approximately linearly with the increased number of passengers. With the taxi-sharing system, the performance using approximation algorithm works best, which can reduce more than 40% trips compared with the case without taxi-sharing system.

Table 1: San Francisco: number of taxi-sharing trips (ten thousand).

# of Passengers	2	4	6	8
Original	2.03	4.09	5.97	8.11
Greedy alg.	1.92	3.25	4.65	6.17
Approximation alg.	1.84	3.02	3.93	5.04

(3) Algorithm Implementation. In this step, two proposed algorithms are implemented. For the greedy algorithm, each trip is selected based on its order of pick-up time. Then this driver will try to pick-up passengers in this continuous process as much as possible. For the approximation algorithm, we first need to find all possible taxi-sharing trips using the deep-first search process. In this process, the topology order needs to be considered. Based on all possible taxi-sharing trips, each trip with the maximum number of waiting passengers is selected in each iteration, until all the passengers are satisfied.

(4) Comparison. To investigate the benefit of our taxi-sharing system, we need to compare with the results without taxi sharing. Actually, the number of passengers is the number of trips needed without taxi-sharing system. It can be regarded as the baseline if we do not use the taxi-sharing system. Here, we denote the original number of trips as "Original" in the figure. The greedy algorithm and the approximation algorithm are denoted as "Greedy Alg." and "Approx. Alg." We can check how many trips can be reduced by the taxi-sharing system.

First, we look at the San Francisco data set. Owing to that there are only about 536 taxis in a large area, all the trips in one day are sparse in space and time. We can combine trajectories of multiple days to one day and check the relationship between the number of passengers and the number of trips needed. We start from all the passengers in one day and include more passengers gradually. The results are shown in Figure 5 and Table 1. We can see the numbers of taxi-sharing trips are increased linearly with the increased number of passengers, no matter whether we use the taxi-sharing system or not. We can also see that with taxi-sharing system, the number of trips is reduced compared with the conventional system, especially using the approximation algorithm. When there are more than 50, 000 passengers, the number of trips can be reduced by more than 20%. If the number of passengers exceeds 80, 000, the reduction is more than 40%. The reason is that when there are more passengers, their trips become denser in space and time. The probability that one passenger has someone to share one taxi is increased. At the same time, we can see that our approximation algorithm performs better than the greedy algorithm. Although the approximation ratio is $O(\log m)$, which is a large number, its performance is great in the real implementation.

Second, for the Porto data set, the trip of each passenger only has the pick-up time, without the timestamp in the delivery process. Thus, we assume that the GPS locations are sampled with about $30 - 50$ seconds according to the distance in the interval. Then we check whether one passenger can share a taxi with other passengers. The results are shown in Figure 6 and Table 2. The results are similar to the San Francisco data set, but the reduction is smaller than the one in San Francisco. The reason is that the size of Porto city is larger than San Francisco. In addition, we can also see from

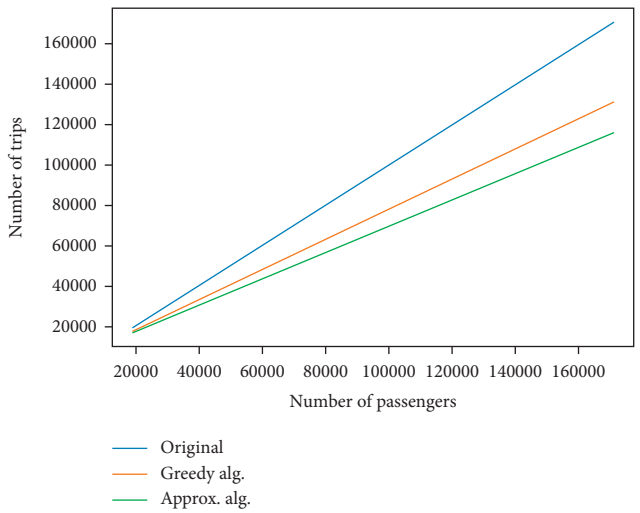


Figure 6: Number of taxi-sharing trips in Porto: the number of trips can be reduced about 30% compared with the conventional systems, where there are more than 170, 000 passengers in one day.

Table 2: Porto: number of taxi-sharing trips (ten thousand).

# of Passengers	2	6	10	14
Original	2.01	6.74	10.96	15.37
Greedy alg.	1.98	5.02	8.21	11.63
Approximation alg.	1.90	4.42	7.43	10.23

the visualization map, the trips in Porto are distributed in a larger area. In the central part, the trips are not denser than in San Francisco.

Some constraints for taxi-sharing trips also influence the results, such as the maximum number of passengers on one taxi and maximum waiting time. The results are similar to the above experiments. The number of taxi-sharing trips is increased with the number of passengers. When the constraints become looser, like more passengers can be shared in one taxi, and more waiting time, the reduction in the number of taxi-sharing trips becomes larger.

6. Conclusion and Future Work

In this work, we have studied the taxi-sharing problem, aiming at minimizing the number of trips needed. First, the formal taxi-sharing model is established. Then the problem is formulated, and we prove this problem is NP-complete. We designed two algorithms to solve this problem with some analysis and optimization for implementation. Extensive experiments on two real-world data sets show the superiority of our algorithms and the advantage of the taxi-sharing system.

In the future, we will continue to focus on this kind of problem. Some detours can be considered in the model. Some constraints can be flexible according to the time and passengers. Besides, we can use the machine learning method to predict the passengers' demand and dispatch taxis to the popular region. We hope to pay more attention to this topic to improve our traffic design and city planning.

Data Availability

The real-world trajectory data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] U.S. Department of Transportation, "Highway statistical 2019," 2019, http://www.fhwa.dot.gov/policyinformation/statistics/2019/.

[2] J. Hardy and L. Liu, "Available forward road capacity detection algorithms to reduce urban traffic congestion," in *Proceedings of the 2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, June 2017.

[3] M. R. Jabbarpour, H. Zarrabi, R. H. Khokhar, S. Shamshirband, and K.-K. R. Choo, "Applications of computational intelligence in vehicle traffic congestion problem: a survey," *Soft Computing*, vol. 22, no. 7, pp. 2299–2320, 2018.

[4] F. Shi, D. Wu, D. I. Arkhipov, Q. Liu, A. C. Regan, and J. A. McCann, "ParkCrowd: reliable crowdsensing for aggregation and dissemination of parking space information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 11, pp. 4032–4044, 2018.

[5] G. Tasseron and K. Martens, "Urban parking space reservation through bottom-up information provision: an agent-based analysis," *Computers, Environment and Urban Systems*, vol. 64, pp. 30–41, 2017.

[6] C. Morency, M. Trépanier, B. Agard, B. Martin, and J. Quashie, "Car sharing system: what transaction datasets reveal on users' behaviors," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 284–289, IEEE, Bellevue, WA, USA, September 2007.

[7] X. Huo, X. Wu, M. Li, N. Zheng, and G. Yu, "The allocation problem of electric car-sharing system: a data-driven approach. Transportation Research Part D:," *Transportation Research Part D: Transport and Environment*, vol. 78, Article ID 102192, 2020.

[8] M. Balac, F. Ciari, and K. W. Axhausen, "Modeling the impact of parking price policy on free-floating carsharing: case study for Zurich," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 207–225, 2017.

[9] Q. Li, F. Liao, H. J. P. Timmermans, H. Huang, and J. Zhou, "Incorporating free-floating car-sharing into an activity-based dynamic user equilibrium model: a demand-side model," *Transportation Research Part B: Methodological*, vol. 107, pp. 102–123, 2018.

[10] M. Nourinejad, S. Zhu, S. Bahrami, and M. J. Roorda, "Vehicle relocation and staff rebalancing in one-way carsharing systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 81, pp. 98–113, 2015.

[11] K. Huang, G. H. D. A. Correia, and K. An, "Solving the station-based one-way carsharing network planning problem with relocations and non-linear demand," *Transportation Research Part C: Emerging Technologies*, vol. 90, pp. 1–17, 2018.

[12] L. Caggiani, R. Camporeale, M. Ottomanelli, and W. Y. Szeto, "A modeling framework for the dynamic management of free-floating bike-sharing systems," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 159–182, 2018.

[13] Y. Du, F. Deng, and F. Liao, "A model framework for discovering the spatio-temporal usage patterns of public free-floating bike-sharing system," *Transportation Research Part C: Emerging Technologies*, vol. 103, pp. 39–55, 2019.

[14] T. Liu, W. Wu, Y. Zhu, and W. Tong, "Predicting taxi demands via an attention-based convolutional recurrent neural network," *Knowledge-Based Systems*, vol. 206, Article ID 106294, 2020.

[15] K. Zhao, D. Khryashchev, and H. Vo, "Predicting taxi and uber demand in cities: approaching the limit of predictability," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, pp. 2723–2736, 2019.

[16] H. Wang and J. Gao, "Distributed human trajectory sensing and partial similarity queries," in *Proceedings of the 2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Sydney, Australia, April 2020.

[17] C. A. Folkestad, N. Hansen, K. Fagerholt, H. Andersson, and G. Pantuso, "Optimal charging and repositioning of electric vehicles in a free-floating carsharing system," *Computers Operations Research*, vol. 113, Article ID 104771, 2020.

[18] M. Zhao, X. Li, J. Yin, J. Cui, L. Yang, and S. An, "An integrated framework for electric vehicle rebalancing and staff relocation in one-way carsharing systems: model formulation and Lagrangian relaxation-based solution approach," *Transportation Research Part B: Methodological*, vol. 117, pp. 542–572, 2018.

[19] D. Santos and E. Xavier, "Dynamic taxi and ridesharing: a framework and heuristics for the optimization problem," in *Proceedings of the 23rd international joint conference on artificial intelligence*, pp. 2885–2891, Beijing, China, August 2013.

[20] C.-C. Tao, "Dynamic taxi-sharing service using intelligent transportation system technologies," in *Proceedings of the International conference on wireless communications, networking and mobile computing*, pp. 3209–3212, Honolulu Hawaii USA, August 2007.

[21] P. Lalos, A. Korres, C. Datsikas, G. Tombras, and K. Peppas, "A framework for dynamic car and taxi pools with the use of positioning systems," in *Proceedings of the Computation world: future computing, service computation, cognitive, adaptive, content, patterns*, pp. 385–391, IEEE, Athens, Greece, November 2009.

[22] D. S. Hochba, "Approximation algorithms for NP-hard problems," *ACM Sigact News*, vol. 28, no. 2, pp. 40–52, 1997.

[23] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971.

[24] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD dataset epfl/mobility," 2009, https://crawdad.org/epfl/mobility/20090224,.

[25] KAGGLE data set ecml/pkdd 15, "Taxi trajectory prediction(1)," Downloaded from https://www.kaggle.com/c/pkdd-15-predict-taxiservice-trajectory-i/data, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9075560, 20 pages
https://doi.org/10.1155/2021/9075560

Research Article

Water Production Problem in Gas Reservoirs: Concepts, Challenges, and Practical Solutions

Ali Akbar Roozshenas,[1] **Hamed Hematpur**,[2] **Reza Abdollahi**,[2] **and Hamid Esfandyari**[3]

[1]*Department of Petroleum Engineering, Petroleum University of Technology, Ahwaz, Iran*
[2]*EOR Research Department, Research Institute of Petroleum Industry, Tehran, Iran*
[3]*Abadan Faculty of Petroleum Engineering, Petroleum University of Technology, Abadan, Iran*

Correspondence should be addressed to Hamid Esfandyari; esfandyari_shirazu@yahoo.com

Received 2 June 2021; Revised 27 June 2021; Accepted 15 July 2021; Published 27 July 2021

Academic Editor: Samuel Yousefi

Gas resources play a key role in nowadays energy supply and provide 24% of the diverse energy portfolio. Water encroachment is one of the main trapping mechanisms in gas reservoirs. It decreases recovery by reduction of reservoir life, limits productivity and efficiency of wells, and elevates safety risks in gas production. The lack of a comprehensive study about water production problems is the primary motivation for this study. Contrary to the serious concern over the standalone investigation of an actual water production case study, less concern is put to deal with the problem comprehensively through an investigation of all potential sources and mechanisms, required methods, and available techniques. This study presents the potential sources of the problem, methods to identify it, and approaches to address it. Firstly, possible sources are described. Secondly, the diagnostic techniques are expressed. Then, practical solutions used in actual cases to overcome problems are elaborated. The solutions include both well- and reservoir-oriented approaches. Finally, all proper strategies are summarized to tackle the water problems in gas fields. The current study comprehensively presents the available methods for water control problems in parallel with conceptual and qualitative comparison. The finding of this study can be very constructive for better understanding of water sources, available diagnostic tools, and solutions for controlling water production in gas reservoirs and, consequently, taking the best decision in real case studies before attempting many water shut-off approaches.

1. Introduction

The decreasing trend of fossil resources discovery in parallel with the expanding demand for energy raises the eminence of techniques and approaches (reservoir-based and well-based) applied to improve the recovery of current resources [1–12]. Many reservoirs are limited with water sources contained in water-bearing rocks called aquifers. Water production in gas reservoirs is more critical than in oil reservoirs. They are divided into two general categories based on aquifer support: volumetric reservoir and water-drive reservoirs. Experiences have indicated that the recovery factor of volumetric reservoirs ranges between 80 and 90 percent due to remarkable pressure drop in the life of reservoirs, while it is significantly lower for the case of water-drive gas reservoirs [13].

More than half of gas reservoirs in the world are associated with aquifers [14]. In cases that the water sources are large with high permeability of the water-bearing formation, water encroaches into the gas zone and, consequently, affects the production mechanism and recoverable amount of reservoir gas. A strong aquifer can remarkably decrease the recovery factor in the range of 30 to 85 percent by trapping the gas phase at higher pressures [15, 16]. As the reservoir fluid is produced, a differential pressure causes the water encroachment from gas/water contact. Large packets of gas

may be bypassed and left behind the encroaching front; thus, the amount of residual gas saturation increment leads to the reduction of ultimate gas recovery [14].

Water production decreases the recovery by reduction of the life of the reservoir [17–19], limits productivity and efficiency of wells, and elevates safety risks in gas production. Water production can cease the well from production by water loading or make the production economically deficient. Produced water should be disposed of using proper methods resulting in a considerable increase in operational costs. It often contains large amounts of salts that require unique disposal methods and cause environmental concerns. Additionally, it may include some radioactive contents that intensify environmental effects. Eventually, even if no sign of water influx is observed during production, proper precautions should be predicted for appropriate reactions in severe conditions.

Until now, each water control method has been studied explicitly. Usually, excess water production problems are tackled regarding the severity of the issue. Generally, water production is categorized into well-based and reservoir-based issues, and well-based and easy-to-treat problems are considered first. The necessity of selecting the most proper water control method leads to the discussion of all methods in a single document. This generality enables the reader to compare different methods and efficiently select the most suitable one, and it is rarely included in previous studies. The main idea of the paper considers different aspects of the water production problem and its solution, presenting practical and field examples. This paper consists of four main parts: (1) water production sources, (2) diagnostic methods, (3) solutions, and (4) summary. In the first section, different types of water drives and various sources responsible for the water production problem in a field or a well have been introduced and discussed. The second part contains different methods to identify the water production, activity degree of an aquifer, and other factors related to water production issues. The third section is composed of two subdivisions that aim to explain solutions to the problem. The first subdivision presents reservoir-based solutions, while the second one focuses on solving the problem based on well-oriented operations. Finally, a summary of all sections has been provided. The schematic of the current approach for investigating water production is presented in Figure 1. This paper does not discuss the details of each method by itself; therefore, it is necessary to study and examine the selected methods more specifically regarding the real condition of the case under study.

2. Water Production Sources

2.1. Water Drive. The natural water drive is mainly categorized as follows:

(1) Limited water drive

(2) Partial water drive

(3) Active water drive

The degree of activity in aquifers differs from one reservoir to another. For instance, the reef flat gas reservoir in the Changxing–Feixianguan Formation, Sichuan basin, China, is severely active because water cut sharply increased immediately after the first water breakthrough. However, Sinian gas reservoir in Weiyuan block behaves moderately [20].

In other types of water drive, the quality of aquifer is lower, and the water-bearing portion of reservoirs does not respond to gas depletion as quickly as active water-drive aquifers. Low permeability, heterogeneity, and other possible barriers in aquifers can be responsible for this behavior. If the aquifer is not very strong and active, it does not react to hydrocarbon depletion quickly, so that the pressure drop becomes more significant and the waterfront delays in advancement toward the gas zone.

As water encroaches into the gas-bearing zone from an active water aquifer, it maintains the average reservoir pressure. In other words, it slows down the rate of pressure drop. Consequently, at abandonment conditions, the amount of gas left in the reservoir increases, and the recovery becomes considerably lower.

In some cases, there is not any aquifer in the reservoir, or it is small enough to be ignored in comparison to the portion of the gas reserve. These reservoirs are called volumetric. As more gas is produced, existing fluids and rocks will exhibit a volume change in response to pressure reduction. Since there is not a limitation for pressure reduction, it leads to larger recovery factors.

Overall, water-drive gas reservoirs are reputed to be less recoverable than volumetric reservoirs and are dependent on various conditions as follows [21]:

(1) The production rate and production approach

(2) The residual saturation of gas

(3) The properties of aquifers

(4) The efficiency of volumetric displacement of water invading the gas reservoir

A group of occurrences can be responsible for water production problems in gas fields. They can be identified from their special characteristics. Besides, these problem sources lead to low reservoir recovery and well performance, environmental impacts, and greater operational costs. They have been discussed as follows.

2.2. Coning. Water coning is one of the crucial types of water production challenges in gas fields. It is described as the mechanism underlying the upward movement of water into perforations of a production well. There are two forces that control the water coning mechanism: gravity force and dynamic flow force [22]. Relatively high drawdown pressures in gas wells may end with the water coning and the reduction of gas saturation around the well leading to the decrease of relative permeability and production of gas.

On the contrary to oil wells, a small number of studies have been performed on water coning in gas wells. Muskat [23] supposed that the physical system of water coning in gas

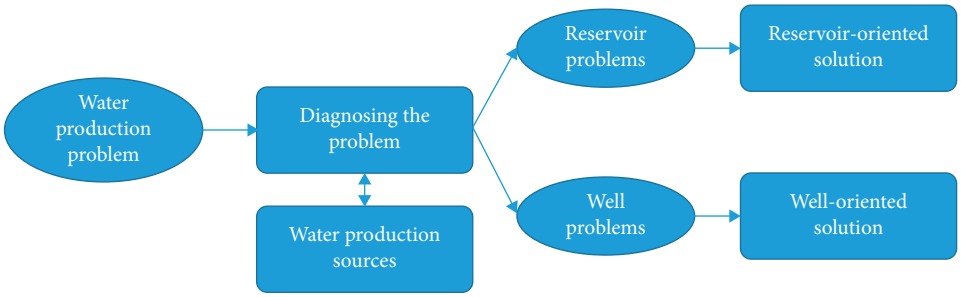


Figure 1: The applied approach for investigating the water production problem.

wells is similar to that in oil wells, and this phenomenon is less severe in gas wells. Trimble and DeRose [24] confirmed Muskat theory with an investigation on the Todhunters Lake Gas field, which is a moderate-to-strong water drive. On the contrary, Armenta and Wojtanowicz [22] and MacMullan and Bassiouni [25] stated that water coning in gas wells is physically different from coning in oil wells. Besides, from a numerical simulation study, Kabir [26] concluded that two parameters, permeability and pay zone thickness, are the most critical factors leading to water coning in gas fields, and other variables such as penetration ratio and production rate have minimal effect on ultimate recovery. Jafari et al. [27] executed a simulation study on one of the Iranian fractured gas reservoirs. They inferred that the well flow rate control is the most important and efficient factor to control water coning, economically and technically.

Moreover, Armenta and Wojtanowicz [22] examined the effect of vertical permeability, non-Darcian flow regime, and perforation density on water coning and the sensitivity of oil/water and gas/water systems to water influx. It was pointed out that as the vertical permeability increases, the water coning becomes more serious, and the water-to-gas ratio increases more rapidly, resulting in an early water breakthrough. Non-Darcian flows also cause further pressure drop near the wellbore that could exaggerate water coning. After all, using a modified analytic model, it was concluded that higher values of perforation density decrease the drawdown pressure and lead to lower water-to-gas ratios during production. In water-drive gas reservoirs, although a well is typically perforated at the top of the interval and produced at a limited rate, this perforation system sometimes lowers gas production rate, and it delays water breakthrough due to water coning [28].

2.3. Cresting. Similar to water coning, cresting happens when viscous forces generated by pressure drawdown defeat gravity forces associated with the density difference of gas and water [25]. This phenomenon is more common in horizontal and highly deviated gas wells. Cresting is more probable at high production rates from horizontal wells and especially near gas/water contact; thus, it is suggested the wells are drilled away from the waterfront [25]. The cresting is intensified by the increase in aquifer size and production rates, and subsequently the recovery decreases [29].

2.4. Fractures. In fractured gas reservoirs, fractures can be small, in size, or very large and deep [30]. When the formation is deeply fractured, the gas-containing zone may be connected to the aquifer; therefore, the water production problem could be extremely severe. The water penetration from fractures toward the producing zone and perforations can dramatically increase the water cut in gas wells. For instance, in Sinian gas reservoir in Weiyuan block, China, fractures are thought to be responsible for water production problems [20]. On the contrary, fractures may not be so deep, which leads to high water cut.

2.5. Water Fingering/Channeling. Sometimes there are highly permeable layers embedded between some other layers having low permeability resulting in high heterogeneity in gas reservoirs. This diversity in transmissibility of rocks leads to more advancement of water in highly permeable layers. In these layers, water advances faster due to more connected pore throats. If a well coincides with one of the mentioned highly permeable strata, an early water breakthrough and a high water-to-gas ratio occur in gas wells. For example, the Redfield storage field in Dallas is a gas reservoir in which heterogeneity causes production and recovery problems [31].

2.6. Liquid Loading. Liquid loading has been known as one of the problematic cases in gas production for many years, but its identification is not an easy task [32]. As the flow velocity of the gas in a well falls, the ability of gas to load liquids decreases.

The necessary information required for a comprehensive examination of the loading problem consists of drilling history, production data, wellbore diagram, completion log, CBL/VDL, or comparable logs necessary to evaluate cementation quality, gas and liquid sample analysis, scale and deposit analysis, and test data as well as data from production control surveys if available [33]. The provided database is analyzed to identify the reason for the liquid loading problem. Common sources of loading problems are as follows [33]:

(1) Subcritical velocity possibly in connection with water traps

(2) Increasing water-gas ratios due to coning

(3) Increasing water-gas ratio due to (selective) water encroachment into the reservoir

(4) Production of reservoir water from water-bearing formations via poor cementation

Mahadevan et al. [34] performed some experiments to investigate the effect of temperature, pressure, and permeability on liquid loading with an evaporation mechanism. They revealed that the liquid loading intensity decreases with the increase in temperature and increases with the pressure enhancement. By comparing two samples with permeability values of 0.01 and 1 md, it was concluded that higher permeability leads to more liquid unloading by evaporation mechanism.

2.7. Substantial Pressure in Water-Invaded Zones. A deviation between the calculated and real pressure values due to substantial pressure gradient exhibiting can be a source of water production problems. Lutes et al. [35] carried out a study on Katy reservoir in Oklahoma, and they found that the problem origin is a strong water drive due to substantial pressure gradient.

3. Diagnostic Methods

Several methods have been presented to identify water production problems, sources of problems, and selection of proper strategies. Agrawal et al. [21] used Carter–Tracy water influx model [36] and material balance equations to study the effect of the production flow rate on the ultimate recovery of a water-drive reservoir. They solved two equations by trial and error simultaneously and obtained a relation between the cumulative gas produced and the P/Z ratio. Then, they plotted cumulative gas production versus P/Z ratio for each production flow rate giving the trend of ultimate recovery with changing flow rates. As the production flow rate increased, the corresponding curve got closer to the illustrative line of no water influx that is representative of the maximum recovery. Afterward, they examined the effect of aquifer permeability, initial pressure, and continuous production on recovery. It was demonstrated that as the permeability of the aquifer increases, the ultimate recovery decreases. Gas recovery is less sensitive to production rate for reasonable production rates as aquifer permeability increases. Water influx reacts so rapidly to pressure changes in the high-permeability gas reservoir. Besides, less recovery was obtained for higher initial pressures [21]. The gas recovery decreased from 81.2% of initial gas in place (IGP) for continuous production to 66.4% IGP for intermittent production. This was caused by an increase in abandonment pressure from 1,552 to 2,721 psia. Finally, they found that the gas recovery was a function of production rate, strength of aquifer and permeability, saturation of residual gas, and volumetric seep efficiency of water invading zone.

Agrawal et al. [21] also proved that a lower abandonment pressure of a reservoir and a higher water influx from an aquifer led to better recoveries. Most of the time, the type of reservoir aquifer is determined by experience and a quick review of pressure/production data. Some type curves have been presented to estimate the degree of activity of aquifer and the recovery of gas reservoirs by Li et al. based on the material balance equation [37].

Knapp et al. [31] performed a multidimensional, two-phase, compressible fluid flow calculation to simulate the reservoir depletion. It was concluded that ultimate recovery in water-drive gas reservoirs is dependent on aquifer strength, production rate, fluid and rock properties, abandonment pressure of the reservoir, and especially reservoir heterogeneity. To study the effect of reservoir heterogeneity on recovery, three hypothetical reservoirs of different heterogeneities from the Redfield gas storage field, Dallas, were selected: (1) completely homogenous, (2) heterogeneous with vertical communicating, and (3) heterogeneous reservoir without vertical communication. Calculations indicated 62.5 percent recovery for homogenous reservoirs and 42.9 percent for reservoirs with vertical communications. The recovery of the noncommunicating system became 20.6 percent because the vertical flow was impeded and water fingering occurred in high permeable regions. It was pointed out that as heterogeneity increases, the recovery exhibits a decrease in water-drive gas reservoirs. In further simulation runs and calculations, Knapp et al. [31] confirmed the idea developed by Agrawal et al. [21] that recovery of gas reservoirs improves with reduction of aquifer strength and increase in gas production rates.

The existence of a substantial pressure gradient was confirmed by Lutes et al. [35] using subsequent one-dimensional numerical solutions based on the solution of partial differential equations. They applied a modified Van Everdingen and Hurst [38] unsteady-state material balance equation and finite difference techniques to describe the behavior of the Katy sand reservoir. This gradient caused an increase in pressure resulting in a reduction of the recovery, and lower pressures ahead of the waterfront influenced the deliverability of wells.

In some cases, it is difficult to estimate the original gas in place and the recovery factor because the conventional P/Z method responds falsely. For instance, in the study of Pepperdine [39] on the Middle Devonian gas fields containing a large aquifer located in Canada, water encroaches into a reef formation and causes some difficulties in routine calculations. Different results were obtained where geological interpretation revealed 1910 BSCF original gas in place, and pressure decline versus cumulative production curve showed 4800 BSCF gas in place. In the third attempt, the unsteady-state water influx equation and material balance equations were solved assuming a linear infinite aquifer resulting in 1800 BSCF gas in place. This value confirms the results achieved by geological interpretation and misleading caused by P/Z curves. Therefore, it was concluded that when there is a remarkable difference between results of geological interpretation and P/Z curves, solving unsteady-state water influx and material balance equations can be helpful. Additionally, a numerical simulation was used to illustrate that overall recovery is affected by the location, distribution, and timing of development wells.

Feng et al. [20] stated the reasons for successful water control of the second member of Xujiahe formation in Zhongba gas reservoir, which is under active edge-water drive as follows: First, some observation wells in the zone of edge-water played a key part considering the water influx direction and water body energy. Second, to hold the balanced production and control water, increasing well density helped in major water influx channel zones. Then, as signs of water influx were observed, they lowered the production rate to control water, stopping the water influx rate and opportunity for further research and control. Eventually, after clearly understanding water influx energy and direction, they utilized wells on water influx channels to drain water to lighten the effect of water influx.

An experimental investigation by Rezaee et al. [40] was conducted to study the reservoir heterogeneity effect on the recovery performance of a gas reservoir. Nitrogen with seven core samples of different heterogeneity degrees was used. The degree of heterogeneity of core samples was characterized by the Dykstra and Parsons [41] coefficient, which varies from 0 (homogenous) to 1 (heterogeneous). The range of samples varied reasonably from 0 to 1 to cover a good variety of samples. Eventually, from experimental results, it was concluded that heterogeneity is not always determinant for the recovery factor. The severity of this effect is clarified when the permeability ratio increases leading to a rise in the Dykstra–Parsons coefficient. As the Dykstra–Parsons coefficient increases, residual saturation in porous media increases too, resulting in a reduction in gas recovery of water-drive gas reservoirs.

Li et al. [37] presented a method to verify the activity level of aquifer and abandonment pressure of gas using pressure production data. The relation between coefficient of water remaining ω and reserve recovery degree R_g was obtained in the following form:

$$\begin{aligned} \ln \omega &= B \ln R_g, \\ \text{or } \omega &= R_g^B, \end{aligned} \tag{1}$$

where ω is defined as

$$\omega = \frac{W_e - W_p}{G * B_{gi}}. \tag{2}$$

For strong water influx, the value of B is small, where it is very large for weak water influx conditions. It was stated that B remains constant if the production strategy does not change during reservoir life. According to (1), if B is equal to 1, ω and R_g are also equally showing a strong aquifer. However, if B goes to infinity, ω becomes approximately zero, denoting the absence of aquifer [37]. Therefore, an aquifer exists for values of B in the range of $1 < B < \infty$. In cases that B is greater than 4, the effect of an aquifer can be ignored so that it can range between 1 and 4 for large aquifers [37]. Li et al. [37] also classified water-drive gas reservoirs into three categories according to the value of B. The aquifer is active if B is in the range of 1–1.5; it is said to be moderate if B is between 1.5 and 2.5; eventually, it is inactive when B ranges from 2.5 to 4.

Neglecting the impact of compressibility, the material balance equation (MBE) of water-drive gas reservoir is

$$\frac{P}{Z}(1 - \omega) = \frac{P_i}{Z_i}\left(1 - \frac{G_p}{G}\right), \tag{3}$$

$$P_r = \frac{1 - R_g}{1 - \omega}, \tag{4}$$

$$P_r = \frac{(P/Z)}{(P_i/Z_i)}, \tag{5}$$

where we have $R_g = (G_p/G)$. By substituting (4) and (5) in the following equation:

$$P_r\left(1 - R_g^B\right) = 1 - R_g, \tag{6}$$

where P_r can be plotted versus R_g to define theoretical curves, which are used to estimate the strength of aquifers (Figure 2).

Considering the effect of compressibility, the material balance equation is in the following form [37]:

$$\frac{P}{Z}\left(1 - C_c \Delta P - \omega\right) = \frac{P_i}{Z_i}\left(1 - \frac{G_p}{G}\right), \tag{7}$$

where we have $C_c = (C_w Swc + C_p/1 - S_{wc})$.

Combining and manipulating the above equations give

$$P_r = \frac{1 - R_g}{1 - C_c \Delta P - \omega}. \tag{8}$$

As discussed below, a suppositional curve was also obtained by Li et al. to calculate the abandonment pressure of a gas reservoir. At each step of production, the gas reserve can be determined using the equation revealed by Agrawal et al. [21]:

$$G_R = G - \frac{\left[\left(E_{va} G B_{gi} S_{gr}/1 - S_{wc}\right) + \left(1 - E_{va}\right) G B_{gi}\right]}{B_{ga}}. \tag{9}$$

By applying $S_{gi} = 1 - S_{wc}$, the above equation can be rearranged to achieve the proceeding equation:

$$\frac{G_R}{G} = 1 - E_{va}\left(\frac{S_{gr}}{S_{gi}} + \frac{1 - E_{va}}{E_{va}}\right)\frac{(P_a/Z_a)}{(P_i/Z_i)}. \tag{10}$$

Considering $E_R = (G_R/G)$ and $P_{ra} = ((P_a/Z_a)/(P_i/Z_i))$, (10) reduces to

$$1 - E_R = \left[1 - \left(1 - \frac{S_{gr}}{S_{gi}}\right)E_{va}\right]P_{ra}. \tag{11}$$

Let $a = 1 - (S_{gr}/S_{gi})$; then, we have

$$P_{ra} = \frac{1 - E_R}{1 - aE_{va}}. \tag{12}$$

Assuming that P_{rK} and E_R are variable, the following equation is achieved:

$$P_r = \frac{1 - R_g}{1 - aE_{va}}, \tag{13}$$

and P_r should be plotted against R_g, both ranging from zero to one (Figure 3). E_{va} and a are known values for each

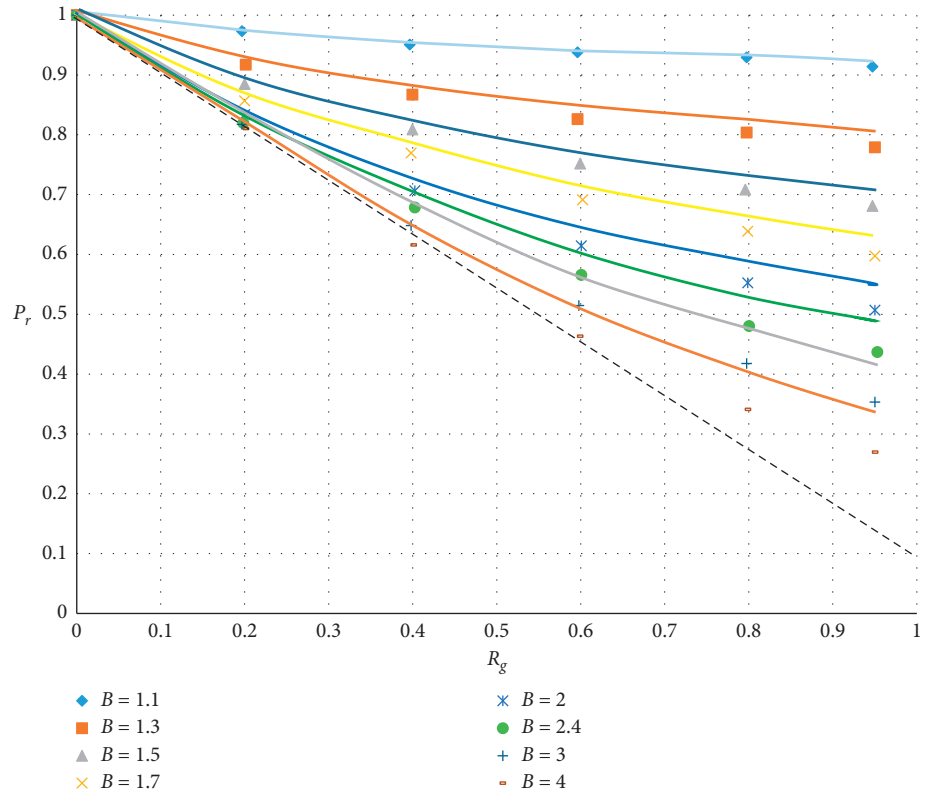


Figure 2: Type curves according to (6) [37].

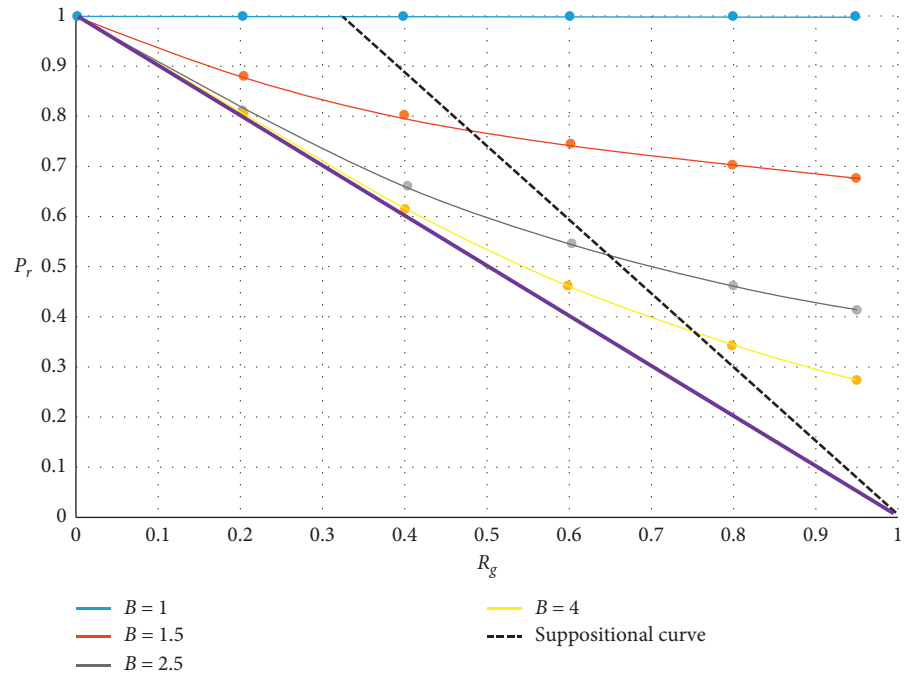


Figure 3: Plots of (6) and (12) with different B values when $a = 0.8$ and $E_{va} = 0.4$ [37].

reservoir. This curve is called suppositional because, in their range, only one point is real. There is a unique point that is obtained when the curve of (6) or (8) bisects the curve generated by (12), which denotes the abandonment pressure of the reservoir for a specific pattern of production. Therefore, the rest of the points are thought to be suppositional.

Feng et al. [20] tried to explain the importance of sufficient data from an aquifer and adequate well numbers that give a better vision of what occurs under the ground. Former studies revealed that typical log-log curves achieve a good prediction of edge-water invasion efficiency [20]. Since the bottom-water-drive model is different from the edge-water, the application of the log-log curve on the bottom-water drive was unsuccessful.

A study has been done by Cheng et al. [42] on the right region of the Amu Darya gas field, which is located at the Turkmenistan border, and they tried to identify the source of water production in a gas field. The gas field has an active bottom aquifer. In the initial days of production of some wells, the mentioned reservoir conditions and their heterogeneities result in the breakthrough of various degrees that restricts their deliverability seriously. Firstly, all candidate reasons for water production were considered. Formation water, operating fluids, and condensate water are probable sources. Besides, three methods were assumed to identify the real source: physical properties and water-to-gas (WGR) ratio for the condition of a particular source; the method of chloride ion conservation for the state of multisource water production.

Physical properties and WGR are vital factors in the identification of water sources, and they are used for single-source water problems. Condensate water has low density, low salinity, and a value of pH considerably less than seven due to acidic dissolution component of natural gas [42]. Operating fluids (drilling and acidizing fluids) contain a large number of additives with high density and high salinity. The drilling fluid pH is as high as strong alkaline solutions and for acidizing fluids is very low [4, 43–45]. The formation water samples of the aquifer indicate that its salinity and density is between the value for condensed water and operating fluids [42].

For WGR, due to high temperature and pressure in the right region of Amu Darya, the modified Mcketta–Wehe chart was used to evaluate the content of condensate water [46]. For the operating fluid, WGR is high at the initial flowback stage and decreases rapidly in the later stages [42]. Afterward, the formation water WGR is more than that of condensate water and constantly increases throughout the reservoir life [42].

Several methods have been developed to recognize the water influx and establish an optimized method for production to get the best deliverability of wells and the highest recovery factor. However, they are not applicable in reservoirs with moderate to high degrees of heterogeneity and uncertainty. Li et al. tried to establish a more functional method using three diagnostic curves and pressure/production data for individual wells [47].

The initial diagnostic curve of water influx was created on the Agarwal–Gardner flowing material balance [48], which is diversely used to evaluate OGIP. Its vertical axis is the productivity index of the pertinent well calculated with pseudo-pressure:

$$\frac{q}{\Delta p_p} = \frac{q}{p_{pi} - p_{pwf}}, \tag{14}$$

where

$$p_p = 2\int_0^p \frac{p}{\mu Z}\mathrm{d}p, \tag{15}$$

where q is the production rate at present, m^3/d; p is the bottom-hole flowing pressure, MPa; μ is the gas viscosity, mPa·s; Z is the dimensionless deviation feature; p_{pi} is the initial pseudo-pressure; and p_{pwf} is the bottom-hole pseudo-pressure. The horizontal axes are $2qt_{ca}p_i/((C_t\mu Z)_i\Delta p_p)$, where

$$t_{ca} = \frac{\left(\mu C_g\right)_i}{q\int_0^t q(t)/\left(\mu_{ave}\, c_{g_{ave}}\right)}, \tag{16}$$

where C_t and C_g are the total and gas compressibility, MPa^{-1}; t_{ca} is the pseudo-material balance time, dimensionless; t is the production time, day; μ_{ave} is the average viscosity at the corresponding time, mPa·s; $c_{g_{ave}}$ is the average gas compressibility at the corresponding time, MPa^{-1}; and the superscript i is initial value.

This indicative curve should be a straight line for a volumetric gas reservoir without any support from an aquifer, and its intersection with horizontal axes gives original gas in place. The three aforementioned periods are shown in Figure 4. It reveals that as water production becomes more severe in gas wells, the diagram deviates from a straight line representing more pressure maintenance by the aquifer.

In the second diagnostic curve, flowing reservoir pressure divided by gas compressibility factor is plotted on vertical axes versus cumulative gas production (Figure 5). Furthermore, the P/Z curve has been frequently used to identify characteristics of water-drive reservoirs. For volumetric reservoirs, the curve is a straight line, and the curve starts to diverge from the straight line for gas reservoirs with aquifer backing. The P/Z method can identify three periods of well life in the following manner: the straight line is representative of the no aquifer influx period, the slight deviation from the straight line accounts for the early aquifer influx period, and the middle-late aquifer influx period is shown when the curve deviates rapidly to the bottom right side.

The last diagnostic method is Blasingame type curves in which normalized production rate $(q\Delta p_p * 10^3 m^3/(\mathrm{d.MPa}))$ is the vertical axis and pseudo-material balance time $t_{ca}d$ is the horizontal axis (Figure 6). These type curves are used to calculate the original gas in place by flowing bottom-hole pressure and production flow rate data. It is obvious in Figure 6 that all type curves with different $r_D = r_e/(r_w e^{-s})$—where r_e is the drainage radius, m; r_w is the wellbore radius, m; and s is the skin factor—converge into a line of negative slope. Blasingame curves are also applicable in the identification of three aquifer influx periods. Li et al. [47] applied the mentioned type curves on gas wells in China. After that, it was suggested that the production from wells in the primary aquifer influx period and middle-late aquifer influx should be condensed simultaneously, and the production from wells in the no

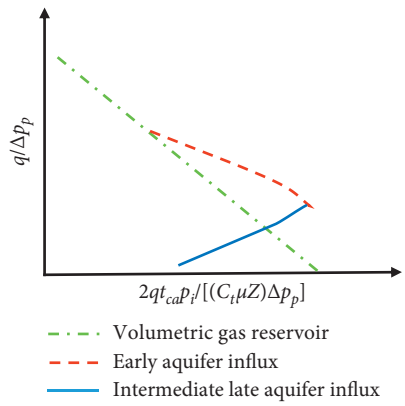


Figure 4: First diagnostic curve [47].

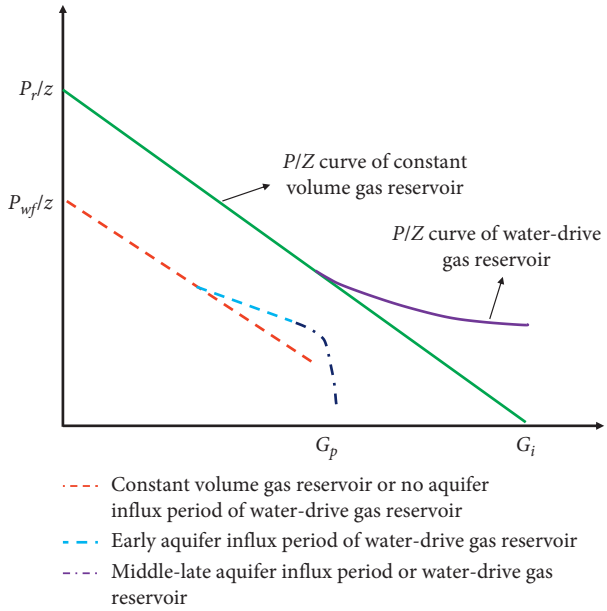


Figure 5: P/Z diagnostic curve [47].

aquifer influx period should be increased to obtain the optimized production.

The liquid loading is one of the most crucial problems in gas-producing wells, and its identification is not an easy task. As the gas rises in the bore hole, the wall becomes wet by liquid droplets. For higher liquid-gas ratios, the thickness of the liquid film increases, and it becomes more stable (assuming the form of regular gas slugs moving through the rising liquid). For natural gas wells, this is usually the end of the natural unloading of liquid [49]. To tackle this problem, the calculation of the critical flow rate can be helpful [50]. These relations are based on the balance between the drag and gravitational forces that give a critical velocity of a falling particle. The critical velocity is tied to the size, shape, and density of gas and liquid; surface tension; and drag coefficient in the following shape:

$$V_c = \frac{1.3\sigma^{0.25}\left(\rho_l - \rho_g\right)^{0.25}}{C_D\rho_g^{0.5}}, \tag{17}$$

where V_c is the critical velocity in ft/s, σ stands for the surface tension in psia, ρ_l and ρ_g are the liquid and gas densities in lb/ft^3, respectively, and C_D indicates the shape factor (dimensionless). Some rearrangements and manipulations on (1) give

$$V_c = \frac{C\left(\rho_l - 0.0031\,P\right)^{0.25}}{(0.0031P)^{0.5}}. \tag{18}$$

Here, C is 5.34 ft/s*lb/ft^3 0.25 for water and 4.02 for condensate. The gas will release liquids, in velocities below the critical velocity, and the accumulated liquid is calculated as follows:

$$q_{gc} = \frac{3.06PV_cA}{TZ}. \tag{19}$$

Here, q_{gc} is the gas flow rate in STBD and A is the conduit cross-sectional area in ft^2.

Perna et al. stated the advantages of this method as follows [32]:

(1) Detection of liquid loading

(2) Making a fast decision by available data

(3) Development of a relationship to compute the critical gas rate as a function of wellhead pressure

Lea et al. stated the critical flow rate in another form as follows [51]:

$$q_c = 1.166 * 520.1D^2\left(\frac{\left(\rho_l zT - 348.3\gamma_g P\right)P^2\sigma}{\left(348.3\gamma_g\right)^2T^3Z^3}\right), \tag{20}$$

where q_c is the critical gas rate, Nm^3/h; T is the in situ temperature, °K; D represents the pipe diameter, in; γ_g stands for the relative gas density (air = 1); ρ_l is the liquid density, kg/m^3; P denotes the in situ pressure, bar; and σ is the liquid-gas surface tension, dyn/cm. In relation to (4), the critical gas rate may be decreased, and so the unloading situations for the liquid enhanced, by [49] the following:

(1) Reduction of the pipe diameter, e.g., installation of a velocity string

(2) Reduction of the liquid density, e.g., foaming of the liquid by use of surfactants

(3) Lowering of pressure, e.g., use of compressors

(4) Increase of temperature

(5) Reduction of the surface tension, e.g., use of surfactants

4. Solutions

4.1. Reservoir-Oriented Solutions. Reservoir-oriented methods are mainly used to obtain better recoveries by working with pressure history and optimization of the overall reservoir flow rate. The main methods used to improve the efficiency of reservoirs are described in the next sections with field examples.

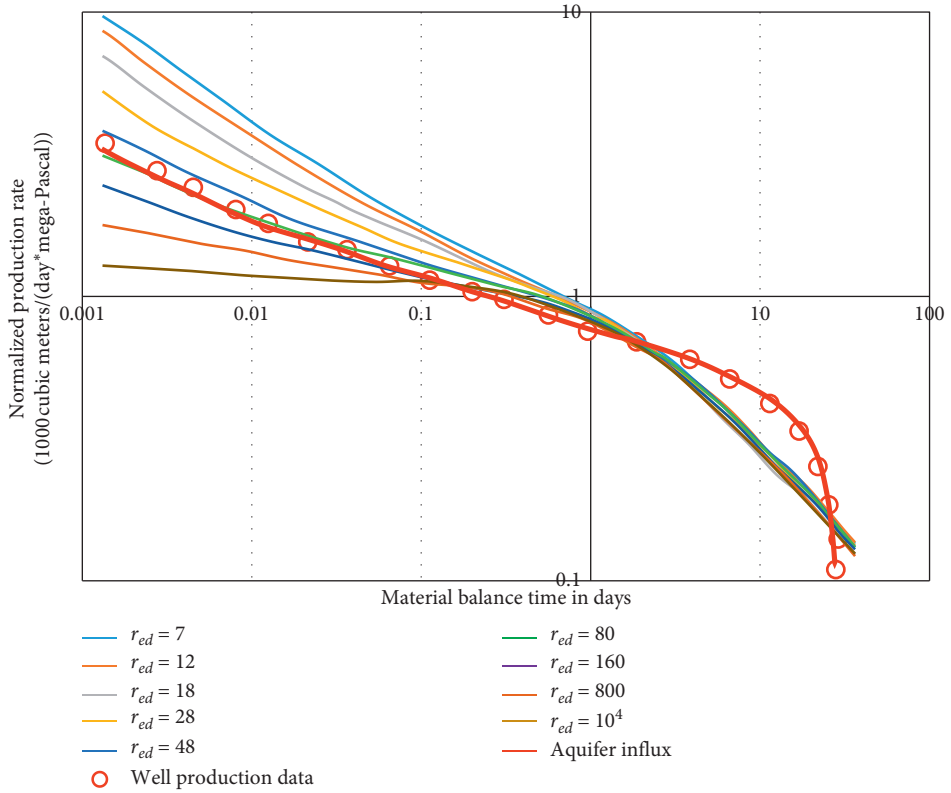


Figure 6: Blasingame type curves [47].

4.1.1. Coproduction Technique. Simultaneous production of water and gas to get a higher ultimate recovery factor is known as the coproduction technique which is the most technical and economical technique in the primary life of a gas reservoir [52, 53]. This technique has been proposed for modest to active water drive gas reservoirs but has the greatest economic potential for reservoirs that are not yet watered out [14]. Lutes et al. [35] pointed out that water production can be uncommercial for strong water drive gas reservoirs. As the reservoir produces, water cut in downdip wells increases, or they become watered out, probably. The production of downdip water enhances the recovery in three aspects. Firstly, it slows down the advance of the waterfront, ending in a delay in watering out of wells. Secondly, this technique reduces reservoir pressure and, finally, decreases the pressure in the swept zone.

Arcaro and Bassiouni [52] operated on a 10300 ft sand gas condensate reservoir in Eugene Island block 305. This reservoir contains six wells. Firstly, all production data and geological evaluations were reviewed. Secondly, volumetric analysis was applied to confirm and determine the water/gas contact. A humble material balance then was used to assess potential water production and recoveries for conventional production (without producing water) and for coproduction techniques. After that, a layered tank model, based on the material balance equation, was established, and it was adjusted with watered-out wells. Eventually, Schilthuis steady-state model [54] was originated to define the water influx. The forecasted recovery for coproduction case became 83 percent compared with 62 percent for conventional production manner, which denotes a rise of 56 BCF (from 274 BCF original gas in place) [52].

4.1.2. Accelerated Blowdown. Accelerated blowdown is a process in which the production rate of gas is increased to improve the ultimate recovery, and it is one of the widely used techniques [53]. Previous studies have demonstrated that it is one of the most appropriate techniques for strong water drive gas reservoirs [21]. When gas is produced, in an accelerated way, the aquifer delays in reacting to depletion and reduction of pressure, resulting in a higher gas recovery. Some situations may affect accelerated blowdown and limit the production program, such as lack of proper surface facilities, low deliverability of wells, economic contracts, and high permeability in the water-bearing zone. Lutes et al. [35] used a modified gas balance at six time steps, assuming that the average pressure of the invaded zone is consistent with residual gas saturation and gas saturation in the invades zone is constant. The recovery enhanced by 13 BCF (from 330 BCF original gas in place) applying an accelerated blowdown method, and 8.8 BCF rose due to reduction of pressure in the invaded zone and pressure improvement of wells.

4.1.3. Optimization of Production Flow Rate. Rezaee et al. [55] stated that increasing the production flow rate does not always lead to increasing the ultimate recovery, but there is

an optimum gas production flow rate that must be determined experimentally. In an experimental work, they used nitrogen and distilled water as two representative fluids. Firstly, core samples were saturated with water; then, nitrogen gas was injected from above to generate connate water saturation. The system was allowed to come to equilibrium for 24 hours. A dimensionless time number was obtained for the upscaling process, relating the experimental flow rate to the actual field gas production rate.

$$\text{Dimensionless time number} = \frac{\text{arrival pressure time to gas} - \text{water contact}}{\text{depletion time}},$$
$$T_{DS} = \frac{C\varnothing\mu L^2 q B_g}{12.656\,\text{kV}_g}. \tag{21}$$

Five core samples with equal dimensionless time numbers had the same recovery factor with an acceptable error. In a similar experiment on one core sample, five different production rates were examined. At first, the recovery factor increased with the production rate, and then it started to reduce. From plotting recoveries versus production flow rates, and curve fitting, the maximum recovery and its pertinent production rate were calculated. Eventually, it can be pointed out that the optimum production rate for each gas reservoir should be determined experimentally. However, heterogeneity and permeability variations are some obstacles of this method in actual reservoirs because all conducted experiments were performed on homogenous core samples.

Sech et al. [29] studied the impact of the gas production rate on the recovery of a reservoir being produced with horizontal and highly deviated gas wells. Large horizontal wells can provide remarkably higher production rates compared with conventional wells. In these wells, the water breakthrough is a severe problem [56, 57]. They found that the recovery reduces with raising of production rate for all permeability values. Moreover, obstacles of accelerating production rate decrease with increasing horizontal permeability because less drawdown pressure is needed to produce at the same rates, so that coning and cresting problems are impeded. In furthers runs, it was demonstrated that, in the lower ratios of vertical to horizontal permeability, the recovery factor is less sensitive to production rate.

4.1.4. CO_2 Injection into the Transition Zone. The use of CO_2 in volumetric drive reservoirs can be extended to strong water-drive reservoirs to enhance the recovery and control the water influx [14]. The density of CO_2 is more than that of natural gas and less than that of water; thus, CO_2 can be accommodated between gas and water zones by continuous injection into the transition zone. CO_2 occupies the gap in the zone and it is expected to significantly decrease water production by changing the water drive mechanism to the complete or partial volumetric mechanism [14]. A stable displacement can be provided by the CO_2 flooding, because it is more viscous and less mobile than methane gas. In fact, the CO_2 expansion in the transition zone maintains the reservoir pressure, slows down the waterfront movement, and sometimes causes the water level to recede toward the aquifer, especially in weak water drive gas reservoirs [58].

Even though the CO_2 injection considerably increases operating costs, the advantages of this technique can be counted as follows [14, 58–60]:

(1) Preventing the water encroachment from aquifers
(2) If the reservoir was invaded, the residual saturation would be from the CO_2 zone
(3) Reduction of water disposal costs
(4) Enhancing the gas and condensate recoveries
(5) Pressure maintenance leading to deliverability preservation of wells
(6) In the environmental aspect, reduction of contaminations by reducing polluted water production and elimination of CO_2 from the atmosphere

Al-Hasami et al. reported that the CO_2 injection after four years of production is more efficient than the injection from the first day of production [59]. Ogolo et al. established a simulation study on a lean gas reservoir with a strong water drive [14]. The CO_2 injection was examined for two conditions: 30 years of production without injection and the CO_2 injection from the onset of production. Seven injection wells and three production wells were used. In the case of CO_2 injection, the recovery showed an increase in two wells by 11 and 17 percent and a decrease in one well by about 12 percent. The condensate recovery increased in the reservoir by 4 percent, and water production significantly decreased in all wells, except for the well having a recovery reduction. The reason for this inappropriate behavior is still unknown. The pressure was also maintained and slightly increased over the initial pressure.

4.1.5. Solving Substantial Pressure Gradient Problem in the Invaded Zone. Lutes et al. [35] used a modified Van Everdingen and Hurst [38] unsteady-state material balance equations to model substantial pressure gradient problems in water-invaded regions. They applied this method on Katy reservoir in Oklahoma and matched the pressure production performance.

4.2. Well-Oriented Solutions. Sometimes problems can be treated in a well-based manner. If the water production source is the water influx, solutions can be stated as follows: First, if the influx has occurred due to water coning, drawdown management can be helpful. Second, if the water influx is from deeper perforation intervals and there is appropriate cementation between intervals, setting a plug can handle the problem. Eventually, cementation should be repaired, if water leaks through the cementation [49].

Liquid loading problems can be handled by consideration of the critical velocity. This requires careful production and controlling of water-gas ratio, condensate-to-gas ratio, and downhole pressure. Some sources of the liquid loading problem are mentioned before. The typical approach to handle the liquid loading is expressed in a simplified way as follows [49]:

(1) At the first signs of liquid loading, reduce the cross section to flow, e.g., by "snubbing" a velocity string.

(2) If the liquid rate is very high (>50 m^3/d), evaluate the gas lift versus pumps.

(3) For smaller liquid rates, too large for the use of plungers (10–50 m^3/d, dependent on the cross section to flow), use surfactants.

  (i) Sticks in wells of shallow depth
  (ii) Liquid batch applications in wells without packer
  (iii) Capillary injection in wells with packer

  The surfactant use requires that there is little condensate and no H_2S.

(4) Otherwise, install plungers in 50–250 m^3/h wells, if there is little risk of the plungers getting stuck. There are 3 types:

  (i) Conventional plungers
  (ii) Two-piece plungers
  (iii) Multiple stages

(5) Monitor the well and lower wellhead pressure as required.

4.2.1. Artificial Lift. When the bottom-hole pressure approaches the reservoir shut-in pressure, the production stops. A general approach to repair the flow is to expel the well to the atmospheric conditions, which is called blowdown [61]. This technique releases a huge amount of methane which is environmentally and economically deficient. In order to decrease mentioned effects, some alternative artificial lift methods are normally used.

4.2.2. Foaming Agent or Surfactant. Compared to other techniques of artificial lift, foaming is one the cheapest methods [62]. According to Figure 7, it works for liquids that consist of at least 50 percent water and does not respond well to liquid hydrocarbons [61]. In shallow wells, the foaming agent is added to wells by operators through the annulus. However, in deep wells, an injection setup is needed (surfactant tank, injection pump, motor valve, power source, and monitoring system). Lisbon and Henry reported a successful application of foaming agents in low permeability gas wells. The water production and gas-to-water ratio were both increased with this method [61].

4.2.3. Velocity Tubing. One choice to defeat the liquid loading is to set up a smaller diameter production tubing or "velocity tubing." Because the smaller diameter increases the velocity and, consequently, the lifting power, it may lower the liquid loading problem. Running a 1 in tubing string inside the production tubing in the northwest of Oklahoma and Texas Panhandle boosted the gas rate by more than 100% in four wells [63].

Ahmad and Zahoor reported a rule of thumb in a water-drive gas reservoir; the lower completion size reduces the abandonment pressure [64]. This rule may not be true in some cases, and lowering the abandonment pressure depends on formation properties, especially permeability, aquifer strength, tubing size, flow rates, surface facilities constraints, and so on. In that case study, the maximum recovery was obtained when tubing size and production rate were brought to their maximum limit without considering any economic and facility constraints. However, the implementation of larger tubing sizes is limited with economic and facility conditions, and in the later ages of some wells, it is compulsory to use smaller tubing sizes for better water unloading.

4.2.4. Plunger Lift Operated Manually or with Smart Well Automation. Plunger lift is an established technique for removing liquids from aging gas wells while minimizing gas losses and methane emissions (Figure 8). The plunger lift system removes liquids from the wellbore, and it helps the well to produce at lower pressures. Production engineers consider the plunger lift to be one of the simplest forms of artificial lift because it uses the own energy of well to remove accumulated liquids and sustain the gas production [65]. This method is simple, but it can increase the chemical costs, and it is less effective if significant quantities of liquid hydrocarbon are present [61]. Brady and Morrow successfully used the plunger lift for 130 low-pressure, tight-sand gas wells with water production problems [66].

4.2.5. Pumps. Currently, it is believed that the pump is an effective and economical way for lifting a large volume of fluid from the deep well under a variety of well conditions. As Figure 9 illustrates, the pump achieves the highest efficiency when the pumped fluid is only liquid [67].

4.2.6. Intelligent Methods using ICVs. Downhole flow and pressure control can be attained through the use of interval control valves (ICVs) which are remotely controllable. Some researchers have indicated the usage and profit of intelligent wells with ICVs like Brouner and Jansen [68], Gai [69], Ajayi

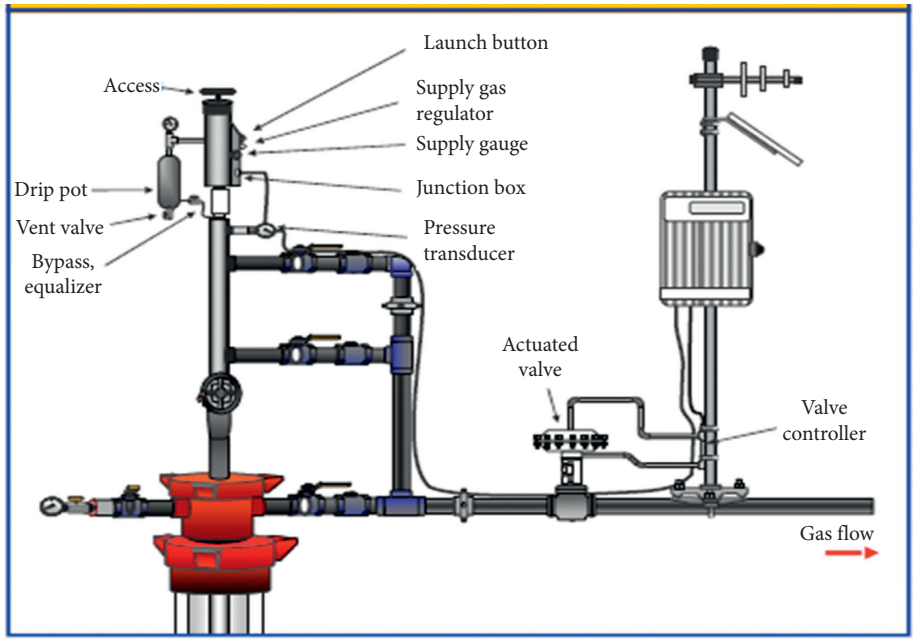


Figure 7: Foaming setup [61].

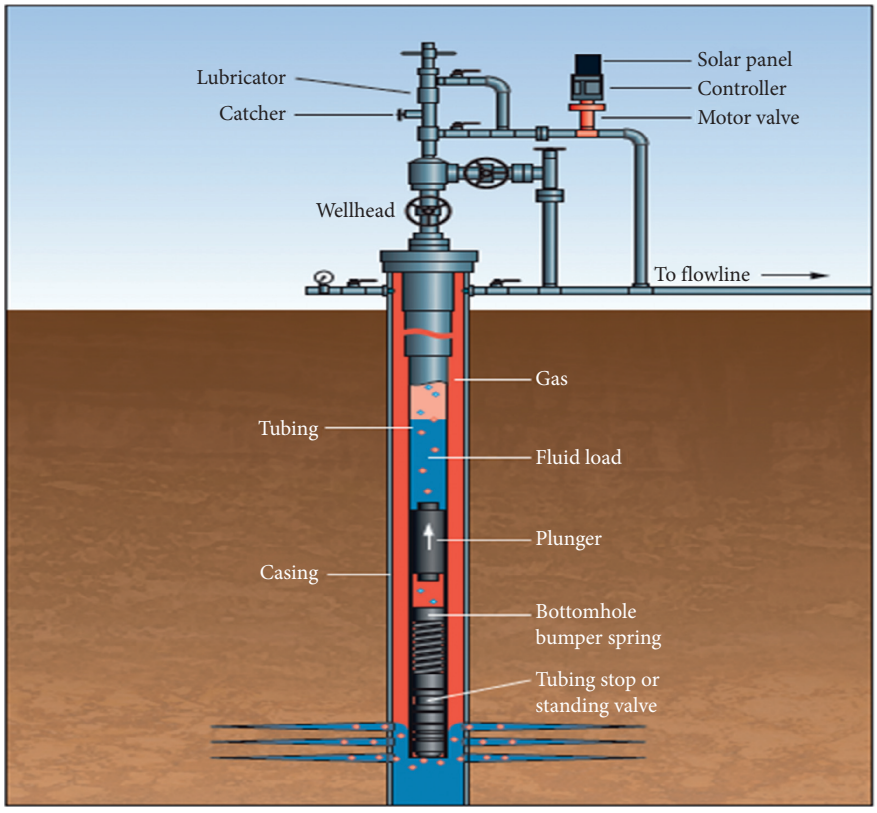


Figure 8: Plunger lift in a typical well.

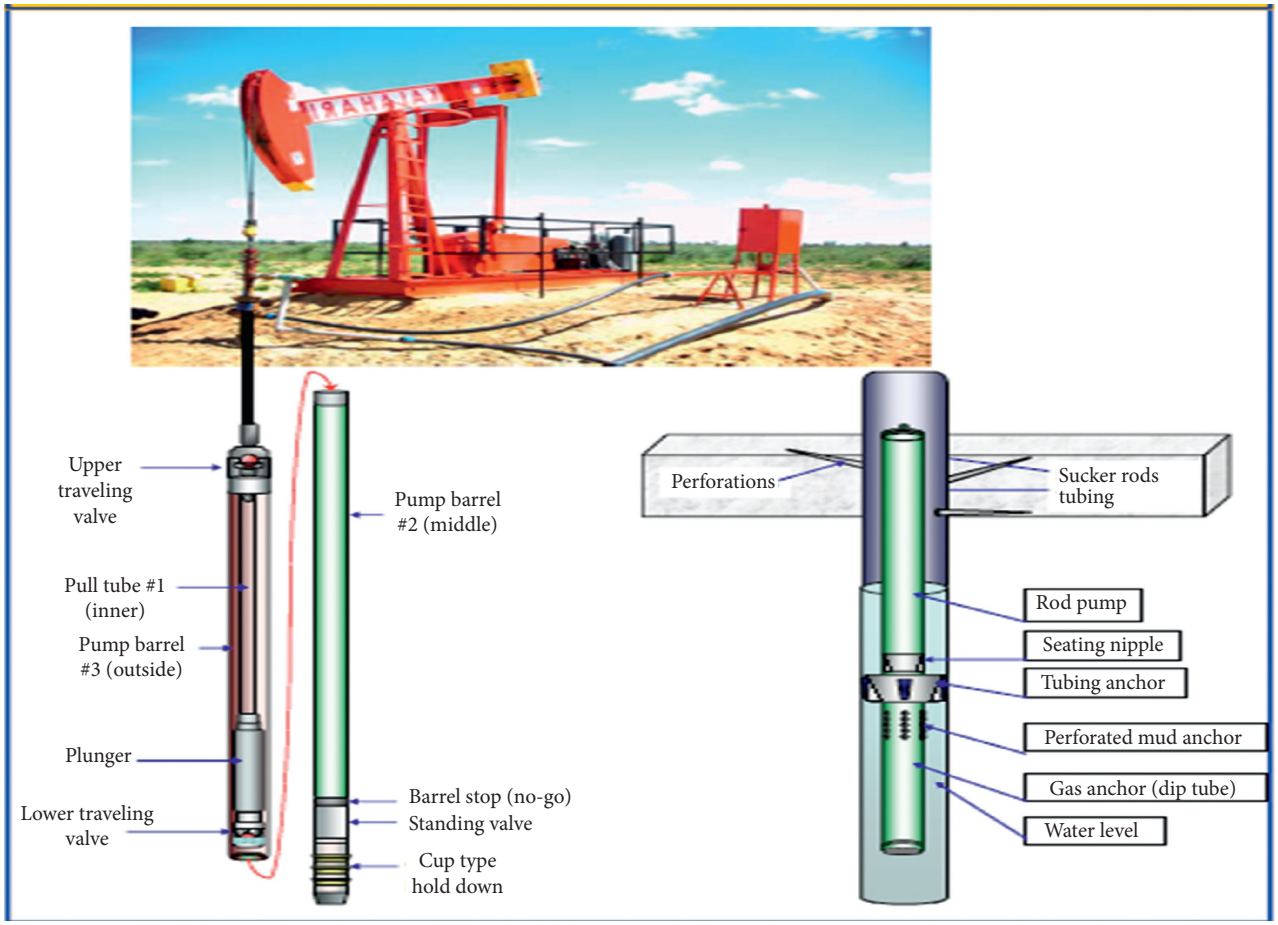


Figure 9: Positive displacement pump and rod pump [67].

and Konopczynski [70], Durlofsky and Aziz [71], and Naus al. [72].

Yamali et al. used a hybrid optimization method to optimize pressure intervals that ICVs could work with [28]. To evaluate the objective function, an approximation method was applied, called proxy model (pressure intervals are independent variables). Approximation methods were generally more cost-effective than simulation methods. Response Surface Methodology (RSM) and experimental designs were utilized to make the proxy model for performance function [28]. Besides, other techniques were tested to ensure that the proxy model concludes a valid objective function. Then, a hybrid genetic algorithm was used to optimize the objective function. Moreover, this method improves the uniqueness of the solution and is less computationally expensive.

4.2.7. Downhole Water Sink (DWS) Technology. The DWS technique uses a dual-completed wellbore in which the lower part is isolated from the upper zone by packers (Figure 10). It controls the water production by redirecting the stream of water to the bottom portion of the borehole, and as a result the performance of the top completion can be maximized. Several studies have been done to evaluate the feasibility of this method [73]. It was figured out that the DWS can boost the recovery and productivity of low-productivity gas reservoirs. The DWS and DGWS (downhole gas/water separation) both give the same recovery, but the DWS accelerates the production and produces sooner.

Armenta and Wojtanowicz stated that the best stage to begin the DWS process is in the early life of a well when the water production begins. Besides, it should not be postponed until the well is completely loaded with water. Moreover, six factors controlling DWS well performance were analyzed using a layer-cake type model to reach the best conditions by Armenta and Wojtanowicz [74].

4.2.8. Reduction of Top Completion Length. Shortening the completion increases the production period and, consequently, increases the recovery. On the other hand, increasing the completion length raises the production rate as well as water production; therefore, more water drainage from the bottom zone is required. Besides, water production should be increased gradually to continue with water invasion.

4.2.9. Downhole Gas/Water Separation (DGWS). One of the technologies introduced to tackle the water problem is downhole gas/water separation (DGWS) (Figure 11). Water and gas are separated, and water is disposed of into a nonproductive layer of the same well at downhole

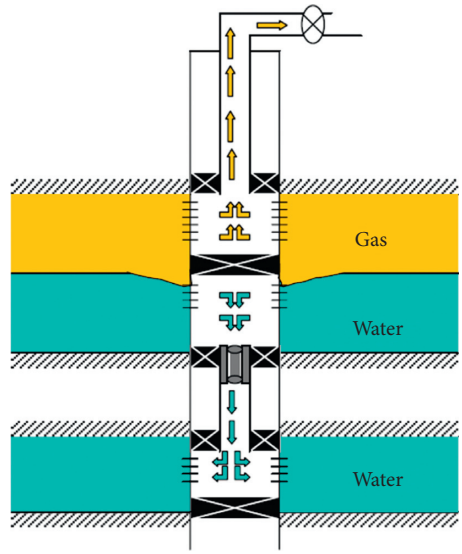


Figure 10: Downhole water sink [73].

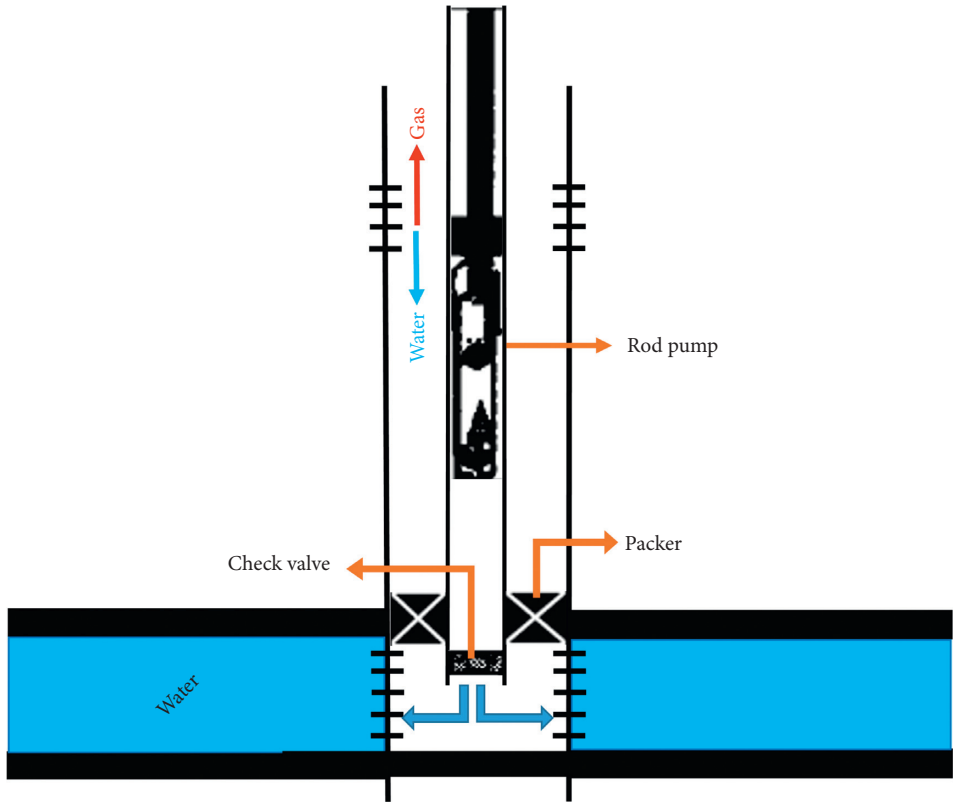


Figure 11: Well completion configuration in a well with the DGWS system [75].

conditions. Water is injected into disposal formations using one of the following pumping agents: modified plunger rod pumps, electrical submersible pumps, progressive cavity pumps, and insert pumps. Since the difference between densities of water and gas is high, separation naturally occurs in the borehole. The DGWS technology can be classified into two main categories: gravity separation and hydrocyclone separation [17].

The appropriate injectivity and disposal zone quality are key factors in choosing a well for the DGWS application, and the candidate reservoir should have sufficient remaining reserve. Furthermore, disposal formations must be isolated from the production zone [17]. The DGWS technology has great potential in gas well dewatering and reservoir development, especially in low-pressure/low-permeability gas reservoirs [17].

A Canadian company installed a special DGWS system that improved the production in comparison to other DGWS systems; moreover, some operating conditions were developed. The production was improved from 555 KSCFD gas and 440 BWPD water to 777 KSCFD and 195 BWPD in a well. Several commercially available systems were tested for injecting water into the lower disposal zone within the production well [75].

After economical and technical analysis, a "by-pass" subconfiguration pump was installed on a well in sandstone reservoir located in Alberta, Canada. The well produced water free for about 18 months. Afterwards, water production could easily be demonstrated by material balance calculations. The determination of water production source was not straightforward due to the complexity of the formation. Installation of this system caused an increase in production. In this case, despite the robust flow characteristics, the flow area was small and the additional pressure drop was caused by the liquid loading. To improve the aforementioned system, the following modifications were also suggested [75]:

(1) Installation of a larger downhole pump such as progressing cavity pump (PCP) or electrical submersible pump (ESP)

(2) Improving the downhole separation efficiency

Gao et al. reviewed 92 wells with the DGWS installation and reported that 61 percent of applications were successful [46].

4.2.10. Formation Cement Squeezing. Because this technique is available and easy to apply, it is one of the primary choices to shut-off problematic perforations, fractures, and channels [76]. Cement slurry is forced under pressure to a specific point in wells [77]. The operation is divided into two specific techniques [77]:

(1) High-pressure squeeze is used when the formation is under a cake of workover fluid. High injecting pressure fractures the formation, resulting in the usage of larger volume of cement

(2) Low-pressure squeeze does not cause the development of more fractures, so it needs less volume of cement compared with the previous technique

Some properties of well and formation, like fractures and injectivity, should be considered before applying cement squeezing. For instance, in gas condensate field of Hassi R'Mel in Algeria, which was suffering from water production through fractures, this technique failed [76]. Some possible reasons for this failure have been suggested: the possibility of formation resistance to injectivity, obstruction of perforations, and some technical reasons. Moreover, the long-term efficiency of this method is restricted in the Tunu field in Indonesia because high-pressure drawdown, pressure variations, and temperature fluctuations ended with the loss of cement from perforations [78].

Moreover, this method may lead to loss of productive zone due to cementing and blocking the problematic interval. This loss is a critical disadvantage of cement squeezing approach.

4.2.11. Gas Lift. One of the candidate methods to remove the water loading is the gas lift technique (Figure 12). The high-pressure gas is injected through the casing, and a gas lift valve allows gas to produce at the surface from the tubing [79]. Water is lifted to the surface by the evaporation mechanism reducing the density of the loaded liquid in wells [79]. To test the feasibility of this method for each well, further economic evaluation is needed [80].

4.2.12. Bridge Plugs. This equipment is used to isolate the water-producing interval in gas wells. It is set with two methods: electric wireline and mechanical methods. Two types of bridge plugs are common: retrievable and permanent. Retrievable plugs can be removed and installed several times, but permanent plugs cannot be used again. Permanent types can be removed by drilling. Bridge plugs are used when the problematic zone is below the producing zone, in other words, when there is no productive interval below the water-producing zone [78]. If there is a gas-producing zone below the problematic interval, using bridge plugs results in the loss of valuable intervals, and less gas is produced [78].

4.2.13. Polymer Treatment. In some cases, it is difficult to distinguish exact boundaries between gas- and water-producing intervals, and the usage of permanent plugs is very risky [81]. Therefore, it is more desirable to use selective barriers that block water and transmit gas by using hydrosoluble polymers [82]. The relative permeability of brine is remarkably reduced due to polymer adsorption on the surface of rock [81]. In addition, the reversible adsorption of polymer aids the long-term efficiency of this method [83] (Figure 13).

The efficiency of this process is severely dependent on the thickness of the absorbing layer in comparison to the pore size [83]. In fractured and high permeable reservoirs, water flows through channels and remarkably reduces polymer efficiency [81, 84]. To tackle this problem, some metal ions (crosslinkers) are used to make a polymer network [85].

Dovan and Hutchins reported an application of a polymer crosslinked with titanium in a gas reservoir in Northern California. The polymer contained potassium bicarbonate to retrieve the gas permeability by evolving CO_2, but both gas and water production reduced. In order to restore the gas permeability, HCL was injected into the formation. It was expected that acid fingers polymer gel, but water production increased again bringing the gas permeability back to its previous state. Therefore, it was concluded that polymer gel could be removed partially or completely by acidizing. In 1988, another well in Northern California was treated with polymer. Ester was injected with crosslinking material and potassium bicarbonate to restore the gas permeability by a chemical reaction (evolving CO_2). The

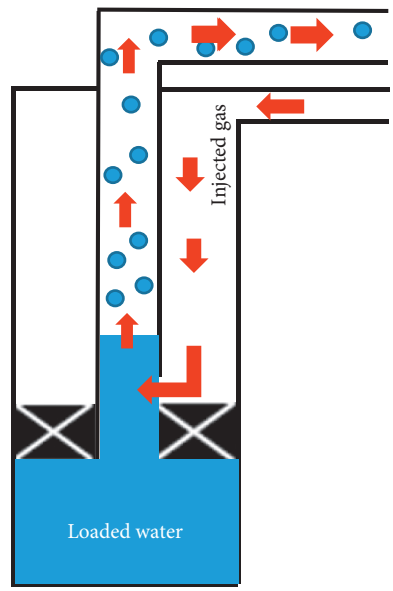


Figure 12: Dewatering of a gas well by gas lift technique [80].

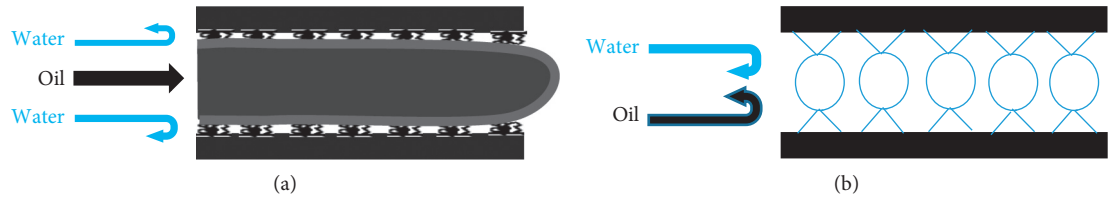


Figure 13: Single polymer treatment and crosslinkers [81]: (a) polymer adsorption; (b) crosslinked polymer.

process was successful, and water production decreased simultaneously with improving the gas relative permeability.

Zaitoun et al. established an experiment to examine two types of polymer for reservoirs with low and high salinities [81]. Polymer A swelled in high salinity and shrank in low salinity, and the opposite was true for polymer B. Polymer A was injected in saltier water than reservoir brine. It was absorbed on the surface of the rock during the injection. Polymer particles swelled when they were exposed to brine, and consequently it blocked the flow of water in the reservoir. The process corresponding to polymer B was exactly the opposite. For the first time in a gas field, in 1986, polymer A was injected into the Cerville-Velaine sandstone reservoir. Water production in wells significantly reduced, and its effect remained for a long time.

A core flood experiment of polymer B was performed by Zaitoun and Pichery to examine polymer injection into the Saint-Clair-sur-Epte reservoir [86]. It was a limestone reservoir associated with a strong bottom aquifer. After the polymer treatment (core scale), the relative permeability of water considerably decreased without any significant change in the relative permeability of gas. In real field injection, water cut of the well remained high for two years. After that, it decreased sharply and remained at low levels for over ten years.

5. Summary and Conclusions

There are several studies in literature revealing various results that are acceptable technically and economically for specific cases of study. In order to choose the best solution, precise study and data analysis should be made like evaluating rock and fluid data, reviewing pressure/production performance, and considering geological aspects. After considering all aspects of reservoir and analyzing the data, the most appropriate method should be used. All mentioned possible sources, diagnostic tools, and solutions for water production problems in gas fields are summarized as follows:

(1) There are various sources for water production problems, including water drive of aquifer, coning, cresting, fractures, water fingering/channeling, liquid loading, and substantial pressure.

(2) To identify the problem source, several approaches have been developed: cumulative gas production versus P/Z plot, numerical simulation (a modified Van Everdingen–Hurst unsteady-state material balance equation), Pi versus Rg plot, suppositional curve in Pi versus Rg, well test log-log curves for water invasion efficiency, Agarwal–Gardner diagnostic curve, and Blasingame type curves.

(3) Once the source has been diagnosed, the appropriate practical solution should be implemented. Practical solutions to tackle the water production problems in gas reservoirs are categorized into two types: reservoir-oriented solutions and well-oriented solutions.

(4) Reservoir-oriented solutions comprise coproduction technique, accelerated blowdown, optimization of production flow rate, CO_2 injection into transition zone, and solving substantial pressure gradient problem.

(5) Well-oriented solutions include artificial lift, foaming agent or surfactant, velocity tubing, plunger lift, pump, intelligent methods using ICVs, downhole water sink (DWS), reduction of top completion length, downhole gas/water separation (DGWS), formation cement squeezing, gas lift, bridge plugs, and polymer treatment.

The results of this research present the basic road map for future studies in dealing with water production problems in gas reservoirs. However, several gaps still exist in the details of mechanisms and solutions. Hence, future research can focus on these existing gaps.

Abbreviations

CBL/VDL: Cement bond logging/variable density logging
IGP: Initial gas in place
BSCF: Billion standard cubic feet
WGR: Water-to-gas
OGIP: Original gas in place
A: Conduit cross-sectional area, ft^2
B: Water influx coefficient
B_{ga}: Gas formation volume factor at abandonment pressure, m^3/std m^3
B_{gi}: Gas formation volume factor at pressure P_i, m^3/std m^3
B_w: Water formation volume factor at pressure P_i, m^3/std m^3
C_D: Shape factor, dimensionless
C_t: Total compressibility
C_g: Total and gas compressibility, MPa^{-1}
$c_{g_{ave}}$: Average gas compressibility at corresponding time, MPa^{-1}
C_c: Compressibility of reservoir volume, MPa^{-1}
C_w: Compressibility of water, MPa^{-1}
C_p: Compressibility of rock, MPa-$^{-1}$
D: Pipe diameter, in
E_{va}: Ultimate volumetric sweep efficiency at abandonment pressure, dimensionless fraction
G_p: Cumulative gas produced, m^3
G_R: Recoverable reserve, m^3
G: Original gas in place, m^3
P: Pressure, MPa
P_i: Initial reservoir pressure, MPa
P_r: Relative pressure, dimensionless
p_{pi}: Initial pseudo-pressure
p_{pwf}: Bottom-hole pseudo-pressure
p: Bottom-hole flowing pressure, MPa
q_c: Critical gas rate, Nm^3/h
q: Production rate at present, m^3/d
q_{gc}: Gas flow rate, STBD
R_g: Reserve recovery degree
r_e: Drainage radius, m
r_w: Wellbore radius, m
s: Skin factor, dimensionless
Swc: Irreducible water saturation, dimensionless fraction
S_{gr}: Residual gas saturation, dimensionless fraction
T: In situ temperature, °K
t_{ca}: Pseudo-material balance time, dimensionless
t: Production time, day
V_c: Critical velocity, ft/s
W_e: Cumulative water influx, m^3
W_p: Cumulative water produced, m^3
Z: Compressibility factor, dimensionless
Z_i: Compressibility factor at pressure P_i, dimensionless
γ_g: Relative gas density
ρ_l: Liquid density, lb/ft^3
ρ_g: Gas density, lb/ft^3
μ: Gas viscosity, mPa·s
μ_{ave}: Average viscosity at corresponding time, mPa.s
ω: Coefficient of water remaining.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] H. Esfandyari, S. R. Shadizadeh, F. Esmaeilzadeh, and A. Davarpanah, "Implications of anionic and natural surfactants to measure wettability alteration in EOR processes," *Fuel*, vol. 278, p. 18392, 2020.

[2] M. Mohajeri, M. Hemmati, and A. S. Shekarabi, "An experimental study on using a nanosurfactant in an EOR process of heavy oil in a fractured micromodel," *Journal of Petroleum Science and Engineering*, vol. 126, pp. 162–173, 2015.

[3] R. Abdollahi and S. R. Shadizadeh, "Experimental investigation of side effect of henna extract as a new and ecofriendly corrosion inhibitor on acid injectivity of calcareous sandstone," *Transport in Porous Media*, vol. 97, pp. 105–118, 2013.

[4] R. Abdollahi and S. R. Shadizadeh, "Effect of acid additives on anticorrosive property of henna in regular mud acid," *Scientia Iranica*, vol. 19, pp. 1665–1671, 2012.

[5] M. Dong, Q. Liu, and A. Li, "Displacement mechanisms of enhanced heavy oil recovery by alkaline flooding in a micromodel," *Particuology*, vol. 10, pp. 298–305, 2012.

[6] H. Amani, "Study of enhanced oil recovery by rhamnolipids in a homogeneous 2D micromodel," *Journal of Petroleum Science and Engineering*, vol. 128, pp. 212–219, 2015.

[7] T. N. C. Dantas, V. C. Santanna, T. T. C. Souza, C. R. S. Lucas, A. A. Dantas Neto, and P. T. P. Aum, "Microemulsions and nanoemulsions applied to well stimulation and enhanced oil recovery (EOR)," *Brazilian Journal of Petroleum and Gas*, vol. 12, pp. 251–265, 2019.

[8] A. Keykhosravi, M. B. Vanani, A. Daryasafar, and C. Aghayari, "Comparative study of different enhanced oil recovery scenarios by silica nanoparticles: an approach to time-dependent wettability alteration in carbonates," *Journal of Molecular Liquids*, vol. 324, p. 115093, 2021.

[9] O. Massarweh and A. S. Abushaikha, "The use of surfactants in enhanced oil recovery: a review of recent advances," *Energy Reports*, vol. 6, pp. 3150–3178, 2020.

[10] J. Foroozesh and S. Kumar, "Nanoparticles behaviors in porous media: application to enhanced oil recovery," *Journal of Molecular Liquids*, vol. 316, p. 113876, 2020.

[11] H. Esfandyari, A. Moghani Rahimi, F. Esmaeilzadeh, A. Davarpanah, and A. H. Mohammadi, "Amphoteric and cationic surfactants for enhancing oil recovery from carbonate oil reservoirs," *Journal of Molecular Liquids*, vol. 322, p. 114518, 2020.

[12] H. Esfandyari, A. Moghani, F. Esmaeilzadeh, and A. Davarpanah, "A laboratory approach to measure carbonate rocks' adsorption density by surfactant and polymer," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5539245, 7 pages, 2021.

[13] R. Abdollahi, M. Nadri, H. Gholghanddashti, M. Safari, and M. Zare Reisabadi, "A stepwise approach for identification of water production mechanisms in gas reservoirs," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 2021.

[14] N. A. Ogolo, J. O. Isebor, and M. O. Onyekonwu, "Feasibility study of improved gas recovery by water influx control in water drive gas reservoirs," in *Proceedings of the SPE Nigeria Annual International Conference and Exhibition*, Lagos, Nigeria, August 2014.

[15] T. M. Geffen, D. R. Parrish, G. W. Haynes, and R. A. Morse, "Efficiency of gas displacement from porous media by liquid flooding," *Journal of Petroleum Technology*, vol. 4, pp. 29–38, 1952.

[16] D. K. Keelan and V. J. Pugh, "Trapped-gas saturations in carbonate formations," *SPE Journal*, vol. 15, pp. 149–160, 1975.

[17] M. F. Radwan, "Feasibility evaluation of using downhole gas-water separation technology in gas reservoirs with bottom water," in *Proceedings of the SPE Middle East Oil & Gas Show and Conference*, Manama, Bahrain, March 2017.

[18] M. Wei, K. Ren, Y. Duan, Q. Chen, and M. Dejam, "Production decline behavior analysis of a vertical well with a natural water influx/waterflood," *Mathematical Problems in Engineering*, vol. 2019, Article ID 1683989, 9 pages, 2019.

[19] I. B. Obot, S. A. Umoren, and N. K. Ankah, "Pyrazine derivatives as green oil field corrosion inhibitors for steel," *Journal of Molecular Liquids*, vol. 277, pp. 749–761, 2019.

[20] X. Feng, B. Zhong, X. Yang, and H. Deng, "Effective water influx control in gas reservoir development: problems and countermeasures," *Natural Gas Industry B*, vol. 2, pp. 240–246, 2015.

[21] R. G. Agarwal, R. Al-Hussainy, and H. J. Ramey, "The importance of water influx in gas reservoirs," *Journal of Petroleum Technology*, vol. 17, pp. 1336–1342, 1965.

[22] M. Armenta and A. Wojtanowicz, "Severity of water coning in gas wells," in *Proceedings of the SPE Gas Technology Symposium*, Calgary, Canada, April 2002.

[23] M. Muskat, "The flow of homogeneous fluids through porous media," *Soil Science*, vol. 46, p. 169, 1938.

[24] A. E. Trimble and W. E. DeRose, "Field application of water-coning theory to todhunters lake gas field," *Journal of Petroleum Technology*, vol. 29, pp. 552–560, 1977.

[25] J. H. McMullan and Z. Bassiouni, "Optimization of gas-well completion and production practices," in *Proceedings of the SPE International Petroleum Conference and Exhibition in Mexico*, Villahermosa, Mexico, February 2000.

[26] C. S. Kabir, "Predicting gas well performance coning water in bottom-water-drive reservoirs," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, San Francisco, CA, USA, October 1983.

[27] I. Jafari, S. Jamshidi, and M. Masihi, "Investigating the mechanism of water inflow in gas wells in fractured gas reservoirs and designing a controlling method," in *Proceedings of the SPE International Production and Operations Conference & Exhibition*, Doha, Qatar, May 2012.

[28] N. Yamali, Q. P. Nguyen, and S. Srinivasan, "Optimum control of unwanted water production in stratified gas reservoirs," in *Proceedings of the Production and Operations Symposium*, Oklahoma City, OK, USA, March 2007.

[29] R. P. Sech, M. D. Jackson, and G. Hampson, "Controls on water cresting in high productivity horizontal gas wells," in *Proceedings of the EUROPEC/EAGE Conference and Exhibition*, London, UK, June 2007.

[30] R. Abdollahi, H. Esfandyari, M. Nadri Pari, and A. Davarpanah, "Conventional diverting techniques and novel fibr-assisted self-diverting system in carbonate reservoir acidizing with successful case studies," *Petroleum Research*, 2021.

[31] R. M. Knapp, J. H. Henderson, J. R. Dempsey, and K. H. Coats, "Calculation of gas recovery upon ultimate depletion of aquifer storage," *Journal of Petroleum Technology*, vol. 20, pp. 1129–1132, 1968.

[32] M. Perna, G. Bartolotto, R. Latronico, and R. Sghair, "Dynamic surveillance templates for reservoir management: diagnostic tools oriented to production optimization," in *Proceedings of the International Petroleum Technology Conference*, Doha, Qatar, December 2009.

[33] E. Adaze, A. Al-Sarkhi, H. M. Badr, and E. Elsaadawy, "Current status of CFD modeling of liquid loading phenomena in gas wells: a literature review," *Journal of Petroleum Exploration and Production Technology*, vol. 9, pp. 1397–1411, 2019.

[34] J. Mahadevan, M. M. Sharma, and Y. C. Yortsos, "Evaporative cleanup of water blocks in gas wells," *SPE Journal*, vol. 12, pp. 209–216, 2007.

[35] J. L. Lutes, C. P. Chiang, R. H. Rossen, and M. M. Brady, "Accelerated blowdown of a strong water-drive gas reservoir," *Journal of Petroleum Technology*, vol. 29, pp. 1533–1538, 1977.

[36] R. D. Carter and G. W. Tracy, "An improved method for calculating water influx," *Transactions AIME*, vol. 219, pp. 415–417, 1960.

[37] M. Li, W. J. Yang, Q. Y. Xiao, S. C. Liu, J. Zhang, and G. M. Peng, "Determination of the aquifer activity level and the recovery of water drive gas reservoirs," in *Proceedings of the North Africa Technical Conference and Exhibition*, Cairo, Egypt, February 2010.

[38] A. F. Van Everdingen and W. Hurst, "The application of the laplace transformation to flow problems in reservoirs," *Journal of Petroleum Technology*, vol. 1, pp. 305–324, 1949.

[39] L. Pepperdine, "Recognition and evaluation of water drive gas reservoirs," in *Proceedings of the Annual Technical Meeting*, Calgary, Canada, June 1978.

[40] M. Rezaee, B. Rostami, and P. Pourafshary, "Heterogeneity effect on non-wetting phase trapping in strong water drive gas reservoirs," *Journal of Natural Gas Science and Engineering*, vol. 14, pp. 185–191, 2013.

[41] H. Dykstra and R. L. Parsons, *The Prediction of Oil Recovery by Waterflooding in Secondary Recovery of Oil in the United States*, API, Washington, DC, USA, 2nd edition, 1950.

[42] Y. Cheng, L. Mu, E. Zhu et al., "Water producing mechanisms of carbonate reservoirs gas wells: a case study of the right bank field of Amu Darya, Turkmenistan," *Petroleum Exploration and Development*, vol. 44, pp. 89–96, 2017.

[43] R. Abdollahi and S. R. Shadizadeh, "The effect of spent acid on carbonate rock wettability during a matrix acidizing treatment," *Petroleum Science and Technology*, vol. 32, pp. 450–454, 2014.

[44] S. Kafashi, R. Taghdimi, and G. Karimi, "Modification of nano clay systems: an approach to stabilizing drilling fluids," *Advanced Materials Research*, vol. 829, pp. 818–824, 2014.

[45] S. Kafashi, M. R. Rasaei, and G. R. Karimi, "Experimental study of nanoclay absorbents and additives' effects on modification of rheological properties of drilling fluids in porous media using glass micromodel," *Journal of Porous Media*, vol. 23, no. 6, pp. 627–639, 2020.

[46] C. Gao, M. Rivero, E. Nakagawa, and G. Sanchez, "Downhole separation technology—past, present and future," *The APPEA Journal*, vol. 47, pp. 283–292, 2007.

[47] Y. Li, B. Li, J. Xia, J. Zhang, K. Guo, and Y. Hu, "New methodology for aquifer influx status classification for single wells in a gas reservoir with aquifer support," *Journal of Natural Gas Geoscience*, vol. 1, pp. 407–411, 2016.

[48] R. G. Agarwal, D. C. Gardner, S. W. Kleinsteiber, and D. D. Fussell, "Analyzing well production data using combined-type-curve and decline-curve analysis concepts," *SPE Reservoir Evaluation & Engineering*, vol. 2, pp. 478–486, 1999.

[49] K. M. Reinicke, G. Hueni, N. Liermann, J. Oppelt, P. Reichetseder, and W. Unverhaun, "Oil and gas, 7. production," in *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH, Weinheim, Germany, 2014.

[50] R. G. Turner, M. G. Hubbard, and A. E. Dukler, "Analysis and prediction of minimum flow rate for the continuous removal of liquids from gas wells," *Journal of Petroleum Technology*, vol. 21, pp. 1475–1482, 1969.

[51] S. C. Lea, G. Landini, and A. D. Walmsley, "Ultrasonic scaler tip performance under various load conditions," *Journal of Clinical Periodontology*, vol. 30, pp. 876–881, 2003.

[52] D. P. Arcaro and Z. Bassiouni, "Technical and economic feasibility of enhanced gas recovery in the eugene island field by use of the coproduction technique," *Journal of Petroleum Technology*, vol. 39, pp. 585–590, 1987.

[53] M. Z. Reza Abdollahi, M. N. Pari, S. M. Motahhari, and M. S. Reisabadi, "Fast approach to manage depletion of naturally fractured carbonate gas reservoir with water rising issue through fracture networks," *Petroleum & Coal Journal*, vol. 63, pp. 387–398, 2021.

[54] R. J. Schilthuis, "Active oil and reservoir energy," *Transactions AIME*, vol. 118, pp. 33–52, 1936.

[55] M. Rezaee, B. Rostami, M. S. Zadeh, and M. Mojarad, "Experimental determination of optimized production rate and its upscaling analysis in strong water drive gas reservoirs," in *Proceedings of the International Petroleum Technology Conference*, Beijing, China, March 2013.

[56] P. J. Lumsden, C. J. Balgobin, D. Bodnar et al., "The kapok field-a step change for trinidad gas developments," in *Proceedings of the SPE Gas Technology Symposium*, Calgary, Canada, April 2002.

[57] R. P. Sutton, S. A. Cox, E. Glynn Williams, R. P. Stoltz, and J. V. Gilbert, "Gas well performance at subcritical rates," in *Proceedings of the SPE Production and Operations Symposium*, Oklahoma City, OK, USA, March 2003.

[58] J. A. C. Lopez, "Gas injection as a method for improved recovery in gas-condensate reservoirs with active support," in *Proceedings of the SPE International Petroleum Conference and Exhibition in Mexico*, Villahermosa, Mexico, February 2000.

[59] A. Al-Hasami, S. Ren, and B. Tohidi, "CO_2 injection for enhanced gas recovery and geo-storage: reservoir simulation and economics," in *SPE Europec/EAGE Annual Conference*, Madrid, Spain, June 2005.

[60] J. P. Aguilar-López, T. Bogaard, and H. H. Gerke, "Dual-permeability model improvements for representation of preferential flow in fractured clays," *Water Resources Research*, vol. 56, p. e2020WR027304, 2020.

[61] T. N. Libson and J. R. Henry, "Case histories: identification of and remedial action for liquid loading in gas wells-intermediate shelf gas play," *Journal of Petroleum Technology*, vol. 32, pp. 685–693, 1980.

[62] H. Hematpur, S. M. Mahmood, N. H. Nasr, and K. A. Elraies, "Foam flow in porous media: concepts, models and challenges," *Journal of Natural Gas Science and Engineering*, vol. 53, pp. 163–180, 2018.

[63] E. J. Hutlas and W. R. Granberry, "Practical approach to removing gas well liquids," *Journal of Petroleum Technology*, vol. 24, pp. 916–922, 1972.

[64] F. Ahmad and M. K. Zahoor, "A paradigm shift: higher recovery from water-drive gas reservoirs producing through wells having bigger size completion," in *Proceedings of the PAPG/SPE Pakistan Section Annual Technical Conference and Exhibition*, Islamabad, Pakistan, November 2016.

[65] W. C. Bugbee, "Artificial lift of water in gas wells," in *Proceedings of the North Africa Technical Conference and Exhibition*, Cairo, Egypt, February 2012.

[66] C. L. Brady and S. J. Morrow, "Economic assessment of artificial lift in low-pressure, tight gas sands in Ochiltree country, Texas," in *Proceedings of the SPE Mid-Continent Gas Symposium*, Amarillo, TX, USA, May 1994.

[67] G. N. Ozan, *Production System for Optimization for Submersible Pump Lifted Wells: A Case Study*, The Middle East Technical University, Ankara, Turkey, 2004.

[68] D. R. Brouwer and J. D. Jansen, "Dynamic optimization of water flooding with smart wells using optimal control theory," in *Proceedings of the European Petroleum Conference*, Aberdeen, UK, October 2002.

[69] H. Gai, "Downhole flow control optimization in the worlds 1st extended reach multilateral well at wytch farm," in *Proceedings of the SPE/IADC Drilling Conference*, Amsterdam, Netherlands, February 2001.

[70] A. Ajayi and M. Konopczynski, "A dynamic optimisation technique for simulation of multi-zone intelligent well systems in a reservoir development," in *Proceedings of the SPE Offshore Europe Oil and Gas Exhibition and Conference*, Aberdeen, UK, September 2003.

[71] L. J. Durlofsky and K. Aziz, "Optimization of smart well control," in *Proceedings of the SPE International Thermal Operations and Heavy Oil Symposium and International*

Horizontal Well Technology Conference, Calgary, Canada, November 2002.

[72] M. M. J. J. Naus, N. Dolle, and J. D. Jansen, "Optimization of commingled production using infinitely variable inflow control valves," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, Houston, TX, USA, September 2004.

[73] M. Armenta and A. K. Wojtanowicz, "Incremental recovery using dual-completed wells in gas reservoirs with bottom water drive: a feasibility study," in *Proceedings of the Canadian International Petroleum Conference*, Calgary, Canada, June 2003.

[74] M. Armenta and A. K. Wojtanowicz, "Operating dual-completed well to increase gas recovery in low productivity gas reservoirs with water production problems," in *Proceedings of the Canadian International Petroleum Conference*, Calgary, Canada, June 2004.

[75] J. R. Nichol and J. Marsh, "Downhole gas/water separation: engineering assessment and field experience," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, San Antonio, TX, USA, October 1997.

[76] M. Boussa, "Production optimization of gas wells: problem of water influx," in *Proceedings of the SPE International Thermal Operations and Heavy Oil Symposium and Western Regional Meeting*, Bakersfield, CA, USA, March 2004.

[77] J. L. Rike, "Squeeze cementing: state of the art," *Journal of Petroleum Technology*, vol. 34, pp. 37–45, 1982.

[78] H. Chaabouni, P. Enkababian, K. S. Chan, P. Cheneviere, P. Falxa, and C. Urbanzik, "Successful innovative water-shutoff operations in low-permeability gas wells," in *Proceedings of the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, March 2007.

[79] A. BenAmara, "Gas lift-past & future," in *Proceedings of the SPE Middle East Artificial Lift Conference and Exhibition*, Manama, Bahrain, November 2016.

[80] A. H. Shar, "Optimization of gas well productivity by controlling water production," in *Proceedings of the SPE/PAPG Annual Technical Conference*, Islamabad, Pakistan, November 2005.

[81] A. Zaitoun, N. Kohler, and Y. Guerrini, "Improved polyacrylamide treatments for water control in producing wells," *Journal of Petroleum Technology*, vol. 43, pp. 862–867, 1991.

[82] D. D. Sparlin and R. W. Hagen, "Controlling water in producing operations," *World Oil*, vol. 199, pp. 137–142, 1984.

[83] A. Zaitoun and N. Kohler, "Two-phase flow through porous media: effect of an adsorbed polymer layer," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, Houston, TX, USA, October 1988.

[84] S.-B. Mohsen, H. Esfandyari, A. Hashemi, A. Reza, and H. Hamed, "Numerical well test analysis of condensate dropout effects in dual-permeability model of naturally fractured gas condensate reservoirs: case studies in the south of Iran," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9916914, 10 pages, 2021.

[85] W. G. Routson, M. Neale, and J. R. Penton, "A new blocking agent for waterflood channeling authors," in *Proceedings of the Fall Meeting of the Society of Petroleum Engineers of AIME*, San Antonio, TX, USA, October 1972.

[86] A. Zaitoun and T. Pichery, "A successful polymer treatment for water coning abatement in gas storage reservoir," in *Proceedings of the SPE Annual Technical Conference and Exhibition*, New Orleans, LA, USA, September 2001.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9615463, 17 pages
https://doi.org/10.1155/2021/9615463

Research Article

Modeling and Solving Scheduling in Overloaded Situations with Weighted Partial MaxSAT

Xiaojuan Liao,[1,2] **Hui Zhang**,[2] **Miyuki Koshimura**,[3] **Rong Huang**,[4] **Wenxin Yu**,[5] **and Fagen Li**[1]

[1]*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China*
[2]*College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu, Sichuan 610059, China*
[3]*Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan*
[4]*College of Information Science and Technology, Donghua University, Shanghai 201620, China*
[5]*School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China*

Correspondence should be addressed to Hui Zhang; zhanghui18@cdut.edu.cn

Received 24 May 2021; Revised 6 July 2021; Accepted 8 July 2021; Published 16 July 2021

Academic Editor: Mohammad Yazdi

In real-time systems, where tasks have timing requirements, once the workload exceeds the system's capacity, missed due dates may cause system overload. In this situation, finding an optimal scheduling that minimizes the cumulative values of late tasks is critical in both theory and practice. Recently, formalizing scheduling problems as a class of generalized problems, such as Satisfiability Modulo Theory (SMT) and Maximum Satisfiability (MaxSAT), has been receiving immense concern. Enlightened by the high efficiency of these satisfiability-based methods, this paper formulates the single-machine scheduling problem of minimizing the total weight of late tasks as a Weighted Partial Maximum (WPM) Satisfiability problem. In the formulation, scheduling features are encoded as rigidly enforced hard clauses and the scheduling objective is treated as a set of weighted soft ones. Then an off-the-shelf WPM solver is exploited to maximize the total weight of the satisfied soft clauses, provided that all the hard clauses are satisfied. Experimental results demonstrate that, compared with the existing satisfiability-based methods, the proposed method significantly improves the efficiency of identifying the optimal schedule. Moreover, we make minor changes to apply the WPM formulation to parallel-machine scheduling, showing that the proposed method is sufficiently flexible and well scalable.

1. Introduction

Real-time systems, which are designed to handle tasks with completion due dates, play an important role in a variety of modern applications, such as robotics [1], pacemakers [2], chemical plants [3], telecommunications [4], and multimedia systems [5]. Under ideal circumstances, a real-time system completes all tasks before their due dates expire. However, in reality, the workload may exceed the system's capacity, leading to missed deadlines [6]. Such a phenomenon is called overload. A classic example is a switch in a communication network which polls its incoming links to forward packets that have arrived since its previous servicing of the link [7]. On each link, different packets may have different processing times, deadlines, and importance values. When packages flood the switch, overload happens. In this situation, designing a suitable scheduling strategy to maximize the total value of forwarded packages is critical to maintain a service's stability.

Generally, scheduling algorithms can be classified as online scheduling and offline scheduling, depending on whether tasks' information is known a priori. In online scheduling, the scheduler receives tasks that arrive over time and must schedule tasks without any knowledge of the future. On the contrary, offline scheduling algorithms aim at solving the problem optimally, provided that all data are

known beforehand [1](note that an offline algorithm does not contradict with a real-time system. An offline scheduling algorithm allows a scheduler to make decision based on the total knowledge of the problem, while a real-time system assigns each task with a specific due date). Although online scheduling is more flexible, in many situations, it is necessary to obtain optimal schedules by offline algorithms, especially in time-critical systems or for the evaluation of heuristics [8]. In this paper, we take an interest in designing an offline scheduling method on a single machine to minimize the total weight of tasks that miss their due dates.

To date, there has been an immense amount of work devoted to characterizing scheduling problems and analyzing the complexity of problems with specific characteristics. When preemption is prohibited, Michael Moore [9] presented an optimal algorithm in polynomial time to minimize the number of late tasks on a single machine, under the assumption that all tasks are released simultaneously without dependency relations. In the standard three-field notation [10], this problem is $1\|\sum U_j$. Adding weights on the criteria and setting precedence constraints may complicate the problem. Richard [11] showed that $1\|\sum w_j U_j$ is binary NP-hard but can be solved by dynamic programming in $O(n\sum p_j)$ time [12], where n is the number of tasks and p_j is the processing time of task τ_j. Garey and Johnson [13] showed that the problem of unit-time tasks subject to precedence constraints, that is, $1|\text{prec}, p_j = 1|\sum U_j$, is NP-hard. Furthermore, Lenstra and Rinnooy Kan [14] proved that, even for chain-like precedence constraints, where each task has at most one immediate predecessor and at most one immediate successor, the problem $1|\text{chains}, p_j = 1|\sum U_j$ is also NP-hard. Another prevailing common knowledge is that usually preemptive problems are not harder than their nonpreemptive counterparts. For example, the problem with release dates $1|r_j|\sum U_j$ is NP-hard [10], while the preemptive version $1|r_j, \text{pmtn}|\sum U_j$ can be solved in $O(n^3k^2)$ [15], where k is the number of distinct release dates. Further complexity results for single-machine scheduling problems are listed by [16] and the scheduling problems with the late work criteria are surveyed by [17].

Algorithms for solving scheduling are generally classified as online scheduling and offline scheduling, depending on whether the full information about tasks is known a priori or not. Online algorithms aim to return high-quality results within reasonable CPU time [18–21], whereas offline algorithms are devoted to optimally solving scheduling problems, given that all data are known beforehand. For large-scale problems that are computationally intractable, finding optimal solutions requires a significantly long computation time and heuristics are proposed to seek for suboptimal solutions within a short computation time [22–24]. Nevertheless, it is still of great significance to design optimization algorithms as a testbed for suboptimal solutions and reap huge benefit when the scheduled application is executed many times [8]. Previous attempts at finding optimal solutions to single-machine scheduling problems are mainly based on dynamic programming and branch-and-bound algorithms, with promising results. For a comprehensive survey, see the work by [25]. In the last decade, formalizing scheduling problems as a class of generalized problems, such as Mathematical Programming (MP) [26–31], Satisfiability Modulo Theory (SMT) [8, 32–34], Boolean Satisfiability (SAT) [35–37], and Partial Maximum (PM) Satisfiability [38], has received considerable attention. Motivated by the significant progress in solving these generalized problems, the formalized scheduling problem can be efficiently addressed with the corresponding solving algorithms.

As a pioneering work in satisfiability formalization, Crawford and Baker [35] first encoded scheduling problems into a SAT problem, paving the way for subsequent work [36] that solved six types of open job-shop scheduling problems. Venugopalan and Oliver [27], Liu et al. [37], and Malik et al. [8] presented optimization frameworks to address task graph scheduling with communication costs based on MIP, SAT, and SMT, respectively. Qi et al. [34] utilized task duplication strategy-based SMT formulation to mitigate the negative impact of the interprocessor communication delay in the task graph scheduling problem. Qamhan et al. [31] presented a new MILP model to schedule a set of tasks on a single-machine subject to nonzero release date, sequence-dependent setup time, and periodic maintenance. The objective of all the above formulations is to minimize the maximum completion time makespan.

To achieve the goal of minimizing the number of late jobs, Ourari et al. [26] designed a mathematical integer programming formulation for single-machine scheduling without preemption, and Hung et al. [29] developed a nonstandard MIP formulation to address the arbitrary preemptive version on parallel machines. The restricted preemptive counterpart on a single machine was solved by Cheng et al. [32], which encoded the problem as a set of first-order formulas that are tackled by an SMT solver called Z3. By running the Z3 solver repeatedly to identify the maximum number of on-time tasks, the optimal schedule could be finally determined. The SMT-based scheduling is sufficiently flexible because it handles various task properties and objectives with very few changes in adaption procedure. For example, when tasks have different importance values and the objective turns to minimizing the total weight of late tasks, only the target constraints need to be modified [33]. Later on, Wang et al. [39] enhanced Cheng's SMT formulation by removing redundant constraints and eliminating successive calls of the Z3 solver. Experiments illustrated that the updated formulation improved the efficiency by more than two orders of magnitude. Recently, Liao et al. [38] encoded the unweighted version of the same scheduling problem into Boolean propositional logic and showed that PM solvers are a competitive alternative to SMT solvers. However, we notice that when encoding scheduling features, PM formulation generates redundant variables and clauses. Such redundancy may create extra calculations and decrease the overall performance. Furthermore, PM formulation presented in [38] is incapable of handling weighted problems where the scheduling goal is to minimize the total weight of late tasks.

In this paper, we present a Weighted Partial MaxSAT (WPM) formulation to optimally solve scheduling in

overloaded situations, with the aim of minimizing the total weight value of late tasks. The minimization objective is equivalent to maximizing the total weight of the tasks meeting their due dates. Confronted with a weighted scheduling problem, we first identify scheduling features that uniquely characterize the problem, facilitating the WPM formulation in the encoding phase. Then, enlightened by the WPM characteristics that satisfy all hard clauses and maximize the total weight of the satisfied soft clauses, we recast the weighted scheduling problem as a WPM problem. Finally, in the problem-solving phase, we exploit the off-the-shelf WPM solver to satisfy all the scheduling features and maximize the total weight of tasks meeting their due dates, thus deriving the optimal schedule from the output of the WPM solver. Specifically, we make the following contributions:

(i) We extend the PM formulation in [38], which was originally designed for scheduling tasks without weights, to adapt to the weighted cases.

(ii) Having noticed that redundant Boolean variables and clauses exist in the previous formulation [38], we develop a more compact encoding to characterize the scheduling problem. Particularly, to denote a task's completion time, the number of Boolean variables generated by [38] is proportional to the number of possible preemptions of the task. In contrast, the novel compact encoding generates only one Boolean variable to represent the task's completion time, no matter how many times the task may be preempted. Theoretical analysis shows the correctness of the compact encoding and experiments demonstrate the substantial advantages over the previous PM encoding [38] and SMT formulation [39], which was enhanced from [33].

(iii) In the WPM encoding, task features are encoded by several separate rules. This means if some of the task features happen to change, only partial rules need to be modified. To demonstrate the flexibility of our formulation, we extend the current WPM encoding to adapt to parallel-machine scheduling with little modification. We believe that the proposed WPM encoding can help users readily and effectively design scheduling for practical systems with low design cost.

Confronted with a weighted scheduling problem, we first identify scheduling features that uniquely characterize the problem, facilitating the WPM formulation in the encoding phase. Then, enlightened by the WPM characteristics that satisfy all hard clauses and maximize the total weight of the satisfied soft clauses, we recast the weighted scheduling problem as a WPM problem. In the WPM formulation, tasks' features are encoded as a set of hard clauses, and the goal of completing tasks before their due dates is transformed into a set of weighted soft clauses. Finally, in the problem-solving phase, we exploit the off-the-shelf WPM solver to satisfy all the scheduling features and maximize the total weight of tasks meeting their due dates, thus deriving the optimal schedule from the output of the WPM solver.

To evaluate the performance of the proposed WPM formulation, we compare it with the state-of-the-art SMT formulation [39]. Our evaluation shows that WPM formulation has a dominant advantage over the SMT-based method in finding out the optimal schedule. We also compare WPM with the latest PM formulation [38] on a special case, where all the tasks have equal weights. We reveal redundancies in PM and discuss how they can be avoided in WPM. Experiments show that our WPM encoding is more compact and more time-efficient than PM for solving the same set of problem instances. Furthermore, we show that the presented approach is sufficiently flexible to adapt to parallel-machine scheduling problems with minor changes.

The remainder of this paper is organized as follows. The scheduling model is described in Section 2, followed by the WPM-based optimization framework in Sections 3. We make theoretical comparisons between WPM and PM formulation [38] in Section 4 and provide experimental results to show the superiority of WPM in Section 5. Section 6 exhibits how to extend the WPM formulation of single-machine scheduling to parallel-machine scheduling. Finally, we conclude this paper in Section 7.

2. Scheduling Model

We adhere to the definition of scheduling problems in previous works [33, 39]. For convenience, the notations used in the model are summarized in Table 1.

The problem involves n tasks $\Gamma = \{\tau_1, \ldots, \tau_n\}$ to be processed. All the tasks request a uniprocessor for execution when they arrive in the system. Each task $\tau_\ell \in \Gamma$ is represented by a 4-tuple $\tau_\ell = (r_\ell, c_\ell, d_\ell, w_\ell)$, where r_ℓ, c_ℓ, d_ℓ, and w_ℓ are all nonnegative integers [2](assuming all parameters to be integers does not make the problem less general since real numbers can be scaled to integers with a few orders of magnitude) representing the release date, the execution time, the due date, and the weight of τ_ℓ, respectively. Naturally, $r_\ell + c_\ell \le d_\ell$. Weight w_ℓ reflects the importance of task τ_ℓ. The larger w_ℓ is, the more important τ_ℓ is. Given due date d_ℓ, if task τ_ℓ is completed at or before d_ℓ, τ_ℓ is on time and weight w_ℓ is obtained by the system. Otherwise, τ_ℓ is late and worthless to the system.

To allow preemption, which indicates that a running task may be interrupted and resumed later, each task $\tau_\ell \in \Gamma$ is split into q_ℓ nonpreemptive subtasks (fragments) and is defined as a chain of indivisible fragments $\langle f_1^\ell, \ldots, f_{q_\ell}^\ell \rangle$. Symbol c_i^ℓ stands for the required execution time of f_i^ℓ. Clearly, $\sum_{i=1}^{q_\ell} c_i^\ell = c_\ell$. For $2 \le i \le q_\ell$, f_i^ℓ can only start to run after f_{i-1}^ℓ is completed.

In practical systems, tasks usually have dependency relations. For example, if task τ_ℓ requires the computed result of τ_k, τ_ℓ cannot start until τ_k is finished. Such a dependency relation between tasks is written as $\tau_k \prec \tau_\ell$, where τ_k is the immediate predecessor of τ_ℓ, and τ_ℓ is the immediate successor of τ_k. Obviously, if $\tau_k \prec \tau_\ell$, the constraint $r_k + c_k \le d_\ell - c_\ell$ should be satisfied; otherwise, τ_ℓ can never

Table 1: Notations and descriptions in scheduling model.

Notation	Description
Γ	Finite set of real-time tasks
n	Number of real-time tasks. $n = \lvert\Gamma\rvert$
τ_ℓ	Task in Γ, where ℓ is its index
r_ℓ	Release date of τ_ℓ
c_ℓ	Execution time of τ_ℓ
d_ℓ	Due date of τ_ℓ
f_i^ℓ	i^{th} fragment of τ_ℓ
q_ℓ	Number of fragments in τ_ℓ
c_i^ℓ	Execution time of f_i^ℓ
$\tau_k \prec \tau_\ell$	τ_ℓ succeeds τ_k
R_{dp}	Set of task pairs that have dependency relations

be completed no matter when its predecessor finishes. The set of task pairs that have dependency relations over Γ is denoted by $R_{\text{dp}} = \{(\tau_k, \tau_\ell)\colon \tau_k, \tau_\ell \in \Gamma, \tau_k \prec \tau_\ell\}$.

A system is defined as overloaded if no scheduling algorithm can meet the due dates of all the tasks that have been submitted to it. This paper focuses on designing an exact method to tackle overloaded single-machine scheduling problems. The scheduling objective is to maximize the total weight of the on-time tasks. In particular, if all tasks have equal weights, then a schedule that maximizes the total weight will be one that maximizes the number of on-time tasks.

3. WPM-Based Optimization Framework

In this section, we provide the WPM formulation for solving the task scheduling problem on a single machine. The overview of the WPM formulation is illustrated in Figure 1.

Feature preprocessing identifies the scheduling problem with two types of constraints, that is, constraints on scheduling features and those on objective. Given a set of tasks $\Gamma = \{\tau_1, \ldots, \tau_n\}$, scheduling features uniquely characterize the problem over Γ and the objective is to seek for a schedule that completes all tasks on time subject to constraints on the scheduling features. In overloaded situations, making all the tasks complete by their due dates is impossible, and the scheduling problem is then treated as an optimization problem, which aims to maximize the total weight of tasks that are completed by their due dates. After feature preprocessing, WPM encoding can be implemented separately on the scheduling features and objective. Specifically, scheduling features, which are intrinsic and determined when tasks are released, are encoded into a set of hard clauses that should be satisfied without exception. On the other hand, the objective in overloaded situations is encoded as a set of weighted soft clauses that are allowed to be unsatisfied. By conjunction of hard clauses with weighted soft clauses, the problem turns to a WPM problem that can be addressed by any off-the-shelf WPM solver. In the problem-solving phase, a WPM solver tries to satisfy all the hard clauses and maximize the total weight of satisfied soft clauses. The output of the WPM solver includes the assignment of all Boolean variables, from which the optimal schedule can be derived.

Section 3.1 shows how to identify fragments' critical time points to characterize the scheduling problem, paving the way for the WPM encoding described in Sections 3.2. Section 3.3 exhibits an example to show how the scheduling problem is addressed with the presented WPM encoding. Details on the problem-solving phase are also exhibited in Section 3.3.

3.1. Feature Preprocessing. Given a task set $\Gamma = \{\tau_1, \ldots, \tau_n\}$, where each task $\tau_\ell \in \Gamma$ is characterized by $(r_\ell, c_\ell, d_\ell, w_\ell)$, $\forall \tau_\ell \in \Gamma$, τ_ℓ cannot start before its release date r_ℓ. Thus, the possible Earliest Start Time (EST) of τ_ℓ, denoted by EST_ℓ, is equal to its release date r_ℓ, and the Earliest Completion Time (ECT) of τ_ℓ, denoted by ECT_ℓ, is $\text{EST}_i + c_i$. $\forall (\tau_k, \tau_\ell) \in R_{\text{dp}}$, EST_ℓ is jointly restricted by r_ℓ and ECT_k. Specifically, if $\text{ECT}_k \le r_\ell$, then $\text{EST}_\ell = r_\ell$; otherwise, $\text{EST}_\ell = \text{ECT}_k$, suggesting that τ_ℓ should wait until τ_k is finished. Typically, let P_ℓ be the set of predecessors of τ_ℓ; then $\text{EST}_\ell = \max\{r_\ell, \max_{\tau_k \in P_\ell}\{\text{ECT}_k\}\}$.

After EST_ℓ for each τ_ℓ in Γ has been determined, EST and ECT of each fragment in τ_ℓ can be calculated as follows:

(i) Earliest Start Time (EST) of f_i^ℓ is denoted by EST_i^ℓ, where $\text{EST}_i^\ell = \text{EST}_\ell + \sum_{u=1}^{i-1} c_u^\ell$. $\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, f_i^ℓ should start at or after EST_i^ℓ, indicating that f_i^ℓ cannot be started before all its previous fragments are finished. Typically, $\text{EST}_1^\ell = \text{EST}_\ell$.

(ii) Earliest Completion Time (ECT) of f_i^ℓ is denoted by ECT_i^ℓ, where $\text{ECT}_i^\ell = \text{EST}_\ell + \sum_{u=1}^{i} c_u^\ell$. No fragment f_i^ℓ can be completed before ECT_i^ℓ. Particularly, $\text{ECT}_i^\ell = \text{EST}_i^\ell + c_i^\ell$.

In addition to EST and ECT, $\forall \tau_\ell \in \Gamma$, each fragment $f_i^\ell \in \tau_\ell$ is characterized by the two following types of time:

(i) Latest Start Time (LST) of f_i^ℓ is denoted by LST_i^ℓ, where $\text{LST}_i^\ell = d_\ell - \sum_{u=i}^{q_\ell} c_u^\ell$. If f_i^ℓ fails to start before LST_i^ℓ, τ_ℓ cannot end by its due date d_ℓ, and thus this task becomes worthless to the system.

(ii) Latest Completion Time (LCT) of f_i^ℓ is denoted by LCT_i^ℓ, where $\text{LCT}_i^\ell = d_\ell - \sum_{u=i+1}^{q_\ell} c_u^\ell$. Particularly, $\text{LCT}_i^\ell = \text{LST}_i^\ell + c_i^\ell$.

Note that, in the above four critical time points, ECT is unrelated to our WPM encoding. Nevertheless, it is indispensable in the previous work [38], which will be discussed in Section 4..

3.2. WPM Formulation. This section introduces how to encode the scheduling problem as a WPM problem. A WPM instance consists of a number of clauses that need to be managed by the WPM solver. To formulate all the necessary constraints that characterize the scheduling model, we introduce the three following Boolean variables:

(i) $\text{sa}_{i,t}^\ell$, which is true if f_i^ℓ starts at time t or later

(ii) $\text{pr}_{i,j}^{\ell,k}$, which is true if f_i^ℓ precedes f_j^k

(iii) eb_ℓ, which is true if τ_ℓ ends by its due date d_ℓ

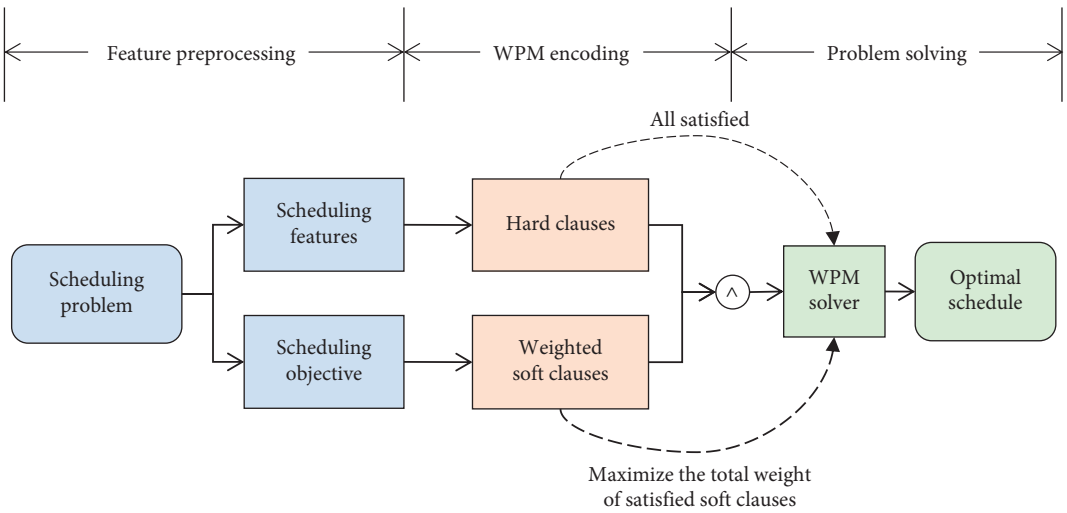


Figure 1: Overview of the WPM-based scheduling method.

Based on the generated Boolean variables, the WPM encoding can be implemented. In what follows, several rules are presented to encode the features of fragments into a set of hard clauses. The main encoding is derived and enhanced from previous work [35, 36, 38]. Logical implication $a \longrightarrow b$ is equivalent to $a \vee b$ in classical logic.

(i) (C1) $\forall \tau_\ell \in \Gamma$, the first fragment of τ_ℓ, that is, f_1^ℓ, starts at or after EST_1^ℓ:

$$\mathrm{sa}_{1,\mathrm{EST}_1^\ell}^\ell. \tag{1}$$

(ii) (C2) $\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell, f_{i+1}^\ell \in \tau_\ell$, f_i^ℓ precedes f_{i+1}^ℓ:

$$\mathrm{pr}_{i,i+1}^{\ell,\ell}. \tag{2}$$

(iii) (C3) $\forall \tau_k, \tau_\ell \in \Gamma$, $\forall f_i^k \in \tau_k$, and $\forall f_j^\ell \in \tau_\ell$, if $k \neq l$, $\tau_k \prec \tau_\ell$, $\tau_\ell \prec \tau_k$, $\mathrm{EST}_i^k < \mathrm{LCT}_j^\ell$ and $\mathrm{EST}_j^\ell < \mathrm{LCT}_i^k$, and then f_i^k and f_j^ℓ may require the processor at the same time. In this condition, f_i^k precedes f_j^ℓ or f_j^ℓ precedes f_i^k:

$$\mathrm{pr}_{i,j}^{k,\ell} \vee \mathrm{pr}_{j,i}^{\ell,k}. \tag{3}$$

(iv) (C4) $\forall (\tau_k, \tau_\ell) \in R_{dp}$, if τ_k fails to be completed by its due date, then τ_ℓ cannot even start at LST_1^ℓ:

$$\mathrm{eb}_k \longrightarrow \mathrm{sa}_{1,\mathrm{LST}_1^\ell+1}^\ell. \tag{4}$$

(v) Furthermore, if $d_k > \mathrm{EST}_1^\ell$, the first fragment of τ_ℓ cannot start until the last fragment of τ_k finishes. That is, if $d_k > \mathrm{EST}_1^\ell$, then the constraint that $f_{q_k}^k$ precedes f_1^ℓ is enforced:

$$\mathrm{pr}_{q_k,1}^{k,\ell}. \tag{5}$$

(vi) (C5)$\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, if f_i^ℓ starts at or after time t, then it starts at or after time $t-1$:

$$\mathrm{sa}_{i,t}^\ell \longrightarrow \mathrm{sa}_{i,t-1}^\ell, \quad \left(\mathrm{EST}_i^\ell + 1 \leq t \leq \mathrm{LST}_i^\ell + 1\right). \tag{6}$$

(vii) (C6) $\forall \tau_\ell \in \Gamma$, if τ_ℓ ends before its due date d_ℓ, then last fragment $f_{q_\ell}^\ell$ must start at or before time $\mathrm{LST}_{q_\ell}^\ell$. In other words, $f_{q_\ell}^\ell$ cannot start at or after time $\mathrm{LST}_{q_\ell}^\ell + 1$:

$$\mathrm{eb}_\ell \longrightarrow \mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell. \tag{7}$$

(viii) (C7) $\forall \tau_k, \tau_\ell \in \Gamma$, $\forall f_i^k \in \tau_k$, and $\forall f_j^\ell \in \tau_\ell$, if f_i^k starts at or after time t and f_j^ℓ follows f_i^k, then f_j^ℓ cannot start until f_i^k is finished. That is, for each $\mathrm{pr}_{i,j}^{k,\ell}$ asserted by (C2) ~ (C4), one clause is generated:

$$\mathrm{sa}_{i,t}^k \wedge \mathrm{pr}_{i,j}^{k,\ell} \longrightarrow \mathrm{sa}_{j,t'}^\ell, \tag{8}$$

where t varies in $[\mathrm{EST}_i^k, \mathrm{LST}_i^k + 1]$ and

$$t' = \begin{cases} \mathrm{LST}_j^\ell + 1, & \text{if } t + c_i^k > \mathrm{LST}_j^\ell, \\ \mathrm{EST}_j^\ell, & \text{if } t + c_i^k < \mathrm{EST}_j^\ell, \\ t + c_i^k, & \text{Otherwise.} \end{cases} \tag{9}$$

This formula reveals the following facts: First, if f_i^k ends after LST_j^ℓ (i.e., $t + c_i^k > \mathrm{LST}_j^\ell$), then f_j^ℓ cannot start at or before LST_j^ℓ. Second, if f_i^k finishes before EST_j^ℓ (i.e., $t + c_i^k < \mathrm{EST}_j^\ell$), then f_j^ℓ starts at or after EST_j^ℓ. Otherwise, f_j^ℓ must start at or after time f_i^k finishes, that is, $t + c_i^k$.

Up to this point, we have encoded the scheduling features as Boolean formulas that can be converted to a set of clauses. Since the scheduling features are intrinsic properties inherent in the tasks and their fragments, such clauses are specified as hard, indicating that all of them must absolutely be satisfied. For convenience, we refer to the set of hard clauses introduced in (C1) ~ (C7) as $\mathscr{C}$.

A task is said to be on time if and only if it is completed before its due date. Thus, the scheduling objective can be directly encoded by the following rule:

(O) Maximizing the sum of the weights of on-time tasks:

$$(\mathrm{eb}_\ell, w_\ell), \quad (1 \le \ell \le n). \tag{10}$$

Equation (10) indicates that if clause eb_ℓ is satisfied (i.e., evaluating to true), then weight w_ℓ is gained; otherwise, the gain is zero. To simplify the discussion, we introduce $\mathcal{O}$ to denote the set of clauses in equation (10). In an overloaded system, not all tasks can be completed before their due dates. To handle such situations, we declare the clauses in $\mathcal{O}$ to be soft, indicating that completing all the tasks by their due dates is a soft constraint. Conjunct with $\mathcal{C}$, the problem is then $\{\mathcal{C}, \mathcal{O}\}$. This leads to a WPM problem, which tries to find an assignment of variables to satisfy all the hard clauses in $\mathcal{C}$ and to maximize the sum of the weights of the satisfied soft clauses in $\mathcal{O}$, that is, to maximize $\sum_{\ell=1}^{n} w_\ell \cdot \mathrm{eb}_\ell$.

3.3. A Pedagogical Example. Let us consider a simple scheduling problem to describe how WPM formulation works. As shown in Figure 2(a), there are a set of real-time tasks $\Gamma = \{\tau_1, \tau_2, \tau_3\}$, where both τ_1 and τ_3 rely on the computed result of τ_2, represented as $\tau_2 \prec \tau_1$ and $\tau_2 \prec \tau_3$; that is, $R_{\mathrm{dp}} = \{(\tau_2, \tau_1), (\tau_2, \tau_3)\}$. The release dates, the execution times, the due dates, and the weights of these tasks are, respectively, defined as $\tau_1 = (0, 2, 3, 1)$, $\tau_2 = (0, 1, 1, 2)$, and $\tau_3 = (0, 1, 2, 3)$. Suppose that τ_1 has two fragments, that is, $\langle f_1^1, f_2^1 \rangle$. τ_2 and τ_3 each have one, denoted by f_1^2 and f_1^3, respectively. The execution time of each fragment is 1.

The refined problem exhibition and the critical time points after feature preprocessing are summarized in Figure 2(b) and Table 2, respectively. Since ECT has nothing to do with the WPM encoding, we omit this entry in Table 2 for conciseness.

The MaxSAT formulation applied to the scheduling problem is shown in Figure 3. Constraint (C1) states that each task starts at or after its EST. Constraints (C2) and (C3) work together to specify the execution sequence of the fragments. (C2) forces all the fragments in a single task to be executed sequentially, and (C3) guarantees no overlap of the execution times of any two fragments in different tasks. In the three tasks, only τ_1 has more than one fragment, and hence constraint (C2) only applies to τ_1, ensuring that f_1^1 precedes f_2^1. Constraint (C3) applies to pairs of fragments f_i^k and $f_j^\ell (k \neq l)$ satisfying $\mathrm{EST}_i^k < \mathrm{LCT}_j^\ell$ and $\mathrm{EST}_j^\ell < \mathrm{LCT}_i^k$. If both conditions are met, then f_i^k and f_j^ℓ may simultaneously occupy the processor, and thus we need to decide in what order to execute them. Constraint (C3) tackles this ordering dilemma, which states that either one can precede the other. Consider f_1^1 and f_1^3. As seen in Table 2, $\mathrm{EST}_1^1 < \mathrm{LCT}_1^3$ and $\mathrm{EST}_1^3 < \mathrm{LCT}_1^1$; hence, we must explicitly specify that f_1^1 precedes f_1^3 or f_1^3 precedes f_1^1; otherwise, the execution time of these two fragments may overlap. Constraint (C4) applies to a situation where tasks have dependency relations. In the example, τ_2 is the predecessor of both τ_1 and τ_3; thus we should give up processing τ_1 and τ_3 if τ_2 misses its due date. In addition, given $d_2 = \mathrm{EST}_1^1$ and $d_2 = \mathrm{EST}_1^3$, there is no need to explicitly specify the execution sequence of the predecessor and successors since τ_2 must have finished before τ_1 and τ_3 start. Constraints (C5) ~ (C7) are partially extracted from a collection of coherence conditions [35] on the introduced variables for all the fragments of all the tasks. Finally, constraint (O) gives the problem's objective, that is, completing the last fragment of each task by its due date.

All the clauses are conjunct with $\wedge$ to form a WPM problem in CNF, where clauses (C1) ~ (C7) are declared hard and those in (O) are soft. Then the CNF formula is input to a WPM solver. The solver's output includes the maximum sum of the weights of the satisfied soft clauses as well as the corresponding assignment of all the Boolean variables.

In the exemplified problem, the assigned truth values of all the Boolean variables are listed in Figure 4, from which the exact start time of each fragment can be derived. Consider f_1^2. $\mathrm{sa}_{1,0}^2 = 1$ and $\mathrm{sa}_{1,1}^2 = 0$ indicate that f_1^2 starts at or after time 0, but it does not start at or after time 1. Then we can readily determine that f_1^2 starts at time 0. A similar reference can be made on f_1^3, which starts at time 1. Both τ_2 and τ_3 can be completed by their due dates. This is ensured by $\mathrm{eb}_2 = 1$ and $\mathrm{eb}_3 = 1$. Now, consider f_1^1. Figure 4 shows that $\mathrm{sa}_{1,1}^1 = 1$ and $\mathrm{sa}_{1,2}^1 = 1$. Since $\mathrm{LST}_1^1 = 1$, that $\mathrm{sa}_{1,2}^1 = 1$ means that f_1^1 cannot start at or before LST_1^1. Similarly, as indicated by the fact that $\mathrm{sa}_{2,3}^1 = 1$, f_2^1 cannot start at or before LST_2^1 either. Thus, we infer that τ_1 cannot finish on time. This is confirmed by $\mathrm{eb}_1 = 0$. As a result, the maximized total weight of the completed on-time tasks is $2 + 3 = 5$, achieved by completing τ_2 and τ_3 by their due dates.

4. Theoretical Discussion

Recall that, in the scheduling model described in Section 2, each task τ_ℓ is represented as a 4-tuple $\tau_\ell = (r_\ell, c_\ell, d_\ell, w_\ell)$. Consider a special case where the weights of all the tasks are equal to a constant w, that is, $\forall \tau_\ell \in \Gamma$, $w_\ell = w$. According to Section 3.2, to encode this problem, a set of weighted soft clauses $\mathcal{O} = \{(\mathrm{eb}_\ell, w)\}$ are introduced, where $1 \le \ell \le n$. Conjunct with hard clauses $\mathcal{C}$, problem $\{\mathcal{C}, \mathcal{O}\}$ becomes a special WPM problem with equal weights, which tries to find an assignment of variables to satisfy all hard clauses in $\mathcal{C}$ and to maximize $w \cdot \sum_{\ell=1}^{n} \mathrm{eb}_\ell$. Obviously, the essence of maximizing $w \cdot \sum_{\ell=1}^{n} \mathrm{eb}_\ell$ is equivalent to maximizing $\sum_{\ell=1}^{n} \mathrm{eb}_\ell$. This is a PM problem encoded by [38], where each task τ_ℓ is modeled as a 3-tuple $\tau_\ell = (r_\ell, c_\ell, d_\ell)$, and the scheduling goal is to maximize the total number of on-time tasks. In this section, we make theoretical comparisons between PM-based [38] and our WPM-based optimization frameworks when all the tasks have equal weights.

4.1. Similarities of PM and WPM Formulations. Essentially, the purpose of solving the scheduling problem is to determine the start time of each fragment to construct an optimal scheduling solution. To describe such crucial information, $\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, a set of Boolean variables $\mathrm{sa}_{i,t}^\ell (\mathrm{EST}_i^\ell \le t \le \mathrm{LST}_i^\ell + 1)$ are introduced to indicate whether f_i^ℓ starts at or after time t. As long as the truth values of sa variables are all assigned, we can readily determine the exact start time of each fragment. To be specific, if we find a certain time $t\prime \in [\mathrm{EST}_i^\ell, \mathrm{LST}_i^\ell]$ such that $\mathrm{sa}_{i,t}^\ell = 1$ for $t \le t'$ and $\mathrm{sa}_{i,t}^\ell =$

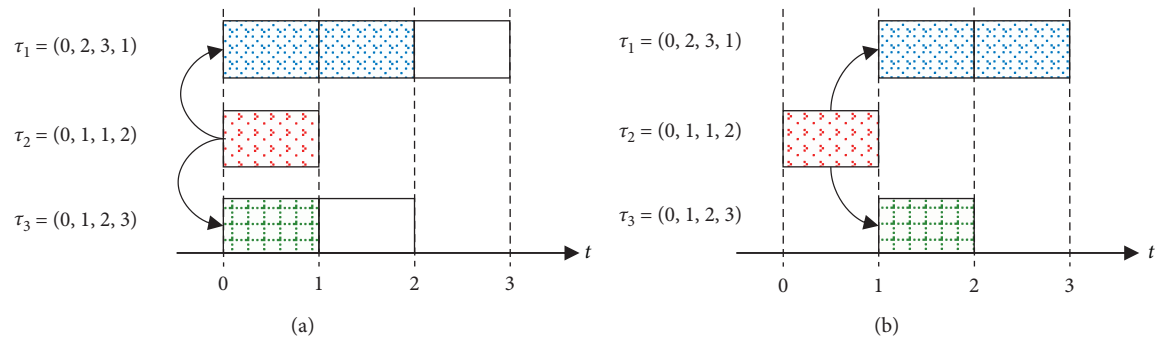


Figure 2: A simple scheduling example. (a) Original problem. (b) Problem after feature preprocessing.

Table 2: Critical time points of each fragment in three tasks.

Task	Fragment	EST	LST	LCT
τ_1	f_1^1	1	1	2
	f_2^1	2	2	3
τ_2	f_1^2	0	0	1
τ_3	f_1^3	1	1	2

$$
\begin{array}{llll}
\text{(C1)} & sa_{1,1}^1 & \wedge & \\
\wedge & sa_{1,0}^2 & \text{(C5)} & \neg sa_{1,2}^1 \vee sa_{1,1}^1 \\
\wedge & sa_{1,1}^3 & \wedge & \neg sa_{2,3}^1 \vee sa_{2,2}^1 \\
\wedge & & \wedge & \neg sa_{1,1}^2 \vee sa_{1,0}^2 \\
\text{(C2)} & pr_{1,2}^{1,1} & \wedge & \neg sa_{1,2}^3 \vee sa_{1,1}^3 \\
\wedge & & \wedge & \\
\text{(C3)} & pr_{1,1}^{1,3} \vee pr_{1,1}^{3,1} & \text{(C6)} & \neg sa_{2,3}^1 \vee \neg eb_1 \\
\wedge & & \wedge & \neg sa_{1,1}^2 \vee \neg eb_2 \\
\text{(C4)} & eb_2 \vee sa_{1,2}^1 & \wedge & \neg sa_{1,2}^3 \vee \neg eb_3 \\
\wedge & eb_2 \vee sa_{1,2}^3 & \wedge &
\end{array}
\qquad
\begin{array}{ll}
\text{(C7)} & \neg sa_{1,1}^1 \vee \neg pr_{1,2}^{1,1} \vee sa_{2,2}^1 \\
\wedge & \neg sa_{1,2}^1 \vee \neg pr_{1,2}^{1,1} \vee sa_{2,3}^1 \\
\wedge & \neg sa_{1,1}^1 \vee \neg pr_{1,1}^{1,3} \vee sa_{1,2}^3 \\
\wedge & \neg sa_{1,2}^1 \vee \neg pr_{1,1}^{1,3} \vee sa_{1,2}^3 \\
\wedge & \neg sa_{1,1}^3 \vee \neg pr_{1,1}^{3,1} \vee sa_{1,2}^1 \\
\wedge & \neg sa_{1,2}^3 \vee \neg pr_{1,1}^{3,1} \vee sa_{1,2}^1 \\
\wedge & \\
\text{(O)} & (eb_1, 1) \\
\wedge & (eb_2, 2) \\
\wedge & (eb_3, 3)
\end{array}
$$

Figure 3: WPM formulation for exemplified problem.

$$
\begin{array}{llll}
eb_1 = 0 & sa_{1,1}^1 = 1 & sa_{1,0}^2 = 1 & pr_{1,2}^{1,1} = 1 \\
eb_2 = 1 & sa_{1,2}^1 = 1 & sa_{1,1}^2 = 0 & pr_{1,1}^{1,3} = 0 \\
eb_3 = 1 & sa_{2,2}^1 = 1 & sa_{1,1}^3 = 1 & pr_{1,1}^{3,1} = 1 \\
 & sa_{2,3}^1 = 1 & sa_{1,2}^3 = 0 &
\end{array}
$$

Figure 4: Assignment of all Boolean variables in the exemplified problem.

0 for $t > t'$, then we can determine that the start time of f_i^ℓ is $t\prime$. If $\text{sa}_{i,t}^\ell$ is constantly true when t varies in $[\text{EST}_i^\ell, \text{LST}_i^\ell + 1]$, we can conclude that τ_ℓ is late. Note that it is impossible to designate $\text{sa}_{i,t}^\ell$ as true while making $\text{sa}_{i,t-1}^\ell$ false, since if f_i^ℓ starts at or after time t, f_i^ℓ must start at or after $t - 1$. This constraint is interpreted by rule (C5) in our encoding. In addition, $\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, the start time of f_i^ℓ should be at or after EST_i^ℓ. Compared with PM, WPM encodes this constraint in a more concise way by (C1). We postpone the discussion later in Section 4.2.

Aside from the inherent properties of a fragment's start time specified in (C1) and (C5), whether a fragment can start at a certain time is constrained by two other factors. The first is the execution order of the fragments, which should be explicitly pointed out in the following two situations: if two fragments belong to the same task, then they should be executed sequentially; if two fragments from different tasks may simultaneously occupy the processor, then either one can precede the other. The execution order of two fragments, f_i^k and f_j^ℓ, is characterized by Boolean variable $\text{pr}_{i,j}^{k,\ell}$, and the constraints specified in the above two cases are interpreted by rules (C2) and (C3), respectively. If the execution order of the two fragments is determined, the relation of their start times should be specified. To be specific, if f_i^k precedes f_j^ℓ, then f_j^ℓ must start after f_i^k finishes. This constraint is guaranteed by rule (C7). Up to this point, all the Boolean variables and clauses introduced by PM and WPM are identical. That is, rules (C2), (C3), (C5), and (C7) are all consistent with those introduced in PM.

4.2. WPM Improvement. As mentioned in Section 4.1, $\forall\tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, the start time of f_i^ℓ cannot be earlier than EST_i^ℓ. PM interprets this constraint by introducing one hard clause $\mathrm{sa}_{i,\mathrm{EST}_i^\ell}^\ell$ for $1 \le i \le q_\ell$; thus, a total of q_ℓ clauses are generated for τ_ℓ. In comparison, as described by (C1) in Section 3.2, WPM generates only one clause $\mathrm{sa}_{1,\mathrm{EST}_1^\ell}^\ell$ for each task $\tau_\ell \in \Gamma$ no matter how many fragments τ_ℓ contains. By combining (C1) with constraints (C2) and (C7), we can easily obtain $\mathrm{sa}_{i,\mathrm{EST}_i^\ell}^\ell = 1$ for $1 \le i \le q_\ell$, which is the same as that declared by PM. Therefore, the updated rule (C1) reduces the number of generated clauses while maintaining the correctness of the encoding.

For tasks with dependency relations, their execution sequence may be explicitly or implicitly specified. In particular, $\forall(\tau_k, \tau_\ell) \in R_{\mathrm{dp}}$, PM indistinguishably generates $\mathrm{pr}_{q_k,1}^{k,\ell}$ for the dependency relation $\tau_k \prec \tau_\ell$. In fact, if the due date of τ_k is no later than the EST of τ_ℓ, that is, $d_k \le \mathrm{EST}_1^\ell$, τ_ℓ naturally starts after τ_k is completed; thus the constraint that is explicitly imposed on the execution sequence is unnecessary and can be omitted safely. WPM formulation refined the encoding in (C4) by introducing $\mathrm{pr}_{q_k,1}^{k,\ell}$ to interpret the constraint that τ_k precedes τ_ℓ only when $d_k > \mathrm{EST}_1^\ell$; thus, redundant variables and clauses are eliminated.

Another factor that affects a fragment's start time is the completion time of a task. In general, if a task is expected to be completed by a particular time, then each fragment of the task should not start later than its LST. The relation of completion time and LST is encoded by PM and WPM in different ways. In PM, for each fragment $f_i^\ell (1 \le \ell \le n, 1 \le i \le q_\ell)$, a set of Boolean variables $\mathrm{eb}_{i,t}^\ell (\mathrm{ECT}_i^\ell - 1 \le t \le \mathrm{LCT}_i^\ell)$ are introduced to indicate whether f_i^ℓ finishes at or before time t. Then a set of hard clauses are generated to restrict the values of the eb variables.

(i) (H1) If f_i^ℓ ends by t, then it ends by time $t+1$:

$$\mathrm{eb}_{i,t}^\ell \longrightarrow \mathrm{eb}_{i,t+1}^\ell, \quad \left(\mathrm{ECT}_i^\ell - 1 \le t \le \mathrm{LCT}_i^\ell - 1\right), \tag{11}$$

(ii) (H2) if τ_ℓ is completed by its due date, then, $\forall f_i^\ell \in \tau_\ell$, f_i^ℓ should finish by LCT_i^ℓ:

$$\mathrm{eb}_{q_\ell,d_\ell}^\ell \longrightarrow \mathrm{eb}_{i,\mathrm{LCT}_i^\ell}^\ell, \quad \left(1 \le i \le q_\ell - 1\right), \tag{12}$$

(iii) (H3) if f_i^ℓ starts at or after time t, then it cannot end before time $t + c_i^\ell - 1$:

$$\mathrm{sa}_{i,t}^\ell \longrightarrow \mathrm{eb}_{i,t+c_i^\ell-1}^\ell, \quad \left(\mathrm{EST}_i^\ell \le t \le \mathrm{LST}_i^\ell + 1\right), \tag{13}$$

In fact, not all the variables and clauses in (H1) ~ (H3) are requisite. Given a task, instead of generating variables to characterize each fragment's completion time, it suffices to create only one variable to declare the task's due date and introduce one clause to clarify how the task's due date both restricts and is restricted by its start time. In particular, $\forall\tau_\ell \in \Gamma$, WPM introduces Boolean variable eb_ℓ to describe whether τ_ℓ can finish by its due date. Then a constraint is set to correlate the due date and the start time of τ_ℓ as follows: if τ_ℓ ends by d_ℓ, then $f_{q_\ell}^\ell$ must start at or before $\mathrm{LST}_{q_\ell}^\ell$. In other words, if $f_{q_\ell}^\ell$ fails to start at or before $\mathrm{LST}_{q_\ell}^\ell$, then it cannot end by d_ℓ. This constraint is encoded by just one clause described in rule (C6), as declared in Section 3.2:

$$\mathrm{eb}_\ell \longrightarrow \mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell. \tag{14}$$

Theorem 1. *Given task $\tau_\ell \in \Gamma$, if $\mathrm{eb}_{q_\ell,d_\ell}^\ell$ in PM (resp., eb_ℓ in WPM) is evaluated as true, then $\mathrm{sa}_{i,LST_i^\ell+1}^\ell$ for $1 \le i \le q_\ell$ in both PM and WPM are false.*

Proof. In PM, according to (H2), $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 1$ makes $\mathrm{eb}_{i,LCT_i^\ell}^\ell = 1$ for $1 \le i \le q_\ell - 1$. This indicates that, $\forall f_i^\ell \in \tau_\ell$, f_i^ℓ should finish by LCT_i^ℓ. Combined with (H3), it can be inferred that $\mathrm{sa}_{i,\mathrm{LST}_{i+1}^\ell}^\ell = 0$ for $1 \le i \le q_\ell$.

In WPM, that $\mathrm{eb}_\ell = 1$ leads to $\mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell = 0$ by (C6). In addition, based on (C2) and (C7), it is clear that $\mathrm{sa}_{q_\ell-1,\mathrm{LST}_{q_\ell-1}^\ell+1}^\ell \longrightarrow \mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell$ holds. Under the condition where $\mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell = 0$, it can be immediately inferred that $\mathrm{sa}_{q_\ell-1,\mathrm{LST}_{q_\ell-1}^\ell+1}^\ell = 0$. By repeatedly substituting (C2) into (C7), we finally obtain $\mathrm{sa}_{i,\mathrm{LST}_{i+1}^\ell}^\ell = 0$ for $1 \le i \le q_\ell$. □

Theorem 2. *Given task $\tau_\ell \in \Gamma$, if $\exists f_i^\ell \in \tau_\ell$ that makes $\mathrm{sa}_{i,LST_i^\ell+1}^\ell$ be evaluated as true, then $\mathrm{eb}_{q_\ell,d_\ell}^\ell$ in PM (resp., eb_ℓ in WPM) is false.*

Proof. In PM, by (H3), $\mathrm{sa}_{i,\mathrm{LST}_i^\ell+1}^\ell = 1$ makes $\mathrm{eb}_{i,\mathrm{LCT}_i^\ell}^\ell = 0$. Combined with (H2), it is obvious that $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 0$.

In WPM, we discuss the two following cases:

(1) If critical f_i^ℓ, which makes $\mathrm{sa}_{i,\mathrm{LST}_i^\ell+1}^\ell = 1$, is the last fragment in τ_ℓ, that is, $i = q_\ell$, it is obvious that $\mathrm{eb}_\ell = 0$ by (C6).

(2) If $1 \le i \le q_\ell - 1$, according to (C2) and (C7), we have $\mathrm{sa}_{i,\mathrm{LST}_i^\ell+1}^\ell \longrightarrow \mathrm{sa}_{i+1,\mathrm{LST}_{i+1}^\ell+1}^\ell$. Since $\mathrm{sa}_{i,\mathrm{LST}_i^\ell+1}^\ell = 1$, it is clear that $\mathrm{sa}_{i+1,\mathrm{LST}_{i+1}^\ell+1}^\ell = 1$. The step of substituting (C2) into (C7) can be repeated until $i + 1 = q_\ell$. Finally, we have $\mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell = 1$. Then, by (C6), $\mathrm{eb}_\ell = 0$ is obtained.

Theorem 1 shows that if $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 1$ in PM (resp., $\mathrm{eb}_\ell = 1$), both PM and WPM guarantee that $\mathrm{sa}_{i,\mathrm{LST}_{i+1}^\ell}^\ell = 0$ for $1 \le i \le q_\ell$. Theorem 2 shows that when there exists a fragment $f_i^\ell \in \tau_\ell$ that leads to $\mathrm{sa}_{i,\mathrm{LST}_i^\ell+1}^\ell = 1$, both PM and WPM come to the same conclusion that τ_ℓ cannot finish by τ_ℓ, pointed out by $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 0$ and $\mathrm{eb}_\ell = 0$ in PM and WPM, respectively. Therefore, the eb variables introduced in PM and WPM have the same impact on the *sa* variables and vice versa. Note that, $\forall f_{q_\ell}^\ell \in \tau_\ell$, if $\mathrm{sa}_{q_\ell,\mathrm{LST}_{q_\ell}^\ell+1}^\ell = 0$, a WPM solver always prefers $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 1$ in PM (resp., $\mathrm{eb}_\ell = 1$ in WPM) to satisfy soft clause $\mathrm{eb}_{q_\ell,d_\ell}^\ell$ (resp., (eb_ℓ, w)). In other words, a WPM solver never prefers $\mathrm{eb}_{q_\ell,d_\ell}^\ell = 0$ in PM (resp., $\mathrm{eb}_\ell = 0$ in WPM) as long as all the hard clauses input to the solver are satisfied. □

Example 1. This example shows the differences between the PM and WPM formulations when dealing with the completion time. Assume that $\tau_1 = (0, 2, 3, 1)$ is a task in Γ,

which contains two fragments $\langle f_1^1, f_2^1 \rangle$. Each fragment has execution time 1. Table 3 shows the Boolean variables and the hard clauses introduced by PM and WPM to characterize the fragments' completion times.

First, we can determine that $\mathrm{ECT}_1^1 = 1$, $\mathrm{LCT}_1^1 = 2$, $\mathrm{ECT}_2^1 = 2$, and $\mathrm{LCT}_2^1 = 3$. To encode τ_1, PM introduces a series of Boolean variables, $\{\mathrm{eb}_{1,0}^1, \mathrm{eb}_{1,1}^1, \mathrm{eb}_{1,2}^1\}$ and $\{\mathrm{eb}_{2,1}^1, \mathrm{eb}_{2,2}^1, \mathrm{eb}_{2,3}^1\}$ for f_1^1 and f_2^1, respectively, representing whether each fragment is completed by time point t, where $\mathrm{ECT}_i^1 - 1 \le t \le \mathrm{LCT}_i^1$ for $1 \le i \le 2$. The relation of $\mathrm{eb}_{i,t}^1$ and $\mathrm{eb}_{i,t+1}^1$ is specified by rule (H1). To make τ_1 finish by its due date, f_1^1 should be completed no later than time 2. This is guaranteed by rule (H2). Finally, the relation between a fragment's start and completion times is indicated by rule (H3). Therefore, to constrain the completion time of each fragment, six Boolean variables and eleven hard clauses are generated by PM. By contrast, WPM introduces only one variable eb_1 to indicate whether τ_1 is completed by its due date. Then a hard clause is generated to constrain τ_1's due date and its last fragment's start time. In this way, both the Boolean variables and clauses generated by the WPM formulation are reduced.

The experimental comparison of these two encodings is demonstrated in Subsection 5.3.

Table 3: Encodings related to *eb* variables in PM and WPM formulations.

Variables		PM [38]	WPM
		$\mathrm{eb}_{1,0}^1$, $\mathrm{eb}_{1,1}^1$, $eb_{1,2}^1$, $eb_{2,1}^1$, $eb_{2,2}^1$, $eb_{2,3}^1$	eb_1
Clauses	(H1)	$\mathrm{eb}_{1,0}^1 \vee \mathrm{eb}_{1,1}^1$, $\mathrm{eb}_{1,1}^1 \vee \mathrm{eb}_{1,2}^1$, $\mathrm{eb}_{2,1}^1 \vee \mathrm{eb}_{2,2}^1$, $\mathrm{eb}_{2,2}^1 \vee \mathrm{eb}_{2,3}^1$	
	(H2)	$\mathrm{eb}_{2,3}^1 \vee \mathrm{eb}_{1,2}^1$	$\mathrm{eb}_1 \vee \mathrm{sa}_{2,3}^1$
	(H3)	$\mathrm{sa}_{1,0}^1 \vee \mathrm{eb}_{1,0}^1$, $\mathrm{sa}_{1,1}^1 \vee \mathrm{eb}_{1,1}^1$, $\mathrm{sa}_{1,2}^1 \vee \mathrm{eb}_{1,2}^1$ $\mathrm{sa}_{2,1}^1 \vee \mathrm{eb}_{2,1}^1$, $\mathrm{sa}_{2,2}^1 \vee \mathrm{eb}_{2,2}^1$, $\mathrm{sa}_{2,3}^1 \vee \mathrm{eb}_{2,3}^1$	

5. Experiments

In this section, we scrutinize the performance comparisons of the presented WPM formulation and two satisfiability-based formulations on a set of randomly generated problems. The experimental design is described in Section 5.1, followed by the evaluation of WPM, SMT [39], and PM formulations [38] in Section 5.2 and Section 5.3. All tests were conducted on a 3.4 GHz Intel E3-1230 processor with 8 GB RAM. The selected solver to evaluate WPM and PM is QMaxSAT [40], which is a SAT-based solver using the CNF encodings of cardinality constraints. The solver for SMT is Z3 [41], which is a high-performance theorem prover chosen by the previous SMT formulations [33, 39].

5.1. Experimental Design. The procedure for generating test instances is similar to [33, 38, 39], with a variety of parameter values. Tasks' release times are randomly generated following a discrete uniform distribution with an arriving rate λ, which represents the number of tasks that arrive during 100 time units. Clearly, a larger λ indicates that more tasks arrive in the system during a specific period of time, thus causing heavier overload. For each task τ_ℓ, the execution time c_ℓ and the number of fragments in τ_ℓ (denoted by q_ℓ) are also randomly generated according to a discrete uniform distribution. The value of deadline d_ℓ is calculated by the formula $d_\ell = r_\ell + sf_\ell \cdot c_\ell$, where sf_ℓ is a slack factor that reflects the tightness of the due date. Weight w_ℓ of task τ_ℓ is randomly chosen from 1 to n in Section 5.2 and is taken as constant 1 in Section 5.3. The number of rule pairs with dependency relations is set 10% to the total number of tasks.

We study the factors that may affect the performance of the satisfiability-based formulations. These factors include the number of tasks n, the task density (determined by λ), the execution time c_ℓ, the number of fragments q_ℓ, and the slack factor sf_ℓ. In the rest of this section, we run different types of experiments to test the possible changes on these parameters. Table 4 provides a summary of the experimental design, where $\mathrm{DU}(i, j)$ means the discrete uniform distribution over integer interval $[i, j]$. For each individual scenario, 100 problem instances are generated. For each instance and solver, we set a time limit of 300 seconds. If the solver fails to output the optimal result of an instance within the time limit, we terminate the procedure and move to the next problem instance.

5.2. Comparison on Weighted Case. This subsection evaluates the behavior of WPM and SMT formulations [39] for maximizing the total weight value of on-time tasks. The main metrics for comparison include (1) the proportion of instances solved within the given time limit and (2) the average computation time spent on solved instances. The performances of these two formulations in different types of scenarios are shown in Figure 5. Each data point is the average computation time of the solved problem instances. A number with an arrow in the figures denotes the proportion of instances solved within the time limit and is omitted if the solver addresses all the 100 instances. When the proportion of the solved instances drops to zero, the corresponding curve is omitted.

As shown in Figure 5, both formulations perform worse with the increasing values of all parameters. Nevertheless, in all the varied scenarios, WPM substantially outperforms SMT in terms of both the average computation time and the proportion of successfully solved instances.

5.3. Comparison on Unweighted Case. A Partial MaxSAT (PM) formulation for optimal scheduling on a single machine was recently presented [38]. The scheduling objective tackled by PM differs from our WPM formulation in that PM aims to maximize the total number of on-time tasks, while WPM tries to maximize the total weight of the on-time tasks. When all the tasks have equal weights, these two objectives become identical, and thus formulations on these two objectives are directly comparable. In this subsection, we compare the SMT, PM, and WPM formulations on the objective of maximizing the number of on-time tasks.

Table 4: Experimental design.

Scenario	n	λ	c_ℓ	q_ℓ	sf_ℓ
S1	$50, 100, \ldots, 300$	10	$DU(1, 30)$	$DU(1, 3)$	$DU(1, 4)$
S2	100	$5, 10, \ldots, 25$	$DU(1, 30)$	$DU(1, 3)$	$DU(1, 4)$
S3	100	10	$DU(1, \alpha)$ $\alpha = 10, 20, \ldots, 100$	$DU(1, 3)$	$DU(1, 4)$
S4	100	10	$DU(1, 30)$	$DU(1, \beta)$ $\beta = 2, 4, \ldots, 12$	$DU(1, 4)$
S5	100	10	$DU(1, 30)$	$DU(1, 3)$	$DU(1, \gamma)$ $\gamma = 1, 2, \ldots, 7$

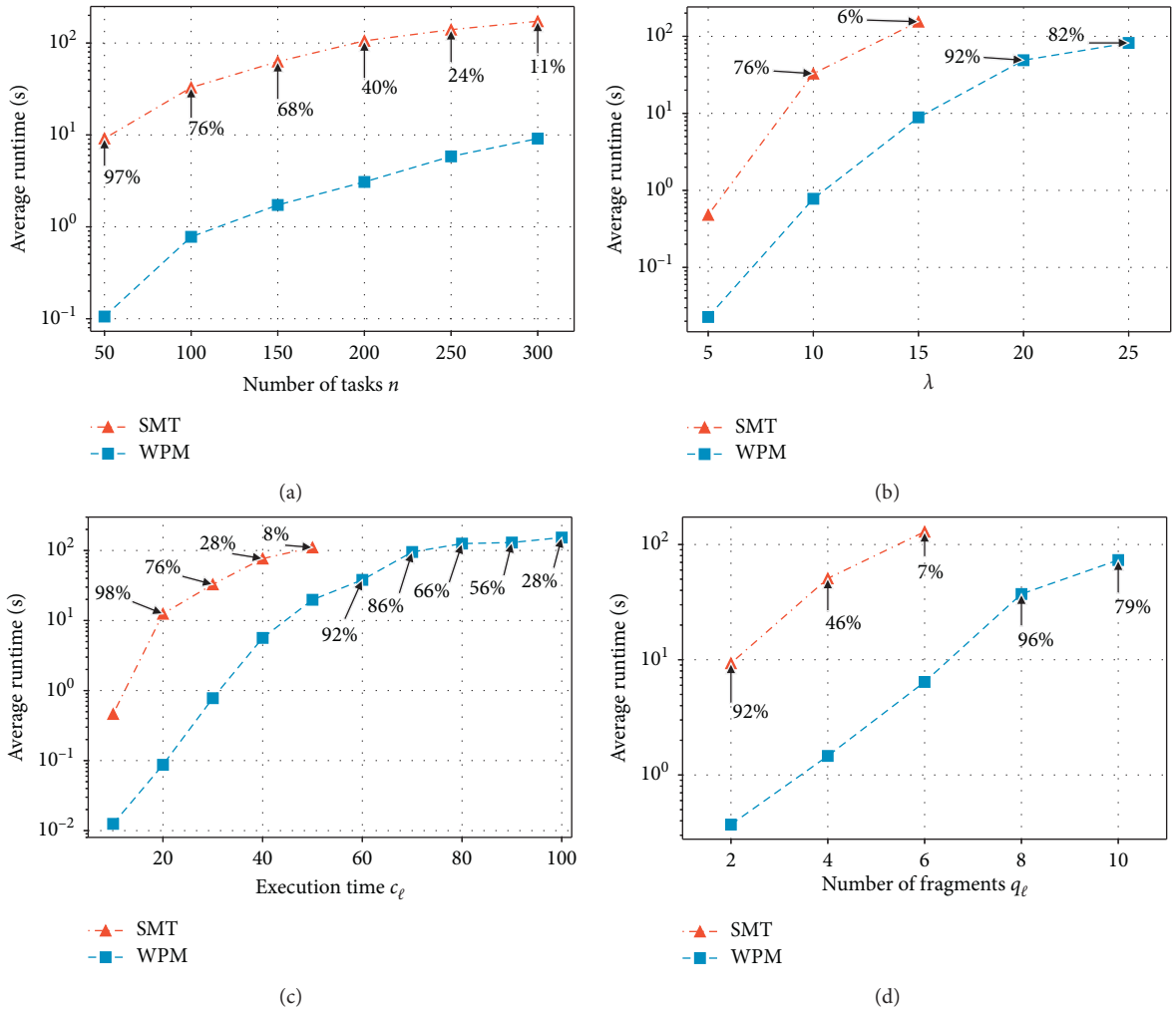
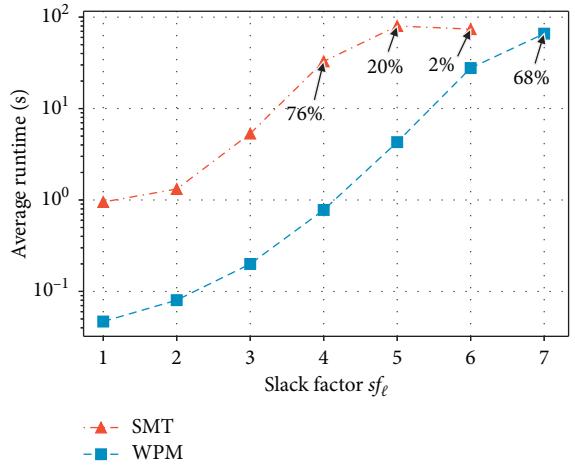


Figure 5: Continued.



(e)

Figure 5: Performance comparison of WPM and SMT for weighted cases in scenarios 1–5. (a) S1: varied number of tasks n; (b) S2: varied task arriving rate λ; (c) S3: varied execution time c_ℓ; (d) S4: varied number of fragments q_ℓ; (e) S5: varied slack factor sf_ℓ.

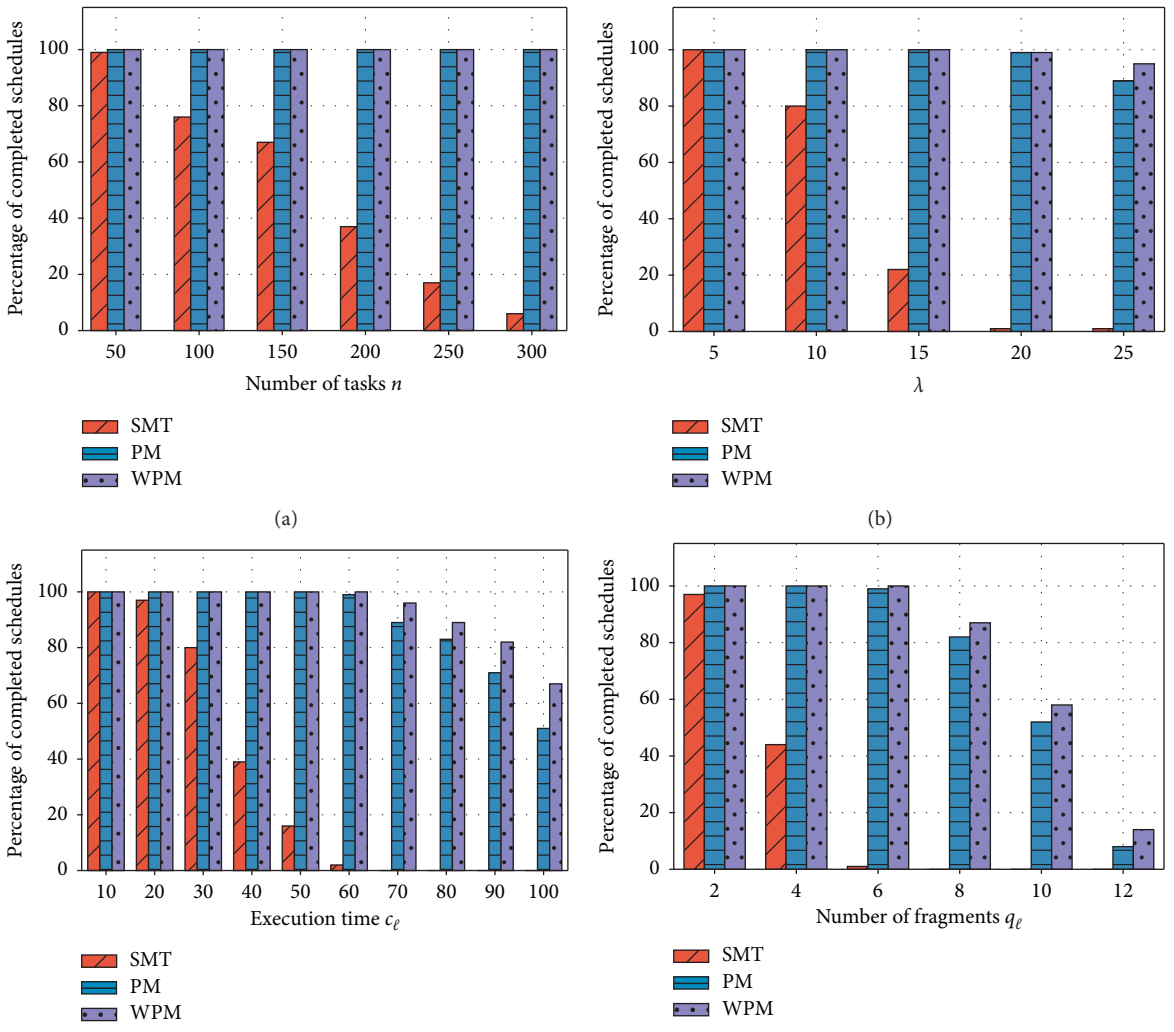
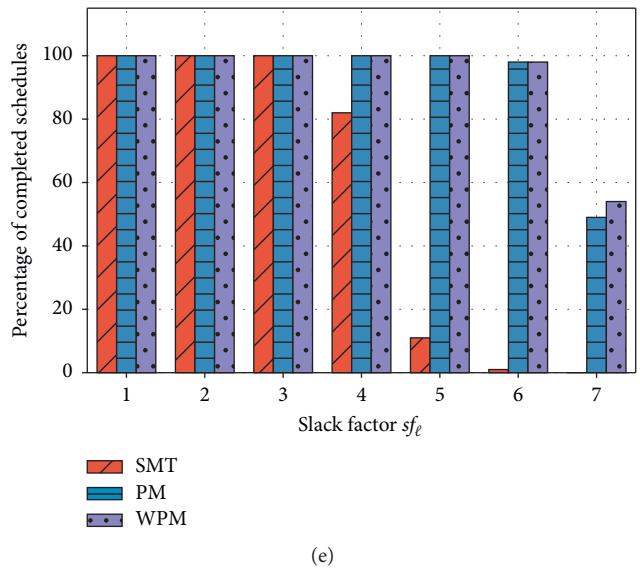


Figure 6: Continued.



(e)

FIGURE 6: Completion percentage of formulations for unweighted cases in scenarios 1–5. (a) S1: varied number of tasks n; (b) S2: varied task arriving rate λ; (c) S3: varied execution time c_ℓ; (d) S4: varied number of fragments q_ℓ; (e) Scenario 5: varied slack factor sf_ℓ.

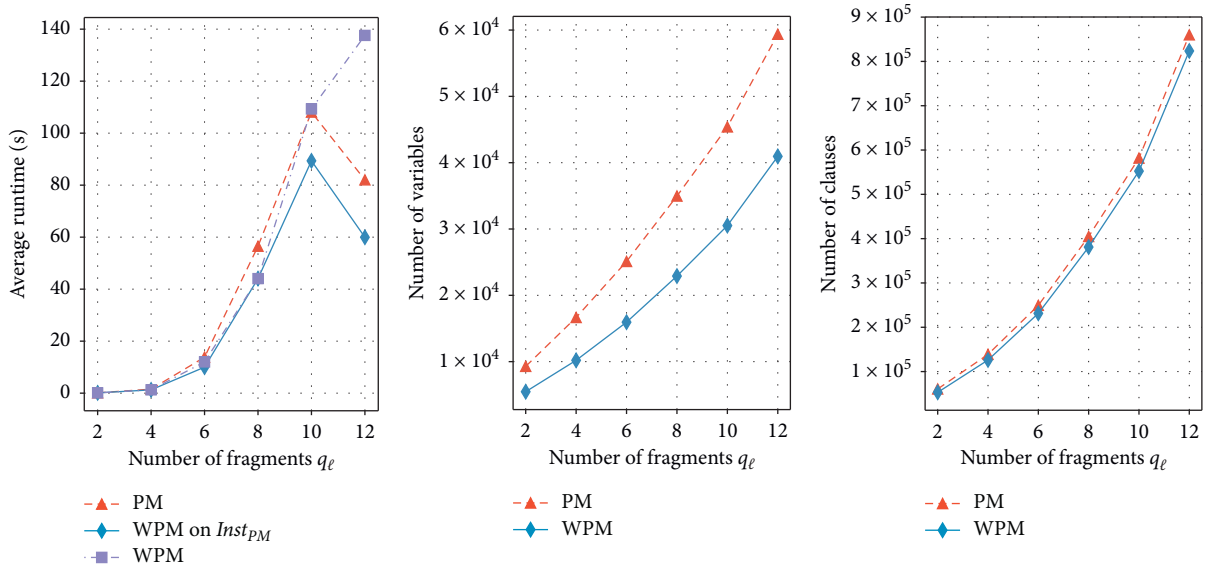


FIGURE 7: Comparison statistics of PM and WPM for unweighted cases in Scenario 4. (a) Average runtime (s) (b) Number of variables. (c) Number of clauses.

The overall results with varied n, λ, c_ℓ, q_ℓ, and sf_ℓ are shown in Figure 6. MaxSAT encodings (both PM and WPM) solved substantially more instances than SMT within the time limit in all cases. As to the performance comparison between PM and WPM, the completion percentage of WPM is always no less than that of PM no matter how the experimental parameters change. To be specific, with the number of tasks n increasing from 50 to 300, both PM and WPM solved all the instances within the time limit (Figure 6(a)). In comparison, as any of the parameters λ, c_ℓ, q_ℓ, and sf_ℓ increases, WPM formulation gradually surpasses PM with more completed schedules (Figures 6(b)–6(e)).

To compare the performances of PM and WPM more clearly, we provide more criteria for evaluation. The first criterion is the average runtime of PM and WPM formulations on their respective completed schedules. Note that when PM and WPM solve different numbers of instances within the time limit, the average runtime on solved instances may not clearly reflect the efficiency of these two formulations. For example, when $q_\ell \sim \mathrm{DU}(1, 12)$, WPM

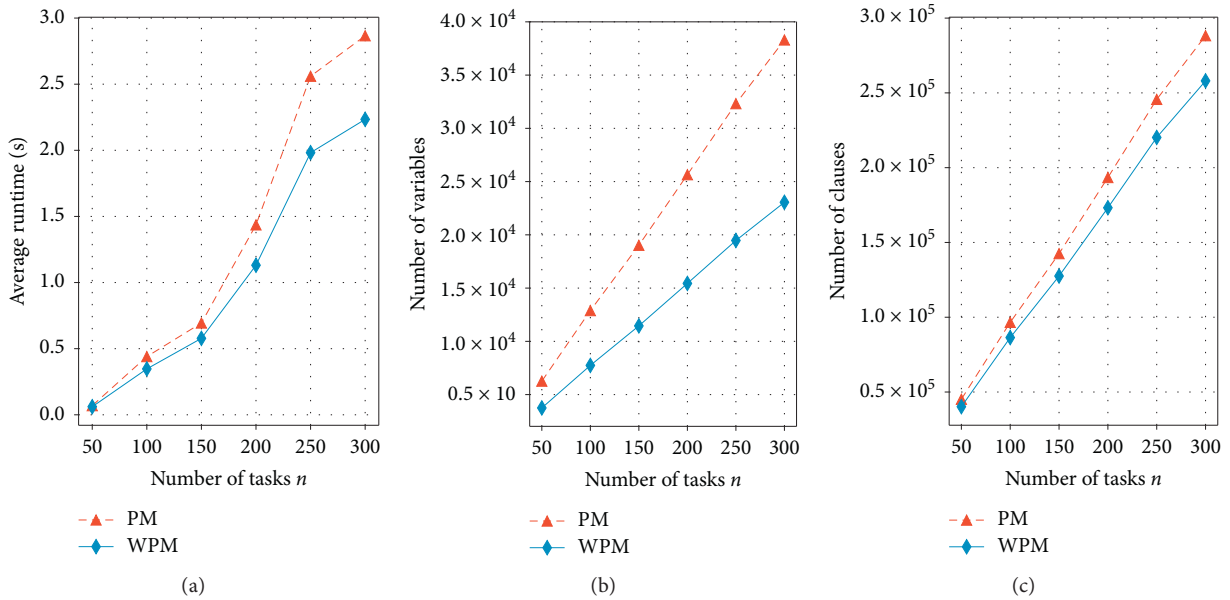


Figure 8: Comparison statistics of PM and WPM for unweighted cases in Scenario 1. (a) Average runtime (s). (b) Number of variables. (c) Number of clauses.

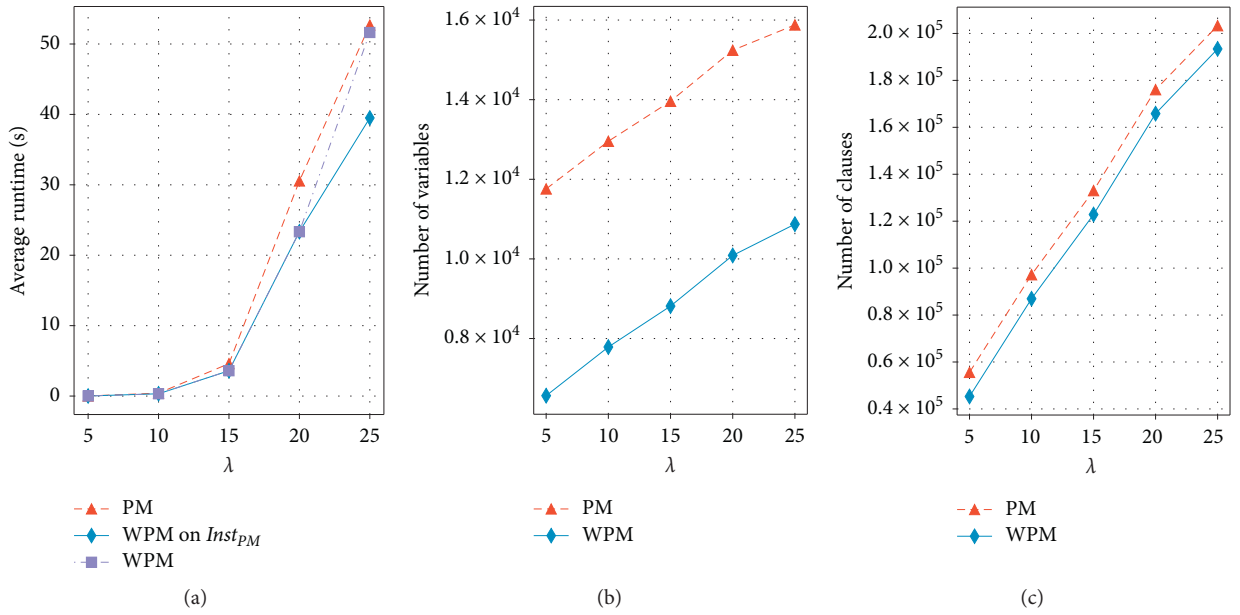


Figure 9: Comparison statistics of PM and WPM for unweighted cases in Scenario 2. (a) Average runtime (s). (b) Number of variables. (c) Number of clauses.

solved more instances than PM (as shown in Figure 6(d)) and meanwhile consumed longer time on solved instances (as shown in Figure 7(a)). In this case, the average runtime is on longer the proper metric to indicate the performance. On the other hand, we notice that, in all the tested scenarios, the instances that were solved by PM could also be addressed by WPM. Therefore, we provide a secondary criterion on runtime, that is, the average runtime of WPM, in solving the same set of instances as PM solved. Furthermore, since the computation time of MaxSAT-based methods is closely related to the numbers of variables and clauses, we collect such information for reference. The comparison results are illustrated in Figures 8–11 . Criteria for evaluating PM and WPM are summarized as follows:

(i) The average runtime of MaxSAT formulations on their respective completed schedules. The corresponding information of PM and WPM is depicted

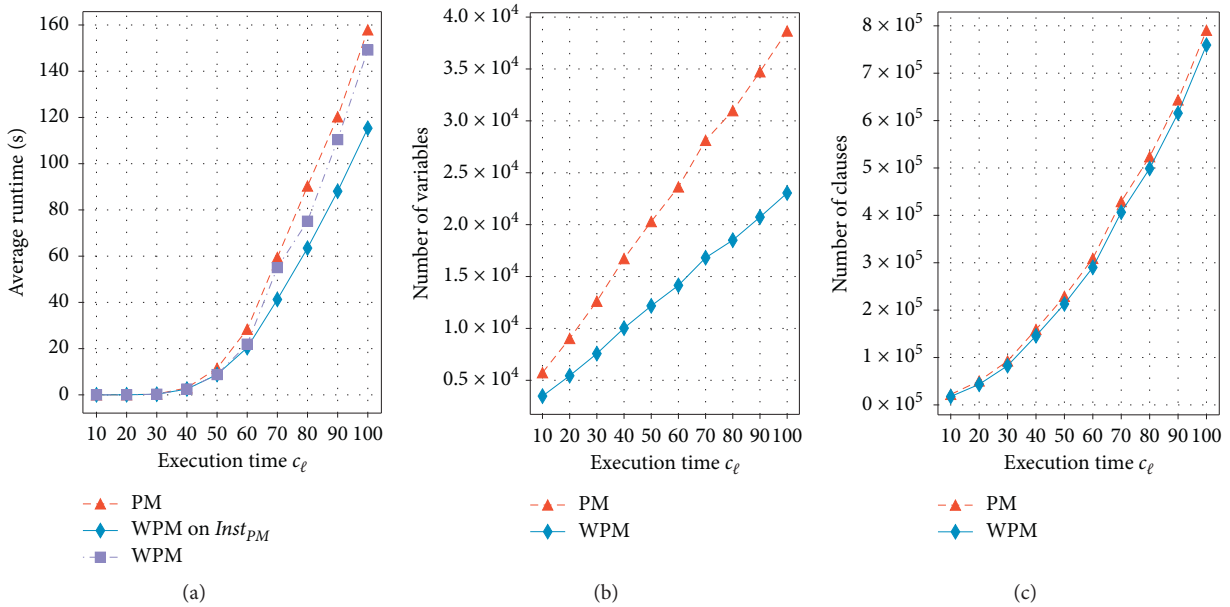


Figure 10: Comparison statistics of PM and WPM for unweighted cases in Scenario 3. (a) Average runtime (s). (b) Number of variables. (c) Number of clauses.

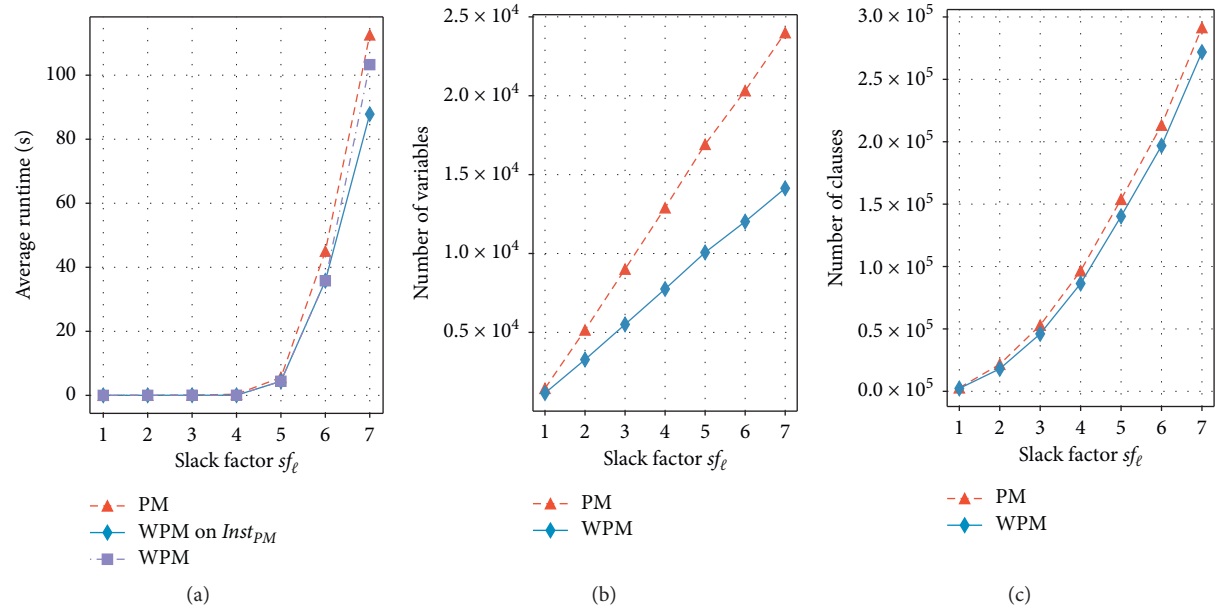


Figure 11: Comparison statistics of PM and WPM for unweighted cases in Scenario 5. (a) Average runtime (s). (b) Number of variables. (c) Number of clauses.

in Figures 8(a), 9(a), 10(a), 7(a), and 11(a) labeled "PM" and "WPM," respectively.

(ii) The average runtime of WPM in solving the same set of instances as PM solved. The corresponding information is depicted with label "WPM on Inst_{PM}." If the percentages of schedules completed by PM and WPM are the same, then "WPM on Inst_{PM}" is omitted (e.g., Figure 7(a)).

(iii) The average number of variables generated by MaxSAT formulations.

(iv) The average number of clauses generated by MaxSAT formulations.

Overall, the WPM formulation is more efficient than PM in both runtime criteria. The only exception appears in Figure 7(a), where the average runtime of WPM and PM on solved instances when $q_\ell \sim \text{DU}(1, 12)$ is around 140 and 80 seconds, respectively. Although the average computation time taken by WPM is longer than that by PM, Figure 6(d) shows that WPM managed to solve 14 instances, while PM solved only 8 within the time limit. In this case, we resort to

the secondary criterion on runtime. As shown in Figure 7(a), WPM consumed merely 60 seconds to solve all the instances completed by PM. Thus, it is correct to say that WPM can make a considerable improvement to the solving time. In addition, WPM achieved more compact encodings by significantly reducing the number of Boolean variables and slightly decreasing the number of clauses. In general, the computation time increases as the numbers of variables and clauses grow. This in turn explains why WPM achieves higher efficiency in our experiment.

6. Adaption for Parallel-Machine Scheduling Problem

While the WPM formulation is designed for single-machine scheduling, the encoding can be extended to parallel identical machines with minor changes. This section shows how to apply the WPM formulation to parallel identical machine scheduling problems.

The parallel-machine scheduling involves processing n tasks $\Gamma = \{\tau_1, \ldots, \tau_n\}$ on m identical machines $\mathscr{M} = \{M_1, M_2, \ldots, M_m\}$. Each machine can handle only one task at a time, and each task cannot be processed in parallel. Other settings of the problem are consistent with the single-machine scheduling problem described in Section 2. Given a set of tasks Γ and that of machines $\mathscr{M}$, the optimization problem is not only to find a schedule that gives the start execution time of each fragment $f_i^\ell \in \tau_\ell$ in Γ but also a mapping for f_i^ℓ to a machine in $\mathscr{M}$.

To indicate on which machine a fragment is executed, $\forall \tau_\ell \in \Gamma$, $\forall f_i^\ell \in \tau_\ell$, and $\forall M_u \in \mathscr{M}$, we extend the single-machine scheduling formulation by introducing one more Boolean variable $b_{i,u}^\ell$. $b_{i,u}^\ell = 1$ if f_i^ℓ is executed by M_u and $b_{i,u}^\ell = 0$ otherwise. Correspondingly, rule (C3) in Section 3.2 is replaced with the two following additional constraints:

(i) (R1) $\forall \tau_\ell \in \Gamma$ and $\forall f_i^\ell \in \tau_\ell$, f_i^ℓ is allocated and executed on a single machine:

$$\bigvee_{u=1,\ldots,m} b_{i,u}^\ell, \quad (1 \le \ell \le n, 1 \le i \le q_\ell). \tag{15}$$

(ii) (R2) $\forall \tau_k, \tau_\ell \in \Gamma$, $\forall f_i^k \in \tau_k$, and $\forall f_j^\ell \in \tau_\ell$, if $k \neq l$, $\tau_k \prec \tau_\ell$, $\tau_\ell \prec \tau_k$, $\mathrm{EST}_i^k < \mathrm{LCT}_j^\ell$ and $\mathrm{EST}_j^\ell < \mathrm{LCT}_i^k$, and then f_i^k and f_j^ℓ may require the processor at the same time. In this condition, f_i^k precedes f_j^ℓ or f_j^ℓ precedes f_i^k:

$$b_{i,u}^k \wedge b_{i,u}^\ell \longrightarrow \mathrm{pr}_{i,j}^{k,\ell} \vee \mathrm{pr}_{j,i}^{\ell,k}. \tag{16}$$

We refer to the set of hard clauses introduced in (R1) and (R2) as $\mathscr{R}$ and rules in (C1) ~ (C7) excluding (C3) as $\mathscr{C}'$. Conjunct with the set of weighted soft clauses $\mathscr{O}$ defined in Section 3.2, the problem is $\{\mathscr{C}', \mathscr{R}, \mathscr{O}\}$, which tries to find an assignment of variables to satisfy all the hard clauses in $\mathscr{C}' \cup \mathscr{R}$ and to maximize the sum of the weights of the satisfied soft clauses in $\mathscr{O}$. Thereby, the parallel identical machine scheduling problem of maximizing the total weight of on-time tasks can be optimally solved with any off-the-shelf weighted Partial MaxSAT solver.

7. Conclusions and Perspective

We concentrated on a Weight Partial MaxSAT (WPM) formulation for optimal scheduling in overloaded situations. The aim is to maximize the (weighted) number of on-time tasks. Motivated by the WPM feature that distinguishes between hard clauses and weighted soft clauses, we encoded the properties of tasks as hard clauses and the goal of completing tasks on time as a set of weighted soft clauses. Then an off-the-shelf WPM solver was employed to satisfy all the hard clauses and maximize the total weight of the satisfied soft clauses. From the output of the WPM solver, the optimal schedule can be obtained.

The WPM formulation's performance was compared with that of the recent SMT and Partial MaxSAT (PM) formulations. First, we compared the performance of WPM and SMT, demonstrating that WPM is significantly superior to SMT. Then, we considered a special case where WPM and PM formulations are directly comparable. Results indicate that the WPM formulation achieves more compact encoding and higher efficiency than PM. Finally, we applied the WPM formulation to parallel identical machines with very little modification, highlighting the flexibility and scalability of the encoding.

The restricted preemptive scheduling model considered in this paper is built upon earlier works [33, 38, 39], where each task is split into several nonpreemptive fragments, and preemptions can only take place at the fragments' boundaries. This preemptive model with fixed points is too rigid to apply in many practical systems. Our future work is to design exact methods to adapt to variations of preemptive models. One example is to execute tasks continuously without interruption for at least a certain portion of time [42–44]. Instead of splitting tasks into fragments in advance, this restricted preemptive model only guarantees the minimum "granularity" of preemption without predefined subtasks and thus is more suitable for real-world applications.

Data Availability

All the datasets are available from the corresponding author upon request.

Disclosure

This paper is an extended version of the authors' paper published in Pacific Rim International Conference on Artificial Intelligence, Yanuca Island, Fiji, 08, 2019 (Springer).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant no. 61806171) and JSPS KAKENHI (Grants nos. JP17K00307 and JP19H04175).

References

[1] C. E. Aguero, N. Koenig, I. Chen et al., "Inside the virtual robotics challenge: simulating real-time robotic disaster response," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 2, pp. 494–506, 2015.

[2] S. B. Er and R. E. Smith, *Method and Apparatus for Monitoring and Displaying Lead Impedance in Real-Time for an Implantable Medical Device: US*, Patent US5891179 A[P], 1999.

[3] D. P. Xenos, M. Cicciotti, G. M. Kopanos et al., "Optimization of a network of compressors in parallel: real time optimization (rto) of compressors in chemical plants - an industrial case study," *Applied Energy*, vol. 144, no. 5, pp. 51–63, 2015.

[4] G. D. Baulier, S. M. Blott, B. L. Branch et al., *Real-time Event Processing System for Telecommunications and Other Applications*, Patent US6496831 B1[P], 2002.

[5] L. Abeni, G. Buttazzo, S. Superiore, and S. Anna, "Integrating multimedia applications in hard real-time systems," in *Proceedings of the IEEE Real-Time Systems Symposium*, Madrid, Spain, December 1998.

[6] S. K Baruah and J. R. Haritsa, "Scheduling for overload in real-time systems," *IEEE Transactions on Computers*, vol. 46, no. 9, pp. 1034–1039, 1997.

[7] S. K. Baruah, J. R. Haritsa, and N. Sharma, "On-line scheduling to maximize task completions," in *Proceedings of the 15th IEEE Real-Time Systems Symposium (RTSS '94)*, pp. 228–236, San Juan, PR, USA, December 1994.

[8] A. Malik, C. Walker, M. OSullivan, and S. Oliver, "Satisfiability modulo theory (SMT) formulation for optimal scheduling of task graphs with communication delay," *Computers and Operations Research*, vol. 89, pp. 113–126, 2018.

[9] J. Michael Moore, "An n job, one machine sequencing algorithm for minimizing the number of late jobs," *Management Science*, vol. 15, no. 1, pp. 102–109, 1968.

[10] R. L. Graham, E. L. Lawler, J. K. Lenstra, and A. H. G. Rinnooy Kan, "Optimization and approximation in deterministic sequencing and scheduling: a survey," *Annals of Discrete Mathematics*, vol. 5, no. 1, pp. 287–326, 1979.

[11] M. Richard, *Karp. Reducibility Among Combinatorial Problems*, pp. 85–103, Plenum Press, New York, NY, USA, 1972.

[12] E. L. Lawler and J. M. Moore, "A functional equation and its application to the characterization of gamma distributions," *Management Science*, vol. 16, no. 1, pp. 77–84, 1969.

[13] M. R Garey and D. S. Johnson, "Scheduling tasks with nonuniform deadlines on two processors," *Journal of the Acm*, vol. 23, no. 3, pp. 461–467, 1976.

[14] J. K. Lenstra and A. H. G. Rinnooy Kan, "Complexity results for scheduling chains on a single machine," *European Journal of Operational Research*, vol. 4, no. 4, pp. 270–275, 1980.

[15] E. L. Lawler, "A dynamic programming algorithm for preemptive scheduling of a single machine to minimize the number of late jobs," *Annals of Operations Research*, vol. 26, no. 1, pp. 125–133, 1990.

[16] D. Nikos and M. Ross: Single Machine Problems. http://www2.informatik.uni-osnabrueck.de/knust/class/dateien/classes/ein_ma/ein_ma/.

[17] M. Sterna, *A Survey of Scheduling Problems with Late Work Criteria*, Omega, Bienne, Switzerland, 2011.

[18] X. Chu and J. Tao, "A competitive online algorithm for minimizing total weighted completion time on uniform machines," *Mathematical Problems in Engineering*, vol. 2020, Article ID 7527862, 9 pages, 2020.

[19] Y. Li, E. Fadda, D. Manerba, R. Tadei, and T. Olivier, "Reinforcement learning algorithms for online single-machine scheduling," in *Proceedings of the 2020 Federated Conference on Computer Science and Information Systems, FedCSIS 2020*, pp. 277–283, Sofia, Bulgaria, September 2020.

[20] W. Li and J. Yuan, "Single-machine online scheduling of jobs with non-delayed processing constraint," *Journal of Combinatorial Optimization*, vol. 41, no. 4, pp. 830–843, 2021.

[21] B. Goldengorin and V. Romanuke, "Online heuristic for the preemptive single machine scheduling problem to minimize the total weighted tardiness," *Computers & Industrial Engineering*, vol. 155, Article ID 107090, 2021.

[22] J. Y. Lee, "A genetic algorithm for a two-machine flowshop with a limited waiting time constraint and sequence-dependent setup times," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8833645, 13 pages, 2020.

[23] Q. Jiang, X. Liao, R. Zhang, and Q. Lin, "Energy-saving production scheduling in a single-machine manufacturing system by improved particle swarm optimization," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8870917, 16 pages, 2020.

[24] H. Wei, S. Li, H. Quan et al., "Unified multi-objective genetic algorithm for energy efficient job shop scheduling," *IEEE Access*, vol. 9, pp. 54542–54557, 2021.

[25] O. A. Muminu and A. O Adewumi, "A survey of single machine scheduling to minimize weighted number of tardy jobs," *Journal of Industrial & Management Optimization*, vol. 10, no. 1, pp. 219–241, 2013.

[26] S. Ourari, C. Briand, and B. Bouzouiac, "A MIP approach for the minimization of the number of late jobs in single machine scheduling," *Journal of Mathematical Modelling and Algorithms*, vol. 1, pp. 1–15, 2009.

[27] S. Venugopalan and S. Oliver, "ILP formulations for optimal task scheduling with communication delays on parallel systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 1, pp. 142–151, 2015.

[28] K. Ying, C. Cheng, S. Lin, and C. Hung, "Comparative analysis of mixed integer programming formulations for single-machine and parallel-machine scheduling problems," *IEEE Access*, vol. 7, pp. 152998–153011, 2019.

[29] H.-C. Hung, M. Bertrand, T. Lin, M. E. Posner, and J.-M. Wei, "Preemptive parallel-machine scheduling problem of maximizing the number of on-time jobs," *Journal of Scheduling*, vol. 22, no. 4, pp. 413–431, 2019.

[30] Z. Yang, H. S. Xiao, R. Guan, Y. Yang, and H. L. Ji, "Task scheduling for multiunit parallel test using mixed-integer linear programming," *Mathematical Problems in Engineering*, vol. 2021, Article ID 3785452, 13 pages, 2021.

[31] M. A. Qamhan, A. A. Qamhan, I. M. Al-Harkan, and Y. A. Alotaibi, "Mathematical modeling and discrete firefly algorithm to optimize scheduling problem with release date, sequence-dependent setup time, and periodic maintenance," *Mathematical Problems in Engineering*, vol. 2019, Article ID 8028759, 16 pages, 2019.

[32] Z. Cheng, H. Zhang, Y. Tan, and Y. Lim, "Scheduling overload for real-time systems using SMT solver," in *Proceedings of the IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/distributed Computing*, pp. 189–194, IEEE, Shanghai, China, June 2016.

[33] Z. Cheng, H. Zhang, Y. Tan, and Y. Lim, "SMT-based scheduling for overloaded real-time systems," *IEICE Transactions on Informations and Systems*, vol. E100-D, no. 5, pp. 1055–1066, 2017.

[34] T. Qi, L.-H. Zhu, J. Lian, L. Zhou, and J.-B. Wei, "An efficient multi-functional duplication-based scheduling framework for multiprocessor systems," *The Journal of Supercomputing*, vol. 76, no. 11, pp. 9142–9167, 2020.

[35] J. M. Crawford and A. B. Baker, "Experimental results on the application of satisfiability algorithms to scheduling problems," in *Proceedings of the Twelfth AAAI National Conference on Artificial Intelligence*, pp. 1092–1097, American Association for Artificial Intelligence, Menlo Park, CA, USA, August 1994.

[36] M. Koshimura, H. Nabeshima, H. Fujita, and R. Hasegawa, "Solving open job-shop scheduling problems by sat encoding," *IEICE Transactions on Informations and Systems*, vol. E93-D, no. 8, pp. 2316–2318, 2010.

[37] W. Liu, Z. Gu, X. Jiang, X. Wu, and Y. Ye, "Satisfiability modulo graph theory for task mapping and scheduling on multiprocessor systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, no. 8, pp. 1382–1389, 2011.

[38] X. Liao, H. Zhang, M. Koshimura, R. Huang, and W. Yu, "Maximum satisfiability formulation for optimal scheduling in overloaded real-time systems," in *Proceedings of the PRICAI 2019: Trends in Artificial Intelligence—16th Pacific Rim International Conference on Artificial Intelligence*, pp. 618–631, Springer, Yanuca Island, Fiji, August 2019.

[39] S. Wang, X. Liao, M. Wang, L. Chang, H. Yang, and T. Wang, "An improved smt-based scheduling for overloaded real-time systems," *Engineering Letters*, vol. 28, no. 1, pp. 112–122, 2020.

[40] M. Koshimura, T. Zhang, H. Fujita, and R. Hasegawa, "Qmaxsat: a partial max-sat solver," *Journal on Satisfiability, Boolean Modeling and Computation*, vol. 8, pp. 95–100, 2012.

[41] L. De Moura and N. Bjørner, "Z3: an efficient SMT solver," in *International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, pp. 337–340, Springer, Berlin, Germany, 2008.

[42] K. Ecker and R. Hirschberg, "Task scheduling with restricted preemptions," in *Arndt Bode, Mike Reeve, and Gottfried Wolf*, pp. 464–475, Springer, Berlin, Germany, 1993.

[43] T. Barański, "Task scheduling with restricted preemptions," in *Proceedings of the 2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 231–238, IEEE, Szczecin, Poland, 2011.

[44] K. Pieńkosz and A. Prus, "Task scheduling with restricted preemptions on two parallel processors," in *Proceedings of the 2015 20th International Conference On Methods and Models in Automation and Robotics (MMAR)*, August 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4893776, 18 pages
https://doi.org/10.1155/2021/4893776

Research Article

Analysis of Environmental Costs' Effect in Green Mining Strategy Using a System Dynamics Approach: A Case Study

Amir Jafarpour [1] **and Siamak Khatami** [2]

[1]*Department of Mining and Metallurgical Engineering, Yazd University, Yazd, Iran*
[2]*Department of Computer Science, Faculty of Information Technology and Electrical Engineering, NTNU, Trondheim, Norway*

Correspondence should be addressed to Amir Jafarpour; jafarpour.a67@gmail.com

Received 1 May 2021; Accepted 21 June 2021; Published 16 July 2021

Academic Editor: Noorbakhsh Amiri Golilarz

Nowadays, one of the most significant problems in mining activities is the significance of analyzing environmental issues along with mining, concentration, and mineral processing operations to achieve the goals of sustainable development. Nevertheless, mine owners refuse to include environmental costs (EC) and consider them unprofitable. Due to the ever-increasing importance of environmental and social topics in recent years, there is a vital need for assessing the EC and its impact on total mining costs and implementing green strategies by the mining managers and engineers. The current study tries to model the mining cost structure by considering the causal relationships between different factors affecting open-pit mining costs to highlight the EC's role. Furthermore, this research evaluates the effectiveness of implementing each possible mining green strategy in a large-scale copper mine using the System Dynamics (SD) approach. In this regard, seven scenarios and a combination of different environmental strategies, including mine reclamation, an environmental strategy for a condensation and processing plant, and environmental mining operations, have been considered for the SD-based economic analysis. The simultaneous use of the green mining strategies for the concentration and processing plant (Scenario 4) shows a high impact on cost reduction in the mining operation.

1. Introduction

Mining is one of the most important large-scale industries in developed and developing countries [1]. Sustainable development and its components, including economic, social, and environmental factors, play a major role in the decision-making problems of the mining projects, such as feasibility assessment, planning, and management of mining operation studies [2, 3]. Several technical and managerial studies evaluated these factors separately. The studies are conducted based on assumptions and variables with unknown values, which should be estimated using the existing data [4, 5]. Attentions to the environment, social, and economic topics, which are the main goals of sustainable development, are of importance in mining activities [6, 7]. Even with the implementation of principles and the framework of engineering, mines have a nature of being environmentally destructive in the absence of green mining construction policies [8]. Also, mines can be considered dangerous to the people who live around the mines because mining operations involve demolition. The reasons for conducting more research focusing on environmental costs (EC) in mines: (A) the environment is an undeniable part of human life; (B) human beings always need natural raw materials to survive, and (C) a balance is needed between basic needs of human beings and their destructive environmental activities.

It is necessary to evaluate the EC's role and its impact on total mining costs. In this regard, using biocompatible mining strategies entitled green mining can reduce the destructive environmental effects of mines [9]. On the other hand, paying excessive attention to environmental problems during mining operations may have an impact on increasing operating costs and reducing the overall profitability of the open-pit mine. The cost components play a significant role in the scope of changes in the economic earnings of a project, and thus their analysis is essential [10]. For this purpose,

economic analysis on the implementation of green mining strategies in mining projects and their impacts on total costs become highly important [11]. This study considers the green mining strategies based on the concept of sustainable development and evaluates the costs resulting from the destructive environmental effects of mining and the implementation of green mining strategies, introduced in the form of sustainable strategies. The output of this study can solve the misconceptions of mining managers because they believe that the EC they spend on mining is not profitable, so mining industry managers are always avoiding the cost of the environmental solutions and even the cost of reclamation. The study also evaluates the effects of the EC of open-pit mining on total mining costs to prove that the nature of the EC is different from what mining engineers and managers have in mind.

Due to the importance of reducing the destructive effects of mining on nature and addressing the existing social problems, this study aims to find answers to the main questions using dynamic modeling: What are the advantages of considering EC in the studied problem? Does the implementation of EC technically and economically lead to achieving sustainable development goals? What is the impact of green mining strategies on sustainable development solutions? It is possible to answer these questions by considering the EC's role in reducing total mining costs and using a dynamic approach to analyze this type of cost in the mining activities. To overcome some shortcomings of static approaches, such as Multi-Criteria Decision-Making (MCDM), this study uses the SD model to reach its objectives, including modeling and formulating causal relationships and assessing the EC's role in green mining activities.

The SD approach has various advantages, including the ability to analyze complex systems, having integrated vision of all parts of the problem, evaluating the effects between the criteria affecting each other and the whole system, presenting simple and accurate analysis for complex problems, and providing an instantaneous analysis of the effects of each system component on the other components and the whole system. This study evaluates the effectiveness of each possible green mining strategy in a real-world case study by modeling the mining cost structure and considering the causal relationships between concepts. To put it precisely, the overarching aim of this study was to model the mining cost structure by considering the causal relationships between its constructive factors to highlight the EC's role using the SD model. In addition, evaluating the effectiveness of each possible mining green strategy in this sector can provide an opportunity for the managers of open-pit metal mines to choose the most appropriate and efficient scenario. Since the SD method is a dynamic analytical-mathematical method and can solve various economic problems, this approach was used in this study. The problem of economic analysis of the environmental costs of mining activities is one of the most difficult dynamic problems in mining engineering for which there is no suitable solution. Therefore, to analyze this type of cost accurately and realistically, it has been evaluated in this study.

The structure of the study is as follows. Section 2 investigates previous studies separately in the fields of environmental cost estimation of open-pit mines and SD applications in the mining sector. In Section 3, the used concepts and methods are defined. Section 4 presents descriptions of the case study and analyzes the outputs of implementation of the SD approach. Section 5 presents the conclusion and summary of the results.

2. Literature Review

This section discusses studies related to the EC of the mining process and researches conducted using a variety of methods on the EC of mining activities focusing on EC and total costs. Also, this section reviews applications of the SD approach, which has been widely used in solving various engineering problems and managerial-strategic decisions.

2.1. Analysis of the Environmental Costs in Mining Activities. The studies about mining costs are very different and numerous. In the last decades, mining studies have been done in the field of reducing costs of operation and environmental costs of mining operations. In the following, this section reviews some researches carried out in this regard.

Elliot and Harris [12] analyzed the cost-benefit mining activities in Mozambique. This study only predicted profit-cost values for the next ten years and calculated the Net Present Value (NPV) of the mine without paying attention to the relationship between the factors affecting the cash flow of the mine. In another study, Byrd and Gildestad [13] analyzed the economic and social impacts of mining activities in Afghanistan. The researchers assessed the effects of mining operations on the environment, taking into account the social components and economic parameters of mining activities. Radev [14] made theoretical studies of the economic evaluation of investment projects in the mining industry. This study focuses only on economic issues, after discussing energy and related costs in mines. Although energy is one of the most significant environmental problems in mines, this research has not investigated the destructive environmental effects of improper use of energy in mines.

Rashidinejad et al. [15] presented an environmental model to optimize the cutoff grade to minimize environmental degradation of open-pit mines (especially, acid mine drainage). This study analyzed the components' effects of sustainable development on economic problems of mineral exploitation using an operations research technique (mathematical model). However, this research investigated only those related criteria to mining cutoff grade and neglected other green mining strategies-related parameters.

Botín and Vergara [16] presented the cost management model for a sustainable economy and a permanent recovery of mining operational costs. Kusi-Sarpong et al. [17] analyzed mining industries using the green supply chain and evaluated the cost of mining operations as one of the most important parameters. This study provided green supply chain practices and developed an approach for the studied

problem using Joint Rough Sets (JRS) and fuzzy TOPSIS. However, the fuzzy MCDM method used in this research does not provide very accurate results. Also, the proposed combined approach cannot analyze the dynamic and complex systems due to the static nature of the method. Narrei and Osanloo [18] provided the optimum cutoff grade concerning reducing adverse environmental impacts of open-pit mines. This study presented the environmental factors affecting the cutoff grade and their role in mining activities. The impact of the EC and the resulting profits on the cutoff grade—first introduced by Lane in 1988—are highlighted in this research. However, this study has not considered the dynamic nature of the studied system in grade calculation.

Xu et al. [19] studied the optimization of the production planning of open-pit metal mines, taking into account ecological costs. This study covered the EC as internal costs and within the mine optimization plan. Xu et al. [20] used the dynamic programming (DP) model to calculate the NPV of the mine for several years and for the best production scheduling. Despite using dynamic sequencing of the geologically optimum push-backs, this research has not investigated the effectiveness of the various environmental and economic factors in the proposed model. Guo et al. [21] predicted the cost of mining capital for open-pit mining projects with an artificial neural network approach. The study used Artificial Neural Network (ANN), Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Tree (CART) estimates of open-pit mining costs. Although these methods can easily estimate the desired calculations, the approaches used in this method cannot calculate the costs accurately. Brescia-Norambuena et al. [22] used stochastic modeling to calculate the cost of underground mines. This study studied productivity, production, and environmental issues. As the random model is a mathematical model, it is difficult to formulate relationships between the effective criteria and analyze the outputs in this method.

Numerous studies focused on analyzing and evaluating costs in mining industries [23–31]. However, these studies have not investigated the EC's role and its direct impact on the total costs of mining operations. In the following, some researches carried out in this regard are reviewed.

2.2. System Dynamics Applications in the Mining Sector. The SD approach can address various problems, including integrated planning and designing the strategies [32], analyzing the economic behavior [33], biological modeling [34], energy optimization [20], Bayesian approach [35], tackling environmental problems [36], and urban transport planning [37]. This, in turn, demonstrates the capability and level of trust of the SD model in the complex problem analysis. These studies show that researchers have used an SD approach to solve different engineering problems. Also, this approach has been used in the mining engineering and mining project's management analysis. The following studies are some of the most important researches focusing on SD approaches' applications in the mining sector. Cooke [38] used the SD approach in the assessment of the Westray mine disaster. This study created a general model for the mine disaster causal structure and analyzed the subsystems, including mine capacity, production, and human resources. Wang et al. [39] developed a dynamic model to plan the development of coal production and sales capacity in China. In this research, Wang et al. considered the amount of carbon dioxide emissions and demand in different scenarios and tried to identify the best plans under each scenario. Moreover, interesting research has been carried out in modeling, analyzing, and evaluating the production process and operating costs using the SD approach. Also, Lagnika et al. [40] examined the possibility of using an SD approach in analyzing mine environmental topics. Although this study aims to integrate environmental management (EM) tools based on Dynamic Simulation (DS) for mining, it has been concluded that the SD can process environmental problems.

Based on previous studies and considering the capability of the SD approach in evaluating complex systems and analyzing the relationships between the effective criteria in each system, the current study uses this approach to address the studied problem. Dynamic analysis, applying the effectiveness of different criteria, the possibility of scenario-making based on the strategies considered by managers and experts, evaluating the effectiveness of each scenario, and the possibility of momentary analysis of the effectiveness of each criterion on the whole system are significant features of the SD approach. Table 1 presents a comparison between the present study and previous research with a focus on the mining sector, environmental and mining costs, as well as using the SD approach. The present study has tried to evaluate the EC's impact on total mining costs considering the causal relationships between different factors affecting open-pit mining costs that have not been done in previous studies. Besides, the role of managing the EC has been examined based on different scenarios resulting from green mining. The previous researchers have not focused on the analysis of the effectiveness of such strategies according to the EC of mining.

Considering that not much research has been done to evaluate the interaction of EC with mining costs, the output of some studies conducted with the approach of environmental cost analysis can be compared with the current research. Botín and Vergara [16] and Mchaina [41] assessed the environmental cost of open-pit mining. The comparison between these works and the current study shows that this study has results close to the real conditions of open-pit mining and the defined scenarios have the necessary efficiency in assessing the EC and their mutual impact on each other and the total costs of mining. The result of Botín and Vergara's [16] study on underground mining has only led to the presentation of a model. In this study, after presenting a list of the underground mining costs, researchers have proposed a simple model based on mathematical relationships to calculate sustainable mining costs, which is a simple model compared to the current study.

The SD model in the present study has better performance than Botín and Vergara's [16] study, due to the extent of causal relationships between the influencing factors and

Table 1: Comparison of the characteristics of the current study with other studies focusing on the analysis of the environmental costs of mining.

Researcher(s)	Year	Mining costs	Environmental costs	System dynamics
Byrd and Gildestad	2001	✓		
Elliot and Harris	2001	✓		
Cooke	2003			✓
Radev	2003	✓	✓	
Rashidinejad et al.	2008	✓	✓	
Dehghani and Ataee-pour	2012	✓		
Dehghani and Ataee-pour	2014	✓		
Botín and Vergara	2015	✓		
Kusi-Sarpong et al.	2015	✓	✓	
Narrei and Osanloo	2015	✓	✓	
Sivakumar et al.	2015		✓	
de Werk et al.	2017	✓		
Lagnika et al.	2017			✓
Akpalu and Normanyo	2017	✓	✓	
Wang et al.	2018	✓		✓
Xu et al.	2018	✓	✓	
Spitz and Trudinger	2019		✓	
Guo et al.	2019	✓		
Brescia-Norambuena et al.	2020	✓	✓	
Current study		✓	✓	✓

the system's complexity. The output of the study of Kusi-Sarpong et al. [17] was based on a fuzzy MCDM method, namely, the fuzzy TOPSIS method, and is not comparable to the dynamic model presented in the current study. Although the TOPSIS method has found many applications, the weight of the criteria used in this method and the completion of the questionnaire (to form a decision matrix) is done based on experts' opinions. So, the probability of error in this method is higher than the SD approach. Also, the effect of factors in the mentioned method is applied asynchronously and separately in the model. The SD approach leads to more desirable outputs because the model is solved dynamically taking into account the simultaneous effect of different factors. Besides, due to the multiplicity of influential factors used in the studied problem, the proposed SD model can better model and solve this problem compared to MCDM methods. Mchaina [41] introduced only a conceptual model of environmental planning, open-pit mine reclamation, and mine closure problems. This study did not provide a comprehensive approach in comparison with the current study due to the lack of a specific mathematical model. In addition to presenting a model based on causal relationships, the present study evaluates the effectiveness of various environmental solutions in reducing the total costs of open-pit mining.

3. Methods

This section has investigated open-pit mining operation, minerals concentrating and processing, and the essence and nature of mining costs. Also, in the following, the theoretical principles of the SD approach are presented.

3.1. Mining Operation and Related Costs. Mining activities are one of the important sources of environmental pollution, and environmental impacts of mining on the area around mining sites can occur premining, during mining operations, or postmining [42, 43]. An overview of open mining operations is imperative to calculate the costs of each part of the mining operation and make apparent its potential environmental degradation effects. In open-pit mining, extraction first creates a bench to perform. Mining specialists design and build up benches to a depth from the mine, and thus extraction of minerals does not have any economic losses. This method is popular in metallic mines and mass deposits [44]. In general, the mining process (from the extraction of crude minerals to the production of pure metal or mineral) in open-pit mines has the following steps:

(i) Harvesting valuable topsoil and transferring it to a suitable location

(ii) Prestripping and overburden removal

(iii) Stripping and extraction

(iv) Transferring and accumulating waste rock and ore

(v) End of mine life, mine closure, and reclamation [45] (see Figure 1)

Mining specialists must consider the mining reclamation as an integral part of the entire mine design (from the early stages of mineral extraction) due to having a high environmental impact in areas where new mines are discovered and extracted [46]. The environmental problems in mines usually include problems related to the tailings dam of the processing plant, dust caused by the activity of drilling, loading, transporting machinery, destructive effects of blasting operations (dust), contamination from the activity of processing and smelting plants, wastewater leakage from the wastewater pipeline from the plant to the tailings dam, etc. [6]. The management for the reduction of these malicious effects imposes minor costs to mining projects. In this

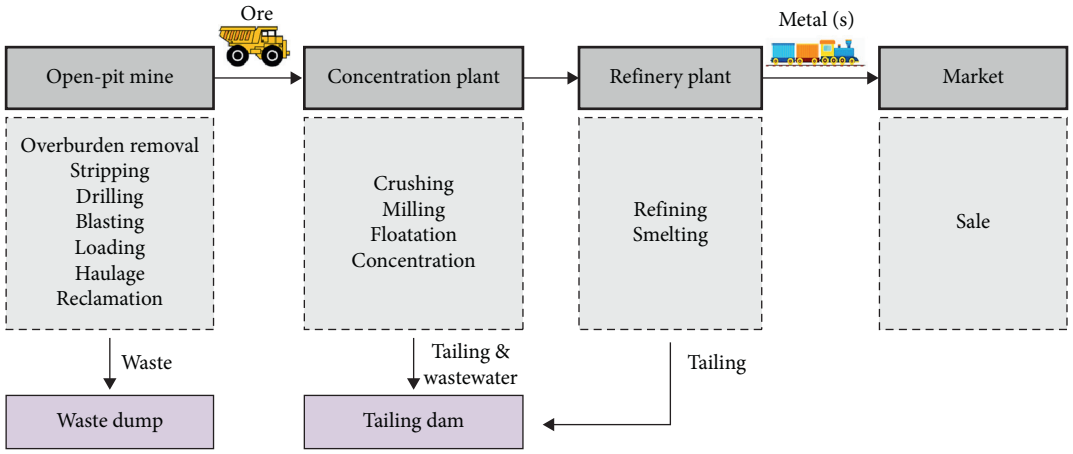


Figure 1: The schematic of mining operation in open-pit mines.

regard, biocompatible approaches can be considered, which are not without costs; however, the cost type of environmental strategies is different from other types of costs [47]. The cost of environmental approaches can be evaluated as positive costs because these approaches can reduce other costs, especially the cost of pollution. However, if mine owners ignore these costs, they must prepare themselves to face more costs to eliminate or control undesirable environmental consequences of mining activities. In fact, environmental fines and the EC caused by mining damages, such as destruction of pastures and forests, pollution of water/air, etc., impose great costs on mine owners. The current study aims to focus on the EC and its effect on the total costs of open-pit mining. To better illustrate the EC's role, the proposed model separated these costs from other mining costs. The mentioned costs in Figure 2 are common in the economic analysis of mining projects. On the other hand, to evaluate more accurately, environmental costs (due to the effects of pollution on the environment and related crimes) and the costs of environmental solutions (methods to reduce the effects of pollution) are considered separately.

Like all of the engineering operations, economic and financial problems are of cardinal importance in mining projects. Therefore, it is necessary to identify the nature and essence of the costs of mining projects and assess their positive/negative effects on the total costs of mining activities. There are different kinds of costs in each mining operation as well as many methods to report these costs [48, 49]. In general, mining specialists divide mining costs in open-pit mines into three classes: capital expenditure, operating costs, and general and administrative costs. In this classification, capital costs refer to the investment required for the mine and mineral processing plant. Operational costs reflect the cost of drilling, blasting, etc., performed per ton. General and administrative costs are reported annually [44, 50]. Although the EC is considered in the above categories, the current study classifies EC in another class due to the extent of the impact of this type of cost, which separates it from other types of costs (see Figure 2). The various intrinsic and exogenous factors affect the mining cost estimation problem, and often these factors are interconnected. Accordingly, the current study focuses on causal relationships between these factors. One of the most appropriate methods to investigate problems with a cause-effect structure is the SD approach used in the present study.

3.2. System Dynamics. The SD was first introduced by Forrester in 1950, and a theory about the structure of the system and a set of tools was used to identify the structure of complex systems and analyze their dynamic behavior [51]. The SD approach can model and simulate complex economic and social systems, which can be used to make policies and strategies for change [37, 52]. In this approach, specialists determine variables affecting the problem and their relationships firstly. Then, this approach analyzes the behavior of these variables under different scenarios with simulation. These scenarios and observed behaviors of the system will be shown visually and graphically with the least cost of the system's future experience. The SD helps managers to make more informed decisions [53]. The principles of SD and basic concepts of this method can be explained in Figure 3.

As can be seen schematically in Figure 3, the SD can be easily described by evaluating how glass is filled with water. Forrester [51] made a good point: "Filling a glass with water is not merely a matter of water flowing into the glass. Rather, there is control over the amount of water. This control is the feedback loop from the water level to the eye to hand to faucet to the water flow and back to the water level. Such closed loops control all actions everywhere." Also, Figure 3 shows a simple feedback system. In the figure, there are two symbols: a stock and a flow. The stock is a collection from different fields of accumulation or integration or level. The flow changes the amount in the stock. The flow is determined by a statement that tells how the flow is controlled by stock value compared to a target. Of course, all of the simple or complex systems include these two types of concepts: stocks and flows. Such a statement, based on the existence of two and only two kinds of variables in a system, is powerful in simplifying the decision-maker's view of the world [51, 53]. This structure expresses the basic truth about all systems.

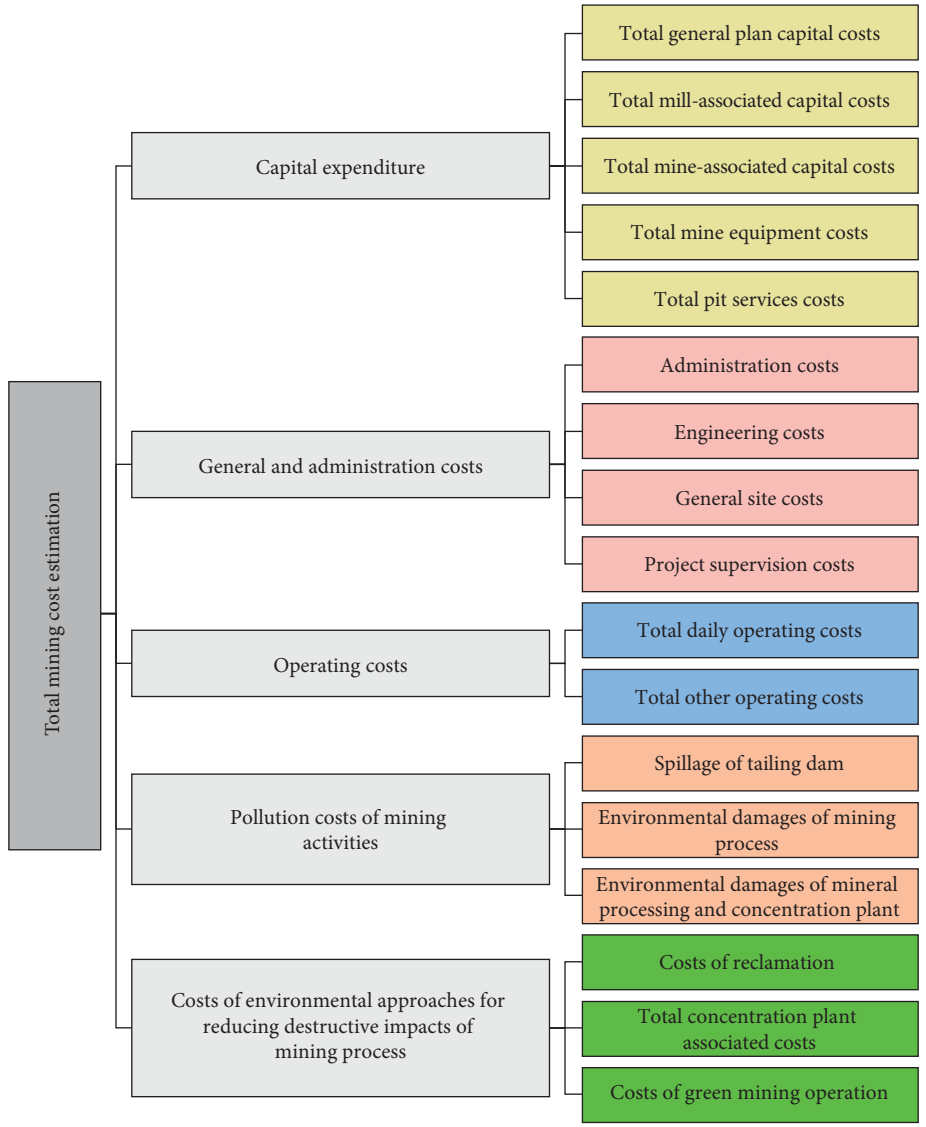


Figure 2: Types of mining extraction and mineral processing costs in open-pit mines.

Specifically, the implementation of the SD approach includes the following steps:

(i) Identification and definition of the problem: in this step, the nature of the problem, key variables affecting the problem, and horizons of interest are determined.

(ii) Determination of the system behavior based on reference patterns: in this step, the past behavior of problem variables as reference patterns is examined and their desired behaviors are determined.

(iii) Explanation of the relationship between variables and plotting of cause-effect and accumulation-flow diagrams: in this step, dynamic assumptions of the system are defined based on early assumptions that are capable of explaining the behavior of the system. Then, based on these hypotheses, key variables, reference patterns, and other data, as well as causal diagrams and accumulation and flow of the system are drawn.

(iv) Development and implementation of a simulation model: in this step, based on the accumulation chart and the flow drawn in the previous step, and the assumptions expressed in the dynamic hypothesis, a simulation model is drawn for the problem. Then, initial conditions to run the simulation as well as initial values of the variables and parameters of the model are determined.

(v) Analysis of the model and designing of policies: in the last step, by implementing a simulation model and examining the model's ability under extreme conditions, the model's behavior is compared with the observed situation from past information and reference patterns to validate the model. Finally, specialists test new structures and rules to see whether desired behaviors are observed in the system by designing different policies and scenarios [54].

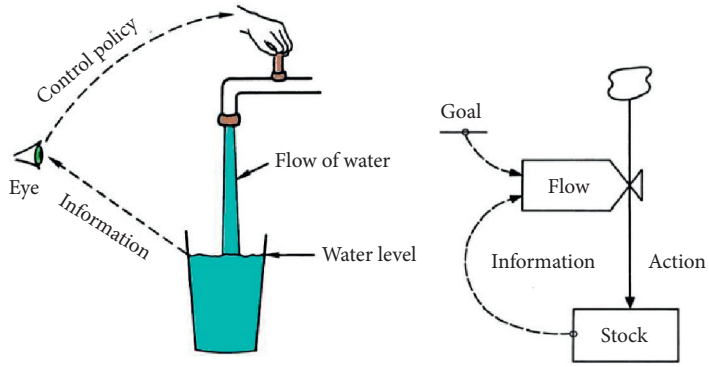


Figure 3: The simplest possible feedback system [51].

There are various software packages to model and simulate dynamic systems, but in the current study, the Vensim software has been used, which is a powerful tool for simulation, model testing, designing, and studying various scenarios in complex systems.

4. Results and Discussion

In this section, the open-pit mine whose information has been used in the research is introduced firstly in the case study section. Then, the second section describes the proposed SD model, and finally, the results of solving this model are analyzed.

4.1. Case Study. The present study uses the information of an active copper complex to evaluate and validate the proposed model. The Sungun copper mine is located in the northwest of Iran, adjacent to a protected area, and is registered as a living biosphere by UNESCO. This mine is active in Eastern Azerbaijan province, 110 km northeast of Tabriz, 75 km northwest of Ahar city, and 15 km north of Varzeqan city (see Figure 4).

The environmental significance of the region is high, which requires environmental assessment and economic analysis of environmental approaches. The deposit of this mine is porphyry copper, and its products are molybdenum and gold, which are extracted by open-pit mining. The geological resource of the deposit of Sungun copper mine is 796 million tons, and the total mineable deposit of the mine is 410 million tons. The average grade of ore (mineable deposit) is 0.6% of copper. The life of the mine is 32 years. So far, 12 years have passed since the mining activity. Thus, the remainder of the mine's life in the current study is 20 years.

The used data in the current study were collected from technical and economic documentaries of Sungun copper mine and published in technical reports about the mine. Also, the main part of data was extracted from annual statistical reports of mining industries of Iran by authors.

4.2. Model Description. At first, the current study created a model based on the relationships between effective factors to apply the SD approach. Then, a computer simulation of the studied system (i.e., environmental costs) has been performed with the implementation of the model by the software. Finally, the results have been evaluated and analyzed. The details of the implementation of this approach have been provided in the following sections.

This study used mine engineering experts' opinions, including three professors at the Urmia University of Technology and Yazd University, and seven experts from the mining engineering organization of Eastern Azerbaijan province, to determine relationships between identified factors and analyze the extent to which EC affect the total cost of open-pit mining. During a meeting, the experts determined the causal relationship between the constructive concepts of the EC, using the brainstorming technique and the Nominal Group Technique (NGT). Finally, this research has modeled a studied system based on the raw opinions of these experts. Also, this study created mathematical formulations of these linguistic relationships for use in Vensim software. Then, a graphical model has been depicted for the results, as shown in Figure 5, using the Vensim software. Graphs related to the extent of the impact of different factors have been obtained by implementing the Vensim based on the raw data. In addition to examining the basic condition (without applying effective factors related to EC), the current study defined different scenarios (see Table 2) to evaluate the impact of existing factors and their various combinations.

Regarding Table 1, this study defined the basic mode without considering the effectiveness of environmental solutions and incurring costs resulting from environmental solutions. Scenario 1 applies the costs of mine reclamation operations, including flattening of the mined land, relocation of vegetable soils, planting of trees compatible with postmining conditions, etc., in the model. In open-pit mining, these solutions are a part of the mining operation and are the last step of mining, which enters the dynamic calculations using the first scenario. Scenario 2 investigates the costs of environmental solutions related to the processing plant and concentrating the extracted ores in the dynamic model. Besides, this scenario considers the costs associated with the tailings dam. Using phytoremediation and wastewater treatment processes from the processing plant is also considered. Scenario 3 is defined based on the effect of environment-friendly mining operations costs on total open-pit mining costs. Also, in this scenario, green blasting is considered a suitable and sustainable solution for

(a) (b)

Figure 4: Location of the studied mine and a view of the mine [55].

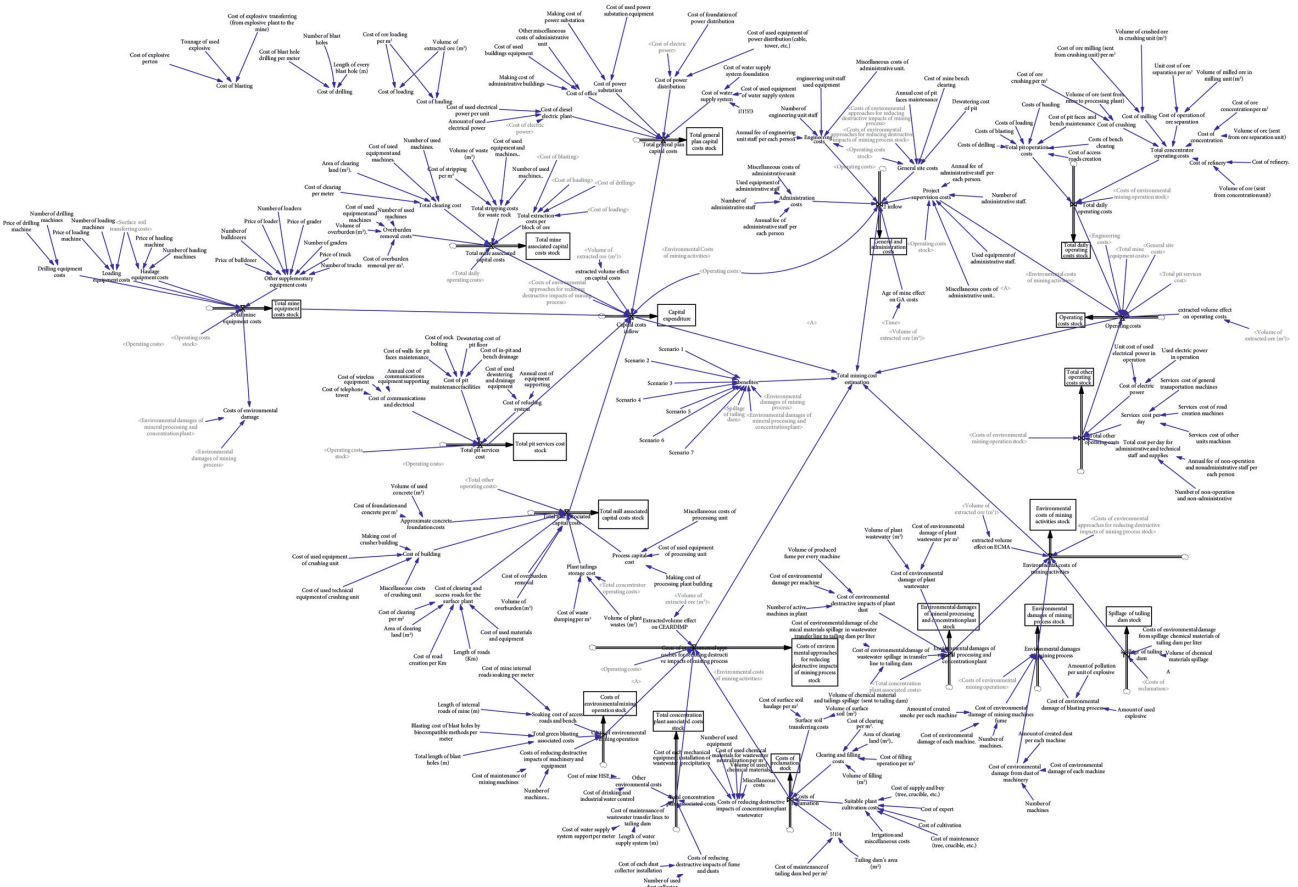


Figure 5: SD model designed in the current study.

biocompatible mining. Scenario 4 is developed based on a combination of Scenarios 1 and 2. This scenario simultaneously applies the impact of mine reclamation and environmental solutions related to the processing plant. Scenario 5 is the simultaneous application of Scenarios 1 and 3 in the SD model. Based on this scenario, costs associated with mine reclamation and biocompatible mining are applied in the model. Scenario 6 is also the simultaneous application of Scenarios 2 and 3, intending to consider the impact of the costs of environmental solutions related to the processing plant and biocompatible mining on total mining costs. Scenario 7 is another hybrid scenario, defined by simultaneous implementation of Scenarios 1, 2, and 3 in the SD model. Regarding this scenario, this study applies all costs related to environmental solutions to mining costs.

Each of the scenarios presented in the current study indicates the implementation of one or more green mining strategies to reduce the EC of the mining operation.

Table 2: The introduced scenarios based on the green mining strategies.

Scenario	Description
Basic mode	Dynamic analysis without considering the impact of environmental approaches on the whole system
Scenario 1	Dynamic analysis by applying the costs associated with mine reclamation operations
Scenario 2	Dynamic analysis by applying environmental costs related to the concentration and processing plant
Scenario 3	Dynamic analysis by applying the costs associated with eco-friendly mining operations
Scenario 4	Dynamic analysis by simultaneously applying the costs associated with mine reclamation and the environmental approach of the concentration and processing plant
Scenario 5	Dynamic analysis by simultaneously imposing costs associated with mine reclamation and eco-friendly mining operations
Scenario 6	Dynamic analysis by simultaneously applying the costs associated with the environmental approach of the concentration and processing plant and eco-friendly mining operations
Scenario 7	Dynamic analysis by simultaneously applying the costs associated with mine reclamation, the environmental approach of the concentration and processing plant and Eco-friendly mining operations

According to financial resources, management can implement one or more strategies, so this study examined the impact of each strategy on the mining activities and EC (total costs) in single and hybrid forms. The output of this study shows that management can make decisions and select any of the presented scenarios. From a mining point of view, each of the proposed scenarios is one of the well-known strategies of the mining engineering department to reduce costs. In the present study, there is a basic mode that determines the current state of the system. Defining these scenarios helps decision-makers to validate the model and evaluate the effectiveness of each scenario simultaneously. In other words, the proposed model can display the impact of implementing each scenario on the basic model and validate the outputs.

All the mathematical relationships among the factors used in Figure 3 were defined based on accounting management relationships. The present study used the experts' opinions to create a significant number of relationships considered in designing the SD model. Due to the great volume of the used relationships, only some of the most important relationships are discussed in the equations (1) to (6). The used parameters in the equations have been described in Table 3.

$$\mathrm{TMC} = (\mathrm{CE} + \mathrm{GA} + \mathrm{OC} + \mathrm{EC} + \mathrm{CEA}) - \mathrm{UEAP}, \tag{1}$$

$$\mathrm{CE} = (0.1 \times \mathrm{TGPC}) + (0.3 \times \mathrm{TMAC}) + (0.2 \times \mathrm{TMIC}) + (0.4 \times \mathrm{TME}), \tag{2}$$

$$\mathrm{GA} = (0.25 \times \mathrm{AC}) + (0.2 \times \mathrm{EC}) + (0.35 \times \mathrm{GSC}) + (0.2 \times \mathrm{PSC}), \tag{3}$$

$$\mathrm{OC} = (0.85 \times \mathrm{TDO}) + (0.15 \times \mathrm{TOO}), \tag{4}$$

$$\mathrm{EC} = (0.2 \times \mathrm{LTD}) + (0.5 \times \mathrm{EDM}) + (0.3 \times \mathrm{ECP}), \tag{5}$$

$$\mathrm{CEAB} = (0.5 \times \mathrm{CR}) + (0.3 \times \mathrm{TPA}) + (0.2 \times \mathrm{EDC}). \tag{6}$$

These equations show the relationships between the various factors used in the SD model. In this study, because the economic issues (i.e., mining costs and environmental costs of mining activities) have been considered, the relationships between different factors of linear type have been considered.

In some relationships, the coefficient α has been used to investigate the impact of environmental approaches, which is the effectiveness coefficient of each one of the environmental approaches and their combination equation (7). The amount for each scenario used to implement the model is shown in Table 4.

$$\mathrm{LTD} = (\mathrm{EDC} \times \mathrm{VCM}) - (\alpha \times \mathrm{CR}), \tag{7}$$

where α is the efficiency coefficient of the environmental approach and its value is presented in Table 4 for each scenario.

Since the purpose of the current study is to examine the total cost of mining activities, the target nodes of the SD model are total mining cost and environmental cost of mining operations. Furthermore, the main advantage of the proposed approach versus other analyzing models is the ability to define different and complex scenarios. In each scenario, the analyst may use different values in different levels for better mapping the model to the real world. Each scenario presented in the present study indicates the implementation of one or more strategies to reduce the EC of the mining operation. These strategies may be defined for potential events, harsher legislation, tax on pollution, and the related risks. The means of the "effectiveness" in the current study is the effect of each strategy that is presented individually or in the combination form of scenarios. The calculation of the effectiveness of each scenario is based on previous studies and experiences of mining and environmental experts in the real-world. For this purpose, the effectiveness of each scenario has been obtained based on the existing capability and capacity of each strategy.

4.3. Analysis of the Results. This section discusses the outputs of the proposed SD model in more detail. Figure 6 depicts the effectiveness of different scenarios using Vensim software.

Given the very large and relatively similar values of the effectiveness of some scenarios and the scale of the vertical axis of the chart (see Figure 6), the curves of the basic mode and Scenarios 1, 3, and, 5 are overlapped. In Figure 6, numerical details were not visible with good quality, and

Table 3: The important parameters used in the economic evaluation of environmental costs of open-pit mining.

Parameter	Description
AC	Administration costs
AFA	Annual fee of administrative staff per person
CE	Capital expenditure
CEA	Costs of environmental approaches for reducing destructive impacts
CR	Costs of reclamation
EC	Environmental costs
ECP	Environmental damages of concentration plant
EDC	Cost of the environmental damage of chemical material spillage while transferring waste-water to a tailing dam per liter
EDM	Environmental damages of the mining process
EDMP	Environmental damages of mineral processing and concentration plant
EGC	Engineering costs
EMO	Costs of the environmental mining operation
EPC	Cost of electric power
GA	General and administration costs
GSC	General site costs
LTD	Leakage of tailing dam
MAU	Miscellaneous costs of the administrative unit
NAS	Number of administrative staff
OC	Operating costs
PSC	Project supervision costs
RDI	Costs of reducing destructive impacts of machinery and equipment
SPC	Suitable plant cultivation costs
SPD	Service cost per day
TCA	Total concentration plant-associated costs
TCO	Total concentrator operating costs
TCP	Total cost per day for administrative and technical staff and supplies
TDO	Total daily operating costs
TGB	Total green blasting-associated costs
TGPC	Total general plan capital costs
TMAC	Total mill-associated capital costs
TMC	Total mining cost estimation
TME	Total mine equipment costs
TMIC	Total mine-associated capital costs
TOO	Total other operating costs
TPA	Total concentration plant-associated costs
TPC	Total pit services cost
TPO	Total pit operation costs
UEA	Equipment used by administrative staff
UEAP	Used environmental approach profit
VCM	The volume of chemical material and tailing spillage
WSR	Cost of water sprinkling at the bed of roads and benches
α	Efficiency coefficient of the environmental approach

Table 4: Coefficient of efficiency of different scenarios of mining based on environmental approaches.

Scenario	Coefficient of efficiency α
Basic mode	0
Scenario 1	0.85
Scenario 2	0.80
Scenario 3	0.60
Scenario 4	0.82
Scenario 5	0.72
Scenario 6	0.70
Scenario 7	0.75

changes were displayed in a straight line, according to the large scale of the numbers and the rounding of the numbers using Vensim software. However, in the following figures, oscillations and behaviors were visible, which were derived from the scenario of each approach individually.

Figure 6 indicates that total mining costs have been reduced significantly due to the effectiveness of scenarios related to environmental issues. Although the implementation of various environmental scenarios in open-pit mines appeared to increase mining costs, the estimation of total mining costs showed the decreased total mining costs. The reason behind this significant reduction in total mining costs is the nonimposition of environmental issues-related costs, such as fines for environmental pollution in and around the mine, changing the ecosystem of the region, destruction of natural resources, etc. If the standards are not met, owners and managers of open-pit mines should pay high environmental fines for the reclamation of the whole mining area and solving the environmental problems.

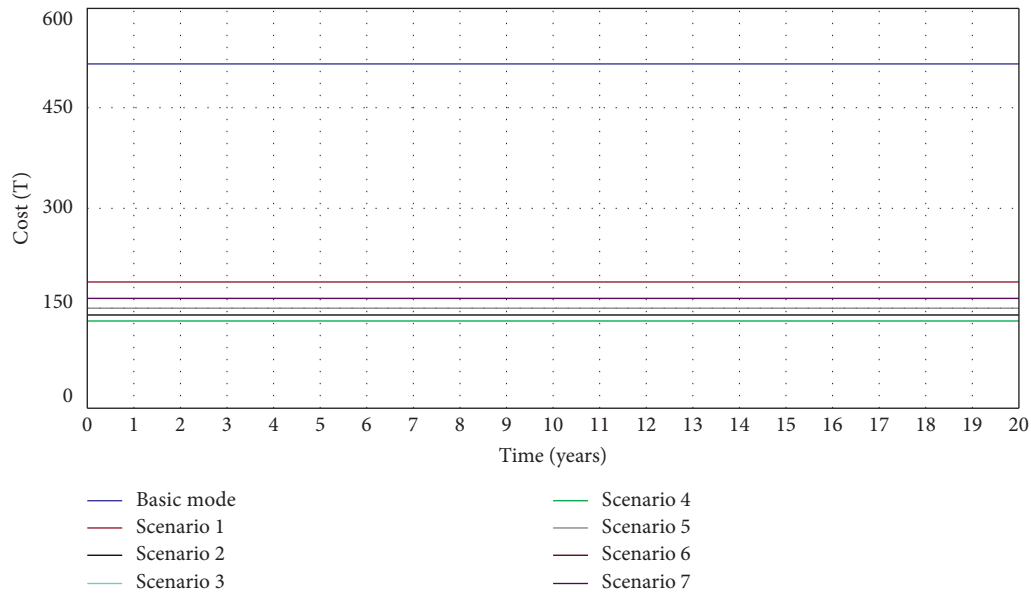


Figure 6: Effectiveness of different scenarios based on the output of the system dynamics model.

In some cases, failure to pay attention to environmental problems leads to mine closure by environmental protection organizations. This figure also shows that the total cost of open-pit mining operations is reduced if low costs related to the application of environmental solutions are spent. Thus, this issue is of cardinal importance for mine managers. Accordingly, the importance of the impact of costs related to environmental solutions, especially in large metal mines, on the total cost of mining is very significant.

On the other hand, implementing appropriate environmental approaches in open-pit mines can help policymakers reach sustainable development purposes. Moreover, implementing these strategies will lead to a reduction in total mining costs. Figure 6 illustrates that a significant amount of the total cost of open-pit mining has been reduced by the implementation of many approaches. Implementing Scenario 4 led to the highest reduction of costs, whereas implementing Scenario 6 had the lowest impact on the cost reduction. Small fluctuations associated with the changes resulting from the impact of various variables on the system are not visible in Figure 6. The reason is that each graph had extremely large numerical scales, and the number of fluctuations was extremely low compared to the values in each graph. Therefore, it is necessary to separately analyze the graphs of different scenarios, in which these oscillations are observed (see Figures 7 to 14).

According to the results of solving the proposed dynamic model using Vensim software, the impact of costs of environmental approaches on the estimation of total mining costs of open-pit mines and other costs of different mining processes is shown in Figure 7. Notably, rounding the outputs by Vensim software has resulted in repeating the numbers in the vertical axis of this figure. Although the general trend in the graphs of Figure 7 was ascending, the graphs had slight fluctuations, and slopes were mild. These fluctuations indicated that EC had a relatively mild growth by increasing the amount of ore extraction (and increasing the environmental damage caused by open-pit mining activities). Moreover, the rate of increase was moderate in total mining costs.

The final section of Figure 6 was relatively flat, which was about the last three years of mining activities; the sign was that total mining costs would remain constant with increasing EC. The greatest rate of decline in total mining costs can be observed in graph (e) of Figure 7, related to Scenario 4, after implementing the mine reclamation operations and the environmental approach of the concentration and processing plant (accompanied by costs). Moreover, implementing Scenario 7 (simultaneous implementation of the three environmental approaches) and Scenario 6 (simultaneous implementation of the environmental approach of the concentration and processing plant and the eco-friendly mining operation) led to the high total cost reduction. This is a significant reduction in the total cost of mining caused by environmental deterrence, which is considered in rules, standards, and environmental guidelines. It highlights the importance of mining-related environmental problems, which are considered unavoidable in open-pit mining activities. This means that by spending fewer costs on environmental solutions, the mine managers can see significant profits during the various years of mining activity.

In modeling using the SD approach, when the loop is formed, there must be a state variable to provide conditions to solve the model. Given that the Vensim software has a causal structure, the need for initial conditions was for the variables in the model to solve the model based on available values. In the current study, all the variables are dependent, and none of them had constant values. Therefore, it is necessary to define the flow of inputs to the system based on different variables and then use the delay function (delay value) to make the model solvable. This method of solving the SD models is considered a new solution approach, and one of the contributions of the present research is to solve complex problems. For example, the equation for project supervision costs is as follows:

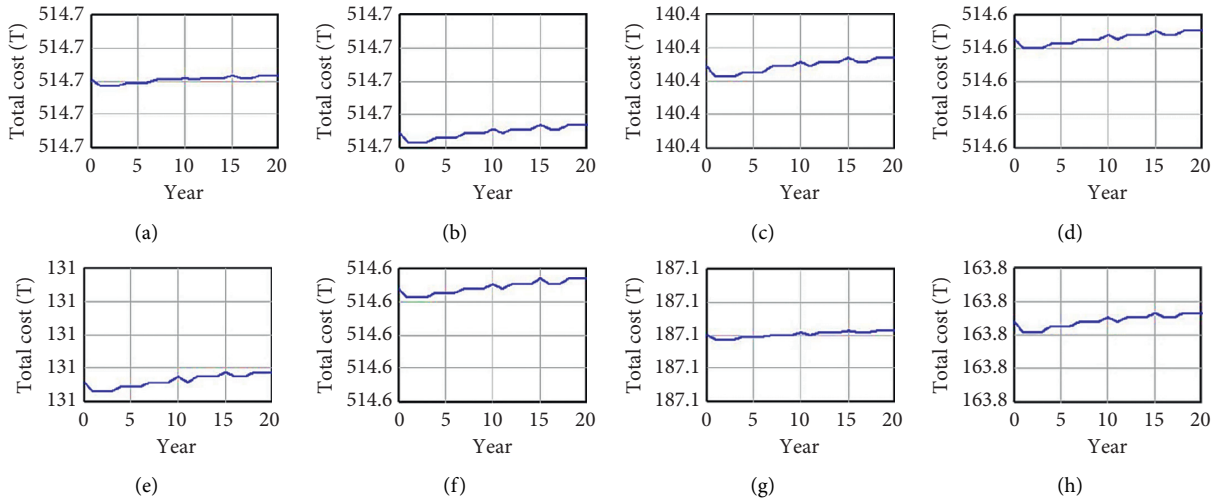


Figure 7: Estimated total costs of mining. (a) Basic mode; (b) Scenario 1; (c) Scenario 2; (d) Scenario 3; (e) Scenario 4; (f) Scenario 5; (g) Scenario 6; (h) Scenario 7.

$$\text{PSC} = (0.65 \times \text{AFA} \times \text{NAS}) + (0.25 \times \text{UEA}) + (0.05 \times \text{MAU}) + (0.05 \times \text{OC}). \quad (8)$$

In the following, $(0.05 \times \text{OC})$ is replaced by equation (9):

$$(\text{OC Stock} - \text{Delay 1}(\text{OC Stock}, 1)) \times 0.05. \quad (9)$$

Moreover, the present study used binary logic to activate each of the variables to analyze the effectiveness of different scenarios. To put it precisely, the binary logic has been used to apply the individual effects of each scenario and analyze the effectiveness of each of the criteria and parameters. This means that by assigning a value of zero or one to each of the scenarios, the outputs associated with each scenario and its impact on the whole system and other components are determined. This study used the values one and zero to activate and inactivate each node or component, respectively. The cost-time graphs for the basic mode (regardless of EC) in the designed dynamic model are shown in Figure 8 by considering the initial values after implementing the model with the flow of actions related to the basic state.

Figure 8 is related to solving the model for the inflow of the variables based on Scenario 1 (the basic mode). The graphs (a), (b), (c), and (e) in Figure 8 are goal-oriented functions. In goal-oriented patterns of SD models, an objective function attempts to achieve the system purpose by increasing or decreasing itself, and there is a constant trend after reaching the goal value. These functions are observed in loops, which seek to equilibrate the function and are called inventory control balancing loops. For this purpose, it is necessary to identify these loops in the model. These loops for graphs (a), (b), (c), and (d) are as follows.

According to graph (a) in Figure 8, the system had a goal-oriented structure in this curve. To this end, the effect of different variables increased total mining costs to the highest level (660 million monetary units in the next ten years of mining). Then, the system was seeking another goal due to the effects of the other variables, which occurred in the final ten years of mining, and the level of costs decreased (649 million monetary units). This cost reduction continued gradually until several steps (step-by-step) due to system retargeting. Retargeting means that after the calculations for each loop and all of the stocks based on a specific goal (each scenario), the system prepares itself to perform other calculations by changing the good value of the resulting value and returning to the base state. This means that by reaching the target circuit based on the causal relationships in the system, the other system continues its calculations for each stock and each current by returning to the starting point of the calculations. This process is done permanently and dynamically while solving the model. Finally, it reached its ultimate goal (600 million monetary units) in the last ten years of mining. The effective variables in the loop related to this goal-oriented pattern are shown in Figure 9, by reviewing the designed model. The greatest impact was on the operating costs of mining, which acted as a limiting factor. Now, by assuming a review of the operating cost, the equation for this factor is as follows:

$$\text{OC} = (\text{TME} + \text{TPC} + \text{TDO} + \text{TOO} + \text{EGC} + \text{GSC} + \text{PSC}) \times (1 - \text{extracted volume effect on OC}). \quad (10)$$

Notably, the extracted volume effect on OC had a negative impact on the system. This structure was such that, up to the middle of the first quartile, c, the positive part overcame the negative sector. In other words, the extracted volume increased year by year, depending on the mine's age, until the extracted volume effect reached 1. The same amount added to the system was removed from the system, causing the inventory or system state to remain constant.

Moreover, according to graph (b) in Figure 8, after increasing the total cost of mining, it remained constant at its final goal (400 million). The loop associated with this graph is shown in Figure 10. The operating costs in this graph were the limiting factor, which controlled the total cost of mining. However, with the assumption that the operating cost was reassessed, the equation for this factor

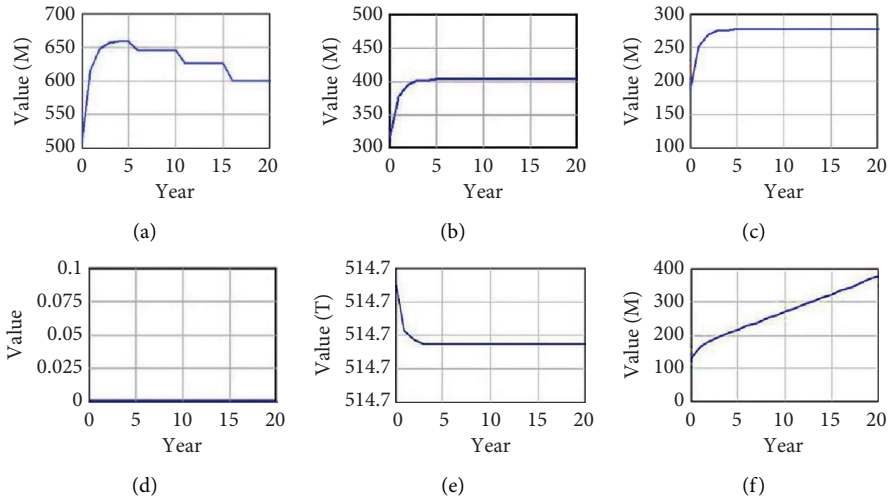


Figure 8: Dynamic model results for basic mode with apply inflow for variables. (a) General and administration costs; (b) costs of environmental approaches for reducing destructive impacts of the mining process; (c) operating costs; (d) profit; (e) environmental costs of mining activities; (f) capital costs.

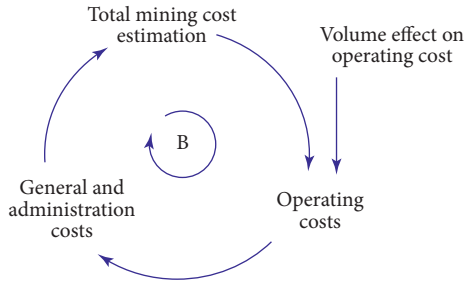


Figure 9: The loop of variables affecting graph (a) in Figure 7 (basic mode).

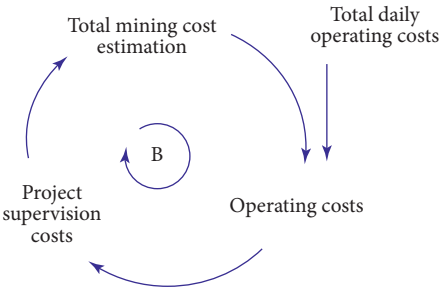


Figure 10: The loop for variables affecting graph (b) in Figure 7 (Scenario 1).

affected by another system was investigated using equation (11):

$$OC = (TPO + TCO + EPC + SPD + TCP) \times (1 - TDO). \quad (11)$$

Total daily operating costs had a negative impact on the system. This structure was such that the positive part overcame the negative part up to the middle of the first quartile of total daily operating costs (TDO) with the year-on-year increase in mining efficiency. When TDO reached 1, the same amount added to the system was removed from it, which caused the inventory or system situation to remain constant.

In addition, graph (c) in Figure 8 shows that if the variables related to environmental approaches were not applied (Scenario 1), the system would have a goal-oriented function, and after reaching its goal value (275 million), it showed a constant and steady trend. The loop associated with these conditions and the variables affecting the whole system is shown in Figure 11. The variables related to EC caused by mining activities were the limiting factor for this whole system goal situation, which was considered a negative cost. By assuming the EC examination, the equation for this factor has been investigated, as equation (12):

$$EC = (EDMP + EDM + LTD) \times (1 - \text{extracted volume effect on EC}). \quad (12)$$

It can be observed that the extracted volume effect on EC had a significant negative impact on the system. This structure was such that the positive part overcame the negative part, up to the middle of the first quartile of Figure 10. In other words, EC increased annually, and when EC reached the value of 1, the same amount added to the system has been removed from which.

Graph (d) in Figure 8 shows that if managers have not implemented the environmental approaches (basic mode), there would be no significant profit from green mining. Notably, the profit is positive in this figure, but it is a very small amount, and the profit for each studied scenario refers to the reduction in environmental costs stemmed from applying its relevant green mining strategies/environmental solutions. In other words, in the proposed model, the profitable variables were related to EC. Therefore, if the environmental variables were not used, there would be no profit to the mine for many years.

Moreover, graph (e) in Figure 8 shows that by applying the current inflow of the environmental approaches to the system,

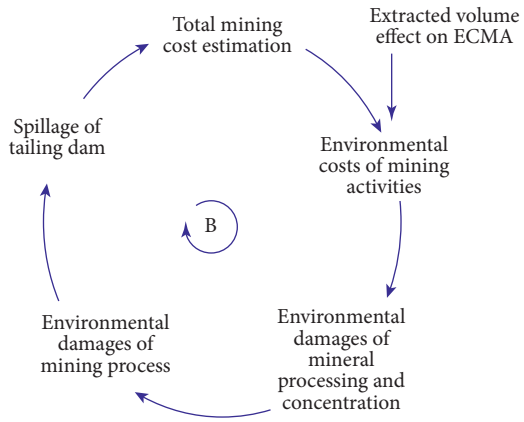


Figure 11: The loop associated with the variables affecting graph (c) in Figure 7 (Scenario 2).

total mining costs had a goal function and were steadily decreasing during the first six years of mining. After total mining costs reached the goal value, they showed a constant trend and remained uniform until the end of the mine's life. The equilibrium loop of this graph is shown in Figure 12. This figure illustrates that the subsets of the costs of environmental approaches were the limiting factors of the system. For analysis, by assuming CEA, the equation for this factor is as follows:

$$\text{CEA} = (\text{EDMP} + \text{EDM} + \text{LTD} + \text{EMO} + \text{TCA} + \text{SPC} + \text{WSR} + \text{TGB} + \text{RDI}) - (1 \times \text{extracted volume effect on CEA}). \quad (13)$$

Accordingly, the extracted volume effect on CEA had a significant negative impact on the system. This structure was such that the positive part overcame the negative part, up to the middle of the first quartile of Figure 11. In other words, extracted volume effect of CEA increased annually, and when this amount reached 1, the same amount added to the system has been removed from which, causing the inventory or system state to remain constant. That means that the volume of ore extraction is directly related to the amount of pollution and environmental degradation. Also, the costs of unprincipled extraction of open-pit mines affect EC and total mining costs.

As shown in graph (f) in Figure 8, the capital inflow was not exponential, but its function was incremental and had a behavior similar to exponential functions. The start point of this graph is exponential and then has a quasilinear trend. However, the whole chart is mostly in an exponential form. Hence, it can be considered with the exponential growth pattern used in dynamic analysis. This graph had an ascending function. Accordingly, the graph showed a strong upward trend in the first few years of operation, and then it had an ascending trend, indicating that total mining costs were moderately ascending. In the final years of the mine's life, total mining costs were proven to be almost constant and had a reduced slope, indicating that its system had been recovered. The loop associated with this variable and its relation to other variables is shown in Figure 13. It showed that the underlying capital expenditure affects the whole system.

In this case, assuming that total mining costs were evaluated, the equation for this factor was investigated as equation (14).

$$\text{TMC} = (\text{OC} + \text{CE} + \text{TGPC} + \text{TME} + \text{TMAC}) + (0.05 \times t \times \text{CE}). \quad (14)$$

Total mining costs (TMC) were observed to be ascending under the influence of several factors. Moreover, the effect of time and the almost constant increasing rate in capital expenditures, which was a small percentage, maintained its upward trend.

Given the similar situation of the cost-time graphs to implement different scenarios (with environmental costs), their dynamic analysis was the same as the one mentioned above. Moreover, by applying various currents related to the environmental variables in the proposed model, there were no changes in various graphs related to each scenario. Therefore, the analyses related to the basic mode of affairs were applied to them. In the following, only graphs with extreme changes have been analyzed (see Figure 14).

Regarding Figure 8, it can be seen that the estimated total costs of mining fluctuate in different scenarios for the studied period, and this, in turn, leads to a change in profit. But as mentioned before, the small fluctuations in Figure 14 are not visible due to the scale of the vertical axis of the diagrams. For this reason, a steady trend is observed. Accordingly, in Figure 14, it can be seen that implementing Scenario 1 did not yield any profit, and it caused about 107 million losses, which was almost uniform throughout the years of mining. However, with implementing Scenario 2, the profit generated by this environmental approach was 375 trillion, and during the years of the mine's life, there was an almost constant trend. Moreover, the profit of implementing Scenario 3 was approximately 25 billion, the profit of implementing Scenario 4 was about 380 trillion, and implementing Scenario 5 had an equivalent value of 30 billion.

By implementing Scenario 6, the profit was about 325 trillion, and implementing Scenario 7 yielded 350 trillion profits for the mine. Notably, implementing Scenario 4 had the most profit. By comparing this result with the graphs in Figures 7 and 8, it can be seen that Scenario 4 had the largest reduction in total mining costs and also was the most profitable scenario. According to this scenario, the simultaneous implementation of the mine reclamation and the environmental approach of the concentration and processing plant in the mine significantly reduced the total costs

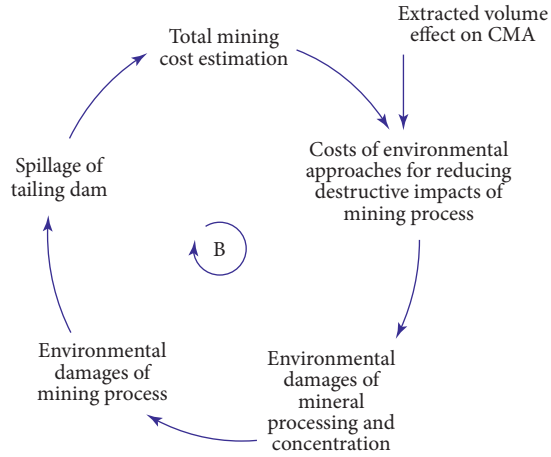


Figure 12: The loop associated with the variables affecting graph (e) in Figure 7 (Scenario 4).

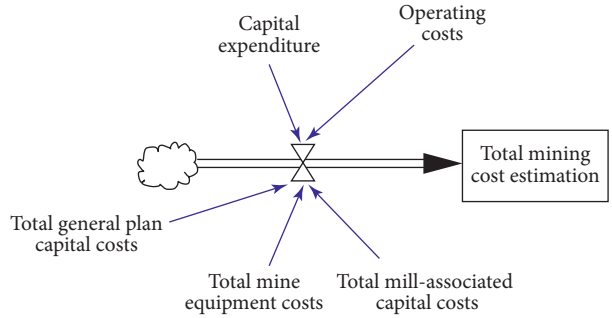


Figure 13: The loop associated with the variables affecting graph (f) in Figure 7 (Scenario 5).

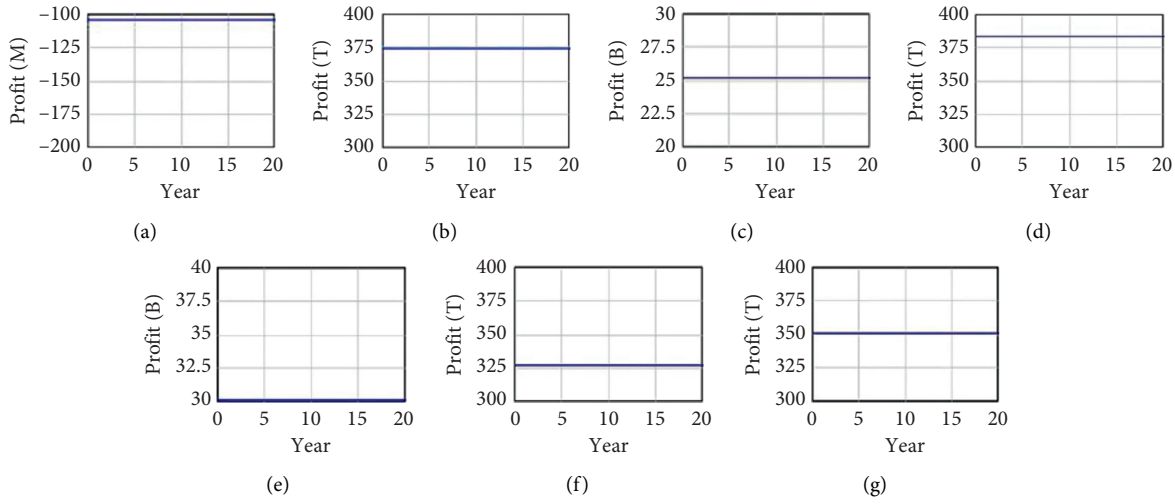


Figure 14: The results of solving the dynamic model for the profit variable from each scenario: (a) Scenario 1; (b) Scenario 2; (c) Scenario 3; (d) Scenario 4; (e) Scenario 5; (f) Scenario 6; (g) Scenario 7.

of open-pit mining. This cost reduction can also be attributed to affecting all mining and postmining activities. Regarding implementing Scenario 4 in the studied mine, in addition to the economic arguments presented in the current study, engineering calculations should be made in this regard. The implementation of Scenario 7 was also desirable and did not exclude environmental degradation related to mining activities, especially in open-pit mines, which is consistent with the goals of mining sustainable development. In addition to the positive features of Scenario 4, Scenario 7 also focuses on eco-friendly mining operations. Due to the legal requirement to implement mine reclamation strategy in open-pit mines, mentioned in Scenarios 4 and 7, at the end of their life, the potential of such mines can be used to

develop the Geotourism industry and make a profit in the postmining period.

On the other hand, the main purpose of this study was to determine the modeling boundaries of the studied problem. To put it precisely, this study was aimed at identifying the cost components in open-pit mining operations and defining the causal relationships between these identified components according to the role of environmental costs. Providing such a system allows the decision-maker to monitor the situation by defining different scenarios and adding other economic variables such as inflation. Considering such variables outside the model's boundaries can lead to a significant positive trend in Figure 14. However, estimating such factors is challenging due to the uncertainty in the economic conditions.

In general, due to the importance of social problems in parallel with economic and environmental conditions, it is necessary to mention that by applying environmental solutions, in addition to reducing the total cost of mining operations, social issues are also somewhat resolved. Regarding social issues, if decision-makers and policymakers pay attention to environmental issues, the migration of the natives of the area and the inhabitants of the villages around the mine will decrease. Besides, due to the reduction of pollution from the mining operation and processing plants using environmental solutions, agriculture in the studied region, especially in the areas around the mine, will not be harmed. Moreover, expanding biocompatible mining activities can also increase employment in this region.

5. Conclusion

Given the importance of sustainability concept in managing today's industries, there is a need to evaluate its constituent components quantitatively. In this regard, the current study tried to analyze the EC's effect of green mining strategies on the total costs of mining activities using the SD approach. Therefore, due to the importance of the total cost of the mine and the EC, this study created scenarios based on green mining strategies. They proposed to emphasize the concept of sustainable development in the mining sector and related subdivisions. To assess the achievement of sustainable goals, the SD model has formulated these senarios based on the mathematical relationships between the effective parameters. The outputs of the proposed SD model indicated that Scenario 4 is the optimal scenario, compared to other defined scenarios. When Scenario 4 was implemented, the system guaranteed the highest profit, about 380 million, because of applying the largest decrease in the total costs of mining. The simultaneous use of the eco-friendly mining operations and the green mining strategy for the concentration and processing plant through Scenario 4 indicates the greatest impact on cost reduction in the mining operation. Using this approach shows the fact that by reducing the environmental damage of mining activities, not only is there no need for mining managers to pay a fine but also there is no need to spend extra funds on applying corrective measures in this regard. This profitability stemmed from considering environmental issues in mining activities that can reduce the total cost of open-pit mining projects.

Data Availability

The data that support the findings of the current study are available from the corresponding author, Amir Jafarpour, upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest..

References

[1] E. Bakhtavar, S. Yousefi, and A. Jafarpour, "Evaluation of shaft locations in underground mines: fuzzy multi-objective optimization by ratio analysis with fuzzy cognitive map weights," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 119, no. 10, pp. 855–864, 2019.

[2] E. T. Asr, R. Kakaie, M. Ataei, and M. R. Tavakoli Mohammadi, "A review of studies on sustainable development in mining life cycle," *Journal of Cleaner Production*, vol. 229, pp. 213–231, 2019.

[3] S. Ghadernejad, A. Jafarpour, and P. Ahmadi, "Application of an integrated decision-making approach based on FDAHP and PROMETHEE for selection of optimal coal seam for mechanization; A case study of the Tazareh coal mine complex, Iran," *International Journal of Mining and Geo-Engineering*, vol. 53, no. 1, pp. 15–23, 2019.

[4] P. A. Dowd, "Risk assessment in reserve estimation and open-pit planning," *Transactions of the Institution of Mining and Metallurgy (Section A: Mining Industry)*, vol. 103, 1994.

[5] F. Pavloudakis, C. Roumpos, and M. Galetakis, "Public acceptance of surface mining projects and the determination of the marginal environmental cost," *International Journal of Mining, Reclamation and Environment*, vol. 26, no. 4, pp. 292–308, 2012.

[6] A. Jafarpour, J. A. Sharif, and A. Eivazi, "Reducing destructive environmental impacts of Sungun copper mine effluents with using of phytoremediation processes," *International Journal of Pure & Applied Bioscience*, vol. 5, no. 2, pp. 43–55, 2017.

[7] M. M. A. Malakoutian and M. Khaksar, "SBM model based productivity evaluation," *ENG Transactions*, vol. 1, pp. 1–9, 2020.

[8] R. Qi, T. Liu, Q. Jia, L. Sun, and J. Liu, "Simulating the sustainable effect of green mining construction policies on coal mining industry of China," *Journal of Cleaner Production*, vol. 226, pp. 392–406, 2019.

[9] T. Dutta, K.-H. Kim, M. Uchimiya et al., "Global demand for rare earth resources and strategies for green mining," *Environmental Research*, vol. 150, pp. 182–190, 2016.

[10] M. Jahangoshai Rezaee, S. Yousefi, M. Eshkevari, M. Valipour, and M. Saberi, "Risk analysis of health, safety and environment in chemical industry integrating linguistic FMEA, fuzzy inference system and fuzzy DEA," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 1, pp. 201–218, 2020.

[11] A. Warhurst and M. L. Noronha, Eds., *Environmental Policy in Mining: Corporate Strategy and Planning*, CRC Press, 1999.

[12] G. Elliot and G. Harris, "A cost-benefit analysis of landmine clearance in Mozambique," *Development Southern Africa*, vol. 18, no. 5, pp. 625–633, 2001.

[13] W. A. Byrd and B. Gildestad, "The Socio-economic impact of mine action in Afghanistan: a cost-benefit analysis," *Afghan Digital Libraries*, 2001.

[14] J. Radev, "Economic analysis of investment projects in mining industry," *Annual of University of Mining and Geology "St. Ivan Rilski" Humanitarian and Economic Sciences*, vol. 46, no. IV, pp. 66–71, 2003.

[15] F. Rashidinejad, M. Osanloo, and B. Rezai, "An environmental oriented model for optimum cut-off grades in open pit mining projects to minimize acid mine drainage," *International Journal of Environmental Science & Technology*, vol. 5, no. 2, pp. 183–194, 2008.

[16] J. A. Botín and M. A. Vergara, "A cost management model for economic sustainability and continuos improvement of mining operations," *Resources Policy*, vol. 46, pp. 212–218, 2015.

[17] S. Kusi-Sarpong, C. Bai, J. Sarkis, and X. Wang, "Green supply chain practices evaluation in the mining industry using a joint rough sets and fuzzy TOPSIS methodology," *Resources Policy*, vol. 46, pp. 86–100, 2015.

[18] S. Narrei and M. Osanloo, "Optimum cut-off grade's calculation in open pit mines with regard to reducing the undesirable environmental impacts," *International Journal of Mining, Reclamation and Environment*, vol. 29, no. 3, pp. 226–242, 2015.

[19] X.-c. Xu, X.-w. Gu, Q. Wang et al., "Production scheduling optimization considering ecological costs for open pit metal mines," *Journal of Cleaner Production*, vol. 180, pp. 210–221, 2018.

[20] Z. Xu, L. Yao, and X. Chen, "Urban water supply system optimization and planning: Bi-objective optimization and system dynamics methods," *Computers & Industrial Engineering*, vol. 142, p. 106373, 2020.

[21] H. Guo, H. Nguyen, D. A. Vu, and X. N. Bui, "Forecasting mining capital cost for open-pit mining projects based on artificial neural network approach," *Resources Policy*, Article ID 101474, 2019.

[22] L. Brescia-Norambuena, D. Pickel, M. González, S. L. Tighe, G. Azúa, and G. Azúa, "Accelerated construction as a new approach for underground-mining pavement: productivity, cost and environmental study through stochastic modeling," *Journal of Cleaner Production*, vol. 251, p. 119605, 2020.

[23] K. Spitz and J. Trudinger, *Mining and the Environment: From Ore to Metal*, CRC Press, Boca Raton, FL, USA, 2019.

[24] R. Sivakumar, D. Kannan, and P. Murugesan, "Green vendor evaluation and selection using AHP and Taguchi loss functions in production outsourcing in mining industry," *Resources Policy*, vol. 46, pp. 64–75, 2015.

[25] H. Dehghani and M. Ataee-pour, "Determination of the effect of operating cost uncertainty on mining project evaluation," *Resources Policy*, vol. 37, no. 1, pp. 109–117, 2012.

[26] H. Dehghani, M. Ataee-pour, and A. Esfahanipour, "Evaluation of the mining projects under economic uncertainties using multidimensional binomial tree," *Resources Policy*, vol. 39, pp. 124–133, 2014.

[27] M. de Werk, B. Ozdemir, B. Ragoub, T. Dunbrack, and M. Kumral, "Cost analysis of material handling systems in open pit mining: case study on an iron ore prefeasibility study," *The Engineering Economist*, vol. 62, no. 4, pp. 369–386, 2017.

[28] E. Bakhtavar, A. Jafarpour, and S. Yousefi, "Optimal production strategy of bimetallic deposits under technical and economic uncertainties using stochastic chance-constrained programming," *Journal of Mining and Environment*, vol. 8, no. 3, pp. 475–485, 2017.

[29] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "Bayesian hierarchical model-based information fusion for degradation analysis considering non-competing relationship," *IEEE Access*, vol. 7, pp. 175222–175227, 2019.

[30] R. Mikaeil, A. Esmailzadeh, S. Aghaei et al., "Evaluating the sawability of rocks by chain-saw machines using the promethee technique," *Rudarsko-geološko-naftni Zbornik*, vol. 36, no. 1, 2021.

[31] S. Shojaei, A. Jafarpour, S. Shojaei, Y. Gyasi-Agyei, and J. Rodrigo-Comino, "Heavy metal uptake by plants from wastewater of different pulp concentrations and contaminated soils," *Journal of Cleaner Production*, vol. 296, p. 126345, 2021.

[32] F. Albrecht, O. Kleine, and E. Abele, "Planning and optimization of changeable production systems by applying an integrated system dynamic and discrete event simulation approach," *Procedia CIRP*, vol. 17, pp. 386–391, 2014.

[33] M. J. Radzicki, "System dynamics and its contribution to economics and economic modeling," *System Dynamics*, pp. 401–415, 2020.

[34] B. Hannon and M. Ruth, "Modeling dynamic biological systems," in *Modeling Dynamic Biological Systems*, pp. 3–28, Springer, Berlin, Germany, 2014.

[35] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "A Bayesian approach for degradation analysis with individual differences," *IEEE Access*, vol. 7, pp. 175033–175040, 2019.

[36] J. Liu, Y. Liu, and X. Wang, "An environmental assessment model of construction and demolition waste based on system dynamics: a case study in Guangzhou," *Environmental Science and Pollution Research*, vol. 27, no. 30, pp. 37237–37259, 2020.

[37] P. Norouzian-Maleki, H. Izadbakhsh, M. Saberi, O. Hussain, M. Jahangoshai Rezaee, and N. GhanbarTehrani, "An integrated approach to system dynamics and data envelopment analysis for determining efficient policies and forecasting travel demand in an urban transport system," *Transportation Letters*, pp. 1–17, 2020.

[38] D. L. Cooke, "A system dynamics analysis of the Westray mine disaster," *System Dynamics Review*, vol. 19, no. 2, pp. 139–166, 2003.

[39] D. Wang, R. Nie, R. Long, R. Shi, and Y. Zhao, "Scenario prediction of China's coal production capacity based on system dynamics model," *Resources, Conservation and Recycling*, vol. 129, pp. 432–442, 2018.

[40] S. B. M. Lagnika, R. Hausler, and M. Glaus, "Modeling or dynamic simulation: a tool for environmental management in mining?," *Journal of Integrative Environmental Sciences*, vol. 14, no. 1, pp. 19–37, 2017.

[41] D. M. Mchaina, "Environmental planning considerations for the decommissioning, closure and reclamation of a mine site," *International Journal of Surface Mining, Reclamation and Environment*, vol. 15, no. 3, pp. 163–176, 2001.

[42] K. A. Hudson-Edwards, H. E. Jamieson, and B. G. Lottermoser, "Mine wastes: past, present, future," *Elements*, vol. 7, no. 6, pp. 375–380, 2011.

[43] J. Abdollahei Sharif, A. Jafarpour, and S. Yousefi, "A hybrid fuzzy MCDM approach to determine an optimal block size in open-pit mine modeling: a case study," *Journal of Mining and Environment*, vol. 11, no. 2, pp. 611–627, 2020.

[44] W. A. Hustrulid, M. Kuchta, and R. K. Martin, *Open Pit Mine Planning and Design, Two Volume Set & CD-ROM Pack*, CRC Press, Berlin, Germany, 2013.

[45] H. L. Hartman, *SME Mining Engineering Handbook*, Society for Mining, Metallurgy, and Exploration, Inc., Littleton, CO, USA, 2nd Edition, 1992.
[46] H. Soltanmohammadi, M. Osanloo, and A. Aghajani Bazzazi, "An analytical approach with a reliable logic and a ranking policy for post-mining land-use determination," *Land Use Policy*, vol. 27, no. 2, pp. 364–372, 2010.
[47] F. J. Golrokh, G. Azeem, and A. Hasan, "Eco-efficiency evaluation in cement industries: DEA malmquist productivity index using optimization models," *ENG Transactions*, vol. 1, pp. 1–8, 2020.
[48] T. F. Torries, "Evaluating mineral projects: applications and misconceptions," Society for Mining Metallurgy and Exploartion, Denver Colorado, USA, (SME), 1998.
[49] T. Isaksson, *Model for Estimation of Time and Cost Based on Risk Evaluation Applied on Tunnel Projects*, Doctoral Dissertation, Byggvetenskap Royal, Institute of Technology Stockholm, Stockholm, Sweden, 2002.
[50] P. A. Dowd and C. Xu, "Financial evaluation of mining projects," in *Proceedings of the APCOM XXV 1995 Conference*, pp. 247–254, Brisbane, Australia, July 1995.
[51] J. W. Forrester, *Some Basic Concepts in System Dynamics*, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, UK, 2009.
[52] N. Choucri, D. Goldsmith, S. Madnick, D. Mistree, J. B. Morrison, and M. Siegel, *Using System Dynamics to Model and Better Understand State Stability*, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007.
[53] J. D. Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*, McGraw-Hill, Boston Raton, FL, USA, 2000.
[54] D. M. Fisher, *Modeling Dynamic Systems: Lessons for a First course*, Teachers and Trainer's Guide, 2005.
[55] www.farsnews.ir.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 8069873, 8 pages
https://doi.org/10.1155/2021/8069873

Research Article

A Novel Ultrawideband Gear-Shaped Dielectric Ring Resonator Antenna

Mehdi Ghorbani and Habib Ghorbaninejad

Department of Electrical Engineering, University of Guilan, Rasht, Iran

Correspondence should be addressed to Mehdi Ghorbani; ghorbani.me@gmail.com

Received 20 May 2021; Accepted 29 June 2021; Published 13 July 2021

Academic Editor: Noorbakhsh Amiri Golilarz

In this study, a novel ultrawideband (UWB) dielectric ring resonator (DRR) antenna has been proposed. DRR antennas include a single monopole antenna in the center of a ground plane and a dielectric with a symmetric structure around the monopole. This structure will lead to ultrawide band antenna. However, it is still possible to enhance the antenna bandwidth. In this study, we combine the DRR structure with an array antenna. The proposed antenna includes a circular array of four triangle resonators, which is rotated around the center of the triangle base to form a gear-shaped ring resonator antenna. In this design, characteristics of all these antennas are combined to enhance the antenna bandwidth including triangular dielectric resonator, circular array antenna, dielectric ring resonator structure, and a quarter-wave electric monopole. Triangular dielectric resonator antennas are wideband and in small size. Ring resonator antennas are inherently ultrawideband. Quarter-wave electric monopole and circular array structure can also enhance antenna bandwidth. This novel shape of the DRR antenna possesses the wider impedance bandwidth compared to similar works. Impedance bandwidth is 150% (5.2–36.1 GHz), and the bandwidth ratio is 1 : 6.9, which is much greater than earlier reports.

1. Introduction

Dielectric resonator antennas (DRA) are very attractive topics due to a great variety of reasons. These antennas' benefits include small size, low cost, high radiation efficiency, and low losses due to the elimination of the conductor loss. Moreover, UWB antennas can widely be used in portable broadband wireless, radar, navigation, GPS, the biomedical system, mobile satellite communications, the direct broadcast system, and telemetry. For wideband applications, a variety of techniques were applied to improve the bandwidth of these antennas. Dielectric ring resonator (DRR) antennas are a kind of DRAs which intrinsically possess wide bandwidth.

In 2002, a research group discovered that the combination of a quarter-wave electric monopole and a ring resonator will lead to wideband structure [1]. Hemispherical, conical, and annular resonators are the most common shapes for design of DRR antenna. In [2], a hemispherical and a conical DRR antenna was investigated and the bandwidth ratio of 4 : 1 was reported. In [3], an annular DRR with the step in outer radius was announced.

In DRR antennas, as will be discussed in more detail in Section 4, the first resonance is due to the monopole in $\lambda/4$, and the next resonance is due to the resonator. In fact, the second resonance is activated by mode TM01δ [4]. There are series of resonances in odd multiples of quarter wavelength. For achieving high bandwidth, plenty of research has been driven toward feed modification and resonator structure. As an illustration, stacked forms of aforementioned shapes can significantly improve impedance bandwidth. A pawn-shaped DRR antenna was designed by attaching conical and hemispherical shapes [5]. In [6], two stacked conical DRR with the same and opposite directions were investigated. In [7], a stacked form of a double annular DRR antenna with different permitivities was investigated. In [8], three different methods of impedance matching, dielectric, and ground plane are applied in order to enhance the antenna

bandwidth. As mentioned before, modification of the feeding is another technique on impedance matching and bandwidth enhancement. Other shapes of feeding include L-shaped microstrip [9], annular-shaped microstrip [10], a tilted modified square-shaped aperture [11], and circular-shaped offset feed [12] for 3-band applications. In [13], the high permittivity and low losses properties of the ceramic material were used for manufacturing the magnetic resonance probes.

In recent studies, more focus is on the feed modification methods to improve antenna features. However, beside this, it is still possible to improve the antenna characteristics using traditional feeding for new structures. In this study, the potential of combination of the DRR antenna and array antenna is demonstrated.

Referenced studies reveal inherent potential of DRR antennas for wideband applications. However, it is possible to improve antenna bandwidth more efficiently. In this regard, based on our knowledge, we have combined and promoted array formation of elements on dielectric ring resonator antenna concept for the first time. Array formation is usually used to enhance antenna gain. However, it can be used to enhance antenna impedance bandwidth. Using these two techniques has concurrently enhanced antenna impedance bandwidth significantly. The impedance bandwidth of the antenna has been outlined and compared with some works of literature in Table 1. This table clearly demonstrates the efficacy of the design. For antenna elements, an isosceles triangular dielectric resonator (TDR) has been opted due to its potential for wideband applications [14].

In this study, a novel shape of ultrawideband DRR antenna has been proposed. The DRR antenna is a gear-shaped DRR antenna, including a circular array of four triangular dielectric resonators. The resonator is located on an extended conducting ground plane and is fed by a quarter-wave monopole probe. The proposed antenna has been simulated with HFSS, and dimensions have been optimized using an extensive parametric study in order to achieve best results.

In this study, in Section 2, antenna configuration will be discussed. In Section 3 (design steps), the effect of antenna parameters on the reflection coefficient chart will be shown and explained. Then, the antenna will be simulated with different softwares and the result will be compared with relevant works in Section 4. In Section 5, theory of the antenna will be elaborated. Finally in Section 6, the whole work will be concluded.

2. Antenna Configuration

Figure 1(a) shows the 3D view of the proposed antenna. The resonator is located on the conducting ground plane and is excited with a coaxial probe as a quarter-wave monopole. The ground plane is square shaped with dimensions of $g \times g$ ($70 \times 70\ \text{mm}^2$). The probe is a 50 ohm coaxial cable with an inner conductor radius of 0.65 mm and height of p.

The resonator includes the dielectric material of RO3010 with permittivity of 10.2 and dissipation factor 0.0035. Figures 1(b) and 1(c) show the DRR configuration and antenna parameters. The proposed DRR is composed of four arrays of triangle elements, which are rotated around the center of the base side of the triangle (center of the coordinate system, and dimensions of these triangles consist the base side (b), the median (a), and central hole radius (r)). Table 2 provides the final dimensions of the proposed antenna.

3. Design Steps

Step 1. Triangle median size (a): Figure 2 shows the antenna reflection coefficient in accordance with the triangle median (a).

The triangle antenna's median is one of the most important parameters in impedance bandwidth. In the first half of the bandwidth, there are two resonances due to monopole (at 6.2 GHz and 15.1 GHz) and two other resonances due to the resonator (at 9.8 GHz and 11.8 GHz). Lower values of the triangle median deteriorate the reflection coefficient at the frequencies around 13 GHz, and its higher values worsen the reflection coefficient at the frequencies around 8 GHz. Therefore, $a = 6.0$ mm is the best value for triangle median.

Step 2. Triangle base size (b): Figure 3 shows the reflection coefficient of the proposed antenna in accordance with the base size of the triangle (b). This figure shows that this parameter does not have a significant effect on the first half of the bandwidth. Also, it shows that the decrease of base size slightly improves antenna bandwidth, while disturbing return loss at some other frequencies including 8, 20, and 32 GHz. Moreover, it moves resonant frequencies in the second half of the bandwidth. On the whole, $b = 2.0$ mm is the best value for the size of the triangle base.

Step 3. Central hole radius (r): at the next step of the design process, the effect of central hole radius of the DRR is investigated. Figure 4 shows the reflection coefficient of the proposed antenna with the central hole radius (r). This figure depicts that an increase in r improves bandwidth, but it increases return loss. Therefore, it disturbs the return loss in some frequencies. Also, it can be seen that the parameter of r does not have a significant effect on the resonant frequencies.

Step 4. DRR antenna height (h): at the next step of the design process, the proposed DRR antenna height is investigated. Figure 5 shows the reflection coefficient of the proposed antenna in accordance with the parameter of h. This figure shows that an increase in the parameter of h reduces impedance bandwidth and simultaneously disturbs return loss in some ranges of frequencies, especially around 13.5 GHz and 21 GHz. For lower values of h, return loss is improved at higher frequencies, in order to improve bandwidth but it disturbs it at lower frequencies, especially around 25 GHz and 33 GHz. Therefore, it restricts the impedance bandwidth. From the figure, $h = 7.0$ mm is the best value for the antenna height.

Table 1: Comparison between the proposed antenna and some previous literatures.

Reference no.	Frequency range (GHz)	Bandwidth (%)	Bandwidth ratio	Permittivity
7	5.6–22.4	126	4 : 1	10
8	1.95–7.05	110	3.6 : 1	10
10	5.5–23.0	122	4.2 : 1	10
11	2.8–15.2	138	5.4 : 1	10
12	3.6–5.5	42	1.5 : 1	36, 4
13	1.8–6.9	117	3.8 : 1	10
15	3.49–7.20	66.7	2.1 : 1	22
PA*	5.2–36.1	150	6.9 : 1	10.2

*PA, the proposed antenna.

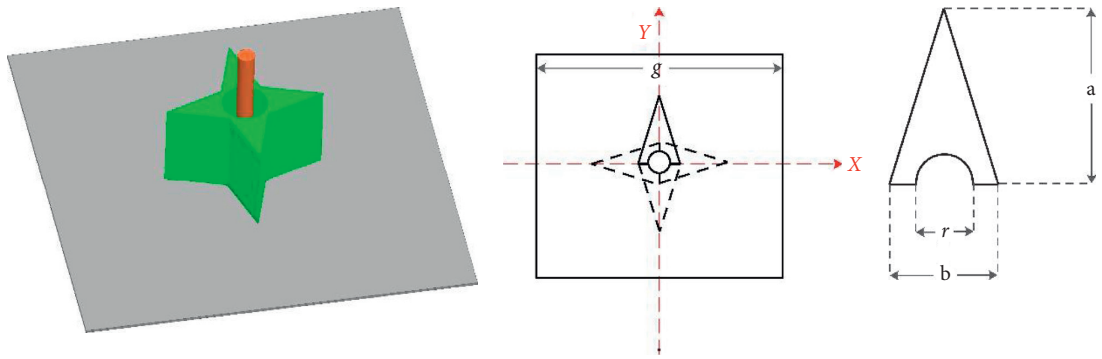


Figure 1: (a) 3D view, (b) DRR configuration, and (c) elements of the proposed DRR.

Table 2: Final dimensions of the proposed antenna (all dimensions in mm).

Parameter	Description	Value (mm)
g	Ground plane width and length	70.0
a	Triangle median	6.0
b	Triangle base	2.0
r	Drilled hole diameter	1.7
h	DRR antenna height	7.0
p	Probe height	12.0
e	Probe radius	0.65

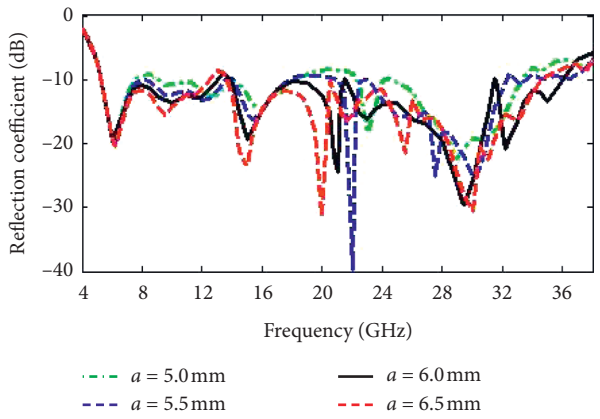


Figure 2: Reflection coefficient of the proposed antenna in accordance with triangle median.

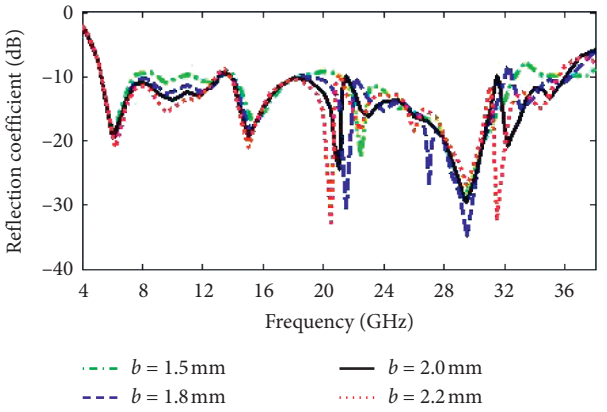


Figure 3: Reflection coefficient of the proposed antenna in accordance with base triangle.

Step 5. Probe height (p): Figure 6 shows the effect of the probe height on antenna return loss. This figure depicts that probe height has a significant effect in almost all resonance frequencies and the reflection coefficient chart. From this figure, it can clearly be seen that the probe height of $p = 12$ mm possesses the best reflection coefficient chart and impedance matching feature.

Step 6. Probe radius (e): Figure 7 shows the effect of the probe diameter on the antenna reflection coefficient. From

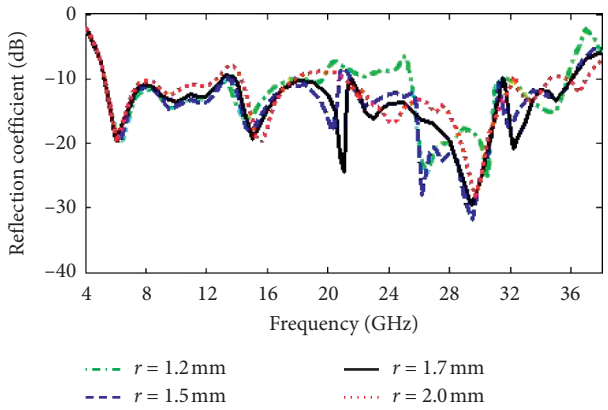


Figure 4: Reflection coefficient of the proposed antenna in accordance with central hole radius.

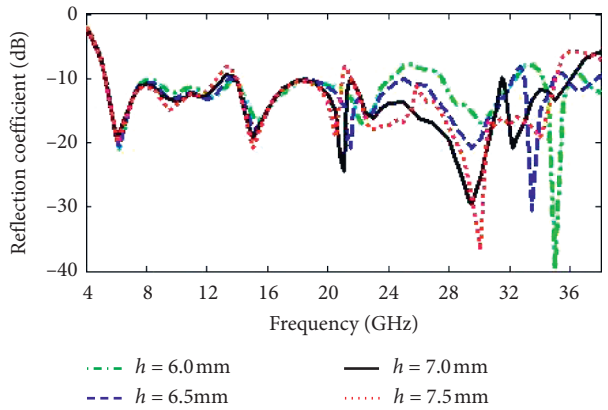


Figure 5: Reflection coefficient of the proposed antenna in accordance with DRR antenna height.

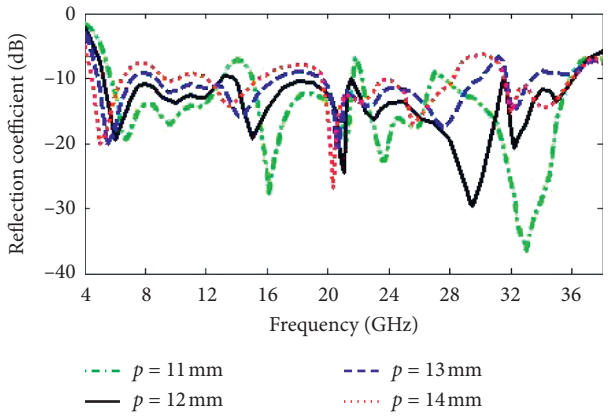


Figure 6: Reflection coefficient of the proposed antenna in accordance with probe height.

this figure, it can clearly be seen that the probe diameter has no effect on the antenna bandwidth in lower values of e.

For all of the values of e, starting and ending points are coincided; however, in some higher frequency ranges, they trespass the criteria of −10 dB of the reflection coefficient chart. But this issue can be solved by other parameters

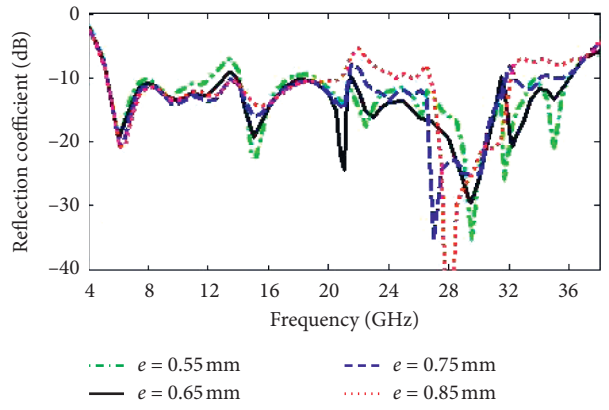


Figure 7: Reflection coefficient of the proposed antenna in accordance with probe diameter.

reoptimization. It can also be seen that almost all frequency resonances remain constant.

4. Results and Comparison

The proposed antenna has been simulated and optimized using HFSS, and a good characteristic has been achieved. The impedance bandwidth of the antenna is 150% (5.2–36.1 GHz). Figure 8 shows the efficiency and the gain of the proposed antenna. Efficiency of the proposed antenna is between 93% and 97% in the whole bandwidth.

Table 1 provides the summary comparison between the proposed antenna and some similar studies. From the table, it can be found that the proposed antenna possesses an excellent ultrabandwidth feature. The radiation pattern is also good in lower frequencies, but it disturbed slightly in higher frequencies.

4.1. Comparison of the Results from HFSS and CST. The proposed antenna has been simulated with HFSS and CST, and the same results have been achieved. Figure 9 shows the reflection coefficient of the antenna simulated with these two softwares, which are based on the method of moment (MoM) and finite element method (FEM), respectively. It can clearly be seen that the same results have been achieved. Results from these two different methods validate the antenna characteristics.

Figure 10 depicts the radiation pattern of the antenna at 10 GHz, 20 GHz, and 30 GHz simulated with HFSS and CST. It can clearly be seen that a good agreement has been achieved.

Figure 11 shows the E-field patterns of the proposed antenna at resonance frequencies. Figure 11(a) (6.2 GHz) shows the main resonance of the monopole at 6.2 GHz ($\lambda/4$). At this frequency TE10 has been activated. Similarly, Figure 11(d) depicts the second resonance of monopole, and the TE mode is activated.

In Figures 11(b), 11(c), and 11(e), electric field distributions inside the DRR have been shown. These figures demonstrate that Ez≠0 and Hz = 0. Therefore, TM modes are activated due to resonator overall dimension and configuration.

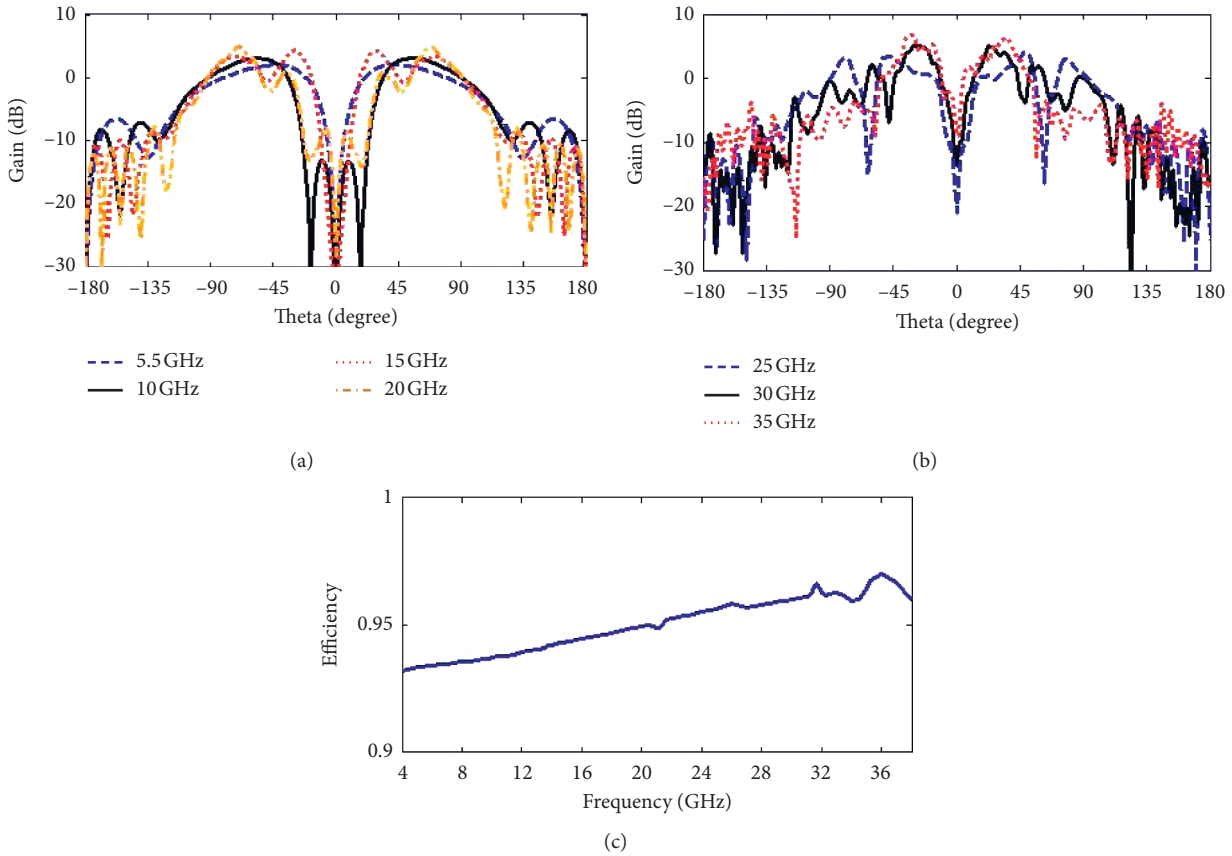


Figure 8: (a), (b) Gain of the proposed antenna. (c) Efficiency.

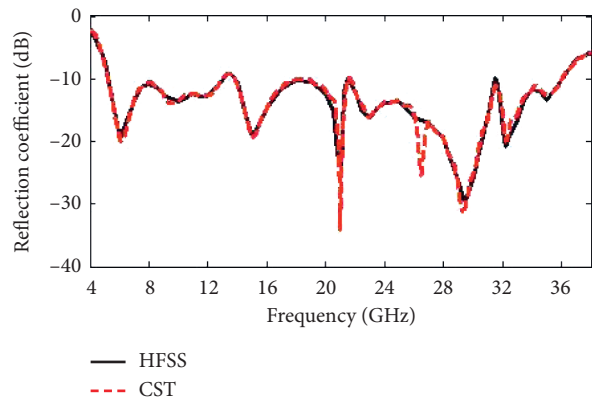


Figure 9: Reflection coefficient of the proposed antenna simulated with HFSS and CST.

Figure 11(g) (29.5 GHz) is due to electric monopole (at $5\lambda/4$), and Figures 11(f) and 11(h) depict the electric field distributions produced by the resonator. These frequencies include both Ez and Hz, and hybrid modes have been activated.

5. Theory

The monopole has resonant frequencies in odd multiples of quarter wavelength. In this case, monopole height is $p = 12$ mm. Therefore, according to equation (1), the resonant frequencies of the monopole are 6.25 GHz, 18.75 GHz, and 31.25 GHz.

$$f_r = \left\{\frac{1}{4}, \frac{3}{4}, \frac{5}{4}, \ldots\right\} * \left(\frac{C}{P}\right), \tag{1}$$

where C is the light velocity in free space. These resonant frequencies are shown in Figure 12(a), red dash line. In Figure 12(a), the reflection coefficient of the antenna without gear resonator has been shown. This figure shows that the antenna resonates at 5.80 GHz, 17.80 GHz, and 30.2 GHz, which are close to the resonance frequencies calculated from

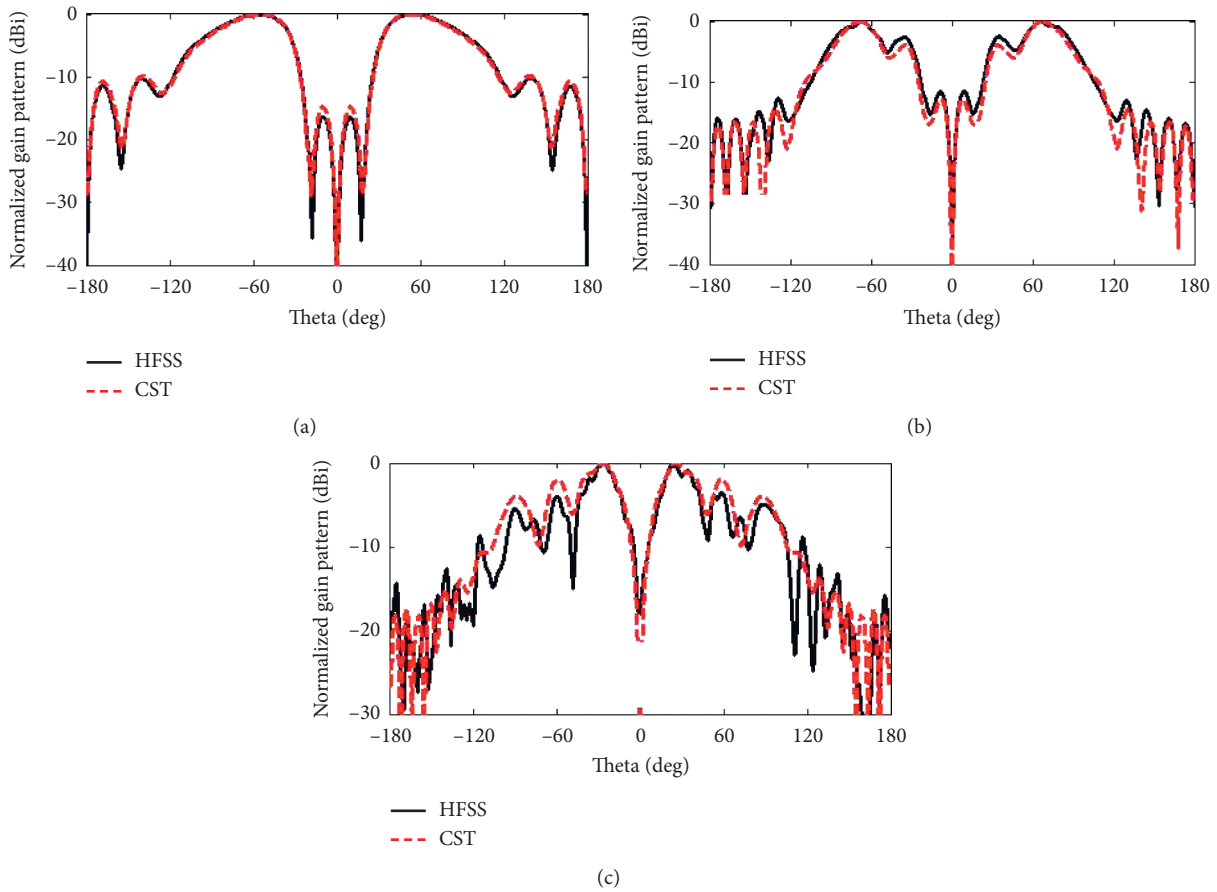


Figure 10: Radiation pattern of the proposed antenna simulated in HFSS and CST at (a) 10 GHz, (b) 20 GHz, and (c) 30 GHz.

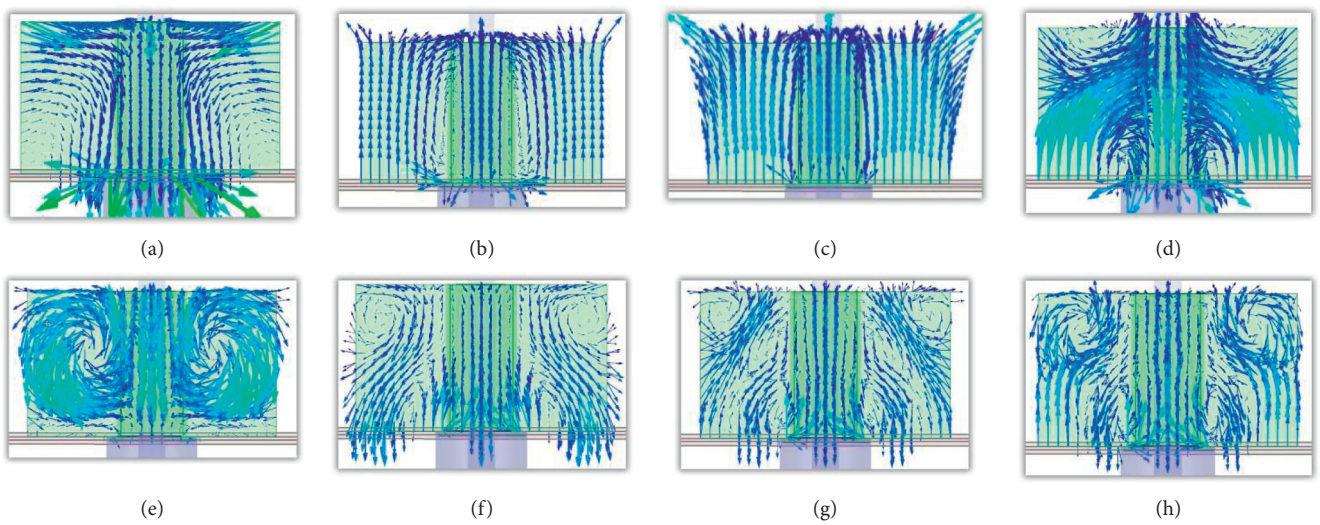


Figure 11: E-field pattern of the antenna at (a) 6.2 GHz, (b) 9.8 GHz, (c) 11.8 GHz, (d) 15.1 GHz, (e) 21 GHz, (f) 22.8 GHz, (g) 29.5 GHz, and (h) 32.2 GHz.

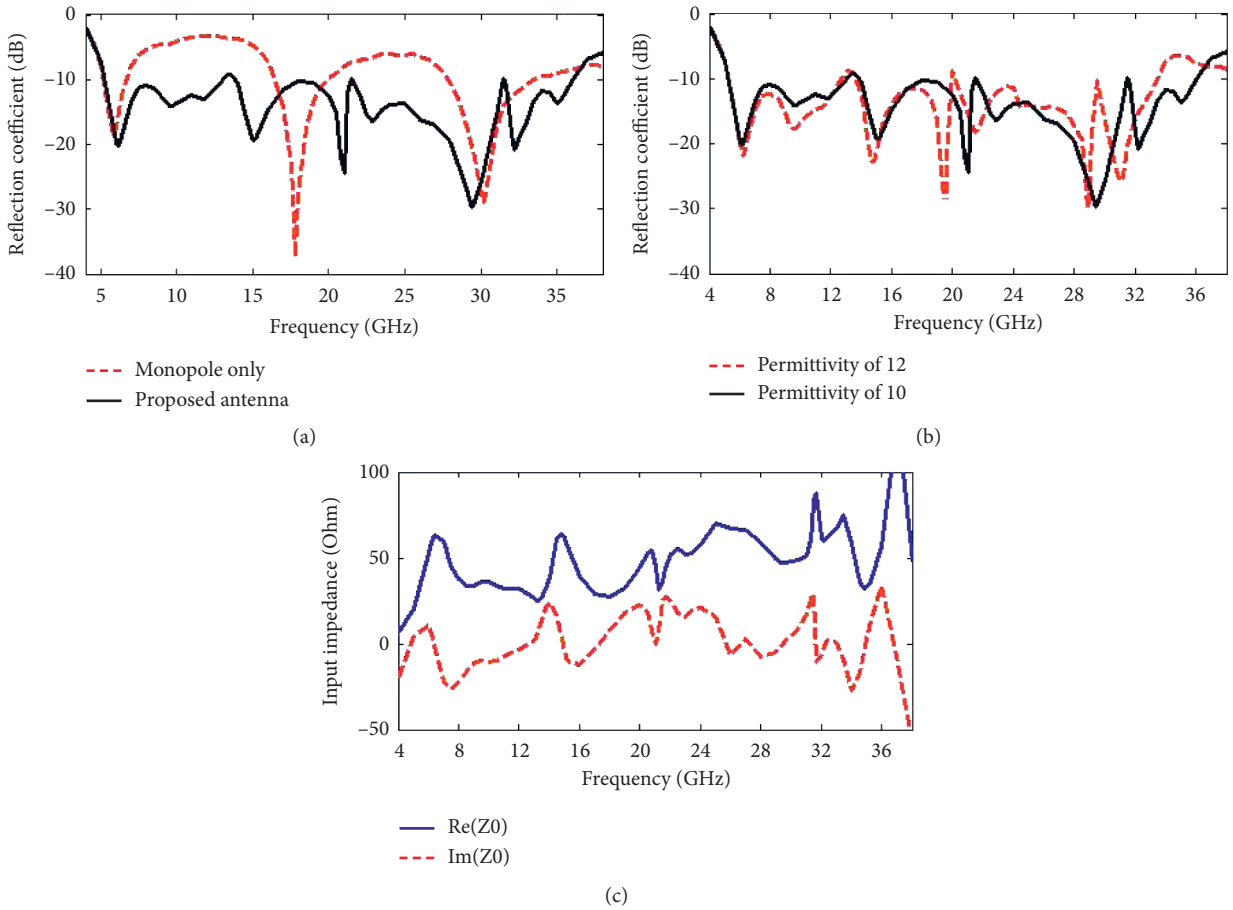


Figure 12: (a) Reflection coefficient of the proposed antenna compared to monopole alone. (b) Reflection coefficient for permittivity of 10.2 and 12. (c) Real and imaginary part of the input impedance.

the equation. These frequencies moved to 6.20 GHz, 15.10 GHz, and 29.5 GHz due to applying the resonator. These reductions, particularly at the second resonance, are from capacitive loading of the resonator. Therefore, these frequencies are expected to remain unchanged by changing the relative permittivity of the resonator. Figure 12(b) shows that these resonance frequencies remain unchanged with relative permittivity, while other resonance frequencies have been decreased with increasing the permittivity of the resonator.

Figure 12(c) shows the real and imaginary part of the characteristic impedance. This figure shows that the imaginary part of the Z_0 at 6.20 GHz, 29.5 GHz are positive, and then, they have not significantly moved after applying the resonator, while it is negative at 15.1 GHz. Therefore, this resonance frequency has been decreased 3.65 GHz, in comparison with the monopole only.

6. Conclusion

In this study, a novel ultrawideband DRR antenna has been proposed. It includes a quarter-wave monopole which resonates in odd multiples of quarter wavelength. Therefore, using a wideband structure can cover these gaps between monopole resonances to enhance antenna bandwidth. Moreover, it is well known that array antennas are used in wideband applications. Therefore, combination of these two techniques can enhance the impedance bandwidth. The last step is to find a suitable shape for array elements. In this regard, a triangular dielectric resonator (TDR) is opted. TDR antennas are small sized and wideband. In general, designing an antenna using all mentioned above has the potential of wideband application.

The proposed antenna consists of a novel gear-shaped DRR which is located on the conducting ground plane and is fed using a coaxial probe as a quarter-wave monopole. The DRR profile consists of a circular array of four triangle resonators. The resonators turn around the monopole and connect to each other from the apex of the base side to form a gear shape. The proposed antenna is designed and optimized using HFSS, and an ultrahigh bandwidth has been achieved. The proposed antenna has an impedance bandwidth of 150% (5.2–36.1 GHz) with 6.94 : 1 bandwidth ratio.

Data Availability

All the data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Ittipiboon, A. Petosa, and S. Thirakoune, "Bandwidth enhancement of a monopole using dielectric antenna resonator loading," in *Proceedings of the ANTEM Conference*, pp. 387–390, Montreal, Canada, August 2002.

[2] D. Guha, B. Gupta, and Y. M. M. Antar, "Hybrid monopole-DRAs using hemispherical/conical-shaped dielectric ring resonators: improved ultrawideband designs," *IEEE Transactions on Antennas and Propagation*, vol. 60, no. 1, 2012.

[3] C. Ozzaim, "Monopole antenna loaded by a stepped-radius dielectric ring resonator for ultrawide bandwidth," *IEEE Antennas and Wireless Propagation Letters*, vol. 10, 2011.

[4] D. Guha, Y. M. M. Antar, A. Ittipiboon, A. Petosa, and D. Lee, "Improved design guidelines for the ultra wideband monopole-dielectric resonator antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 5, pp. 373–376, 2006.

[5] D. Guha, B. Gupta, and Y. M. M. Antar, "New pawn-shaped dielectric ring resonator loaded hybrid monopole antenna for improved ultrawide bandwidth," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, 2009.

[6] C. Ozzaim, F. Üstüner, and N. Tarim, "Stacked conical ring dielectric resonator antenna excited by a monopole for improved ultrawide bandwidth," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 3, 2013.

[7] Y. X. Guo, Y. F. Ruan, and X. Q. Shi, "Wide-band stacked double annular-ring dielectric resonator antenna at the end-fire mode operation," *IEEE Transactions on Antennas and Propagation*, vol. 53, no. 10, 2005.

[8] M. N. Jazi and T. A. Denidni, "Design and implementation of an ultrawideband hybrid skirt monopole dielectric resonator antenna," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, 2008.

[9] W. Chang and Z. Feng, "Investigation of a novel wideband feeding technique for dielectric ring resonator antennas," *IEEE Antennas and Wireless Propagation Letters*, vol. 8, 2009.

[10] R. K. Chaudhary, R. Kumar, and K. V. Srivastava, "Wideband ring dielectric resonator antenna with annular-shaped microstrip feed," *IEEE Antennas and Wireless Propagation Letters*, vol. 12, pp. 595–598, 2013.

[11] D. Pathak, S. K. Sharma, and V. S. Kushwah, "Investigation on circularly polarized ring dielectric resonator antenna for dual-band wireless applications," *Progress In Electromagnetics Research M*, vol. 62, pp. 123–130, 2017.

[12] K. Tulasi, K. Jyothi, B. L. V. Kumari et al., "A novel offset feed annular ring dielectric resonator antenna for bandwidth enhancement," *International Journal of Engineering and Advanced Technology*, vol. 94 pages, 2020.

[13] M. Moussu, R. Abdeddaim, M. Dubois et al., "A semi-analytical model of high permittivity dielectric ring resonators for magnetic resonance imaging," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 8, 2020.

[14] M. Ghorbani, M. Khalily, H. Ghorbaninejad, P. Xiao, and R. Tafazolli, "Bandwidth enhancement and radiation characteristics improvement of triangular dielectric resonator antenna," in *2019 International Symposium on Networks, Computers and Communications (ISNCC)*, Istanbul, Turkey, June 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4759461, 10 pages
https://doi.org/10.1155/2021/4759461

Research Article

Analysis and Measurement of Parameters of Quality of Life in Informal Settlements Surrounding of Tehran Metropolis

Yousef Ghaderpour,[1] **Mohammad Reza Arasteh Taleshmekaiil**,[2] **Bahareh Rouki**,[1] **Mona Mohemsaz**,[3] **Mohammad Hossein Azimi**,[1] **and Ali Sadeghpour**[4]

[1]*Faculty of Architectural and Urbanism Engineering, Qazvin Islamic Azad University, Qazvin, Iran*
[2]*Department of Urbanism, Faculty of Marine Science and Technology, Islamic Azad University, Tehran North Branch, Tehran, Iran*
[3]*Faculty of Urban and Regional Planning, Science and Research Branch, Islamic Azad University, Tehran, Iran*
[4]*Department of Civil Engineering, Varamin Branch, Islamic Azad University, Varamin, Iran*

Correspondence should be addressed to Mohammad Reza Arasteh Taleshmekaiil; mohamadreza.arasteh@gmail.com

Received 14 April 2021; Revised 4 June 2021; Accepted 12 June 2021; Published 24 June 2021

Academic Editor: Noorbakhsh Amiri Golilarz

Quality of life is a concept used to describe the development of the welfare of society, and improving the quality of life is one of the most important goals of governments around the world. Informal settlements, as one of the most vulnerable urban areas, are always in unfavorable conditions in terms of various social, economic, and physical indicators. Therefore, the main purpose of this study is to investigate the status of various quality of life indicators in informal settlements around the metropolis of Tehran in Iran. This research is descriptive, analytical, and correlational based on primary data in which the initial data are collected through the documentary and survey (field) method using a questionnaire. The collected data were processed by SPSS and SmartPLS software. Based on the results of the *t*-test, quality of life indices in informal settlements around the metropolis of Tehran, on a Likert scale with an average of 23.2, is at a low level. Also, the results of the PSL path modelling and multivariate regression test show that increasing the quality of social, economic, physical, service, and access indicators, in terms of residents, has the greatest impact on improving the quality of life of this urban fabric and security index. It has the least impact. There is also a significant relationship between indicators, and the mental perspective of citizens changes under social, service, economic, and physical conditions.

1. Introduction

The 20th and 21st centuries have seen the rapid growth of urbanization and the sudden increase in slums, with only 13% of the world's population living in cities at the beginning of the industrial age [1]. While, now, by 2020, the world's population has reached 7 billion and 817 million people [2] in which only 3.1 billion live in rural areas [3]. This has led to excessive attention on urban areas and the accumulation of wealth and services in these settlements, while the rural areas have become more deprived. As a result of this escalation of economic, social, and welfare inequalities between urban and rural areas that have occurred in the last century, the city environment has become more attractive to immigrants in a way that many of these immigrants prefer living in the slums of the city rather than living in rural areas. In fact, due to the high cost of housing and financial inability, as well as the low level of technical skills in immigrants, most of them inadvertently go to areas of the city that have minimal facilities, services, and environmental quality [4].

Many people continue to live in such conditions because they are trapped in poverty and are unable to move towards a formal urban lifestyle due to the competitive market forces that exist in the city [5]. This the start point for developing outskirt urban textures that lack any environmental services and quality per se. These settlements grow at the outskirt of metropolises under the title of informal settlements.

Informal settlements have been recognized as a global challenge for decades. In countries such as India, Egypt, South Korea, and Iran, such settlements represent a method of urbanization [6, 7]. According to the studies of the Ministry of Housing and Urban Development, studies have been conducted so far about informal settlements in 77 cities in the form of 851 informal neighbourhoods with an area of 52,443 and a population of more than 5,900,000 people. Tehran, Mashhad, Zahedan, Tabriz, and Isfahan are the cities with the highest number of informal settlements, respectively, and in need of improvement [8]. Despite the long history of urbanization and recognition by government policies, living conditions in these areas are very difficult [9]. Despite years of efforts to strengthen the urban status, in the current situation, these neighbourhoods are commonly identified with severe insecurity and vulnerability in the social, economic, and physical sectors. Residents of these urban neighbourhoods also struggle with basic services such as water, health, electricity and environmental hazards, and air pollution [6].

Inadequate urban environmental conditions and their ongoing consequences on quality of life in informal settlements have become the most controversial aspect of cities. Cities and towns in developing countries such as Iran are growing rapidly [10].

In informal urban settlements, the rate and scale of physical growth and development and population density have exceeded acceptable standards of public health, physical infrastructure development, environmental safety, and sustainable economic growth. Therefore, the quality of life is significantly reduced compared to other areas [11].

Therefore, a comprehensive study and assessment of living conditions should be conducted with scientific criteria to plan for the improvement of living conditions based on a comprehensive cognitive. In the meantime, the criterion of quality of life can indicate the achievement of such a goal because it is a comprehensive concept that covers various aspects of human life [12]. Quality of life is one of the most important issues facing the world today. Also, improving the quality of life is one of the most important goals of different governments and states [13].

In the past few decades, paying attention to the quality of life and improving it in urban and rural settlements has been one of the main goals of human development programs and has continually influenced the policies of countries around the world [14, 15].

The term "quality of life" is used to assess the general well-being of individuals in societies, but its meaning is very complex and very comprehensive and varies according to the time and beliefs of the individuals [11]. Quality of life is a concept to show how to meet human needs and a criterion for understanding satisfaction or dissatisfaction with various aspects of life [16].

The World Health Organization defines the quality of life as people's perception of their position in life in the context of the culture and value systems in which they live and according to their goals, standards, expectations, and concerns [17]. According to Ghasemi et al., the quality of life depends on how people live, feel, and understand daily life. This includes aspects such as health, education, housing [18], employment, and participation in decision-making [18–20]. Dovey points out that quality of life is a term that has emerged as a concept of living conditions, health and physical safety, and mental and social ability [21]. Richards and Mutsonziwa consider the level of quality of life as how each individual or group views the society [22].

As mentioned, the world's urban population is about twice that of the rural population, and it is predicted that, by 2030, this gap will increase even more in developing countries [23]. One of the consequences of these population shifts and urbanization is the increase of urban poverty and the physical development of cities [24]. The emergence this phenomenon is informal settlements with the main feature of urban poverty and informal development of the city body [25]. According to the United Nations estimations, around 1 billion people, 14 per cent of the world's population, live in these suburban areas, which can be doubled by 2030 [26, 27].

Ono introduces these settlements in Nairobi as amazing spaces with planned growth and high population density, which accommodates most of the poor [28]. Verma finds living conditions in Indian slums very difficult. According to him, quality of life is summarized in a wide range of fields such as international development, health, political sciences, urban environment, education, leisure, and social affiliation [29]. Dovey et al. believe that a wide range of social, economic, and political complexities will lead to the creation of informal settlements around metropolitan areas [30]. These settlements are generally defined as the rapid, unauthorized, and organic growth of new urban neighbourhoods. These settlements have always been considered as an attraction for migration and affordable housing in the face of a lower quality of life [31, 32].

All these approaches are in line with achieving sustainable urban development. Quality of life, in the category of sustainable development, emphasizes continuity, coordination, and stability of development, including the economy, environment, culture, living conditions, and other factors [33]. Quality of life findings can be used to identify past policy strategies and design future planning policies [34]. Such design has some fundamental structures and it can be done using many artificial intelligence approaches in intelligent systems [35–42], which we cannot review all of them as it goes beyond the scope of this research. Santos and Martinez point out that local social participation in quality of life studies can be an important support for setting long-term policies and goals [43]. The issue of quality of life has entered the urban planning and development literature since the 1960s. This view was further explored by geographers such as Harvey and Smith [44, 45].

In Iran, as in other middle-income developing countries, rapid urbanization in the last five decades has led to the expansion of informal settlements in areas around metropolitan areas [46]. Among the big cities of Iran, the metropolis of Tehran is facing more informal settlements and self-growth areas in its outskirts or middle, rather than other cities of the country due to various factors such as being political-administrative central of the country and the concentration of the most economic, industrial, and service activities in it.

The emergence and growth of informal settlements in Iran is the result of government policies since 1980. During this period, incomplete efforts have been made to improve the conditions of these settlements. However, the main question is what indicators are important in improving the conditions of informal settlements and why? The day-to-day growth of these areas and the increasing complexity of the social, economic, physical, service, and security situation show that we are paying attention to the different dimensions of these urban contexts.

The main approach of this study is to investigate the quality of life in these settlements and the impact of quality of life indicators of improving the situation in these areas. The varieties of indicators used in urban quality of life studies are innumerable. This is obvious given the multidimensionality of the concept of quality of urban life. It is generally assumed that each indicator reflects the size and importance of a particular dimension of the quality of urban life. This assumption is based on the fact that the quality of urban life can be separated into a set of components or dimensions, and if these components are properly combined, a value for the quality of urban life can be obtained.

In this study, various indicators of quality of life have been studied, and finally, according to the type of context studied, the 5 main physical, access and service, security, social, and economic indicators in 13 components (Figure 1) have been selected. These indicators of Islamshahr are considered as an informal settlement around the metropolis of Tehran in Iran. As a result, it seems that, in order to formulate an appropriate strategy, we need more work and research in various fields in these urban areas. Therefore, the present study attempts to measure the criteria and indicators related to quality of life at the desired neighbourhoods (Figure 1) and specify the lack or suitableness of each of the indicators.

The indicators presented in Figure 1 have been examined based on the current situation and the indicators presented in previous studies, and finally, by matching all the indicators, 5 indicators in 13 criteria have been selected. These indicators are the basis for compiling a questionnaire in the form of related questions on which the collection of study data has been done.

2. Materials and Methods

The study, to develop criteria and indicators of quality of urban life and upgrading these criteria in informal urban contexts, requires a comprehensive approach to planning and study that while identifying and classifying criteria and indicators of quality of urban life, there needs to be a causal relationship between these indicators.

The research methodology in the present study is descriptive, analytical, and correlational and the data collection method is a documentary and survey in which the survey part has been done through a questionnaire method.

To assess the quality of life in informal and peripheral settlements around the metropolis of Tehran, with emphasis on Mozaffariyeh, Mianabad, and Ziaabad areas in Islamshahr, five main dimensions of physical, economic, social, services (accessibility), and security have been used. Each of these dimensions has subcategories that have been explored. The sampling method was simple and random and the sample size was estimated using the Cochrane methodology.

According to the census of the Statistics Center of Iran in 2016, the population in the three mentioned neighbourhoods of Mianabad, Ziaabad, and Mozaffariyeh are about 53234 people [8]. This population is considered as a statistical population in which a statistical sample of 381 people was selected and a questionnaire was distributed among them.

Collecting and preprocessing of data is a very crucial step in any statistical or analytical analysis [47–52]. The data obtained from the questionnaire were analysed using SPSS software and statistical analysis of PLS. In the section of descriptive statistics, abundance, frequency percentage, mean, etc., eere used, and in the inferential statistics section, due to the conversion of research variables into quasi-distance variables, in order to calculate the importance or existence of any feature, numbers of 5, 4, 3, 2, and 1 were, respectively, assigned to very high, high, medium, little, and very little options; then, the score of each option was calculated by multiplying the frequency by the score of each option. Then, with the option's scores, the score of each question was obtained, where average scores were calculated by dividing the sum of the points by the number of people who answered to that option. Given that the average score of each question is a number between 1 and 5, so this criterion was used to assess the importance of questions or options. Then, some parametric tests such as t-test and multivariate regression in SPSS software environment have been applied [53] and PLS path modelling has been used with the help of SmartPLS software [42, 53–58]. Cronbach's alpha test was used to calculate the reliability of the questionnaire questions in which the reliability score is equal to 0.89 and at an acceptable level [59–65].

3. Results and Discussion

Between 1992 and 2007, with the development of Tehran metropolis and the filling of vacant lands, the influx of capital and special social groups into the almost untouched suburban lands of the southern and southwestern part of the metropolis of Tehran began; physical disorders peaked and construction of uncoordinated textures of sheds, informal urban textures, annexation in villages, and self-growth cities was observed in abundance in Tehran.

As of now, despite serious measures to prevent illegal construction, the destruction of gardens and land occupation has continued. Due to the indecision of landowners and lack of appropriate action by the municipality, a significant part of the area was separated and was built in the form of disordered and self-growth textures and sheds which cause the development of the urban area in an attached or detached from around Tehran as much as possible.

Since the quality of life is a relative issue and the individual characteristics and living conditions of each person have a great impact on the quality of life of each person, part of the questionnaire is dedicated to descriptive information. The descriptive statistics of the research include age, gender, marital

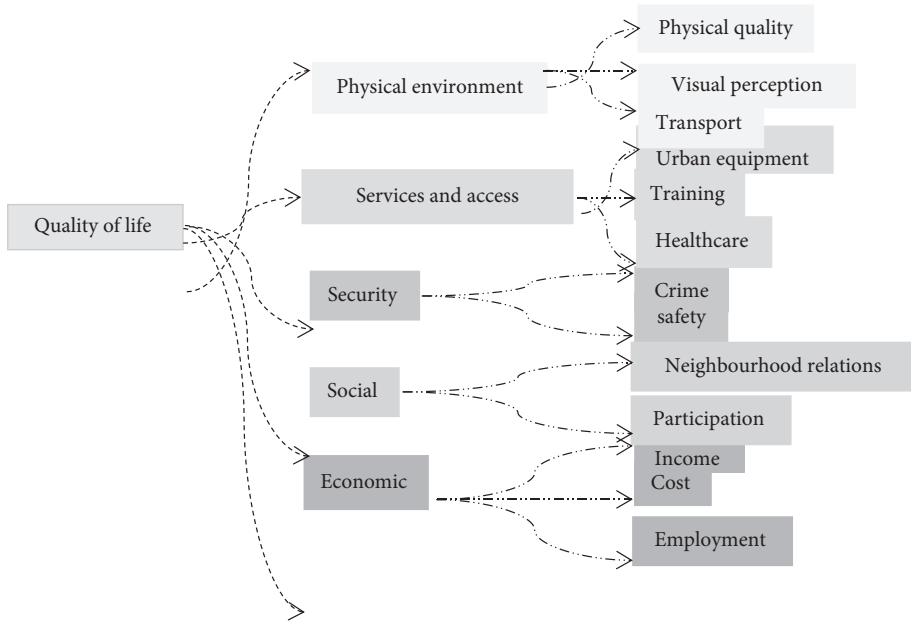


Figure 1: Research indicators.

status, education, homeownership, length of residence in the neighbourhood, age of the residential house, type of vehicle, and reason for living in these contexts, which is shown in Table 1.

Here, we have examined in detail some of the descriptive information and personal characteristics that are more important in the discussion of the quality of life.

In terms of age, the statistical sample was 4.1% of 12–18 age range, 45.5% of 19–35 age range, 20.9% of 36–45 age range, 21.3% of 46–60 age, and 74% were older than 60 years.

In terms of gender, 62.4% were male and 37.6% were female. In terms of job difference, 2.3% of the samples were employees, 40.2% were self-employed, 14.7% were retired, 5.7% were student, 25.5% were housewives, and 11.7% were unemployed.

Also, the residence period in the neighbourhood of the sample size is 3.6% between 1–5 years, 7.4% between 5–10 years, 12.8% between 10–15, 21% between 15–20 years, and 56.2% over 20 years in these areas.

In terms of the residential house age, the status of respondents was as follows: age of their homes of 2.8% was between 1–5 years old, 47% was between 5–10, 8.1% was between 10–15, 12.3% was between 15–20 years, and 72.1% was more than 20 years and in terms of ownership, and 82% of respondents were landlords and 18% were tenants.

Also, the average of each quality of life indicators is shown in Table 2.

3.1. Measuring the Dimensions of Quality of Life from the Perspective of Informal Residents. In this research, the average of the sample is compared with the value of 3, which is considered the average value. If the average of each index is significant and has a significant difference to the value of ($P < 0.05$), it indicates that the level of satisfaction with the quality of life is below average.

To assess the level of quality of life in the intended contexts around the metropolis of Tehran, indicators, including physical, economic, social, access services, and security have been used. The one-sample test was used to assess the status of each indicator and the level of satisfaction of each. The *T*-test was also taken for each of the subindicators and their sum was analysed. Table 3 summarizes the obtained results.

As can be seen in Table 3, the quality of life in three neighbourhoods of Mianabad, Ziaabad, and Mozaffariyeh is low based on the results of the *T*-test. Also, in terms of indicators and subindicators, the state of satisfaction with physical, economic, social, access services, and security features has been significant and shows a low quality of life in all aspects of quality of life.

The results show that, in all factors affecting the quality of life, the actual mean value is greater than the assumed mean; in other words, the average level of life satisfaction in the dysfunctional urban context is considered as the *T* value. Since, in informal settlements, renovations or appropriate constructions are rarely done (mostly due to the poor economic situation of the residents of these neighbourhoods), so most of the texture of these neighbourhoods is worn out and based on the average, minimum satisfaction is in the physical dimension. Also, due to the residency of low-income people and immigrants and vulnerable groups (addicts) in these neighbourhoods, the level of economic satisfaction is very low.

However, on the contrary, due to residents pressured by the government through various laws and different organizations, as well as their social distance from urban life and low level of access to services, there is a high level of solidarity and unity between them, and the level of social solidarity is high in this area.

Since, in these neighbourhoods, services are not provided by the government and the municipality, due to the

Table 1: Research descriptive data.

Indicator	Percent	Indicator	Percent
Vehicle type		Job	
Subway	12	Employee	2.3
Bus	45.2	Freelance	40.2
Taxi	15.7	Retired	14.7
Private car	9.4	Student	5.6
Bicycles and motorcycles	7.3	Housewife	25.5
Other	10.4	Unemployed	11.7
Duration of residence		Age	
1–5 years	3.6	12–18 years	4.1
5–10years	7.8	19–35 years	45.9
10–15 years	12.4	36–45 years	20.9
15–20 years	21	46–60 years	21.3
20 and up	56.2	60 and up	7.8
Marital status		Gender	
Married	69.6	Man	62.4
Single	30.4	Female	37.6
Oldness of residential building		Education	
1–5 years	28	Elementary school	30.7
5–10 years	11.7	Guidance	19.3
10–15 years	18.1	Diploma	35.2
15–20 years	12.3	Bachelor	10.2
20 and up	29.1	Master's degree and higher	4.6
Housing ownership		Reason for living in the neighborhood	
Owner	82	House price	25.4
Rental	18	Proximity to work	23.3
		Proximity to relatives	30.2
		Other	21.1

Table 2: Average level of satisfaction in quality of life indicators.

Indicators	Very little (%)	Low (%)	Medium (%)	Much (%)	Very much (%)
Security	35.20	26.30	30.90	7.60	—
Physical environment	12	40	32.50	10.30	5.20
Accessibility and services	33	41.10	17.60	8.30	—
Economic	60.80	31.30	7.90	—	—
Social	12.40	18.90	37.50	16.70	14.50

Table 3: Obtained results from T-test.

Quality of life indicators	Average	Test value	T	Significant level
Physical environment	2.62	3	−13.318	0.000
Economic	1.56		−34.561	0.000
Social	3.71		16.532	0.000
Services and accessibility	1.68		−62.868	0.000
Security	2.03		−45.786	0.000
All indices	2.33		−36.521	0.000

informal recognition, and also insecure places, due to the presence of criminals and addicts, and the level of satisfaction in terms of access services and security is low.

3.2. Determining the Share of Important Indicators in the Quality of Life (Using Multivariate Regression). One of the important defaults of linear regression is related to the distribution and dispersion of variable data in which data of the research variables should be normal or close to normal [66–70]. The Kolmogorov–Smirnov test is used to solve this problem. The results of the Kolmogorov–Smirnov test in Table 4 show that the distribution of research data is normal and has a normal statistical proportion so that the statistical value of this test for the data of this research is equal to 0.792.

R, which is sometimes known as the correlation coefficient, fluctuates between zero and one. The closer the value of this correlation is to one, the stronger the correlation is between the independent variables and the dependent variables. According to Table 5, R value is equal to 0.921, which indicates an acceptable and high value. R square is the coefficient for determining the research indicators, and the

Table 4: Kolmogorov–Smirnov test to check the normality of the data.

Kolmogorov–Smirnov test	
Kolmogorov–Smirnov Z	0.792
Significant level	0.566
Result	Normal data distribution

Table 5: Statistical summary of model fitting.

Coefficient of determination	R square	Correlation coefficient (R)
0.866	0.868	0.921

value of this coefficient also fluctuates between zero and one, and the closer it is to the value of one, the more it indicates that the independent variables were able to explain the amount of the dependent variable.

Therefore, since the coefficient of determination is equal to 0.868, it shows that approximately 0.866% of the total changes in the level of satisfaction with the quality of life in informal contexts around the metropolis of Tehran can be determined with the five dimensions of this research.

We have followed fair comparison rules to reach unbiased results and comparisons regardless of the computing environment [71–77]. The sig value of the analysis of variance column was used to check the correctness of the regression model fitting. As shown in Table 6, the value of the significance level is equal to zero, and since the value obtained is less than 0.05, it follows that the used regression model is significant and a good predictor of the dependent variable which is quality of life.

Since the scale of independent variables is often composed of different units in regression analysis, the standardized regression coefficients help the researcher to determine the relative contribution of each independent variable in explaining the changes of the dependent variable by considering other specific independent variables.

In this test, the higher the beta coefficient of a variable, so is its role in predicting changes in the dependent variable. According to the significance test of the coefficients, it can be concluded that all the dimensions of the research have a positive effect on the quality of life. Also, standard coefficients have been used to compare the effect of independent variables on the dependent variable which is quality of life.

Based on these obtained coefficients, the share of each of the considered dimensions to measure the quality of life in the informal settlement of informal neighbourhoods around Tehran was the highest, respectively, on the social dimension with a standard coefficient of 0.48, economic dimension with a standard coefficient of 0.443, physical dimension with a standard coefficient of 0.327, service and access dimension with a standard coefficient of 0.328, and security dimension with a standard coefficient of 0.233.

ANOVA analysis also shows whether the regression model can significantly and appropriately predict the changes of the dependent variable and according to Table 6; the obtained sig value is less than 0.05, which indicates that the used model is a good predictor of the quality of life's variable.

3.3. Measuring the Impact of Indicators on the Quality of Life (PLS Route Modelling). Smart PLS as an alternative to multiple regression and covariance-based structural equation modelling. PLS is often used as a method of component-based structural equations (SEM) or variance-based structural equations (VEM). The results of the PLS analysis in Figure 2 are based on a set of independent variables (quality of life indicators) that affect the dependent variable (quality of life). The results are shown as a regression model to assess the quality of life through a set of indicators. The results show the effect coefficients of the dependent independent variables.

It is very vital to have a clear conceptual model and determine the best way for measuring and assessment of the qualitative data [42, 55, 78–84.] The conceptual model tested in the standard mode or PLS algorithm and the path coefficient are presented in Figure 2. The numbers shown on the path of the structures together are called the path coefficient. This number represents the standardized beta in the regression or correlation coefficient of two factors and is presented to examine the degree of the direct impact of one variable on another variable. The numbers displayed on the path between operations and indicators represent the factor load in reflective models.

According to Figure 2, which shows the status of the research, the measurement model in the standard mode, and according to the values of path coefficients that represent the standardized beta in the regression or correlation coefficient of two instruments, we can measure the effect of each of the independent variables which are the five dimensions of quality of life in this study on the dependent variable; the impact of these dimensions can also be ranked.

Based on the internal or structural model of the research, the social dimension of quality of life with a coefficient of 0.451 had the greatest impact on quality of life in the studied area. Then, the economic dimension with a path coefficient of 0.436, the economic dimension with a path coefficient of 0.305, and then services and access dimension with a path coefficient of 0.232 were placed in the next ranks, and finally, the security dimension with a path coefficient of 0.215 had less impact on quality of life in the studied area.

Table 6: Standardized regression coefficients of independent variables on the dependent variable.

Indicators	Analysis of variance ANOVA		T	Impact factor beta	Significant level
	F	Sig			
Physical environment			15.475	0.327	0.000
Economic			16.791	0.443	0.000
Social	586.939	0.000	17.105	0.488	0.000
Services and accessibility			12.498	0.238	0.000
Security			11.982	0.233	0.000

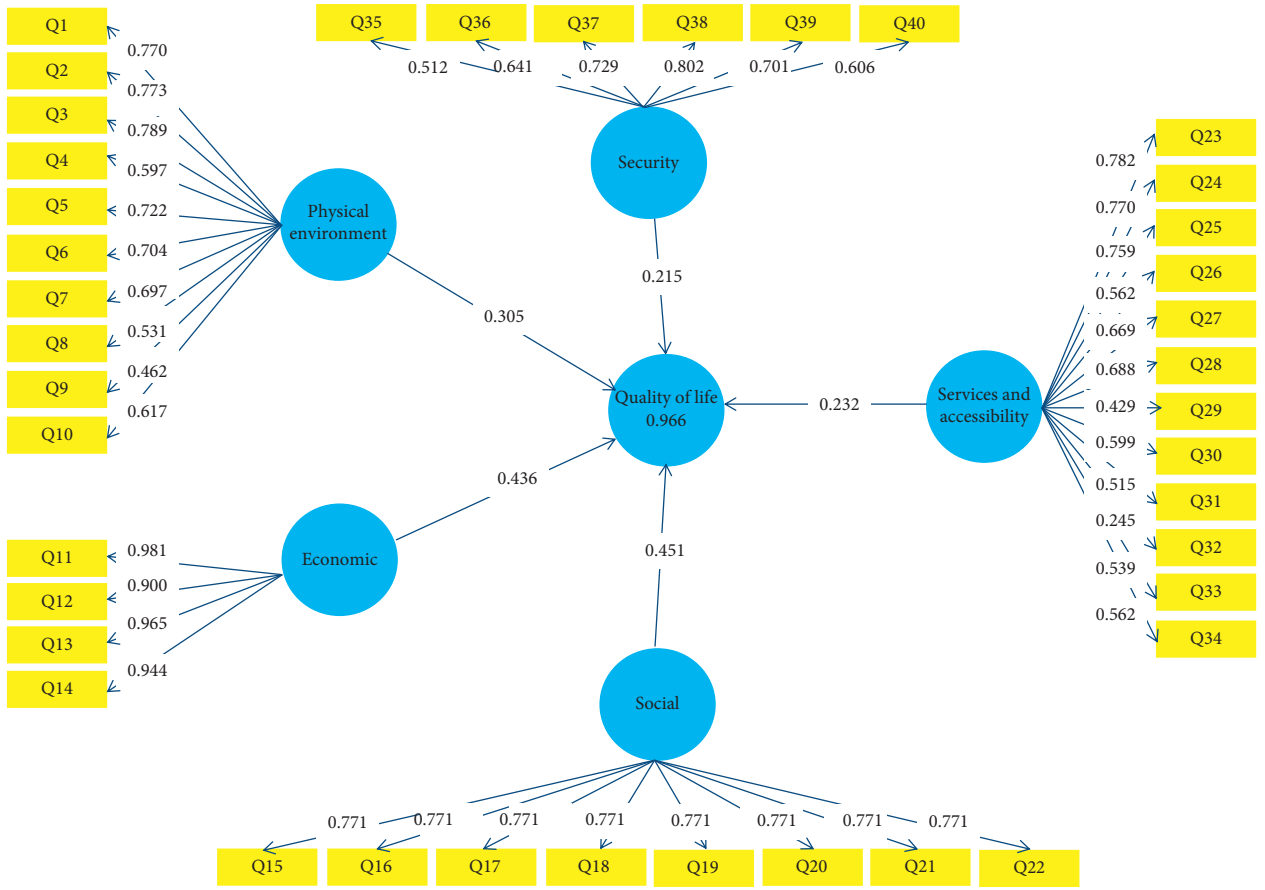


Figure 2: Standard research: the measurement model.

4. Conclusions

According to the results of the research, the quality of life in informal and marginal urban areas is not very favorable so that the average quality of life indicators in these urban neighborhoods is equal to 2.34 in the Likert range (1–5), which is lower than the average. Also, the results of the T-test show that the level of satisfaction with the quality of life in these urban areas is in an unfavorable situation.

T-test on indicators and subindicators demonstrate that the level of satisfaction with all aspects of quality of life is low. Also, based on the results of the regression test and based on the obtained beta coefficients, the share of each of the dimensions intended to measure the quality of life in informal settlements around the metropolis of Tehran, respectively, are a social dimension with a standard coefficient of 0.488, the economic dimension with a standard coefficient of 0.443, the physical dimension with the standard coefficient of 0.23, the service and access dimension with the standard coefficient of 0.238, and the security dimension with the standard coefficient of 0.233.

Also, based on the results obtained from the structural PLS model, the social dimension of quality of life with a path coefficient of 0.451 had the greatest impact on quality of life in the studied area. The economic dimension with a path coefficient of 0.436, the physical dimension with a path coefficient of 0.305, and then services and access dimension with a path coefficient of 0.232 were placed in the next ranks; finally, the security dimension with a path coefficient of 0.215 had less impact on quality of life in the studied area.

Therefore, by recognizing the impact of each of these components and their prioritization and planning to

improve these components, the level of life satisfaction in the studied areas can be increased.

By examining this model in detail, the most important factors that have reduced the quality of life in this neighborhood can be related to economic and access services' indicators, which with an average of 1.65 and 1.68, respectively, had the lowest level of satisfaction among residents. Most of them work in informal and false parts of the city, such as peddling, which has led to low household income and lack of job security.

Regarding the service and access indicators in these neighborhoods, due to its informal settlement, they are not in a favorable situation in terms of having facilities and municipal services.

The sum of these factors has caused the relative deprivation of neighborhood residents from urban facilities and services. However, the highest level of quality of life indicators is related to the social index, which with an average of 3.71 in the Likert spectrum has favorable conditions compared to other indicators, in which one of the important factors, in this case, is due to the long history of people living in this neighborhood; immigration and low immigration of this neighborhood has caused the people of the neighborhood to be highly dependent on each other.

According to the materials presented in the previous sections, in the study areas around the metropolis of Tehran, there are dysfunctional urban contexts, which are also considered informal settlements.

Also, due to the high participation capacity at the neighbourhood level, due to the low level of household income, some cases lack titled deed, weak infrastructure and municipal services and lack of licenses for construction and improvement or renovation of houses, lack of sufficient democratic institution for reflection of citizens' opinions to the responsible institutions, insufficient trust in managing the unfavorable physical condition of buildings, etc. The possibility of improving the quality of life in these urban contexts is facing major problems.

Now that a general understanding of the characteristics of these urban contexts as well as the level of quality of life indicators in this informal settlement context has been determined, clearly, it does not make sense to abandon it in the hope of long-term economic plans, in which their realization is in a haze of ambiguity due to their own complexity or waiting for these areas to reorganize on their own.

Therefore, it is hoped that the authorities will take steps to organize and solve the problem of this settlement by using the internal potentials of these contexts and relying on public participation. Therefore, the most important strategies that can be mentioned about improving the situation of these neighborhoods are as follows:

The most important solution and the first step in improving the situation of this settlement and increasing the quality of life indicators in these areas should be considered as building trust between the residents of this settlement and urban management so that the involvement of personal and group participation of residents in activities such as renovation and empowerment carried out by urban organizations can be increased.

The city administration must first allocate a certain investment to improve the physical and quality of public spaces in order to gain the trust of the residents of this neighborhood towards the municipality and urban management and, in the future, create a good opportunity to improve other spaces. In fact, increasing the quality of public spaces in this place provides a good intention of trust and interaction for residents towards urban management and the implementation of empowerment and renovation projects.

Create adequate support for receiving bank facilities and reviewing all laws and regulations restricting the access of low-income groups to official facilities and credits and applying appropriate initiatives with the lack of ownership and income fluctuations of the informal sector.

Empowering the neighborhood by creating jobs for residents and preventing them from being attracted to mendacious jobs, one of the most important strategies of it, is the development of the local economy.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] A. Das, "Slum upgrading with community-managed microfinance: towards progressive planning in Indonesia," *Habitat International*, vol. 47, pp. 256–266, 2015.

[2] W. P. Clock, "7.8 billion people (2020)-worldometer," 2020, http://https//;%20www.worldometers.info.

[3] Food and Agriculture Organization, "Food and agriculture status 2017," Food and Agriculture Organization, Rome, Italy, 2017.

[4] R. Neuwirth, *Shadow Cities: A Billion Squatters, a New Urban World*, 1st Edition ed. Routledge, Milton Park, UK, 2016.

[5] K. Okurut, "Household demand for sanitation improvements in low-income informal settlements: a case of East African cities," *Habitat International*, vol. 44, pp. 332–338, 2014.

[6] A. Bhide, "Informal settlements, the emerging response to COVID and the imperative of transforming the narrative," *Social and Economic Development*, vol. 22, no. 2, 2020.

[7] D. Roy and N. Menon, "Survey based socio-economic data from slums in Bangalore," *Scientific Data*, vol. 5, 2018.

[8] A. Engineers, *Preparing a Plan for Organizing Informal Settlements in Tehran and Measures to Empower Their Communities with an Emphasis on Urban Improvement*, Ministry of Roads and Urbanism, Baku, Azerbaijan, 2017.

[9] J. Montoya and A. Zumelzu, "Indicators for evaluating sustainability in Bogota's informal settlements: definition and validation," *Sustainable Cities and Society*, vol. 53, 2020.

[10] M. Mirbehesht, S. Mohamad, and A. Modiri, "Measuring the amount of housing in the neighborhood of Eslam Abad deh Vanak (district 2 of Tehran) with the characteristics of informal housing every day (ANP)," *Social Research*, vol. 19, pp. 107–132, 2014.

[11] F. L. Amao, "Housing quality in informal settlements and urban upgrading in ibadan, Nigeria (A case study of apete in ibadan)," 2012.

[12] S. Ebadi, Y. Shahriyar, and M. Ghravi, "Assessing the quality of life in informal settlements (Study sample: east area of Sabzevar Kaleidgah)," *New Attitudes in Human Geography*, vol. 3, no. 6, 2015.

[13] D. Hasanzadeh, "Quality of working life and its impact on productivity," *Management Quarterly*, vol. 123, pp. 17–31, 2008.

[14] I. Bache, "Measuring quality of life for public policy: an idea whose time has come? agenda-setting dynamics in the European union," *European Public Policy*, vol. 20, no. 1, pp. 21–38, 2013.

[15] M. Mirghafourvand, M. A. Jafarabadi, N. Tavananezhad, and M. Karkhane, "Predictors of health-related quality of life in Iranian women of reproductive age," *Applied Research in Quality of Life*, vol. 11, no. 3, pp. 723–737, 2016.

[16] M. Pourtaheri and A. Fatahi, "Assessment quality of life (QOL) in rural areas (Case study: northern Khaveh village, Lorestan province)," *Human Geography-Research*, vol. 43, no. 76, pp. 13–31, 2011.

[17] W. H. Organization, "Health statistics and information systems," 2019.

[18] P. Theofilou, "Quality of life: definition and measurement," *Europe's Journal of Psychology*, vol. 9, no. 1, pp. 150–162, 2013.

[19] S. R. Ghasemi, "Health-related quality of life in informal settlements in Kermanshah, Islamic republic of Iran: role of poverty and perception of family socioeconomic status," *Eastern Mediterranean Health Journal*, vol. 25, no. 11, 2019.

[20] S. Zangane Shahraki, D. Hassanzadeh, and Z. Salary, "Moghadam Evaluation of the quality of life in informal settlements, Tehran metropolitan (case study: Islamabad Salehabad)," *Human Geography-Research*, vol. 64, pp. 177–196, 2014.

[21] K. Dovey, "Informal settlement as a mode of production," in *The new Companion to Urban Design*, A. L.-S. Tridib Banerjee, Ed., Routledge, London, UK, 2019.

[22] R. Richards and K. Mutsonziwa, "Measuring quality of life in informal settlements in South Africa," *Social Indicators Research*, vol. 81, pp. 375–388, 2007.

[23] UN-HABITAT, *State of the World's Cities 2010/2011: Bridging the Urban Divide*, Earthscan, London, UK, 2011.

[24] UN-HABITAT, "Analytical perspective of pro-poor slum upgrading frameworks," 2006.

[25] K. Irandoost and R. Tavallaei, "Defining indicators of informal settlements in Iran," *Urban–Regional Studies Research*, vol. 6, no. 21, pp. 43–60, 2014.

[26] UN-HABITAT, *State of the World's Cities*, The Millennium Development Goals and urban sustainability, London, UK, 2006.

[27] UN-Habitat, "The challenge of slums: global report on human settlements 2003," *Management of Environmental Quality. An International Journal*, vol. 15, pp. 337-338, 2004.

[28] H. Ono, "Understanding the development patterns of informal settlements in Nairobi," *Japan Architectural Review*, vol. 13, no. 3, pp. 337-338, 2020.

[29] K. Verma, "Quality of life in informal settlement: jaipur," *International Journal on Emerging Technologies*, vol. 8, no. 1, pp. 586–591, 2017.

[30] K. Dovey, I. Chatterjee, and T. Shafique, "Towards a morphogenesis of informal settlements," *Habitat International*, vol. 104, 2020.

[31] A. Gilbert, "The return of the slum," *International Journal of Urban and Regional Research*, vol. 31, no. 4, pp. 697–713, 2007.

[32] A. Roy, "The 21st-century metropolis," *Regional Studies*, vol. 43, no. 6, pp. 819–830, 2009.

[33] X. Cong, S. Li, and Y. Gong, "Research on sustainable development ability and spatial-temporal differentiation of urban human settlements in China and Japan based on SDGs, taking dalian and kobe as examples," *Hindawi*, vol. 2021, no. 1, 22 pages, Article ID 8876021, 2021.

[34] Y. J. Lee, "Subjective quality of life measurement in Taipei," *Building and Environment*, vol. 43, no. 7, pp. 1205–1215, 2008.

[35] J. Xia, "Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[36] Y. Zhang, "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 22, 2020.

[37] M. Wang, "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, 2017.

[38] X. Zhao, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing*, vol. 24, pp. 585–596, 2014.

[39] L. Shen, "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

[40] L. Hu, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[41] C. Li, "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[42] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Transactions*, vol. 79, pp. 202–216, 2018.

[43] L. D. Santos, "Monitoring urban quality of life: the porto experience," *Social Indicators Research*, vol. 80, no. 2, pp. 411–425, 2007.

[44] S. Shahrokhi and M. Chour, "Survey of quality of life indicators in Jiroft city," *Space Economics and Rural Development*, vol. 12, no. 2, pp. 19–40, 2015.

[45] A. Fatahi and A. Paydar, "Quality of life and human development," 2013.

[46] M. J. M. Tilaki, M. H. Marzbali, and A. Abdullah, "Ariffin challenges of the informal settlements in developing countries' cities: a case study of Iran," *World Applied Science*, vol. 12, no. 2, pp. 160–169, 2001.

[47] C. Yu, "SGOA: annealing-behaved grasshopper optimizer for global tasks," *Engineering with Computers*, vol. 28, p. 1, 2021.

[48] J. Hu, "Orthogonal learning covariance matrix for defects of grey wolf optimizer: Insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, Article ID 106684, 2021.

[49] H. Chen, "Multi-population differential evolution-assisted Harris hawks optimization: framework and case studies," *Future Generation Computer Systems*, vol. 111, pp. 175–198, 2020.

[50] H.-L. Chen, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson′ s disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[51] D. Zhao, "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowledge-Based Systems*, vol. 216, Article ID 106510, 2020.

[52] X. Zhao, "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Computational Biology and Chemistry*, vol. 78, pp. 481–490, 2019.

[53] M. M. A. Malakoutian and M. Khaksar, "SBM model based productivity evaluation," *ENG Transactions*, vol. 1, no. 1, 2020.

[54] A. Addeh and M. Iri, "Brain tumor type classification using deep features of MRI images and optimized RBFNN," *ENG Transactions*, vol. 2, no. 1, pp. 1–7, 2021.

[55] N. A. Golilarz, "A new automatic method for control chart patterns recognition based on ConvNet and harris hawks meta heuristic optimization algorithm," *IEEE Access*, vol. 7, pp. 149398–149405, 2019.

[56] M. Taleghani and A. Taleghani, "Identification and ranking of factors affecting the implementation of knowledge management engineering based on TOPSIS technique," *ENG Transactions*, vol. 1, no. 1, 2020.

[57] S. Kabir, "Uncertainty-aware dynamic reliability analysis framework for complex systems," *IEEE Access*, vol. 6, pp. 29499–29515, 2018.

[58] X. Xu and H.-L. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Computing*, vol. 18, no. 4, pp. 797–807, 2014.

[59] M. R. Arasteh Taleshmekaiil, "Investigating the effective factors of renewable energy development in tehran metropolis," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6636955, 13 pages, 2021.

[60] A. Sadeghpour and G. Ozay, "Evaluation of reinforced concrete frames designed based on previous Iranian seismic codes," *Arabian Journal for Science and Engineering*, vol. 45, pp. 8069–8085, 2020.

[61] A. Sadeghpour and G. Ozay, "Evaluation of seismic design parameters for reinforced concrete frames retrofitted using eccentric steel bracings," *Environmental Engineering*, vol. 6, pp. 173–178, 2020.

[62] J. Guo, "Bayesian hierarchical model-based information fusion for degradation analysis considering non-competing relationship," *IEEE Access*, vol. 7, pp. 175222–175227, 2019.

[63] J. Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *ENG Transactions*, vol. 1, pp. 1–8, 2020.

[64] J. Guo, "A Bayesian approach for degradation analysis with individual differences," *IEEE Access*, vol. 7, pp. 175033–175040, 2019.

[65] M. R. Hakan Arslan, "Zeynep mutlum, maximum sound transmission Loss in multi-chamber reactive silencers: are two chambers enough?" *ENG Transactions*, vol. 2, pp. 1–15, 2021.

[66] W. Zhu, "Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine," *IEEE Access*, vol. 8, pp. 61107–61123, 2020.

[67] G. Liu, "Predicting cervical hyperextension injury: a covariance guided sine cosine support vector machine," *IEEE Access*, vol. 8, pp. 46895–46908, 2020.

[68] Y. Wei, "Predicting entrepreneurial intention of students: an extreme learning machine with Gaussian barebone harris hawks optimizer," *IEEE Access*, vol. 8, pp. 76841–76855, 2020.

[69] H. Tang, "Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers," *IEEE Access*, vol. 8, pp. 35546–35562, 2020.

[70] A. Lin, "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy K-Nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.

[71] Y. Fan, "Rationalized fruit fly optimization with sine cosine algorithm: a comprehensive analysis," *Expert Systems with Applications*, vol. 157, Article ID 113486, 2020.

[72] E. Rodríguez-Esparza, "An efficient Harris hawks-inspired image segmentation method," *Expert Systems with Applications*, vol. 155, Article ID 113428, 2020.

[73] S. Jiao, "Orthogonally adapted Harris hawks optimization for parameter estimation of photovoltaic models," *Energy*, vol. 203, Article ID 117804, 2020.

[74] Z. Xu, "Orthogonally-designed adapted grasshopper optimization: a comprehensive analysis," *Expert Systems with Applications*, vol. 150, Article ID 113282, 2020.

[75] A. Abbassi, "Parameters identification of photovoltaic cell models using enhanced exploratory salp chains-based approach," *Energy*, vol. 198, Article ID 117333, 2020.

[76] N. A. Golilarz, H. Gao, and H. Demirel, "Satellite image denoising with harris hawks meta heuristic optimization algorithm and improved adaptive generalized Gaussian distribution threshold function," *IEEE Access*, vol. 7, pp. 57459–57468, 2019.

[77] R. U. Khan, "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[78] J. Tu, "Evolutionary biogeography-based whale optimization methods with communication structure: towards measuring the balance," *Knowledge-Based Systems*, vol. 212, Article ID 106642, 2021.

[79] Y. Xu, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.

[80] H. Yu, "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: method and analysis," *Engineering with Computers*, vol. 32, pp. 1–29, 2020.

[81] W. Shan, "Double adaptive weights for stabilization of moth flame optimizer: balance analysis, engineering cases, and medical diagnosis," *Knowledge-Based Systems*, vol. 13, Article ID 106728, 2020.

[82] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing Journal*, vol. 88, 2020.

[83] Y. Zhang, "Boosted binary Harris hawks optimizer and feature selection," *Engineering with Computers*, vol. 43, pp. 1–30, 2020.

[84] M. Naseri, "Evaluation of groundwater resources potential using analytical hierarchy process (AHP) and remote sensing geographic information system (GIS), a case study: Garmsar's catchment basin," *Computational Research Progress in Applied Science & Engineering, CRPASE: Transactions of Civil and Environmental Engineering*, vol. 7, pp. 1–8, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6815802, 16 pages
https://doi.org/10.1155/2021/6815802

Research Article

Applying Bayesian Optimization for Machine Learning Models in Predicting the Surface Roughness in Single-Point Diamond Turning Polycarbonate

Van-Hai Nguyen,[1,2] **Tien-Thinh Le**,[1,2] **Hoanh-Son Truong**,[3] **Minh Vuong Le**,[4] **Van-Luc Ngo**,[1,2] **Anh Tuan Nguyen**,[5] **and Huu Quang Nguyen**[5]

[1]*Faculty of Mechanical Engineering and Mechatronics, PHENIKAA University, Yen Nghia, Ha Dong, Hanoi 12116, Vietnam*
[2]*PHENIKAA Research and Technology Institute (PRATI), A&A Green Phoenix Group JSC, No. 167 Hoang Ngan, Trung Hoa, Cau Giay, Hanoi 11313, Vietnam*
[3]*School of Mechanical Engineering, Hanoi University of Science and Technology, Dai Co Viet Street 1, Hanoi, Vietnam*
[4]*Faculty of Engineering, Vietnam National University of Agriculture, Gia Lam, Hanoi 100000, Vietnam*
[5]*Faculty of Mechanical Engineering, University of Economics and Technology for Industries, 456 Minh Khai Str., Hai Ba Trung District, Hanoi 100000, Vietnam*

Correspondence should be addressed to Van-Hai Nguyen; hai.nguyenvan1@phenikaa-uni.edu.vn and Tien-Thinh Le; thinh.letien@phenikaa-uni.edu.vn

Received 4 May 2021; Revised 31 May 2021; Accepted 11 June 2021; Published 18 June 2021

Academic Editor: Samuel Yousefi

This paper deals with the prediction of surface roughness in manufacturing polycarbonate (PC) by applying Bayesian optimization for machine learning models. The input variables of ultraprecision turning—namely, feed rate, depth of cut, spindle speed, and vibration of the *X*-, *Y*-, and *Z*-axis—are the main factors affecting surface quality. In this research, six machine learning- (ML-) based models—artificial neural network (ANN), Cat Boost Regression (CAT), Support Vector Machine (SVR), Gradient Boosting Regression (GBR), Decision Tree Regression (DTR), and Extreme Gradient Boosting Regression (XGB)—were applied to predict the surface roughness (Ra). The predictive performance of the baseline models was quantitatively assessed through error metrics: root means square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). The overall results indicate that the XGB and CAT models predict Ra with the greatest accuracy. In improving baseline models such as XGB and CAT, the Bayesian optimization (BO) is next used to determine their best hyperparameters, and the results indicate that XGB is the best model according to the evaluation metrics. Results have shown that the performance of the models has been improved significantly with BO. For example, the values of RMSE and MAE of XGB have decreased from 0.0076 to 0.0047 and from 0.0063 to 0.0027, respectively, for the training dataset. Using the testing dataset, the values of RMSE and MAE of XGB have decreased from 0.4033 to 0.2512 and from 0.2845 to 0.2225, respectively. Moreover, the vibrations of the *X*, *Y*, and *Z* axes and feed rate are the most significant feature in predicting the results, which is in high accordance with the literature. We find that, in a specified value domain, the vibration of the axes has a greater influence on the surface quality than does the cutting condition.

1. Introduction

The elements used in optical applications are mostly made of glass, crystalline materials, polymers, or plastic materials. Properties such as degree of transparency, refractive index, and spectrum are the deciding factors for the choice of optical materials [1]. In recent years, polymer optics have become increasingly widely used. They offer advantages over traditional glass materials in different applications, due to their lightweight, low cost, high impact resistance, and flexibility in complex geometric design, despite their lower optical quality in comparison with glass materials [2]. Recently, various methods have been developed to mass-manufacture polymers for optics application such as

compression molding of the polymer followed by multistep grinding to achieve the desired surface finish, with the workpiece being beveled to prevent surface damage from sharp edges, and high-pressure injection into a mold of the desired geometry. After the mold is filled, the liquid polymer is allowed to cool and solidify. The workpiece is then polished to the desired surface finish [3, 4]. In addition, a manufacturing process method improves flexibility and cost-effectiveness in the production of small batches in diamond turning. This is often a significant gain compared to glass optics, facilitating a great deal of freedom in optical style. A variety of polymers are used for optical applications, such as polystyrene, PMMA, polyurethane, and PC. In terms of mechanical properties, PC offers a range of benefits over its rivals, especially for lighting applications [5].

Recently, there has been a dramatic increase in the breadth of applications for ultraprecision machining using single-crystal diamond tools [6]—particularly in the fields of manufacturing optical or magnetic parts, such as magnetic discs, polygonal mirrors, copier drums, and different lens shapes. As a result, the technique has successfully been expanded to machining various soft materials, creating mirror-like surfaces with submicrometer geometrical precision using ultraprecision machining and a single-crystal diamond tool [7]. Previous research indicates that surface quality from diamond turning is influenced by many factors such as machine vibration, chatter, tool wear, tool geometry, built-up edge, chip striking, and lubrication. Indeed, the surface roughness of the diamond turning machine parts plays a crucial role [6, 7]. There are several difficulties when studying the surface roughness of diamond turning parts. One of the main obstacles is that machining is a nonlinear process, governed by several parameters such as tool geometry, the interaction between parts, cutting parameters, machine vibration, and material properties.

Determination of surface quality is an important step in the manufacturing of any machine. Designers of modern production systems and machines are always aiming to better control surface quality, using better computational means with new algorithms. This has led researchers to seek to better understand the prediction of quality [8]. Various mathematical models have been proposed in the literature for predicting surface roughness. For instance, in a series of works, Krolczyk et al. [9, 10] have constructed second-order polynomial prediction functions for predicting the surface roughness and tool life in the dry machining of duplex stainless steel. In such a mathematical model, the influence of different parameters, namely, cutting speed, feed, and depth of cut, has been revealed based on the Student's t-test (comparison of two mean values of populations with Gaussian distributions and homogeneous variances). Krolczyk et al. [10] have exposed that the feed rate was the main influencing factor on the surface roughness. In another study, Kuntoğlu et al. [11] have also employed response surface methodology to seek the optimum cutting conditions, analysis of vibration, and surface roughness under different cutting speeds, feed rates, and cutting edge angles in turning of AISI 5140 steel. It is interesting to notice that the model developed in [11] can predict both surface roughness and vibration during the turning of AISI 5140 within an accepted range of error of 10%. They also found that the feed rate was the most affecting parameter in increasing the surface roughness, which is in accordance with Krolczyk et al. [10] in turning steel. For machining of alloys, Gupta et al. [12] have constructed an optimization procedure to estimate the machining responses of nickel-based superalloy. The optimization problem has been conducted using a combination of response surface methodology, particle swarm optimization, and teaching learning-based optimization techniques. Different parameters have been considered including cutting speed of machining, feed rate, and cutting tool angle, whereas the machining responses were cutting force input, the potential of tool wear, surface roughness, and the length of tool-chip contact.

Recent research has suggested that machine learning may accurately predict surface roughness in the turning process. Eser et al. [13] have estimated the surface roughness of AA6061 alloy in milling using artificial neural networks and response surface methodology. It should be noticed that in such a study, the impact of the cutting parameters on the prediction has been characterized by using variance analysis [13]. In another work, Elangovan et al. [8] built a Multiple Linear Regression (MLR) model to predict surface roughness on the basis of input parameters: feed rate, depth of cut, spindle speed, flank wear, and vibration signal. The artificial neural network (ANN) method has been applied to predict the roughness for different cutting parameters (cutting speed, depth of cut, and feed rate) [14–17]. Özgören et al. [18] have employed the ANN technique to predict the power and torque values obtained from a beta-type Stirling engine. The best ANN's architectures have been determined such as 5-13-9-1 and 5-13-7-1, respectively, by using Levenberg–Marquardt learning algorithm [18]. In recent years, besides ANN modeling, some new ML methods have been introduced: Support Vector Regression (SVR), Gradient Boosting Regression (GBR), Linear Regression (LR), and Random Forest Regression (RFR) [19–21]. Pimenov et al. [22] have tested different machine learning models such as random forest, standard multilayer perceptrons, regression trees, and radial-based functions, for the prediction of surface roughness deviations from face milling machining processes. It is worth noticing that, in [22], the final machine learning prediction model has been developed in an automatic real-time manner for the machining processes. In addition, comparative research into ML models has been carried out to find the best in output prediction. The results for the SVR, polynomial regression, and ANN models indicated that ANN performs best in predicting lifetime but also worst in predicting cutting force and Ra [21]. This shows that not only does the regression of various machine learning models yield different results, but also a specific model has a certain effect on one type of output.

Therefore, this research sets out to investigate the application of SVR, Cat Boost Regressor (CAT), XG Boosting Regressor (XGB), Decision Tree Regressor (DTR), GBR, and ANN on surface roughness prediction. The paper is organized as follows. Section 2.1 introduces the database used in this study, while Section 2.2 outlines the research

methodology. Section 2 gives a brief overview of all the machine learning models used in this work. Finally, Section 3 provides the results and discussion thereof.

2. Materials and Methods

2.1. Data Collection and Analysis. In this study, the results of surface roughness measurements for 35 experiment runs are harvested from the available literature (Bolat [23], published in open access mode). In Bolat [23], the PC sheet was cut into a workpiece whose diameter was 30 mm and thickness was 10 mm. All workpieces were numbered as shown in Figure 1(a). Then, the workpiece was mounted inside a fixture, and the material of the fixture was not affected by temperature changes (shown in Figure 1(b)). During the machining process, the mixture between Kerosene and air was used as cutting lubrication.

The monocrystalline diamond tool S95843 was mounted on a tool holder (Figure 1(c)). According to the tool numbering for monocrystalline of C0.5mLG, the cutting tool has properties such that the type of radius is controlled waviness tool, noise radius is 0.5 mm, the top rank angle is 0°, and front clearance angle is 10°. To implement the experiments, the Precitech Freeform 700 U four-axis diamond turning machine tool was used for 35 experiment runs with feed rate, spindle speed, and depth of cut being 1–12 μm/rev, 1000–2250 rpm, and 3–50 μm, respectively, as shown in Table 1.

The surface roughness prediction dataset consists of two subdatasets. These subdatasets were combined to make the prediction data. Table 1 shows all these 35 measurement results, with seven attributes. These attributes contain information about six independent variables—feed rate, cut depth, spindle speed, and respective vibrations of X-, Y-, and Z-axis—and a dependent variable: surface roughness.

To measure the roughness of the finishing workpiece, the white light interferometer method was used by the Zygo NewView 5000 device with a vertical resolution of 0.1–0.2 nm. The surface roughness of the finished workpiece was measured in terms of three different positions. The measured values of surface roughness shown in Table 1 are root mean squared of three measured results obtained from the interferometer measurements.

2.2. Methodology. The data used in this study have been extracted using exploratory data analysis (EDA). The resulting dataset has been standardized by the $\log(1 + x)$ function to ensure a consistent scale and distribution for all variables. The input data are divided into training and testing datasets, respectively, representing 78% and 22%, corresponding to 27 trials for training and 8 trials for testing, respectively. The training of the models consisted of six regression methods: multilayer perceptron neural network (MLP-NN), SVR, CAT, XGB, DTR, and GBR. The testing dataset was used to validate the model. To assess the performance of the proposed model, various error metrics—RMSE, MAE, and R^2—were employed.

Figure 2 presents the flowchart of the model used in this paper: dataset extraction, feature selection, and different combinations of the dataset made to predict surface roughness. The figure also shows how to fine-tune the model parameters by using the Bayesian optimization algorithm to seek the best parameter for each model and then to determine the best model to predict Ra. Finally, the results and observations are discussed in Section 3.

3. Machine Learning Algorithms

3.1. Artificial Neural Network (ANN). ANNs are complicated computational models inspired by biological neural networks, which are capable of regression, classification, and pattern recognition. There are different ANN-type algorithms, such as backpropagation neural networks [24], probabilistic neural networks [25], convolutional neural networks [26], time-recurrent neural networks [27], and long short-term memory networks [28]. With straightforward and original characteristics, the multilayer perceptron (MLP) model has been chosen in this study. MLP consists of three main layers, fully connected: the input layer, hidden layer, and output layer [29]. Thanks to its properties, MLP has been used to predict tool wear flank and surface roughness [30–32].

The main advantages of the ANN model are the capability to work with any type of input data (complete or incomplete) [33]; the information is stored on the entire network instead of a database and the capability of parallel computing, which helps to reduce the computational time. On the other hand, some drawbacks of this approach need to be mentioned such as the hardware dependence, which requires a lot of computer resources when the input data is large. In addition, the behavior and the duration of the network are hard to control, which requires a lot of trials.

3.2. Support Vector Machine (SVM). Support Vector Machine (SVM) theory was developed by Cortes and Vapnik [34], and a version of an SVM for regression was launched in 1997 [35]. Support Vector Machine consists of different branches in which Support Vector Regression (SVR) is an important application. Many ML algorithms follow the principle of empirical error minimization, while SVR follows the principle of structural risk minimization across a restricted range of learning patterns, so it can obtain better generalization [36]. SVR is a distinguished analytical tool. It is applicable because it uses linear improvement techniques to seek out optimal solutions to nonlinear prediction issues in relation to higher-dimensional features. Therefore, it has been widely used for forecasting in the fields of finance, agriculture, hydrology, the environment, etc., and especially, in mechanical machining [21, 44, 45]. Consequently, SVR is an appropriate model to predict Ra in ultraprecision machining PC for applications in optics.

In terms of advantages, SVM can work really well with high-dimensional input space and is relatively memory efficient. In terms of drawbacks, unlike ANN, SVM is not

(a) (b) (c) (d)

Figure 1: Experiment setup reproduced with permission from Bolat [23]: (a) numbered PC workpieces, (b) fixture for the placement of PC workpieces, (c) mounted monocrystalline diamond on the cutting tool (front view), and (d) mounted assembly fixture and workpiece on the Precitech Freeform 700 U four-axis machine tool.

suitable for large datasets; it does not perform well with any type of data (for example, data set with more noise).

3.3. Cat Boost Regression (CAT). Most popular implementations of gradient boosting use decision trees as base predictors. It is convenient to use decision trees for numerical features, but, in practice, many datasets embody categorical features that are vital for prediction. CAT is a novel gradient boosting technology, developed by Yandex. It is associated with improved implementation within the gradient boosting tree algorithmic framework. This framework relies on a symmetrical decision tree algorithmic rule with few parameters, support for categorical variables, and high accuracy [46]. CAT improves the accuracy of the algorithm and its generalizability [47]. It has been successfully applied in many fields such as weather forecasting, media popularity prediction, evapotranspiration, and biomass [48, 49]. It is for this reason that the model is applied here to predict the performance of ultraprecision machining.

The advantages of gradient boosting approaches are their predictive accuracy compared to other machine learning models. These types of approaches have a lot of flexibility, which means that they can be optimized using different loss functions or several hyperparameter tuning options that make the function fit flexible. They can work directly with the input data, which means that no data preprocessing is required. In terms of drawbacks, gradient boosting approaches are usually computationally expensive. The minimization of errors in these approaches can cause overfitting, and the influence of parameters is quite heavy on the behavior of the approach.

3.4. Decision Tree Regression (DTR). Decision trees (for classification and regression) are classic ML algorithms. As a group, their learning ability is not outstanding, but they are well known for their generalizability and feature filtering. When used for regression tasks, they are called regression trees [50]. As the number of iterations increases, the model continues to learn. The training is stopped when triggered by hyperparameters such as the number of selected features, the maximum depth of the tree, and the minimum sample size of branches.

Compared to other algorithms that require data preprocessing, the decision tree algorithm requires less effort in

Table 1: Surface roughness measurements for experiment (reproduced with permission from Bolat [23]).

Trial number	Feed rate (μm/rev)	Depth of cut (μm)	Spindle speed (rpm)	*VX* (mgRMS)	*VY* (mgRMS)	*VZ* (mgRMS)	Ra (nm)	Type
1	5	40	2000	28.07	45.97	41.69	17.90	Training dataset
2	9	22.5	1000	27.90	52.76	37.75	15.0	
3	5	22.5	2000	27.86	44.58	38.87	16.50	
4	9	22.5	2000	29.90	61.70	46.74	33.70	
5	9	5	1500	26.97	43.16	37.54	19.70	
6	5	5	2000	25.35	49.91	35.04	9.40	
7	1	40	1000	28.74	44.8	39.13	10.10	
8	1	5	2000	25.17	45.21	32.67	2.70	
9	1	40	2000	26.54	56.23	35.81	9.50	
10	5	5	1500	27.09	52.08	36.90	12.0	
11	1	22.5	2000	26.45	52.67	36.93	6.40	
12	1	40	1500	28.75	52.60	35.59	9.20	
13	5	40	1000	32.77	66.69	38.66	12.80	
14	9	5	1000	26.06	54.75	35.74	11.50	
15	1	22.5	1000	25.88	48.86	33.23	8.30	
16	1	5	1000	28.14	54.58	37.88	4.70	
17	5	40	1500	26.01	50.85	40.10	13.60	
18	9	40	1000	25.64	48.62	33.21	25.90	
19	1	22.5	1500	25.25	44.57	33.09	4.30	
20	9	5	2000	24.65	50.64	34.67	22.10	
21	2	3	1250	6.90	28.61	59.82	47.17	
22	7	22	1000	18.10	28.63	59.83	46.42	
23	2	6	2250	6.30	28.76	60.88	49.09	
24	7	30	1750	23.0	37.26	78.91	54.13	
25	12	40	1250	52.10	28.59	59.90	46.48	
26	2	50	1000	5.80	28.92	60.0	46.70	
27	7	20	2250	14.40	28.91	59.77	46.93	
28	5	22.5	1000	26.72	52.41	37.35	12.20	Testing dataset
29	1	5	1500	25.53	46.66	33.04	3.70	
30	5	5	1000	28.91	58.81	46.34	10.20	
31	5	22.5	1500	25.15	46.91	33.72	8.50	
32	9	40	2000	36.22	74.97	45.09	46.80	
33	9	40	1500	25.61	48.27	33.20	21.60	
34	9	22.5	1500	28.301	59.02	46.32	26.40	
35	12	10	1750	47.70	28.16	58.97	47.48	

this process. A decision tree does not require normalization or scaling of data. The decision tree algorithm can also work with incomplete data. In terms of disadvantages, a small change in data requires a lot of changes in the structure of the decision tree, which can eventually cause instability. The computational time of the decision tree is often expensive, especially when training the model.

3.5. Gradient Boosting Regression (GBR). Gradient Boosted Trees (GBTs) are a set of DTs, whose results are a combination of predictions of base models. DT-based ensembles like GBT have often been used in regression and classification problems, as they perform well [51]. GBT is an iterative algorithm, which means that each tree can take account of the error in the previous one. The final result of the GBT is the mean of the predicted results from all trees.

3.6. Extreme Gradient Boosting Regression (XGB). XGB is the most popular ML algorithm, developed in 2015. Regardless of the data type, it is well known to provide better solutions than other ML algorithms, because of its rapidity, efficiency, and scalability [52, 53]. It has been the focus of research in various fields [54–56]. In particular, in mechanical machining [52, 57, 58], XGB is a good choice to predict tool wear and surface roughness.

XGB is used for supervising learning problems, where we use the training data (with multiple features) x_i to predict a target variable y_i. Before we learn about trees specifically, let us first review the basic elements in supervised learning. Below are the analytical formulas for the regression math on which XGB is based [52, 59].

Let $D = \{(x_i, y_i)\}$ be a dataset composed of n samples and m features $|D| = n, x_i \in R^m, y_i \in R$. A number of k-additive functions $f_k(x)$ are used in tree ensemble models to evaluate the function $\varnothing(x_i)$. This function can be expressed as [52]

$$\widehat{y}_i = \varnothing(x_i) = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in F, \tag{1}$$

where k represents the number of trees, x_i is the i^{th} iteration of the training process, and f_k denotes the decision rules of

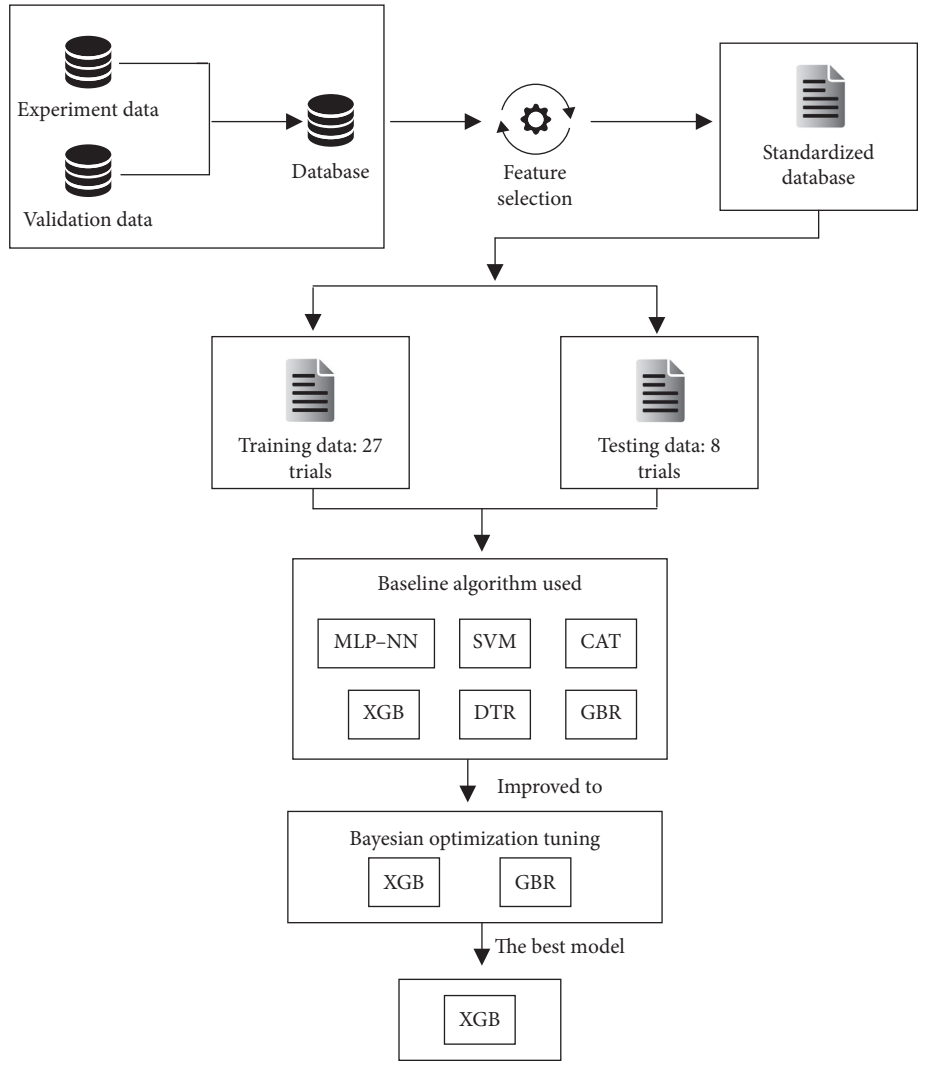


Figure 2: Overall methodology employed in this study.

the tree and weight of leaf score. The regression space F can be expressed as follows [52]:

$$F = \left\{ f_x = w_{q(x)} \right\}, \quad q: R^m \longrightarrow T,\ w \in R^T, \tag{2}$$

with q, T, and w representing the tree structure, the number of leaf nodes, and the corresponding weight, respectively. The errors of the model can be minimized using a regularized objective function, as shown as follows [52]:

$$L(\varnothing) = \sum_i \psi\left(y_i, F_K\left(x_i\right)\right) + \sum_i \Omega\left(f_k\right), \tag{3}$$

with ψ being a differentiable convex loss function, in terms of regression type, which can be the mean squared error function, while Ω is the regularization term that penalizes the complexity of the model to avoid overfitting, defined as [52]

$$\Omega\left(f_k\right) = \gamma T + 0.5\lambda w^2. \tag{4}$$

γ denotes the complexity, and λ is a constant coefficient. Since XGB is an additive algorithm, the prediction $F_K(x_i)$ of the i^{th} instance at the k^{th} iteration can be expressed as follows [52]:

$$L(\varnothing) = \sum_i \psi\left(y_i, F_{K-1}\left(x_i\right) + f_k\left(x_i\right)\right) + \sum_i \Omega\left(f_k\right). \tag{5}$$

The main objective of XGB is to determine an additive function f_k that minimizes the objective function using the gradient descent optimization algorithm.

3.7. Bayesian Algorithm Optimization. In XGB and CAT, the main goal of hyperparameter optimization (i.e., tuning) is to minimize the objective function defined in equation (5). There are two popular hyperparameter optimization methods: random search (RS) and Bayesian optimization (BO). In random search, the hyperparameters are randomly chosen from the predefined search domain, and the searching is independent of the previous boosting result [60, 61]. The main advantage of RS is that it can be applied to high-dimensional problems. Bayesian optimization can be considered a probability approach, using probability

theories to optimize the hyperparameters [62]. In this work, we have chosen Bayesian optimization because of its performance, which has been demonstrated in previous studies in the literature [63, 64]. For simplicity's sake, in the rest of the paper, we use XGB_opt and CAT_opt to denote the XGB and CAT models built by using BO, respectively.

Before GBR and XGB training, we need to initialize the model's hyperparameters. However, the selection of network hyperparameters, based on experience or RS such as GridSearchCV and RandomizedSearchCV, currently requires a large number of attempts. On the other hand, to optimize the model's performance and reduce computational time, the hyperparameter screening process must be optimized—this is the main purpose of using BO. The Bayesian optimization framework utilizes historical data to optimize the search domain and constantly predict the posterior piece of information [65]. In particular, suppose that we have a functional relation between the hyperparameters and loss function:

$$p^* = \operatorname{argmin}_{p \in P} \operatorname{loss}(p), \tag{6}$$

where p is the set of all hyperparameters in which p is the set of hyperparameter combinations, p^* is the optimal parameter combination obtained from the final optimization, and $\operatorname{loss}(p)$ is the objective function.

In the proposed model, the hyperparameters are maximum tree depth (D), number of nodes in each tree (γ), number of trees (K), learning rate (η), regularization parameter (λ), and number of samples (N), as introduced in equations (1)–(4). The loss function is defined by the RMSE as [65]

$$\operatorname{loss}\left(p_j\right) = \sqrt{\frac{\sum_{i=1}^{N}\left(\widehat{y}_i\left(p_j\right) - y_i\right)^2}{N}}, \tag{7}$$

where p_j is the j^{th} hyperparameter combination, y is the true value, and $\widehat{y}_i(p_j)$ is the model output results obtained using the j^{th} hyperparameter combination p_j.

The next step of BO is to construct the dataset $D = \{(x_i, y_i)\}$, where x_i is the i^{th} set of hyperparameters and y_i is the corresponding error of the model output [65]:

$$y_i = \operatorname{loss}(p_i). \tag{8}$$

The posterior probability $P(y|x, D)$ is denoted as D. A Gaussian distribution is applied to the alternative model M whose mean and variance are denoted as μ and K, respectively. The specific functional expression M is obtained by fitting the dataset D [65]:

$$p(\operatorname{loss}) = G(\operatorname{loss}; \mu, K). \tag{9}$$

Based on M, the next observation is calculated using the acquisition function $\alpha(p)$ [65]:

$$p^* = \arg\max \alpha(P, p(y|x)). \tag{10}$$

In the Bayesian decision theory, the capture function works by calculating the expected loss corresponding to the hyperparameter space p loss. In each iteration, the dataset D is updated by receiving the parameters and losses from the previous one. The main characteristic of BO is model construction based on historical data to optimize the hyperparameters for each model [66].

3.8. Performance Assessment Criteria. Three statistical metrics have been used in this study to assess the performance of the proposed AI model in predicting surface roughness: root means square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These metrics are defined as follows [67–70]:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{N}\left(\widehat{y}_i - y_i\right)^2}{N}}, \\ \text{MAE} &= \frac{1}{N}\sum_{i=1}^{N}\left|\widehat{y}_i - y_i\right|, \\ R^2 &= 1 - \frac{\sum_{i=1}^{N}\left(\widehat{y}_i - y_i\right)^2}{\sum_{i=1}^{N}\left(\widehat{y}_i - \overline{y}\right)^2}, \end{aligned} \tag{11}$$

where y_i and $\widehat{y}_i$ are the measured and predicted values, respectively, and N is the total number of predicted data points. Higher values of R^2 represent the better performance of the model. On the other hand, the better performance of the model is reflected by lower values of RMSE and MAE.

When contrasting the values of metrics, we will prioritize the evaluation criteria selected in RMSE, because this is a more suitable method than MAE when the model error follows a normal distribution. Moreover, RMSE has a distinct advantage over MAE in that RMSE avoids using absolute value, which is highly undesirable in many mathematical calculations [71]. Therefore, if comparing the prediction accuracy of various regression models, then RMSE is a better choice as it is simple to calculate and differentiable. Moreover, a higher value of R^2 is considered desirable.

4. Results and Discussion

4.1. Prediction Accuracy of Various Baseline Models. The analyzed performance of MLP-NN, SVR, CAT, XGB, DTR, and GBT baseline regression models in terms of Ra prediction for diamond ultraturning is reported in this section. Table 2 shows the results of the various models acting on the training and testing datasets, sorted in ascending order of RMSE for the testing dataset. It can be seen that the predicted Ra varies considerably from one model to another. Using the testing dataset, XGB exhibits the best performance, in terms of all error metrics—this model yields the highest value of R^2 and the smallest values of RMSE and MAE. With the training dataset, XGB also exhibits similar performance, with the highest R^2. The DTR model performs best in MAE with the training dataset and best in RMSE with the testing dataset. On the contrary, DTR exhibits the worst performance out of all models, in terms of both the training and testing datasets.

Table 2: Accuracy statistics of ML regression models using training and testing datasets.

Model	Training dataset			Testing dataset		
	RMSE	MAE	R^2	RMSE	MAE	R^2
XGB	0.0076	0.0063	0.9999	0.4033	0.2845	0.7227
CAT	0.1594	0.1224	0.9565	0.4704	0.3503	0.6229
GBT	0.0018	0.0015	0.9995	0.4812	0.3440	0.5981
MLP-NN	0.3249	0.2663	0.8194	0.4829	0.41115	0.6025
SVR	0.3964	0.3249	0.7313	0.5032	0.4334	0.5684
DTR	0.1056	0.0720	0.9809	0.5712	0.3537	0.4927

Notes: root mean square error, RMSE; mean absolute error, MAE; coefficient of determination, R^2; multilayer perceptron neural network, MLP-NN; Support Vector Regression, SVR; Cat Boost Regression, CAT; Decision Tree Regression, DTR; Gradient Boosting Regression, GBR; Extreme Gradient Boosting Regression, XGB.

In this work, a comparison between different activation functions for the MLP-NN model was performed with initial parameters, layers (32, 16), optimizer ("Adam"), and activation ("relu", "identity", "sigmoid", "tanh", and "logistic"). Table 3 shows the results of using different activation functions acting on the assessment criteria using training and testing datasets, respectively. It should be noticed that parametric study was only conducted on activation functions in the present study.

As shown in Table 3, the "identity" activation function exhibits the best performance, in terms of all error metrics—this model yields the highest value of R^2 and the smallest values of RMSE and MAE.

Figure 3 illustrates the performance of the six models on metrics on the training dataset. Each color of the bars corresponds to one metric. The ranking of the models is marked above each bar. As can be seen from Figure 3, there is a slight difference between the rankings in relation to the training and testing datasets. With the training dataset, the accuracy rankings with respect to the metrics of RMSE are GBT, XGB, DTR, CAT, MLP-NN, and SVR, while with metrics of MAE, the order is GBT, XGB, DTR, CAT, MLP-NN, and SVR. The rankings with metrics of R^2 are XGB, GBT, DTR, CAT, MLP-NN, and SVR.

However, Figure 4 shows that there is a small difference in terms of performance ranking when using the testing dataset. Furthermore, the performance ranking varies as a function of error metrics. The accuracy ranking with RMSE metrics in increasing order is XGB, CAT, GBT, MLP-NN, SVR, and DTR, and that of MAE in increasing order is XGB, GBT, CAT, DTR, MLP –NN, and SVR, while ranking results with R^2 metrics are XGB, CAT, MLP-NN, GBT, SVR, and DTR. Considering the performance using both the training and testing datasets, it can be concluded that XGB yielded the best performance whereas DTR exhibited the worst performance out of these models.

The line and scatter plots of the measured Ra and the predicted values found by the six ML baseline models with the training and testing dataset are presented in Figures 5 and 6 for each trial, respectively. The different colors of the scattered points in the figures represent the values predicted by different baseline models. In the training and testing datasets, the values predicted by the XGB model are the closest to the original values of Ra. The value predicted by the CAT model does not closely correspond to the original measured values of Ra.

However, as discussed above, RMSE is the preferred criterion for selecting better models, in relation to the testing

Table 3: Comparison MLP-NN models under different activation functions.

Activation functions	Training dataset			Testing dataset		
	RMSE	MAE	R^2	RMSE	MAE	R^2
"relu"	0.7349	0.5711	0.0765	0.6913	0.5649	0.1855
"identity"	**0.3249**	**0.2663**	**0.8194**	**0.4829**	**0.4112**	**0.6025**
"sigmoid"	0.7457	0.6447	0.0490	0.7453	0.6257	0.0533
"tanh"	0.8092	0.6965	−0.1198	0.8301	0.7007	−0.1744
"logistic"	0.7656	0.6417	−0.0023	0.7621	0.6678	0.0101

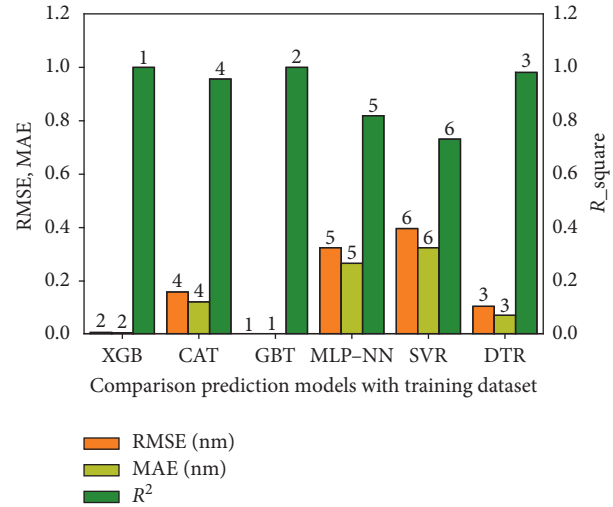


Figure 3: Regression of six baseline models' performance in predicting surface roughness (Ra)—results from 27 trials.

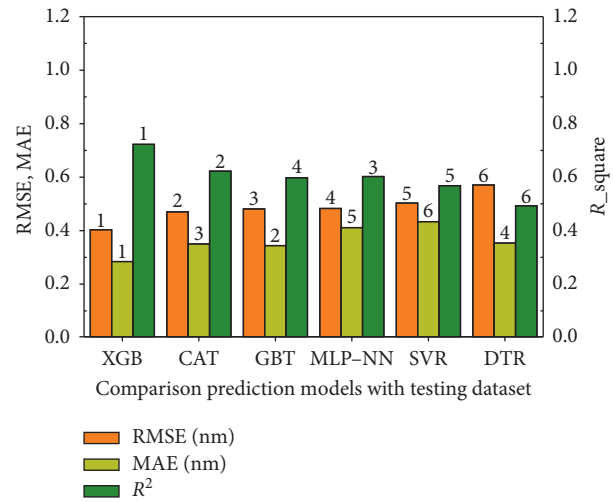


Figure 4: Regression of six baseline models' performance in predicting surface roughness (Ra)—results from 8 trials.

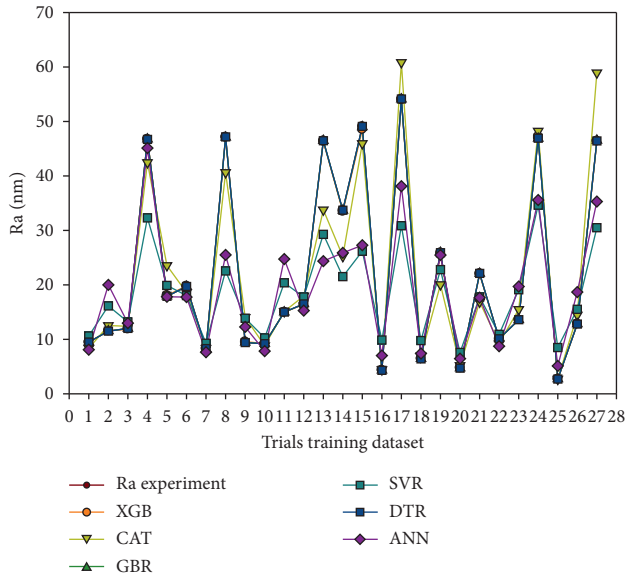


Figure 5: Comparison of Ra experiments and prediction of Ra by six deep learning models on the training dataset.

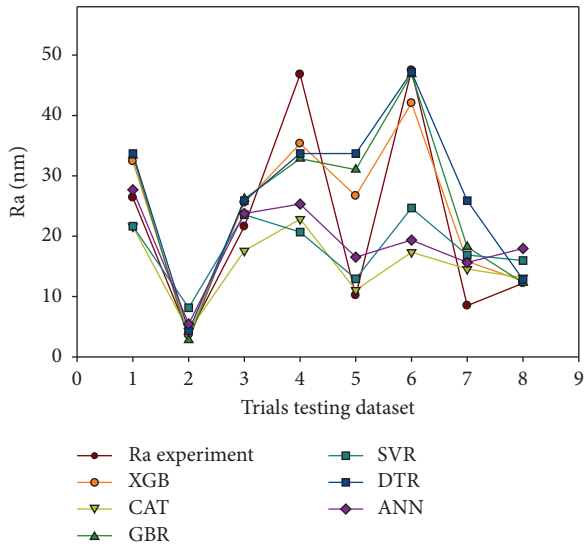


Figure 6: Comparison of Ra experiment and prediction of Ra by six deep learning models on the testing dataset.

dataset. Therefore, to continue to improve the prediction accuracy for Ra, we choose 2 models: XGB and CAT, for the next procedure, in keeping with the approach illustrated in Figure 2.

4.2. Description of the Optimization Problems. The XGB and CAT methods require tuning hyperparameters to prevent overfitting and improve model performance. Table 4 presents the hyperparameters of the XGB and CAT models. As mentioned above, XGB and CAT belong to regression with a large number of hyperparameters. The value of the parameter is crucial and so must be carefully selected. However, to date, only heuristic methods have been put forward.

As discussed in Section 2, the overfitting in XGB and CAT is negated by optimizing their hyperparameters using the Bayesian optimization approach. To save computational time, in this paper, we focus only on the hyperparameters that have a significant effect on the model performance, as found in previous studies.

The two optimization problems (one for XGB and one for CAT) of this study are described in Table 4. For XGB, the decision variables are learning_rate, max_depth, subsample, colsample_bytree, reg_alpha, max_leaves, gamma, and min_child_weight. On the other hand, for CAT, the decision variables are learning_rate, depth, bagging_temperature, and num_leaves. The inferior and superior bounds of these variables are also given in Table 4. In this work, the objective of the optimization problems is to minimize the value of RMSE between predicted and experimental data points.

4.3. Hyperparameter Tuning with Bayesian Optimization. The settings of the Bayesian optimization algorithm search domain are derived from historical data and also from initial tests [72–74]. In terms of other hyperparameters such as "*n*_estimators" (number of boosted trees), "min_child_samples" (minimum number of data points needed in a leaf), and "subsample_for_bin" (number of samples for constructing bins), default settings in Python have been applied [54]. Table 5 presents the evolution of the optimization procedure for XGB, for instance, whereas Table 6 indicates the best value found for each hyperparameter of the two methods XGB and CAT, respectively.

As discussed above, the objective of this study is to find the best model for predicting surface roughness. Therefore, XGB and CAT have been selected for performance comparison, with the Bayesian optimization algorithm being applied to each model. Table 7 demonstrates the comparison models under the metrics of RMSE, MAE, and R^2 found by XGB and CAT, based on their hyperparameters in the best in the BO and the default in the baseline models. By comparing, we found the effect of hyperparameters on the metrics of the models. The correlation exhibits a promising RMSE and MAE in XGB_opt, achieving the lowest values shown in Figures 7 and 8 . Moreover, the metric of R^2 significantly increases from 0.9999 to 1 in performance on the training dataset, and for the testing dataset, it increases from 0.7227 to 0.8924.

The line and scatter plots of the measured Ra and the values predicted by the optimized models XGB_opt and CAT_opt on the training and testing datasets are presented in Figures 9 and 10 for each trial, respectively. The different colors of the scattered points in the figures represent the values predicted by different baseline models. In the training and testing dataset, the predictions of the XGB_opt model are the closest to the original measured values of Ra. The value predicted by the CAT model does not closely correspond to the original values of Ra measured in both datasets.

However, as discussed above, RMSE is the preferred criterion for selecting better models in relation to the training dataset. Therefore, to continue to improve the prediction accuracy for Ra, we choose 2 models: XGB and

Table 4: Description of the optimization problems in this study.

Parameter	For XGB	For CAT
Decision variables	learning_rate max_depth subsample colsample_bytree reg_alpha max_leaves gamma min_child_weight	learning_rate depth bagging_temperature num_leaves
Objective	Minimize(test-RMSE)	Minimize(test-RMSE)
Bounds of decision variables	$0.001 \le$ learning_rate ≤ 0.3 $5 \le$ max_depth ≤ 30 $0.5 \le$ subsample ≤ 1 $0.3 \le$ colsample_bytree ≤ 1 $0.005 \le$ reg_alpha ≤ 0.02 $0 \le$ max_leaves ≤ 0.02 $0 \le$ gamma ≤ 1 $1 \le$ min_child_weight ≤ 10	$0.001 \le$ learning_rate ≤ 0.5 $5 \le$ depth ≤ 8 $3 \le$ bagging_temperature ≤ 10 $30 \le$ num_leaves ≤ 150

Table 5: Evolution of optimization procedure for XGB (first 50 iterations only).

Iter	Target	Colsample_bytree	Gamma	Learning_rate	Max_depth	Min_child_weight	Reg_alpha	Subsample
1	−7.056	0.8801	0.8997	0.2827	20.34	4.861	0.006904	0.5405
2	−6.516	0.5613	0.4843	0.09498	6.732	5.114	0.007405	0.9069
3	−16.69	0.5848	0.2136	0.1965	19.36	9.066	0.01777	0.6896
4	−8.5	0.9966	0.8161	0.1771	5.423	7.812	0.008559	0.9149
5	−8.822	0.7103	0.598	0.1231	6.284	6.038	0.007801	0.9096
6	−5.852	0.5	0.3518	0.06219	7.254	4.037	0.006944	0.9037
7	−9.343	0.5	0.5106	0.01885	5.633	3.34	0.005851	0.8316
8	−6.15	0.5	0.2702	0.1017	8.293	4.807	0.007843	0.9567
9	−7.601	1	1	0.3	7.727	4.384	0.02	0.5
10	−6.5	1	1	0.3	20.87	2.587	0.005	0.5
11	−6.26	1	1	0.3	22.83	3.751	0.005	0.5
12	−25.55	0.5	0	0.001	23.13	1.845	0.02	1
13	−7.601	1	1	0.3	22.05	4.527	0.005	0.5
14	−6.26	1	1	0.3	19.62	3.248	0.005	0.5
15	−7.601	1	1	0.3	23.95	4.842	0.005	0.5
16	−25.48	0.5	0	0.001	8.813	3.053	0.005	1
17	−5.912	0.5	0.1875	0.2799	7.554	4.955	0.008345	1
18	−25.44	0.5	0.8634	0.001	8.342	5.686	0.005	1
19	−6.418	0.5345	0.2921	0.2898	8.177	4.709	0.01122	0.9844
20	−11.41	0.6047	0.03536	0.01905	7.271	4.588	0.01701	0.6337
21	−5.658	0.5	0.73	0.3	7.259	4.412	0.005	1
22	−6.839	0.9758	0.9844	0.2983	20.45	3.433	0.005328	0.5055
23	−25.48	0.531	0.8686	0.001	6.918	3.876	0.005	1
24	−6.541	0.5	0.4384	0.2496	7.453	4.4	0.00543	0.9837
25	−16.64	0.8722	0.1358	0.05604	13.76	9.028	0.01675	0.5171
26	−5.996	0.6539	0.2812	0.1042	7.178	4.032	0.01673	0.919
27	−6.889	0.7965	0.4768	0.2854	7.075	5.049	0.01465	0.9937
28	−5.911	0.8249	0.2209	0.1936	24.47	1.008	0.01477	0.6331
29	−16.66	0.6999	0.2054	0.2231	16.34	7.16	0.01825	0.5339
30	−6.5	1	1	0.3	20.18	2.931	0.005	0.5
31	−6.559	0.5	0.7239	0.3	7.301	4.792	0.005	0.9979
32	−6.974	0.5	0.08312	0.23	7.39	3.94	0.005782	0.9513
33	−6.219	0.7449	0.6745	0.2195	19.98	3.314	0.01041	0.6499
34	−6.614	0.8682	0.3464	0.03563	14.74	2.282	0.006153	0.8187
35	−6.252	0.8297	0.9535	0.2535	14.33	3.153	0.007105	0.6695
36	−11.64	0.6522	0.5504	0.02992	13.11	6.753	0.01481	0.7805
37	−11.8	0.8327	0.5598	0.02797	15.28	5.888	0.009776	0.6274
38	−6.562	0.7739	0.2751	0.293	17.17	4.494	0.01663	0.8818

Table 5: Continued.

Iter	Target	Colsample_bytree	Gamma	Learning_rate	Max_depth	Min_child_weight	Reg_alpha	Subsample
39	−6.339	0.8632	0.9508	0.1429	21.26	1.244	0.01272	0.9479
40	−5.643	0.8481	0.6627	0.1492	14.53	2.736	0.006655	0.741
41	−5.792	0.5397	0.2942	0.09671	7.252	4.123	0.01874	0.8521
42	−5.838	0.8593	0.7312	0.1051	14.02	2.605	0.005175	0.7117
43	−6.016	0.6624	0.2504	0.1153	8.384	3.601	0.00867	0.7995
44	−6.313	0.8624	0.2579	0.2642	13.42	1.179	0.01796	0.9567
45	−5.629	0.7182	0.3777	0.122	8.18	3.918	0.009016	0.7433
46	−5.981	0.7067	0.9652	0.3	14.33	2.576	0.02	0.9897
47	−9.125	0.8218	0.602	0.2969	18.08	6.542	0.0195	0.777
48	−6.014	0.9996	0.9764	0.3	14.38	2.526	0.02	0.5146
49	−16.74	0.7799	0.6707	0.007666	25.28	1.548	0.007938	0.7785
50	−6.792	0.5	0.7475	0.3	14.31	2.629	0.02	0.5203

Table 6: Optimized hyperparameters for XGBoost and CAT.

Algorithm	Library	Hyperparameters tuned	Domain	Optimized value
XGB	XGBoost	*n*_estimators	100	100
		learning_rate	(0.001, 0.3)	0.1220
		max_depth	(5, 30)	9
		subsample	(0.5, 1)	0.7433
		colsample_bytree	(0.3, 1)	0.7182
		reg_alpha	(0.005, 0.02)	0.0090
		max_leaves	(0, 0.02)	0
		gamma	(0, 1)	0.3778
		min_child_weight	(1, 10)	3.9175
CAT	catboost	iterations	(1000, 2000)	1000
		depth	(5, 8)	6.1114
		learning_rate	(0.001, 0.5)	0.4669
		bagging_temperature	(3, 10)	6.8176
		num_leaves	(30, 150)	138

Table 7: Statistical accuracy of ML model regression performing 27 trials on the training dataset and 8 trials on the testing dataset.

Model	Training dataset			Testing dataset		
	RMSE	MAE	R^2	RMSE	MAE	R^2
XGB	0.0076	0.0063	0.9999	0.4033	0.2845	0.7227
XGB bayes_opt	0.0047	0.0027	1	0.2512	0.2225	0.8924
CAT	0.1594	0.1224	0.9565	0.4704	0.3503	0.6229
CAT bayes_opt	0.0792	0.0579	0.9893	0.3940	0.2870	0.7355

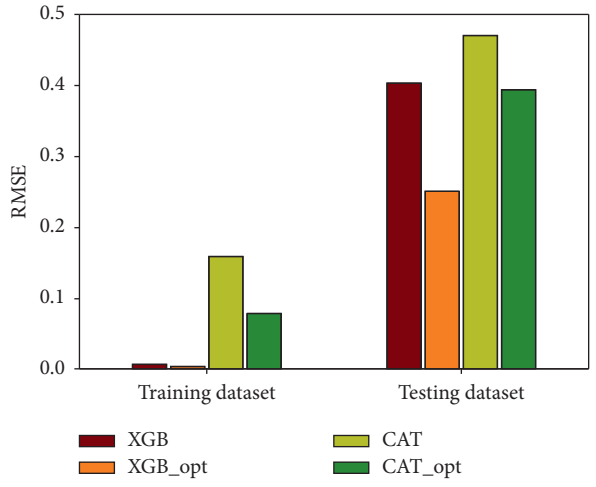


Figure 7: Comparison as a function of RMSE.

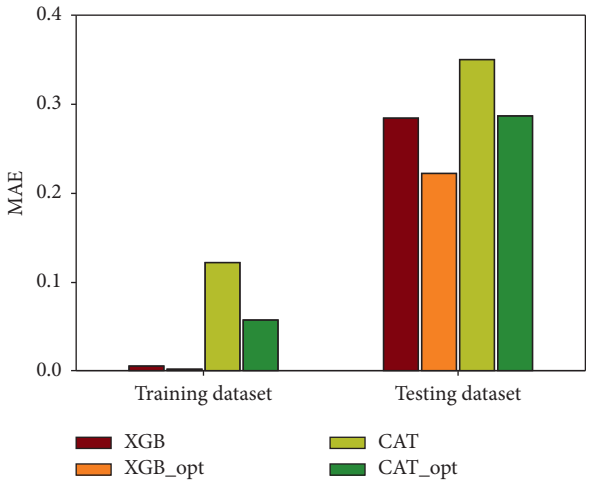


Figure 8: Comparison as a function of MAE.

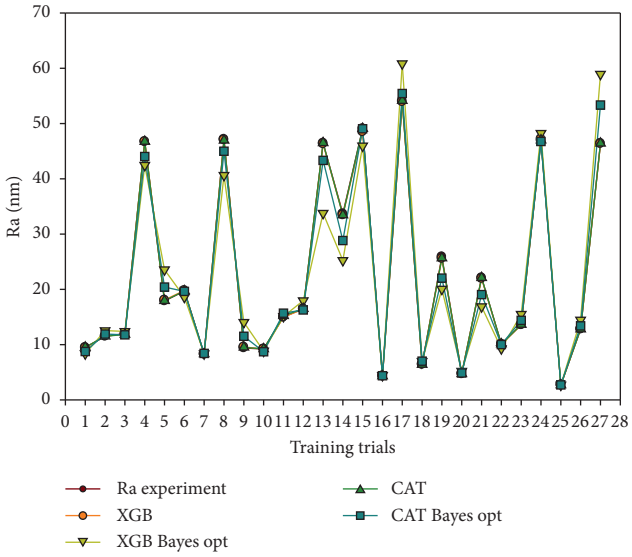


Figure 9: Comparison of Ra experiments and prediction of Ra by baseline and their Bayesian optimization of XGB and CAT machine learning models on the training dataset.

CAT, for the next procedure, in keeping with the approach illustrated in Figure 2.

5. Discussion

With the rise of modern technology, in the production process of PC for optical applications by single-point diamond turning, the equipment maintains a stable operating state, so the production process is also carried out stably. Therefore, for given strip specifications, most turning processes will obtain relatively stable datasets without wide variability. The key point of prediction models is to improve prediction performance. The main goal of this paper is to determine the optimal predictive model in relation to surface roughness, by comparing the performances of different

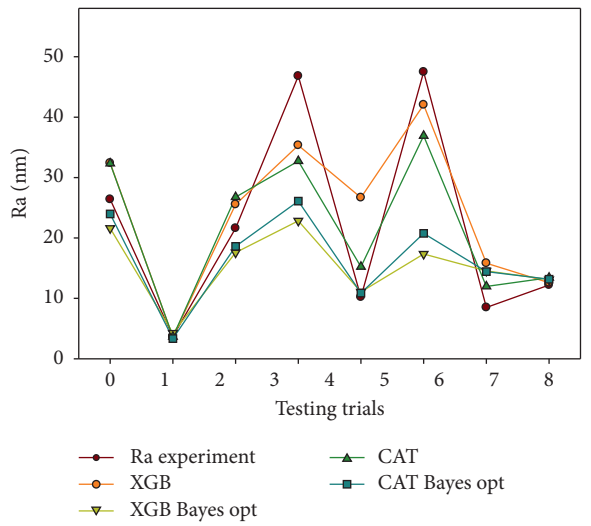


Figure 10: Comparison of Ra experiments and prediction of Ra by baseline and their Bayesian optimization of XGB and CAT machine learning models on the testing dataset.

models. In addition, in order to preserve the principal characteristics of the considered models, in this paper, we have not coupled those models with other optimization algorithms. Then, we performed hyperparameter tuning by using BO to find the best model.

By executing the XGB, we can determine which features contribute most heavily to the result. In short, the importance of each feature can be found and plotted. The feature importance graphs for each dataset are plotted in Figure 11. From Figure 11, we can clearly see that Z-axis vibration is the most important feature which has contributed to the prediction of the results, followed by the Y-axis vibration and X-axis vibration. This result is in accordance with Kara and Bayraktar [75], in which experiments showed that the surface roughness increased while increasing vibration of the machining tool (i.e., increase of the cutting speed). For instance, in Kara and Bayraktar [75], the cutting speed exhibited 42.14% of the influence on the surface roughness using the analysis of variance. However, it is worth noticing that the contrary is not confirmed (low value of vibration does not imply good surface quality) Bolat [23]. Also shown in Figure 11, the feed rate is a crucial variable highly affecting the surface roughness of the material. Kara and Bayraktar [75] and Krolczyk et al. [10] have also found similar results in their works. It has been obtained in these studies such that a low value of feed rate provides a small value of surface roughness and vice versa. An important point should be noticed such that there is an influence of feed rate on the surface roughness and energy consumption. Such consideration has not been conducted in the present study. Moreover, another limitation of the present study is that the lack of confirmation tests. Such procedures should be carried in order to assess the performance of the prediction model. Finally, analysis of variance should be processed in further studies to reveal how the input variables influence each other and the surface roughness. Such information is crucial to control the turning process.

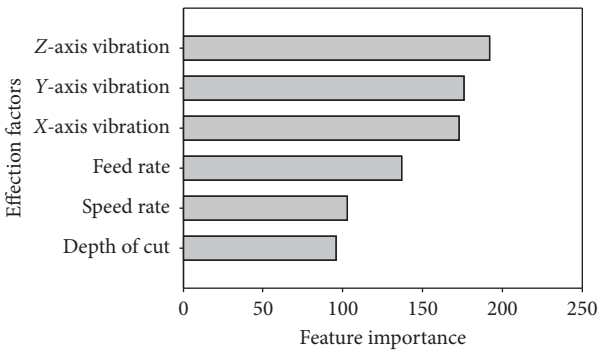


Figure 11: Importance of each contribution feature to the result.

6. Conclusions

In this work, six machine learning models—namely, ANN, SVR, CAT, DRT, GBT, and XGB—have been applied to predict surface roughness in the SPDT process. We have compared the performances of the six models using experimental data. Three quality assessment metrics (RMSE, MAE, and R^2) have been used to evaluate the performance of each model. Subsequently, hyperparameter optimization using BO has been applied. The results are summarized as follows:

(1) The result comparison has shown that there is little difference in terms of ranking when using testing RMSE. For example, the values of RMSE for XGB, CAT, GBT, GBT, MLP-NN, SVR, and DTR are 0.4033, 0.4704, 0.4812, 0.4829, 0.5032, and 0.5712, respectively. However, it has been concluded that overall, XGB, and CAT perform better, followed by GBR, SVR, and DTR, respectively

(2) The results in relation to hyperparameters using BO stability show inconsistency in the evaluation metrics. XGB_opt performs best with RMSE, MAE, and R^2 in both the training and testing datasets. It can be seen that using BO has improved significantly the performance of the 2 considered models, especially CAT. Indeed, the value of R^2 when using CAT has increased from 0.9565 to 0.9893 for the training dataset and from 0.6229 to 0.7355 for the testing dataset. The value of MAE when using CAT has decreased from 0.1224 to 0.0579 for the training dataset and from 0.3503 to 0.2870 for the testing dataset. The value of RMSE has decreased from 0.1594 to 0.0792 for the training dataset and from 0.4704 to 0.3940 for the testing dataset

(3) Comprehensively considering the prediction accuracy of the six models, XGB appeared to be the best model for predicting surface roughness. Indeed, this model has shown strong performance in terms of different qualitative estimators. For example, XGB exhibited the highest values in terms of R^2 for both training and testing datasets. In terms of other estimators such as RMSE and MAE, XBG also produced one of the best values. In addition, XGB also performs best when combined with Bayesian optimization algorithm

(4) We can clearly see that machine vibration is the most important feature in contributing to the prediction of the results—especially, vibration on the Z-axis, followed by the Y-axis vibration and X-axis vibration. The least important feature is the depth of cut

(5) Overall, out of six initially proposed models, we have succeeded to filter out the 2 best models based on different qualitative estimators such as RMSE, MAE, and R^2 for predicting the surface roughness in turning polycarbonate for optical application. In addition, we have improved the performance of these two best models using the Bayesian optimization algorithm

Certainly, further studies should be conducted to enhance the results of the present paper. From an experimental point of view, confirmation tests should be carried out in order to test the performance of the proposed method in real conditions. Besides, working parameters such as feed rate, cut depth, spindle speed, and cutting speed should be related to the surface roughness by an explicit equation, providing an easy and direct application by engineers/researchers in practice. This means that the proposed machine learning model should be employed to derive such an equation in further research studies. From a computational point of view, uncertainty quantification should also be performed, in order to propagate the variability of the experimental database. Besides, the number of data points for training and testing the models was not large, which may yield some unexpected effect in the results. For such a reason, more data points should be collected in further studies, or experimental tests should be conducted by the researchers' group.

Nomenclature

PC: Polycarbonate
ML: Machine learning
BO: Bayesian optimization
ANN: Artificial neural network
MLP: Multilayer perceptron
CAT: Cat Boost Regression
CAT_opt: CAT built with BO
SVR: Support Vector Regression
GBT: Gradient boosting trees
GBR: Gradient Boosting Regression
DTR: Decision Tree Regression
XGB: Extreme Gradient Boosting Regression
XGB_opt: XGB built with BO
EDA: Exploratory data analysis
RMSE: Root mean square error
MAE: Mean absolute error
R^2: Correlation coefficient
VX: Vibration of X-axis (mgRMS)
VY: Vibration of Y-axis (mgRMS)
VZ: Vibration of Z-axis (mgRMS).

Data Availability

The Excel data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] V. Sudarsan, "Optical materials," in *Functional Materials: Preparation, Processing and Applications*, pp. 285–322, Elsevier, London, UK, 2012.

[2] M. Roeder, T. Guenther, and A. Zimmermann, "Review on fabrication technologies for optical mold inserts," *Micromachines*, vol. 10, no. 4, p. 233, 2019.

[3] Y. Jaluria, *Advanced Materials Processing and Manufacturing*, Springer, Berlin, Germany, 2018.

[4] C. Gerhard, *Optics Manufacturing: Components and Systems*, CRC Press, Boca Raton, FL, USA, 2017.

[5] H. de Brouwer, J. van den Bogerd, and J. Hoover, "Color stability of polycarbonate for optical applications," *European Polymer Journal*, vol. 71, pp. 558–566, 2015.

[6] C. He and W. Zong, "Influencing factors and theoretical models for the surface topography in diamond turning process: a review," *Micromachines*, vol. 10, no. 5, p. 288, 2019.

[7] V. Saini, D. Sharma, S. Kalla, and T. Chouhan, "Optimisation of process parameter in ultra-precision diamond turning of polycarbonate material," in *Proceedings of the International Conference on Manufacturing Excellence MANFEX*, Noida, India, 2012.

[8] M. Elangovan, N. R. Sakthivel, S. Saravanamurugan, B. B. Nair, and V. Sugumaran, "Machine learning approach to the prediction of surface roughness using statistical features of vibration signal acquired in turning," *Procedia Computer Science*, vol. 50, pp. 282–288, 2015.

[9] G. Krolczyk, M. Gajek, and S. Legutko, "Predicting the tool life in the dry machining of duplex stainless steel," *Eksploatacja I Niezawodnosc—Maintenance and Reliability*, vol. 15, no. 1, pp. 62–65, 2013.

[10] G. Krolczyk, S. Legutko, and M. Gajek, "Predicting the surface roughness in the dry machining of duplex stainless steel (DSS)," *Metalurgija-Sisak Then Zagreb-*, vol. 52, pp. 259–262, 2013.

[11] M. Kuntoğlu, A. Aslan, D. Y. Pimenov, K. Giasin, T. Mikolajczyk, and S. Sharma, "Modeling of cutting parameters and tool geometry for multi-criteria optimization of surface roughness and vibration via response surface methodology in turning of AISI 5140 steel," *Materials (Basel, Switzerland)*, vol. 13, p. 4242, 2020.

[12] M. K. Gupta, M. Mia, C. I. Pruncu et al., "Parametric optimization and process capability analysis for machining of nickel-based superalloy," *The International Journal of Advanced Manufacturing Technology*, vol. 102, no. 9–12, pp. 3995–4009, 2019.

[13] A. Eser, E. Aşkar Ayyıldız, M. Ayyıldız, and F. Kara, "Artificial intelligence-based surface roughness estimation modelling for milling of AA6061 alloy," *Advances in Materials Science and Engineering*, vol. 2021, Article ID 5576600, 10 pages, 2021.

[14] M. S. Alajmi and A. M. Almeshal, "Prediction and optimization of surface roughness in a turning process using the ANFIS-QPSO method," *Materials*, vol. 13, no. 13, p. 2986, 2020.

[15] A. T. Abbas, M. Alata, A. E. Ragab, M. M. El Rayes, and E. A. El Danaf, "Prediction model of cutting parameters for turning high strength steel grade-H: comparative study of regression model versus ANFIS," *Advances in Materials Science and Engineering*, vol. 2017, Article ID 2759020, 12 pages, 2017.

[16] S. K. Pal and D. Chakraborty, "Surface roughness prediction in turning using artificial neural network," *Neural Computing and Applications*, vol. 14, no. 4, pp. 319–324, 2005.

[17] M. Nalbant, H. Gokkaya, and İ. Toktaş, "Comparison of regression and artificial neural network models for surface roughness prediction with the cutting parameters in CNC turning," *Modelling and Simulation in Engineering*, vol. 2007, Article ID 092717, 14 pages, 2007.

[18] Y. Ö. Özgören, S. Çetinkaya, S. Sarıdemir, A. Çiçek, and F. Kara, "Artificial neural network based modelling of performance of a beta-type Stirling engine," *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, vol. 227, no. 3, pp. 166–177, 2013.

[19] N. K. Vuong, Y. Xue, S. Liu, Y. Zhou, and M. Wu, "Predicting surface roughness and flank wear in turning processes," in *Proceedings of the 2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*, pp. 1–6, IEEE, Detroit, MI, USA, 2020.

[20] T. Zhou, L. He, J. Wu, F. Du, and Z. Zou, "Prediction of surface roughness of 304 stainless steel and multi-objective optimization of cutting parameters based on GA-GBRT," *Applied Sciences*, vol. 9, no. 18, p. 3684, 2019.

[21] Z. Jurkovic, G. Cukor, M. Brezocnik, and T. Brajkovic, "A comparison of machine learning methods for cutting parameters prediction in high speed turning process," *Journal of Intelligent Manufacturing*, vol. 29, no. 8, pp. 1683–1693, 2018.

[22] D. Y. Pimenov, A. Bustillo, and T. Mikolajczyk, "Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth," *Journal of Intelligent Manufacturing*, vol. 29, no. 5, pp. 1045–1061, 2018.

[23] M. Bolat, *Machining of Polycarbonate for Optical Applications*, Middle East Technical University, Ankara, Turkey, 2013.

[24] H. Thanh Duong, H. Chi Phan, T.-T. Le, and N. Duc Bui, "Optimization design of rectangular concrete-filled steel tube short columns with balancing composite motion optimization and data-driven model," *Structures*, vol. 28, pp. 757–765, 2020.

[25] F. Morin and Y. Bengio, "Hierarchical probabilistic neural network language model," in *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistic*, pp. 246–252, Bridgetown, Barbados, January 2005.

[26] T.-T. Le, V.-H. Nguyen, and M. V. Le, "Development of deep learning model for the recognition of cracks on concrete surfaces," *Applied Computational Intelligence and Soft Computing*, vol. 2021, Article ID 8858545, 10 pages, 2021.

[27] A. C. Tsoi and A. Back, "Discrete time recurrent neural network architectures: a unifying review," *Neurocomputing*, vol. 15, no. 3-4, pp. 183–223, 1997.

[28] J. Cheng, L. Dong, and M. Lapata, "Long short-term memory-networks for machine reading," 2016, http://arxiv.org/abs/11601.06733.

[29] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 4, pp. 809–821, 2015.

[30] T.-T. Le, "Prediction of tensile strength of polymer carbon nanotube composites using practical machine learning method:," *Journal of Composite Materials*, vol. 55, no. 6, 2020.

[31] T.-T. Le, "Practical hybrid machine learning approach for estimation of ultimate load of elliptical concrete-filled steel tubular columns under axial loading," *Advances in Civil Engineering*, vol. 2020, Article ID 8832522, 19 pages, 2020.

[32] T.-T. Le, "Practical machine learning-based prediction model for axial capacity of square CFST columns," *Mechanics of Advanced Materials and Structures*, pp. 1–16, 2020.

[33] N. X. Ho and T.-T. Le, "Effects of variability in experimental database on machine-learning-based prediction of ultimate load of circular concrete-filled steel tubes," *Measurement*, vol. 176, Article ID 109198, 2021.

[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[35] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.

[36] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[37] H. Yang, L. Chan, and I. King, "Support vector machine regression for volatile stock market prediction," in *Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning*, pp. 391–396, Springer, Guimaraes, Portugal, November 2002.

[38] D. Gupta, M. Pratama, Z. Ma, J. Li, and M. Prasad, "Financial time series forecasting using twin support vector regression," *PLoS One*, vol. 14, Article ID e0211402, 2019.

[39] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *The Journal of Finance and Data Science*, vol. 4, no. 3, pp. 183–201, 2018.

[40] V. Anandhi and R. Manicka Chezian, "Support vector regression to forecast the demand and supply of pulpwood," *International Journal of Future Computer and Communication*, vol. 2, pp. 266–269, 2013.

[41] S. Kumar, V. Kumar, and R. K. Sharma, "Rice yield forecasting using support vector machine," *International Journal of Recent Technology and Engineering*, vol. 8, no. 4, pp. 2588–2593, 2019.

[42] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Application of support vector regression for modeling low flow time series," *KSCE Journal of Civil Engineering*, vol. 23, no. 2, pp. 923–934, 2019.

[43] A. Suárez Sánchez, P. J. García Nieto, P. Riesgo Fernández, J. J. del Coz Díaz, and F. J. Iglesias-Rodríguez, "Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain)," *Mathematical and Computer Modelling*, vol. 54, no. 5-6, pp. 1453–1466, 2011.

[44] S. Saravanamurugan, S. Thiyagu, N. R. Sakthivel, and B. B. Nair, "Chatter prediction in boring process using machine learning technique," *International Journal of Manufacturing Research*, vol. 12, no. 4, pp. 405–422, 2017.

[45] T.-T. Le and M. V. Le, "Development of user-friendly kernel-based Gaussian process regression model for prediction of load-bearing capacity of square concrete-filled steel tubular members," *Materials and Structures*, vol. 54, no. 2, p. 59, 2021.

[46] P. W. Khan, Y.-C. Byun, S.-J. Lee, D.-H. Kang, J.-Y. Kang, and H.-S. Park, "Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources," *Energies*, vol. 13, no. 18, p. 4870, 2020.

[47] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost: gradient boosting with categorical features support," 2018, http://arxiv.org/abs/1810.11363.

[48] G. Huang, L. Wu, X. Ma et al., "Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions," *Journal of Hydrology*, vol. 574, pp. 1029–1041, 2019.

[49] Y. Zhang, J. Ma, S. Liang, X. Li, and M. Li, "An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products," *Remote Sensing*, vol. 12, no. 24, p. 4015, 2020.

[50] O. Z. Maimon and L. Rokach, *Data Mining with Decision Trees: Theory and Applications*, World Scientific, Singapore, 2014.

[51] H. C. Phan, T.-T. Le, N. D. Bui, H. T. Duong, and T. D. Pham, "An empirical model for bending capacity of defected pipe combined with axial load," *International Journal of Pressure Vessels and Piping*, vol. 191, Article ID 104368, 2021.

[52] M. S. Alajmi and A. M. Almeshal, "Predicting the tool wear of a drilling process using novel machine learning XGBoost-SDA," *Materials*, vol. 13, no. 21, p. 4952, 2020.

[53] S. Ramraj, N. Uzir, R. Sunil, and S. Banerjee, "Experimenting XGBoost algorithm for prediction and classification of different datasets," *International Journal of Control Theory and Applications*, vol. 9, pp. 651–662, 2016.

[54] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, CA, USA, August 2016.

[55] S. Cerna, C. Guyeux, G. Royer, C. Chevallier, and G. Plumerel, "Predicting fire brigades operational breakdowns: a real case study," *Mathematics*, vol. 8, no. 8, p. 1383, 2020.

[56] K. D. Kankanamge, Y. R. Witharanage, C. S. Withanage, M. Hansini, D. Lakmal, and U. Thayasivam, "Taxi trip travel time prediction with isolated XGBoost regression," in *Proceedings of the 2019 Moratuwa Engineering Research Conference (MERCon)*, pp. 54–59, IEEE, Moratuwa, Sri Lanka, 2019.

[57] Y. Li, X. Meng, Z. Zhang, and G. Song, "A remaining useful life prediction method considering the dimension optimization and the iterative speed," *IEEE Access*, vol. 7, pp. 180383–180394, 2019.

[58] R. S. Peres, J. Barata, P. Leitao, and G. Garcia, "Multistage quality control using machine learning in the automotive industry," *IEEE Access*, vol. 7, pp. 79908–79916, 2019.

[59] L. Munkhdalai, T. Munkhdalai, O.-E. Namsrai, J. Lee, and K. Ryu, "An empirical comparison of machine-learning methods on bank client credit assessments," *Sustainability*, vol. 11, no. 3, p. 699, 2019.

[60] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–285, 2012.

[61] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS 2011)*, Granada, Spain, 2011.

[62] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: a python library for model selection and hyper-parameter optimization," *Computational Science & Discovery*, vol. 8, no. 1, 2015.

[63] C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Auto-WEKA: combined selection and hyperparameter

optimization of classification algorithms," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 847–855, Chicago, IL, USA, 2013.

[64] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Proceedings of the International Conference on Learning and Intelligent Optimization*, pp. 507–523, Springer, Athens, Greece, 2011.

[65] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization, Chaos," *Solitons & Fractals*, vol. 142, Article ID 110511, 2020.

[66] K. Zhang, L. Zheng, Z. Liu, and N. Jia, "A deep learning based multitask model for network-wide traffic speed prediction," *Neurocomputing*, vol. 396, pp. 438–450, 2020.

[67] T.-T. Le and H. C. Phan, "Prediction of ultimate load of rectangular CFST columns using interpretable machine learning method," *Advances in Civil Engineering*, vol. 2020, Article ID 8855069, 16 pages, 2020.

[68] T.-T. Le and M. V. Le, "Nanoscale effect investigation for effective bulk modulus of particulate polymer nanocomposites using micromechanical framework," *Advances in Materials Science and Engineering*, vol. 2021, Article ID 1563845, 13 pages, 2021.

[69] T.-T. Le, "Probabilistic modeling of surface effects in nanoreinforced materials," *Computational Materials Science*, vol. 186, Article ID 109987, 2021.

[70] T.-T. Le, "Probabilistic investigation of the effect of stochastic imperfect interfaces in nanocomposites," *Mechanics of Materials*, vol. 151, Article ID 103608, 2020.

[71] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[72] W. Y. Loh, "Classification and regression trees," *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

[73] J. Elith, J. R. Leathwick, and T. Hastie, "A working guide to boosted regression trees," *Journal of Animal Ecology*, vol. 77, no. 4, pp. 802–813, 2008.

[74] G. G. Moisen, E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards Jr., "Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods," *Ecological Modelling*, vol. 199, no. 2, pp. 176–187, 2006.

[75] F. Kara and F. Bayraktar, "Investigation of the effect on surface roughness of cryogenic process applied to cutting tool," *International Journal of Analytical, Experimental and Finite Element Analysis*, vol. 7, no. 2, pp. 19–27, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9921383, 25 pages
https://doi.org/10.1155/2021/9921383

Research Article

Development and Optimization for a New Planar Spring Using Finite Element Method, Deep Feedforward Neural Networks, and Water Cycle Algorithm

Ngoc Le Chau,[1] Hieu Giang Le,[1] Van Anh Dang,[2] and Thanh-Phong Dao [3,4]

[1]*Faculty of Mechanical Engineering, Ho Chi Minh University City of Technology and Education, Ho Chi Minh City, Vietnam*
[2]*Faculty of Mechanical Engineering, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam*
[3]*Division of Computational Mechatronics, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Vietnam*
[4]*Faculty of Electrical & Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam*

Correspondence should be addressed to Thanh-Phong Dao; daothanhphong@tdtu.edu.vn

Received 17 March 2021; Revised 5 May 2021; Accepted 31 May 2021; Published 14 June 2021

Academic Editor: Noorbakhsh Amiri Golilarz

The gravity balance mechanism plays a vital role in maintaining the equilibrium for robots and assistive devices. The purpose of this paper was to optimize the geometry of a planar spring, which is an essential element of the gravity balance mechanism. To implement the optimization process, a hybrid method is proposed by combining the finite element method, the deep feedforward neural network, and the water cycle algorithm. Firstly, datasets are collected using the finite element method with a full experiment design. Secondly, the output datasets are normalized to eliminate the effects of the difference of units. Thirdly, the deep feedforward neural network is then employed to build the approximate models for the strain energy, deformation, and stress of the planar spring. Finally, the water cycle algorithm is used to optimize the dimensions of the planar spring. The results found that the optimal geometries of the spring include the length of 45 mm, the thickness of 1.029 mm, the width of 9 mm, and the radius of 0.3 mm. Besides, the predicted results determined that the strain energy, the deformation, and the stress are 0.01123 mJ, 33.666 mm, and 79.050 MPa, respectively. The errors between the predicted result and the verifying results for the strain energy, the deformation, and the stress are about 1.87%, 1.69%, and 3.06%, respectively.

1. Introduction

A device is balanced when it can maintain equilibrium in any configuration or position without the need for external forces or actuators [1]. Generally, when a device achieves a gravity balance, it can move like in a gravity-less environment [2, 3]. For a device to reach equilibrium, scientists have researched balancers to compensate for the gravity of machines. Gravity balances can be classified in the following ways: active balance and passive balance, and gravity balancer mechanisms can adjust loads or cannot adjust loads [4]. Besides, some other mechanisms employ counterweights, springs, or coupling of counterweights and springs [2]. Among them, a gravity balance mechanism with springs is of great interest for research by scientists because of its suitability in applications and simple adjustment [5, 6]. These types of mechanisms are capable of adjusting loads via using traditional springs [7, 8].

Previously, a gravity balance mechanism with adjusted loads was designed by the combination of a compliant spring and a torsion spring [9]. Besides, a compliant rotary joint for the gravity balance mechanism was designed by creating a torque that balances the torque of the load [10]. The spring and the joint were built based on advantages of compliant mechanisms; however, the stiffness of the spring is not high [11].

Although scientists have had great success in studying many of the different types of mechanisms, a large amount of energy is still required during the adjustment process. Therefore, the present study proposes a new gravity balance

mechanism according to the principle of compliant mechanisms. Compliant mechanism is selected to create the planar spring (PS) because it is a monolithic mechanism, which offers less lubricant, no friction, and minimal cost of manufacturing [12]. The gravity balance mechanism is a combination of a torsion spring and a planar spring. This combination allows the spring to easily adjust its stiffness when adjusting the load. In the proposed balancing mechanism, the PS is a very important detail. The PS is deformed during the working process, and it creates a strain energy and elastic force to compensate for the gravity generated by the mass. In order to work well, the strain energy of the PS should offer as large as possible and the stiffness of the PS should be ensured to create an elastic force that balances with the gravity due to the mass. In addition, the stress concentration must be less than the permissible stress.

In order to serve for a practical application, the PS should concern a full set of performances, including deformation, strain energy, and stress. However, these desirable properties have mutual contradictions, and these properties are very sensitive to PS geometrical dimensions. Therefore, optimization of the PS is necessary to balance its properties. In this study, an optimization process is performed to maximize or minimize one or more properties of the PS. It is noted that the proposed PS is designed based on the principle of compliant mechanisms. It is therefore difficult to build precisely mathematical models that show the relationship between deformation, stress, and strain energy with its geometrical factors. Therefore, in this study, a hybrid method of finite element method (FEM), deep feedforward neural network (DFNN), and water cycle algorithm (WCA) is aimed to build surrogate models and optimize the geometry of the PS.

Nowadays, the FEM is a widely used method to solve complex arithmetic problems [13, 14]. This method is capable of accurately predicting the phenomena and behavior of complex problems with little cost and time. Until now, the FEM has been used effectively to simulate the behavior of compliant mechanism [15, 16], healthy human intervertebral discs [17], and double-row tapered roller bearing [18]. Therefore, this study adopts the FEM to predict the behaviors of the PS. The DFNN is extended to model the performances [19]. The DFNN has the ability to approximate the relationships between the independent input variables to the output responses of highly nonlinearly complex models. However, the accuracy of the approximate model depends greatly on the structure of the DFNN [20]. Therefore, before building the approximate models, the structure optimization of the DFNN needs to be performed. This process is to find the training function, the number of hidden layers, the number of nodes in each hidden layer, and the data division rate that best matches the dataset. The proposed PS is desired to achieve the maximum strain energy but must ensure the required stiffness and durability when working. This is an optimization problem with many constraints. The WCA is capable of solving the optimization with multiple constraints [21, 22]. Therefore, the WCA is used to optimize the geometry of the PS. Although there have been studies on many different gravity balance mechanisms, a full consideration of the performances of the spring has not been considered yet.

In summary, the present paper is aimed to optimize the geometry factors of a planar spring, which is used for the gravity compensation mechanism. To perform the optimization process, the FEM method is used to simulate and collect data. Next, regression models are built by the DFNN. Before using the DFNN, the structure of the DFNN is selected by optimizing the controllable parameters of the DFNN by using the Taguchi method. Finally, the WCA is applied to optimize the geometry of the planar spring.

2. Mechanical Design

2.1. Design of Gravity Balancing Mechanism. Figure 1 shows a design of a gravity balance mechanism. It consists of a bar that carries an object. The bar is connected to the planar spring (k_2) through the pulley and cable. The planar spring is combined with the pulley and cable to create a zero-free-length spring. At the hinge joint, a torsion spring with stiffness k_1 is considered as a rotation spring. During the working process, when the bar rotates clockwise by angle φ, the torsion spring is also deformed by an angle φ, and the planar spring is stretched by an amount equal to l.

Then, the moment balance equation of the mechanical system is expressed as follows:

$$(0.5m_1 + m_2)Lg \sin\varphi = k_1\varphi + k_2 l, \tag{1}$$

where m_1 is the mass of the bar, m_2 is the mass of the object, L is the length of the bar, k_1 is the stiffness of torsion spring, k_2 is the stiffness of a spring of zero free length, and l is deformation of a spring of zero free length.

$$l = 2a^2 \sin\varphi, \tag{2}$$

where a is the distance from the rotating center to the connection point of a spring of zero free length.

In the proposed design, the bar is made of aluminum alloy with $m_1 = 0.5$ kg, the mass m_2 of the object allows to be adjusted in the range from 0.5 kg to 3 kg. The distance a is 65 mm and the length L of the bar is 400 mm. To ensure working requirements, the stiffness of the torsion spring is selected as 200 N/mm and the stiffness of the planar spring is adjustable in the range from 0.325 N/mm to 1.468 N/mm. Based on the required stiffness, a planar spring is designed in Section 2.2.

2.2. Design Planar Spring. To ensure an adjustable stiffness in the range of 0.325 N/mm to 1.468 N/mm, deformation of the spring is required from 0 to 30 degrees, and the planar spring is designed with 31 component springs that are connected in series, as shown in Figure 2. The component springs are formed by arranging the leaf springs in parallel and series styles, as shown in Figure 3. The geometrical parameters of the PS include the thickness t, length L, width w, radius r, height H, and width a.

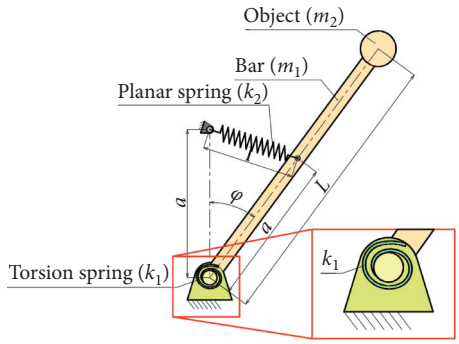


Figure 1: Structure of gravity balance mechanism.

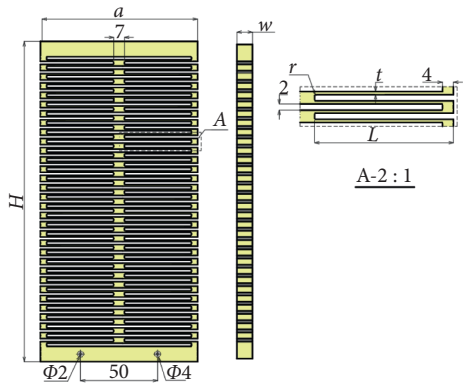


Figure 2: Structure of planar spring.

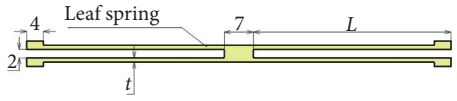


Figure 3: Structure of component spring.

3. Formulation of Optimization Problem

In order to offer an efficient work for the gravity balance mechanism, the performances of the PS should be enhanced because the efficiency of the overall mechanism is highly dependent on the properties of the PS. Meanwhile, the performances of the PS are very sensitive to the geometrical dimensions of the leaf springs, as given in Figures 2 and 3. The main design parameters include the length L, width w, thickness t, and radius r of the leaf spring. Therefore, if there is an expectation to improve the characteristics of the spring, the geometry of the leaf springs should be optimized. In this gravity balance mechanism, the spring should be met the following requirements:

(i) The generated deformation has to be large enough so that the bar rotates at an angle from 0 to 30 degrees, and creates a moment for balancing.

(ii) The generated stress must be less than the yield strength of the material. In this study, a maximum load of 3 kg and a stiffness of planar spring of 1.436 N/mm are utilized.

(iii) The strain energy is as large as possible so that the PS can do highly efficient work.

(iv) The fatigue should also be considered. It is remarked that the performances of the PS often contradict each other. Therefore, the question arises as to how to balance these properties. So, the WCA optimization algorithm is applied to solve the optimization with multiple constraints.

3.1. Design Variables. Based on the proposed design in Figures 2 and 3, the deformation, stress, and strain energy of the PS depend on the leaf springs' properties. The leaf spring's properties are very sensitive to length L, width w, thickness t, and radius r. Therefore, these dimensions are chosen as the design variables. The limited dimensions of the design variables are selected based on the machining technology capabilities, the desired size of the system, and the designer's experience. These are represented as follows:

$$\begin{cases} 45\,\text{mm} \le L \le 55\,\text{mm}, \\ 1\,\text{mm} \le t \le 1.4\,\text{mm}, \\ 9\,\text{mm} \le w \le 14\,\text{mm}, \\ 0.1\,\text{mm} \le r \le 0.3\,\text{mm}. \end{cases} \tag{3}$$

3.2. Objective Function. As stated in the formulation of optimization problem above, the PS should posses a large strain energy to store and release a good elastic deformation. Therefore, the strain energy $f_1(\mathrm{X})$ is chosen as the objective function of the optimization process.

3.3. Constraint Functions. In order to ensure the gravity balance mechanism safely, the PS must create a large enough deformation so that the mechanism can work in the range of 0–30 degrees. In addition, the deformation of the PS should be sufficient to create a moment that is equal to the moment generated by the mass. In addition, the stress of the PS should be less than the yield strength when the mechanism is operating at maximum load to ensure that the spring works safely. Hence, the deformation $g_1(\mathbf{X})$ and stress $g_2(\mathbf{X})$ are selected as the constraints of the optimization process.

The optimization problem for the PS is briefly stated as follows:

Find: $\mathbf{X} = [r, l, t, w]$

$$\max f_1(\mathbf{X}). \tag{4}$$

Subject to constraints,

$$\begin{cases} g_1(\mathbf{X}) = 33.646\,\text{mm}, \\ g_2(\mathbf{X}) \le 120\,\text{MPa}. \end{cases} \tag{5}$$

4. Proposed Hybrid Optimization Approach

As discussed in Section 3, in order to ensure the efficient working of the gravity balancing mechanism, the PS must have three properties (deformation, stress, and strain energy). However, these properties have mutual contradiction. Hence, a hybridization of experiment design, FEM, DFNN, and WCA is developed. First, the experimental design is built by full factor experiment for collecting data. Then, the DFNN was used to establish the approximate models. Before establishing an approximate model, the structure of the DFNN needs to be optimized to get the best structure of the DFNN. Finally, the WCA is employed to optimize the geometrical sizes of the spring. This method is carried out through 4 stages, as shown in Figure 4.

4.1. Stage 1: Mechanical Design. The mechanical design process is performed by the following steps:

Step 1: identify the problem

When the working load is changed, the balance condition will be broken. Besides, when the load is adjusted, the gravity balance mechanism needs to be changed to maintain the equilibrium condition. One of the popular methods for adjusting the gravity balance mechanism is an adjustment of the stiffness of the spring. In this study, the proposed planar spring is designed to adjust easily the number of active leaf springs. So, the stiffness of PS can be adjusted without extra energy. To meet work requirements of the gravity balance mechanism, the design and optimization of the PS is essential.

Step 2: original design

Based on the requirements of the problem, the original gravity balance mechanism is designed, as shown in Figure 1. The principle of this mechanism is presented in Section 2.1. This study proposes the structure of PS as shown in Figure 2.

Step 3: define design variables, objective functions, and constraints

The PS is designed to meet technical requirements. A deformation must be large enough, the stress must be less than the yield strength, and the strain energy is large. The objective function, constraint, and design variable are presented in Section 3.

4.2. Stage 2: General Numerical Data. The purpose of this stage is to create datasets to build the approximate models, and the datasets were generated by the numerical simulation in ANSYS 18.2 software. The sequence of steps is as follows:

Step 4: experiment design

Experimental design is the statistical technique that is widely used in product development. To accurately evaluate the effect of each design variable on the performance of a product, the full factor experiment is used.

Step 5: simulate and collect data

Simulation is a technique of predicting the behavior of the structure. It is applied in engineering to reduce the costs of experimentation. A 3D model is built, and the boundary conditions and load are set up. The simulation process is performed to collect data.

Step 6: normalize data

The properties of a planar spring have different units; to avoid the effect of the different units on the optimal result, the output response should be normalized. This normalization only changes the value of the response, but does not change the nature of the data. Moreover, this process can evaluate fairly. In this study, the properties of PS need to be standardized to unify the units. The standardization process is performed by the following formula:

$$z_i = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)}, \tag{6}$$

where z_i is the normalized value and y_i is the real value.

4.3. Stage 3: Optimize the Structure of the DFNN. To approximate the relationship between the geometry sizes and the output characteristics of the PS, the DFNN is applied to create regression models. The working diagram of the DFNN is shown in Figure 5. Net input (z_i^j) of node j_{th} of layer i_{th} is calculated according to the following formula:

$$z_i^j = \sum_{i=1}^{n} \left(w_{i,j}^j a_{i-1}^j \right) + b_i^j, \tag{7}$$

where w_i^j is the weighted input of the node j_{th} of the hidden layer i_{th}, b_i^j is a constant value called bias, and a_{i-1}^j output of node j_{th} of layer $(i-1)_{th}$.

The net input of node j_{th} of layer i_{th} is then passed through an activation function to produce the output of node j_{th} of layer i_{th}, which is then transmitted to other neurons.

$$a_i^j = g\left(z_i^j\right), \tag{8}$$

where g is the active function.

The approximate models are greatly dependent on the structure of the DFNN and the actual data. To get the exact approximate models, the structure of the DFNN is optimized in this study. The optimization process for the structure of the DFNN is performed as follows:

Step 7: define the objective function of the DFNN structure.

The accuracy of the approximate model is usually assessed through coefficient of determination (R^2) and mean squared error (MSE). In practice, if the value of MSE is smaller, the value of R^2 will be larger and the predicted model is more accurate. Therefore, in this study, MSE is selected as the objective function to optimize the structure of the DFNN.

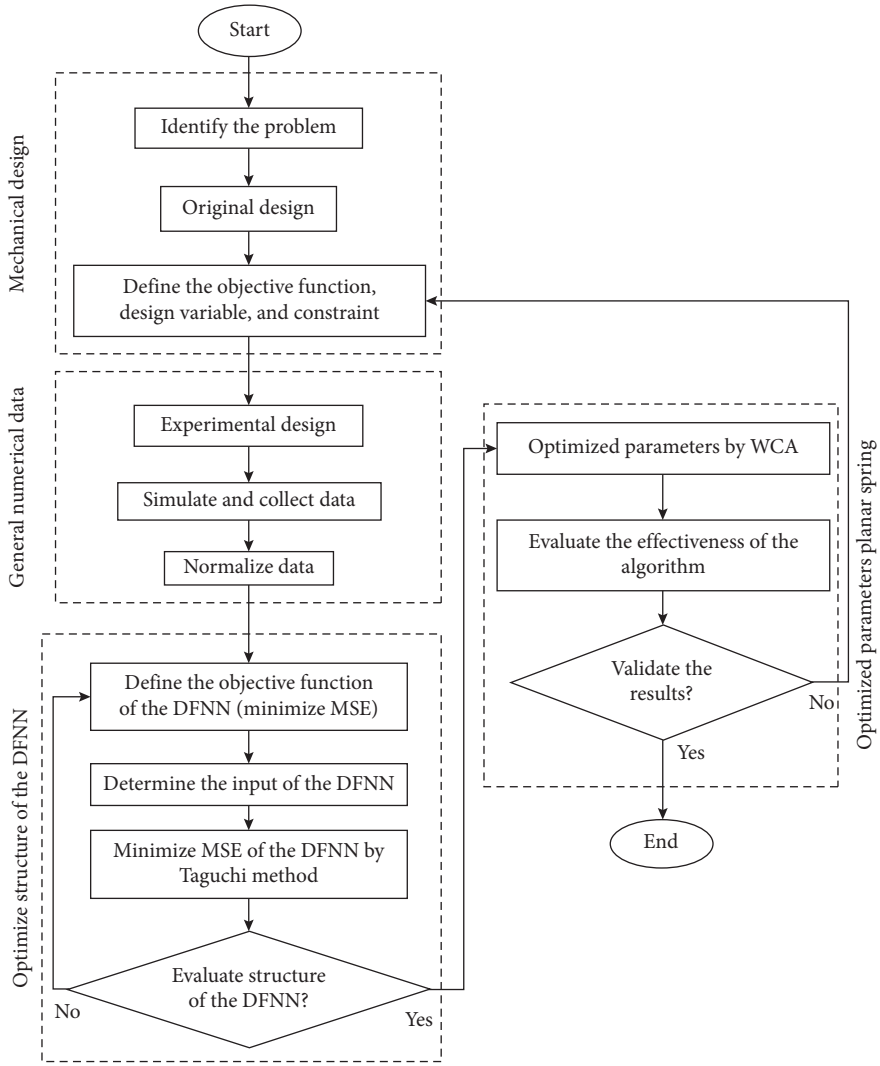


Figure 4: Flowchart of the hybrid approach.

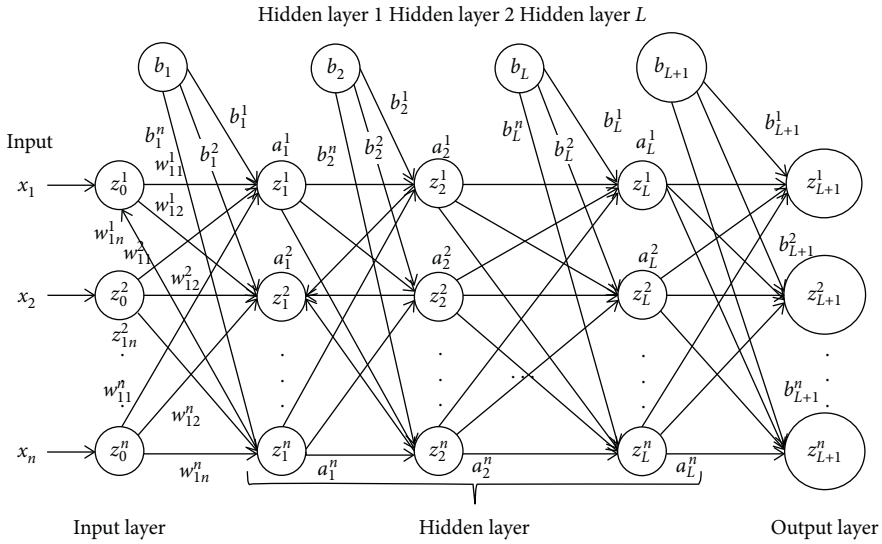


Figure 5: Structure of deep feedforward neural network.

The MSE and R^2 are calculated by using the following formula:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2,$$
$$R^2 = 1 - \frac{\text{residual sum of squares}}{\text{total sum of squares}}, \tag{9}$$

where y_i is the ith observed value $\hat{y}_i$is the ith predicted value and n is the experiment number.

Step 8: select the input parameters of the DFNN structure

A basic structure of the DFNN includes the number of hidden layers, the number of nodes in each hidden layer, training function, activation function, bias coefficient, ratio of division of data for training, validation, and testing. However, in this study, the number of hidden layers, the number of nodes in each hidden layer, the transfer function and the ratio of division of data are chosen as controllable variables.

Normally, researchers usually use training functions: trainlm, trainbr, and trainscg. However, trainbr only uses 2 datasets: training and validating while the other transfer functions use 3 datasets: training, validating, and testing. Therefore, to evaluate equity, this study proposes to choose 3 training functions: trainlb, traincgs, and trainscg.

The number of nodes in hidden layers can be selected differently based on the position of the hidden layer and the number of nodes of the input layer. According to Chen [23], the number of nodes in the hidden layer can be calculated as follows:

$$\begin{aligned}\text{number node of hidden layer } 1 &= 2x + 1,\\ \text{number node of hidden layer } i+1 &= n(2x+1),\end{aligned} \tag{10}$$

where x is the number of input variables, and n is the number of output variables. In this study, $x = 4$ and $n = 1$, so the number of nodes in each hidden layer is selected around 9.

According to Seo et al. [20], the accuracy of the approximate models can be improved as the number of hidden layers increases. However, the complexity of the algorithm also increases. The authors also indicate that when the number of hidden layers is equal or larger than 4, the accuracy increased slightly. For the data that have a small signal to noise, the approximate model has a good accuracy as well when using 2 or 3 hidden layers. Therefore, in this study, the number of hidden layers is selected as 2, 3, and 4. The ratio of division of data for training, validating, and testing is chosen with the ratios of 60 : 20 : 20, 70 : 15 : 15, and 80 : 10 : 10.

Step 9: collect data and optimize the structure of the DFNN

Taguchi is often used to improve a product's quality [24]. Nowadays, the Taguchi method is widely used to solve technical problems, especially for the single-objective problem with discrete variables [25, 26]. Hence, it is chosen to optimize the structure of the DFNN. The Taguchi method uses two major tools, including signal-to-noise ratio (S/N) and orthogonal array. There are three quality characteristics: "nominal is best," "smaller is better," and "larger is better." Their mathematical formula can be presented as follows:

Nominal is the best:

$$\text{S/N} = 10\log_{10}\left[\left(\frac{\overline{y}}{s}\right)^2\right]. \tag{11}$$

The smaller is better:

$$\text{S/N} = -10\log_{10}\left(\frac{\sum_{i=1}^{N} y_i^2}{N}\right). \tag{12}$$

The bigger is better:

$$\text{S/N} = -10\log_{10}\left(\frac{\sum_{i=1}^{N}\left(1/y_i^2\right)}{N}\right), \tag{13}$$

where y_i is responses, and N is repetition of i^{th} experiment.

The Taguchi method used orthogonal arrays to reduce the number of experiments required but ensured that design variables are evaluated independently. The purpose of this step is to find the structure of the DFNN that best matches the existing dataset by minimizing MSE by using equation (12).

Step 10: evaluate the value of MSE

After optimizing the structure of the DFNN, the value of MSE has to be evaluated to find the most suitable structure for the dataset. If the value of MSE is not satisfied, it means that the most suitable structure has not been found, then go back to step 8. If the value of MSE is satisfied, then go to the next step.

4.4. Stage 4: Optimization Using the WCA. The WCA is an optimal algorithm inspired by the water cycle. It was developed by Eskandar [21] and widely used to solve technical problems [27–29]. However, this algorithm has not been applied to solve the problem of optimizing strain energy of the PS. The flowchart of the WCA is presented, as shown in Figure 6.

Step 11: choose the initial parameters of the WCA

Initial rain drops (N_{pop}—initial population), number of rivers and sea (N_{sr}), maximum distance between streams and rivers, between rivers and sea $d_{\max}$, and the maximum iteration must be preselected. N_{sr} and N_{raindrop} are calculated as follows:

$$\begin{aligned}N_{\text{sr}} &= \text{number of rivers} + \text{sea},\\ N_{\text{raindrop}} &= N_{\text{pop}} - N_{\text{sr}},\end{aligned} \tag{14}$$

where N_{sr} is the number of sea and rivers, N_{pop} is the number of the initial population, and N_{raindrop} is the number of streams.

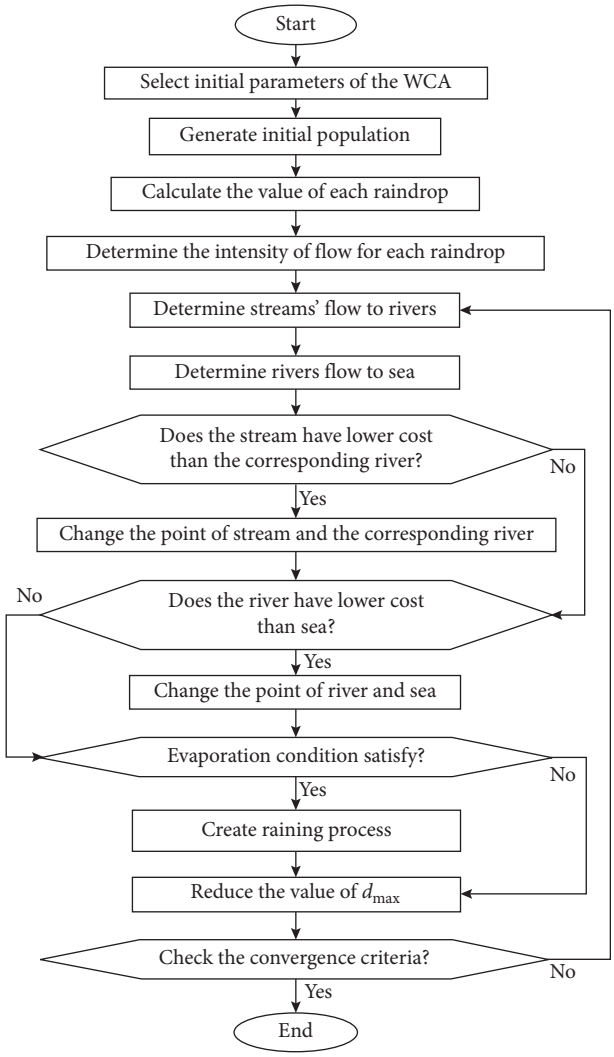


Figure 6: Flowchart of the WCA.

Step 12: generate random initial population

Initial population is generated by random variables as follows:

$$\text{population of raindrop} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_{N_{\text{var}}}^1 \\ x_1^2 & x_2^2 & \cdots & x_{N_{\text{var}}}^2 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_1^{N_{\text{pop}}} & x_2^{N_{\text{pop}}} & \cdots & x_{N_{\text{var}}}^{N_{\text{pop}}} \end{bmatrix}, \tag{15}$$

where x_i^j is the value of the variable i_{th} of the raindrop j_{th}, i is the number of variables, j is the number of the initial population, and N_{var} is the number of variables.

Step 13: calculate the value of the initial raindrop

In order to select raindrops representing the sea and rivers, the value of the initial raindrop needs to be calculated and they are calculated according to the following formula:

$$P_i = \text{cost}_i = f(\mathbf{X}_i), \quad i = 1, 2, 3, \ldots, N_{\text{pop}}, \tag{16}$$

where P_i is the initial raindrop value. i is raindrop.

Step 14: determine the intensity of flow for rivers and sea

The raindrops flow to the rivers or sea depending on the intensity of the flow. The intensity of flow for rivers and sea are calculated by the following formula:

$$NS_n = \text{round}\left\{\left|\frac{P_n}{\sum_{i=1}^{N_{sr}} P_i}\right| \cdot N_{\text{raindrop}}\right\}, n = 1, 2, \ldots, N_{\text{sr}}, \tag{17}$$

where NS_n is the number of streams flowing directly into the rivers or sea, P_n is the value of sea or rivers n_{th}

Step 15: the streams flow to the rivers

During the streams' flow to rivers, the positions of streams are continuously updated. The new position for the stream is determined as follows:

$$X_{\text{stream}}^{i+1} = X_{\text{stream}}^i + \text{rand} \times C \times \left(X_{\text{river}}^i - X_{\text{stream}}^i\right), \tag{18}$$

where rand is a random number that is chosen between 0 and 1, X_{stream}^i is the location of the stream i_{th}, X_{river}^i is the location of the river i_{th}, and C is a value between 1 and 2.

Step 16: the rivers flow to the sea

Similar to a stream, when rivers flow to the sea, the position of the river is always updated, and the new position of the river is determined as follows:

$$X_{\text{river}}^{i+1} = X_{\text{river}}^i + \text{rand} \times C \times \left(X_{\text{sea}}^i - X_{\text{river}}^i\right), \tag{19}$$

where rand is a random number that is chosen between 0 and 1, X_{river}^i is the location of the stream i_{th}, X_{sea}^i is the location of the river i_{th}, and C is a value between 1 and 2.

Step 17: exchange positions of the river with a stream

To offer the best solution, the new cost of a stream is updated. If the solution given by a stream is better than its connecting river, the positions of the river and stream are exchanged

Step 18: exchange positions of the sea with a river

Similar to streams, if the solution of the river is better than the sea, the positions of the sea and the river will be swapped

Step 19: check the evaporation condition

Evaporation is one factor that prevents the algorithm from being optimized locally. As seen in nature, water from ponds, lakes, rivers, streams, seas, and the ground evaporates into clouds. Clouds fly up high then condense in the cold air to form water particles. Water particles fall down and form rain. Raindrops will form

new streams that flow into rivers and the sea. In the WCA, the condition evaporation is determined as follows:

$$\left|X^i_{\text{sea}} - X^i_{\text{river}}\right| < d_{\max}, \quad i = 1, 2, 3, \ldots, N_{sr} - 1, \tag{20}$$

where $d_{\max}$ is small number (close to zero),

When the distance from the river to sea is less than $d_{\max}$, it means the river has flowed into the sea. Then, the evaporation process is applied. After the evaporation is completed, rain will begin to create a new cycle.

Step 20: reduce the value of $d_{\max}$

In the WCA, the value of $d_{\max}$ controls the search intensity near the sea. If the value of $d_{\max}$ is large, then the search density will reduce while a small value for $d_{\max}$ will increase the search density near the sea. Therefore, the value for $d_{\max}$ has to adaptively decrease as:

$$d^{i+1}_{\max} = d^i_{\max} - \frac{d^i_{\max}}{\max \text{iteration}}. \tag{21}$$

Step 21: check the convergence criteria

Like other optimization algorithms, the stop condition of the WCA algorithm can be based on the maximum number of iterations, the CPU time, or the error between the last two results less than a certain tolerance value. If the stopping condition is satisfied, the optimization will be finished, otherwise return to step 5.

5. Results and Discussion

5.1. Collecting and Processing Data. To perform simulation and collect data, firstly, the experimental design was constructed by a full experimental design. The four design variables of PS include r, L, t, and w, which are divided into three levels, as shown in Table 1. A matrix of experimental designs is formed, as seen in Table 2. Then, 3D models of the PS are drawn by Inventor software as shown in Figure 2. Next, the simulation for data collection is performed through ANSYS 18.2.

In this study, a nonlinear FEM is applied for the simulation process. The simulation process is set up as follows: the boundary condition and loads are given as shown in Figure 7. The PS is fixed with two holes, and a force of 10.935 N is applied to the top surface of the PS. Al T63-7075 is selected for the PS. The mechanical properties of the material are given in Table 2. The mesh is divided by using the sizing method. The element size greatly affects the behavior of the PS. So, before the simulation, an exploratory meshing was performed to ensure the good results at the lowest computational cost. The results show that the smaller the mesh size, the larger the processing time and memory used for the simulation. However, if the mesh size is 1 mm, the deformation, stress, and strain energy change very little. Hence, a mesh size of 1 mm is chosen for this study. The result meshes have 101190 elements and 632792 nodes. The quality of the mesh greatly affects the simulation results. Especially, the PS has large deformation. Therefore, in this study, two mesh quality assessment standards are used at the same time. Skewness standard and Orthogonal Quality standard are utilized as two criteria to evaluate the quality of meshing. The quality of the mesh is assessed according to Skewness criteria. The results indicate that the average value of the meshing metric is about 0.21141, and the standard deviation of the meshing metric is approximately 0.194, as shown in Figure 8. According to the Skewness criteria, this value proved that the accuracy of meshing models is excellent. The quality of the mesh is assessed according to Orthogonal Quality criteria. The results indicate that the average value of the meshing metric is about 0.93498, and the standard deviation of the meshing metric is approximately 0.0751, as given in Figure 8. According to the Orthogonal Quality criteria, this value proved that the accuracy of meshing models is very good.

In this study, strain energy is considered an objective function; therefore, strain energy is a factor that needs data collection. In addition, deformation and stress are considered constraints so they also need to be collected as data. The results of data collection are shown in Table 3.

The properties of the PS have different units. Therefore, in this study, datasets are normalized to eliminate the units of measurement for data. This normalization makes the evaluation process fair. Data are normalized using (6). Normalization results are presented in Table 4.

5.2. Optimization of the DFNN Structure. As presented in Section 4, the DFNN is used to approximate the relationship between the design variable and the output characteristics. The accuracy of the approximated model depends on the structure of the DFNN. In this study, four factors of the DFNN structure are selected: training function, number of hidden layers, number of nodes in each hidden layer, and the division ratio of the dataset to create the most accurate approximated model possible. The factors are divided with 3 levels as illustrated in Table 5 and the orthogonal array L_9 is used to establish the experimental plan as shown in Table 6.

As shown in Section 4, normally, the approximation model's accuracy is measured through two indices MSE and R^2. However, when the MSE is small, the R^2 is large and the approximation model has high accuracy. Therefore, in this study, MSE is selected as the objective function for the structural optimization of the DFNN. The approximate results depend on a weighted input that selects random as equation (7). So, in order to optimize the structure of the DFNN, each experiment is performed 10 times to collect data for MSE. MSE data collection results for deformation, stress, and strain energy are presented in Tables 7–9.

Based on MSE results obtained for deformation, stress, and strain energy, the *S/N* calculation and analysis for each experiment are performed, then ranking process for each input parameter of the DFNN structure are performed.

The S/N analysis results for deformations (Figure 9) show that, the training function "Trainlm" has the smallest error and the function "Trainscg" gives the largest error. The structure of the DFNN consists of 2 hidden layers for the

Table 1: Design variables with three levels.

Variables	Level 1	Level 2	Level 3
t (mm)	1.0	1.2	1.4
L (mm)	45	50	55
w (mm)	9	11.5	14
r (mm)	0.1	0.2	0.3

Table 2: Mechanical properties of the material.

Density	Yield strength (MPa)	Poisson's ratio	Young modulus (MPa)
2810 kg/m^3	503	0.33	71700

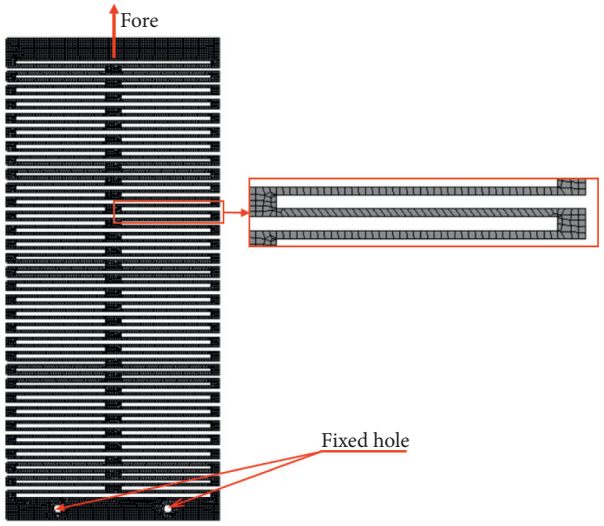


Figure 7: Meshing model and boundary conditions.

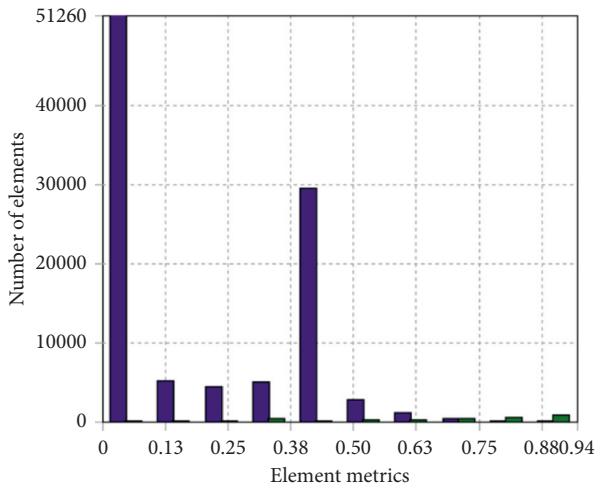


Figure 8: Evaluating the quality of the meshing.

smallest MSE value and 4 hidden layers with the largest error. The results found that when the number of nodes is 7 nodes, the error of the DFNN model is the smallest. Meanwhile, the when the number of nodes is 9 nodes, the error of the DFNN model is the largest. Besides, when the data are divided into the ratio of 60:20:20, the error of the DFNN model is the smallest; while the data ratio is 80:10:10, the error of the DFNN model is the largest. The analysis results also show that the training function in the DFNN structure has the greatest impact on the accuracy of the approximate model and the number of nodes in each hidden layer has the least effect on the accuracy of the equivalent model as shown in Table 10.

For the stress dataset, Figure 10 shows that the training function "Trainlm" gives the largest S/N value and the training function "Trainscg" the minimum S/N value. Thus, the accuracy of the approximate model is best achieved using the training function "Trainlm." At the same time, the number of hidden layers is 2, the number of nodes in a hidden layer is 7, and the data division ratio 70 : 15 : 15 gives the largest S/N. This means that the approximate model has the highest accuracy when the structure has 2 hidden layers, the number of nodes in each hidden layer is 7, and the data division ratio is 70 : 15 : 15. Table 11 also shows that, with data for stress, the number of layers hidden in the DFNN structure has the most influence on the accuracy of the approximate model.

Figure 11 depicts the calculation and analysis S/N results for the strain energy dataset. The results show the DFNN structure with the training function of "Trainlm," the number of hidden layers of 2, the number of nodes in each hidden layer of 7, and the data division ratio of 60 : 20 : 20 give the best accurate approximate model. Table 12 shows that the number of hidden layers is ranked the highest. This means that among the four factors of the DFNN structure, the number of hidden layers has the highest influence on the accuracy of the approximate model.

Table 13 shows that for the deformation dataset, the optimal structure of the DFNN is: The training function is "trainlm," the number of hidden layers is 2, the number of nodes in hidden layers is 7, and the data division ratio is 60 : 20 : 20. For the stress dataset, the optimal structure of the DFNN is: The training function is "trainlm," the number of hidden layers is 2, the number of nodes in hidden layers is 11, and the data division ratio is 70 : 15 : 15. For the strain energy dataset, the optimal structure of the DFNN is: The training function is "trainlm," the number of hidden layers is 2, the number of nodes in hidden layers is 9, and the data division ratio is 60 : 20 : 20.

To evaluate the effectiveness of the structure of the DFNN, the appropriateness of the approximate model built by the DFNN and the linear model is compared with each other.

Table 3: Experimental design and simulation data.

No.	t (mm)	L (mm)	w (mm)	r (mm)	Deformation (mm)	Stress (MPa)	Strain energy (mJ)
1	1	45	9	0.1	37.648	143.88	0.01136
2	1	45	9	0.2	37.359	126.56	0.011931
3	1	45	9	0.3	37.181	91.524	0.01208
4	1	45	11.5	0.1	29.158	110.5	0.0075272
5	1	45	11.5	0.2	28.933	97.469	0.0073316
6	1	45	11.5	0.3	28.794	70.283	0.0074207
7	1	45	14	0.1	23.743	89.134	0.004959
8	1	45	14	0.2	23.559	78.896	0.0052279
9	1	45	14	0.3	23.445	56.8	0.0052896
10	1	50	9	0.1	52.955	135.96	0.015283
11	1	50	9	0.2	52.482	183.47	0.015191
12	1	50	9	0.3	52.271	103.04	0.01534
13	1	50	11.5	0.1	41.035	105.25	0.010162
14	1	50	11.5	0.2	40.666	141.99	0.01011
15	1	50	11.5	0.3	40.502	79.32	0.010206
16	1	50	14	0.1	33.423	85.454	0.0067236
17	1	50	14	0.2	33.121	115.14	0.0066931
18	1	50	14	0.3	32.987	64.151	0.0067545
19	1	55	9	0.1	72.04	175.37	0.018836
20	1	55	9	0.2	71.362	203.9	0.018837
21	1	55	9	0.3	71.012	114.56	0.019014
22	1	55	11.5	0.1	55.856	135.11	0.012552
23	1	55	11.5	0.2	55.326	158.02	0.012564
24	1	55	11.5	0.3	55.053	88.312	0.012679
25	1	55	14	0.1	45.508	109.33	0.008325
26	1	55	14	0.2	45.075	128.3	0.0083395
27	1	55	14	0.3	44.852	71.513	0.008414
28	1.2	45	9	0.1	22.22	96.941	0.0065149
29	1.2	45	9	0.2	22.13	66.215	0.0063295
30	1.2	45	9	0.3	21.933	65.468	0.0062087
31	1.2	45	11.5	0.1	17.205	74.773	0.0040134
32	1.2	45	11.5	0.2	17.135	51.129	0.003901
33	1.2	45	11.5	0.3	16.981	50.323	0.003827
34	1.2	45	14	0.1	14.007	60.582	0.0028689
35	1.2	45	14	0.2	13.95	41.457	0.0027898
36	1.2	45	14	0.3	13.824	40.669	0.0027374
37	1.2	50	9	0.1	31.178	108.37	0.0082735
38	1.2	50	9	0.2	31.011	90.501	0.0080522
39	1.2	50	9	0.3	30.784	73.985	0.0079012
40	1.2	50	11.5	0.1	24.156	83.785	0.0055193
41	1.2	50	11.5	0.2	24.025	69.716	0.0053739
42	1.2	50	11.5	0.3	23.848	57.009	0.005274
43	1.2	50	14	0.1	19.671	67.908	0.0036635
44	1.2	50	14	0.2	19.565	56.337	0.003569
45	1.2	50	14	0.3	19.419	46.096	0.0035029
46	1.2	55	9	0.1	42.206	125.93	0.010211
47	1.2	55	9	0.2	42.035	82.775	0.010003
48	1.2	55	9	0.3	41.738	81.78	0.0098
49	1.2	55	11.5	0.1	32.717	97.869	0.0068261
50	1.2	55	11.5	0.2	32.584	64.17	0.0066893
51	1.2	55	11.5	0.3	32.353	63.11	0.0065549
52	1.2	55	14	0.1	26.652	79.533	0.0045419
53	1.2	55	14	0.2	26.543	52.124	0.004453
54	1.2	55	14	0.3	26.354	51.099	0.0043642
55	1.4	45	9	0.1	14.31	70.515	0.0028812
56	1.4	45	9	0.2	14.21	71.532	0.0023315
57	1.4	45	9	0.3	14.106	66.273	0.0021754
58	1.4	45	11.5	0.1	11.078	54.279	0.0017708
59	1.4	45	11.5	0.2	11	55.096	0.0014267
60	1.4	45	11.5	0.3	10.919	51.038	0.0013319

Table 3: Continued.

No.	t (mm)	L (mm)	w (mm)	r (mm)	Deformation (mm)	Stress (MPa)	Strain energy (mJ)
61	1.4	45	14	0.1	9.0172	43.868	0.0012658
62	1.4	45	14	0.2	8.9536	44.557	0.0010189
63	1.4	45	14	0.3	8.8876	41.251	0.00095013
64	1.4	50	9	0.1	20.005	79.204	0.0036771
65	1.4	50	9	0.2	19.885	73.889	0.002852
66	1.4	50	9	0.3	19.747	74.814	0.0027744
67	1.4	50	11.5	0.1	15.496	61.129	0.0024474
68	1.4	50	11.5	0.2	15.403	57.082	0.0018876
69	1.4	50	11.5	0.3	15.295	57.768	0.0018403
70	1.4	50	14	0.1	12.618	49.426	0.0016241
71	1.4	50	14	0.2	12.541	46.18	0.0012489
72	1.4	50	14	0.3	12.453	46.705	0.0012188
73	1.4	55	9	0.1	27.052	88.311	0.0038068
74	1.4	55	9	0.2	26.893	75.666	0.0035977
75	1.4	55	9	0.3	26.726	81.191	0.0034145
76	1.4	55	11.5	0.1	20.967	68.245	0.0025272
77	1.4	55	11.5	0.2	20.843	58.445	0.002389
78	1.4	55	11.5	0.3	20.713	62.752	0.0022659
79	1.4	55	14	0.1	17.078	55.247	0.0016786
80	1.4	55	14	0.2	16.976	47.321	0.0015868
81	1.4	55	14	0.3	16.87	50.783	0.0015041

Figure 12 compares the suitability of the proposed model with a linear model. Figure 12(a) shows that the predicted points of the proposed model are more suitable with the experimental point than the linear model for the deformation dataset. The predicted points of the proposed model are distributed around the good fitness line. Similarly, the proposed model has better fitness between the predicted points and the experimental points than the linear model for stress and strain energy datasets, as shown in Figures 12(b) and 12(c).

In addition, to evaluate the accuracy of the proposed model, the value R^2 and MSE of the proposed model are also compared with the value of the linear model. Table 14 compares the R^2 value of the proposed model with the linear model for the 4 datasets: model, training, testing, and validation. The results indicate that the proposed model has much better R^2 than the linear model. Table 15 also shows that the MSE of the proposed model is smaller than the error of the linear model for all three sets of deformation, stress, and strain energy data.

As illustrated in Figure 13(a), the MSE values of the training dataset of the deformation continuously decreased at the 9th epoch, but the MSE values of the testing and validating dataset still remained stable. As depicted in Figure 13(b), the MSE values of the training datasets of the stress still decreased, but the MSE values of the testing and validating datasets remained stable at 9th epoch. As depicted in Figure 13(c), the MSE values of the training datasets of the strain energy still decreased, but the MSE values of the testing and validating datasets remained stable at 8th epoch.

Figure 14 demonstrates that the errors of the three training, evaluation, and test datasets of all three sets of deformation, stress, and strain energy datasets are almost zero.

Figure 15 also demonstrates that the appropriateness of the approximate models is constructed by the DFNN through the value of R^2 for the datasets.

5.3. Optimal Results. After optimizing the structure of the DFNN, an alternative model using the DNFF was established. Then, the WCA algorithm was carried out using Matlab R2018a to optimize the parameters of the planar spring. The parameters of the WCA were initialized as initial rain drops of 50, number of rivers and sea of 4, maximum distance between streams and rivers, between river and sea of 10^{-5}, and the maximum iteration of 2000.

The optimal results were generated as Table 16. The evaporation-rate-based water cycle algorithm (ER-WCA) was developed based on the WCA to speed up convergence [29]. However, with the proposed PS, the WCA has higher convergence rate and stability than the ER-WCA. The results in Table 16 show that the optimal mean strain energy is 0.01123 when optimized by the WCA. It is higher than the ER-WCA optimal value. The standard deviation of energy when optimized by the WCA is also smaller than when optimized by the ER-WCA. The average search time also shows that the WCA has a shorter search time than the ER-WCA. The standard deviation of the search time of the WCA is also smaller than the ER-WCA. The optimal results show that the optimal parameters of the planar spring t, L, w, and r is 1.029 mm, 45 mm, 9 mm, 0.3 mm, respectively.

6. Verifications

Optimized results are used to build the 3D model. The model was then used for finite element analysis. Finite element analysis results are presented in Table 17 and Figure 16. FEM results show with optimal parameters. Characteristics of planar spring with strain energy, stress, deformation, and life are 0.01102 mJ, 34.236 mm, 81.475 MPa, and 299 million cycles, respectively. Comparison between the prediction results with the FEM results shows that the strain energy error is 1.87%, the strain error is 1.69%, and the stress error is 3.06%. With

Table 4: Data normalization results.

No.	t (mm)	L (mm)	w (mm)	r (mm)	Z_{Y1}	Z_{Y2}	Z_{Y3}
1	1	45	9	0.1	0.455413	0.6323	0.576281
2	1	45	9	0.2	0.450836	0.526193	0.607891
3	1	45	9	0.3	0.448018	0.311552	0.61614
4	1	45	11.5	0.1	0.320976	0.427805	0.364101
5	1	45	11.5	0.2	0.317413	0.347973	0.353273
6	1	45	11.5	0.3	0.315212	0.181424	0.358205
7	1	45	14	0.1	0.235231	0.296911	0.221928
8	1	45	14	0.2	0.232317	0.23419	0.236814
9	1	45	14	0.3	0.230512	0.098823	0.240229
10	1	50	9	0.1	0.697795	0.58378	0.793455
11	1	50	9	0.2	0.690305	0.87484	0.788362
12	1	50	9	0.3	0.686964	0.382103	0.796611
13	1	50	11.5	0.1	0.509045	0.395642	0.509961
14	1	50	11.5	0.2	0.503202	0.620722	0.507082
15	1	50	11.5	0.3	0.500605	0.236787	0.512397
16	1	50	14	0.1	0.388511	0.274366	0.319614
17	1	50	14	0.2	0.383729	0.456231	0.317926
18	1	50	14	0.3	0.381607	0.143857	0.321325
19	1	55	9	0.1	1	0.825217	0.990146
20	1	55	9	0.2	0.989264	1	0.990201
21	1	55	9	0.3	0.983722	0.452677	1
22	1	55	11.5	0.1	0.743731	0.578573	0.642269
23	1	55	11.5	0.2	0.735339	0.718926	0.642934
24	1	55	11.5	0.3	0.731016	0.291875	0.6493
25	1	55	14	0.1	0.579873	0.420637	0.408266
26	1	55	14	0.2	0.573017	0.536853	0.409069
27	1	55	14	0.3	0.569486	0.188959	0.413193
28	1.2	45	9	0.1	0.211115	0.344738	0.308061
29	1.2	45	9	0.2	0.20969	0.156502	0.297797
30	1.2	45	9	0.3	0.20657	0.151926	0.29111
31	1.2	45	11.5	0.1	0.131704	0.208931	0.16958
32	1.2	45	11.5	0.2	0.130595	0.064081	0.163358
33	1.2	45	11.5	0.3	0.128157	0.059143	0.159261
34	1.2	45	14	0.1	0.081064	0.121993	0.106221
35	1.2	45	14	0.2	0.080162	0.004828	0.101843
36	1.2	45	14	0.3	0.078166	0	0.098942
37	1.2	50	9	0.1	0.352962	0.414756	0.405415
38	1.2	50	9	0.2	0.350318	0.305285	0.393164
39	1.2	50	9	0.3	0.346723	0.204103	0.384805
40	1.2	50	11.5	0.1	0.241771	0.264141	0.252945
41	1.2	50	11.5	0.2	0.239696	0.17795	0.244896
42	1.2	50	11.5	0.3	0.236894	0.100104	0.239366
43	1.2	50	14	0.1	0.170752	0.166874	0.15021
44	1.2	50	14	0.2	0.169074	0.095987	0.144978
45	1.2	50	14	0.3	0.166762	0.033247	0.141319
46	1.2	55	9	0.1	0.527587	0.522333	0.512674
47	1.2	55	9	0.2	0.524879	0.257953	0.501159
48	1.2	55	9	0.3	0.520177	0.251858	0.489921
49	1.2	55	11.5	0.1	0.377332	0.350424	0.325289
50	1.2	55	11.5	0.2	0.375226	0.143974	0.317715
51	1.2	55	11.5	0.3	0.371568	0.13748	0.310275
52	1.2	55	14	0.1	0.281294	0.238092	0.198837
53	1.2	55	14	0.2	0.279568	0.070177	0.193916
54	1.2	55	14	0.3	0.276575	0.063897	0.189
55	1.4	45	9	0.1	0.085862	0.182845	0.106902
56	1.4	45	9	0.2	0.084279	0.189076	0.076471
57	1.4	45	9	0.3	0.082632	0.156857	0.06783
58	1.4	45	11.5	0.1	0.034684	0.083379	0.045432
59	1.4	45	11.5	0.2	0.033449	0.088384	0.026382
60	1.4	45	11.5	0.3	0.032167	0.063523	0.021134

Table 4: Continued.

No.	t (mm)	L (mm)	w (mm)	r (mm)	Z_{Y1}	Z_{Y2}	Z_{Y3}
61	1.4	45	14	0.1	0.002052	0.019598	0.017475
62	1.4	45	14	0.2	0.001045	0.023819	0.003807
63	1.4	45	14	0.3	0	0.003565	0
64	1.4	50	9	0.1	0.176041	0.236076	0.150963
65	1.4	50	9	0.2	0.174141	0.203515	0.105286
66	1.4	50	9	0.3	0.171955	0.209182	0.10099
67	1.4	50	11.5	0.1	0.104642	0.125344	0.082888
68	1.4	50	11.5	0.2	0.103169	0.100551	0.051898
69	1.4	50	11.5	0.3	0.101459	0.104753	0.049279
70	1.4	50	14	0.1	0.05907	0.053648	0.03731
71	1.4	50	14	0.2	0.057851	0.033762	0.01654
72	1.4	50	14	0.3	0.056457	0.036978	0.014873
73	1.4	55	9	0.1	0.287628	0.291869	0.158143
74	1.4	55	9	0.2	0.28511	0.214402	0.146567
75	1.4	55	9	0.3	0.282466	0.248249	0.136425
76	1.4	55	11.5	0.1	0.191274	0.168938	0.087305
77	1.4	55	11.5	0.2	0.18931	0.108901	0.079655
78	1.4	55	11.5	0.3	0.187252	0.135287	0.07284
79	1.4	55	14	0.1	0.129693	0.089309	0.040327
80	1.4	55	14	0.2	0.128077	0.040752	0.035245
81	1.4	55	14	0.3	0.126399	0.061961	0.030667

Table 5: DFNN parameters with 3 levels.

Variable	Level 1	Level 2	Level 3
Training function	Trainlm	Traincgb	Trainscg
Number of hidden layers	2	3	4
Number of nodes	7	9	11
Divide data	60 : 20 : 20	70 : 15 : 15	80 : 10 : 10

Table 6: Experimental design using L_9.

No.	Training function	Number of hidden layers	Number of nodes	Divide data
1	trainlm	2	7	60 : 20 : 20
2	trainlm	3	9	70 : 15 : 15
3	trainlm	4	11	80 : 10 : 10
4	traincgb	2	9	80 : 10 : 10
5	traincgb	3	11	60 : 20 : 20
6	traincgb	4	7	70 : 15 : 15
7	trainscg	2	11	70 : 15 : 15
8	trainscg	3	7	80 : 10 : 10
9	trainscg	4	9	60 : 20 : 20

Table 7: MSE results for deformation.

No.	MSE									
	1	2	3	4	5	6	7	8	9	10
1	0.0003	0.0000	0.0002	0.0006	0.0009	0.0002	0.0001	0.0001	0.0004	0.0002
2	0.0031	0.0072	0.0001	0.0004	0.0000	0.0007	0.0019	0.0048	0.0086	0.0006
3	0.0166	0.0029	0.0006	0.0008	0.0039	0.0058	0.0040	0.0026	0.0057	0.0013
4	0.0108	0.0036	0.0030	0.0010	0.0038	0.0150	0.0005	0.0131	0.0036	0.0110
5	0.0043	0.0038	0.0047	0.0058	0.0014	0.0079	0.0059	0.0158	0.0039	0.0061
6	0.0027	0.0110	0.0164	0.0069	0.0012	0.0093	0.0033	0.0064	0.0312	0.0104
7	0.0088	0.0018	0.0015	0.0032	0.0075	0.0045	0.0037	0.0061	0.0025	0.0057
8	0.0130	0.0239	0.0017	0.0141	0.0276	0.0028	0.0113	0.0034	0.0031	0.0106
9	0.0103	0.0121	0.0146	0.0134	0.0222	0.0109	0.0091	0.0068	0.0205	0.0075

Table 8: MSE results for stress.

No.	MSE									
	1	2	3	4	5	6	7	8	9	10
1	0.0051	0.0023	0.0181	0.0022	0.0116	0.0063	0.0035	0.0023	0.0053	0.0039
2	0.0165	0.0046	0.0100	0.0075	0.0143	0.0084	0.0140	0.0056	0.0034	0.0032
3	0.0091	0.0103	0.0091	0.0009	0.0171	0.0078	0.0032	0.0104	0.0302	0.0036
4	0.0090	0.0076	0.0069	0.0150	0.0068	0.0087	0.0157	0.0043	0.0088	0.0086
5	0.0066	0.0129	0.0104	0.0082	0.0066	0.0155	0.0079	0.0075	0.0073	0.0090
6	0.0144	0.0106	0.0109	0.0127	0.0166	0.0088	0.0153	0.0132	0.0114	0.0179
7	0.0015	0.0115	0.0115	0.0024	0.0086	0.0118	0.0078	0.0049	0.0096	0.0146
8	0.0542	0.0154	0.0151	0.0037	0.0125	0.0138	0.0375	0.0162	0.0303	0.0137
9	0.0160	0.0124	0.0110	0.0356	0.0048	0.0106	0.0173	0.0270	0.0100	0.0283

Table 9: MSE results for strain energy.

No.	MSE									
	1	2	3	4	5	6	7	8	9	10
1	0.0061	0.0001	0.0004	0.0003	0.0004	0.0001	0.0003	0.0037	0.0014	0.0016
2	0.0222	0.0037	0.0086	0.0007	0.0032	0.0024	0.0035	0.0164	0.0082	0.0008
3	0.0013	0.0011	0.0203	0.0011	0.0175	0.0019	0.0296	0.0005	0.0010	0.0034
4	0.0152	0.0060	0.0007	0.0093	0.0016	0.0072	0.0178	0.0049	0.0054	0.0036
5	0.0045	0.0150	0.0039	0.0067	0.0071	0.0058	0.0058	0.0018	0.0078	0.0051
6	0.0048	0.0043	0.0049	0.0058	0.0075	0.0103	0.0037	0.0107	0.0093	0.0043
7	0.0119	0.0118	0.0027	0.0143	0.0046	0.0025	0.0086	0.0039	0.0051	0.0067
8	0.0062	0.0047	0.0015	0.0063	0.0015	0.0020	0.0423	0.0073	0.0061	0.0273
9	0.0043	0.0034	0.0355	0.0061	0.0158	0.0357	0.0091	0.0047	0.0200	0.0104

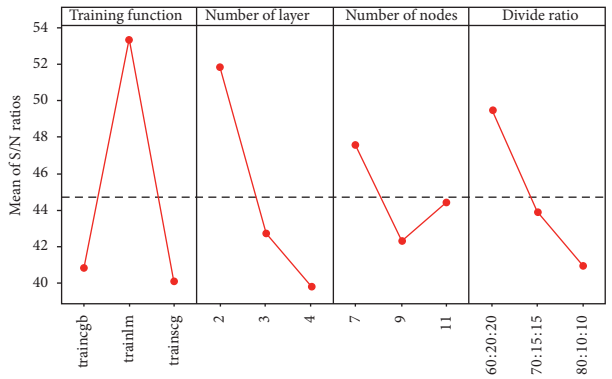


Figure 9: Results analysis S/N for deformation.

Table 10: Results of rank for deformation.

Level	Training function	Number of hidden layers	Number of nodes	Divide data
1	40.84	51.82	47.60	49.46
2	53.31	42.68	42.27	43.84
3	40.07	39.72	44.34	40.91
Delta	13.24	12.10	5.33	8.54
Rank	1	2	4	3

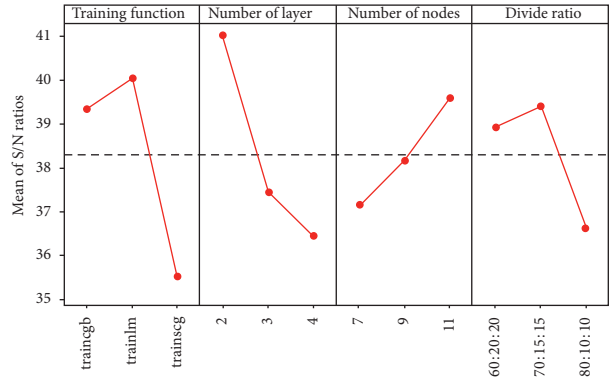


Figure 10: Results analysis S/N for stress.

Table 11: Results of rank for stress.

Level	Training function	Number of hidden layers	Number of nodes	Divide data
1	39.34	41.01	37.16	38.90
2	40.03	37.44	38.16	39.37
3	35.52	36.44	39.57	36.62
Delta	4.51	4.58	2.41	2.75
Rank	2	1	4	3

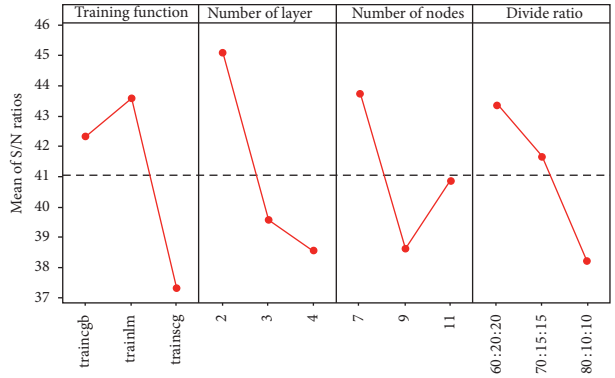


Figure 11: Results analysis S/N for strain energy.

Table 12: Results of rank for strain energy.

Level	Training function	Number of hidden layers	Number of nodes	Divide data
1	42.32	45.08	43.74	43.34
2	43.56	39.59	38.62	41.66
3	37.31	38.52	40.83	38.19
Delta	6.25	6.56	5.12	5.14
Rank	2	1	4	3

Table 13: The optimal structure of DFNN.

Function	Training function	Number of hidden layers	Number of nodes	Divide data
Deformation	trainlm	2	7	60 : 20 : 20
Stress	trainlm	2	11	70 : 15 : 15
Strain energy	trainlm	2	7	60 : 20 : 20

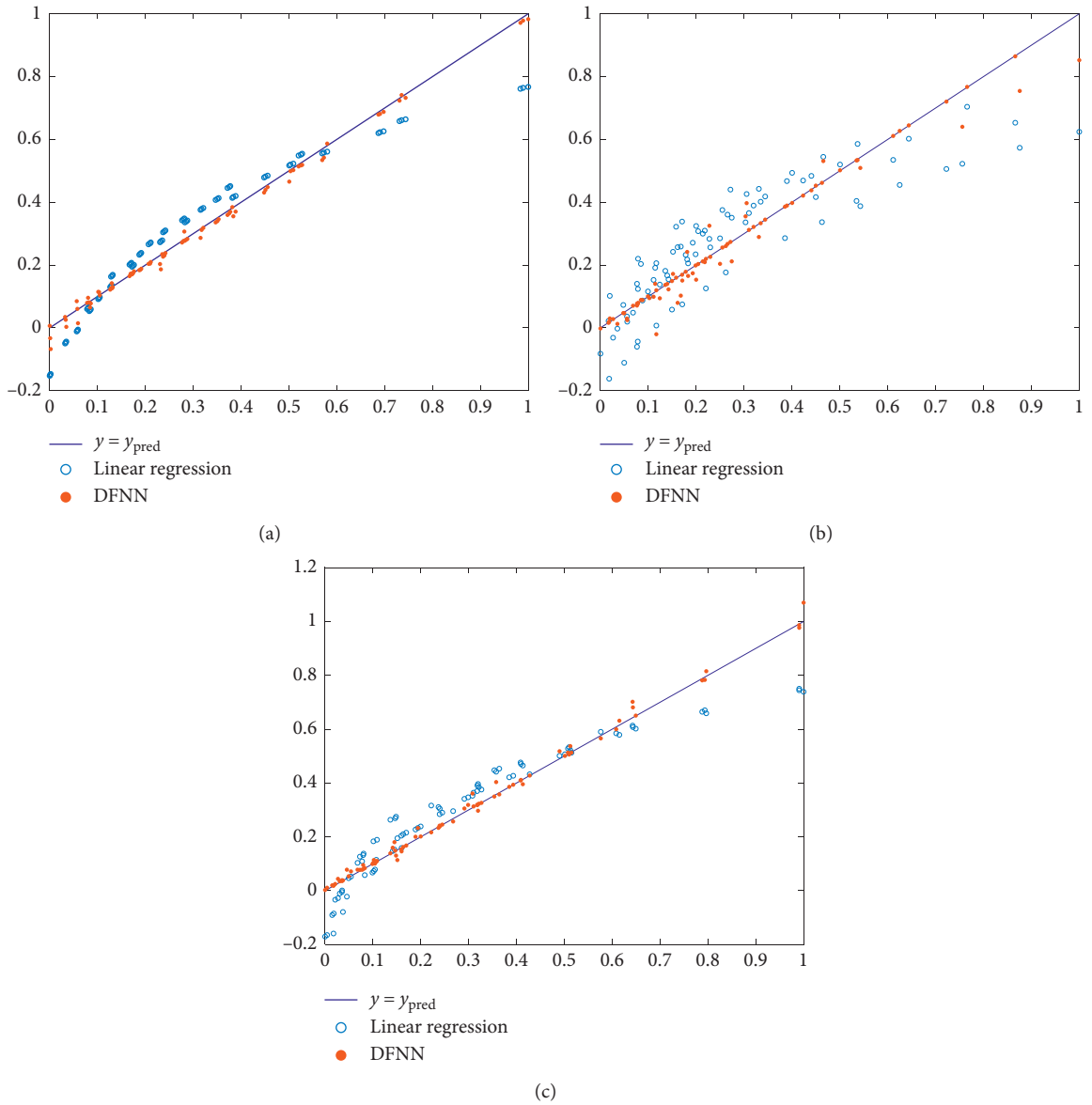


Figure 12: The good fitness of model (a) for deformation, (b) for stress, and (c) for strain energy.

Table 14: Compare the value R^2 of the proposed model and the linear model.

Response	Model data		Training data		Validating data		Testing data	
	DFNN	Linear	DFNN	Linear	DFNN	Linear	DFNN	Linear
Deformation	0.99	0.91	0.99	0.91	1.0	0.92	0.99	0.96
Stress	0.97	0.76	0.90	0.73	0.95	0.78	0.95	0.82
Strain energy	0.99	0.89	0.99	0.97	1.0	0.89	0.99	0.93

Table 15: Compare the value MSE of the proposed model and the linear model.

Response	Model data		Training data		Validating data		Testing data	
	DFNN	Linear	DFNN	Linear	DFNN	Linear	DFNN	Linear
Deformation	$2.82E-4$	0.0050	$8.02E-4$	0.0053	$4.14E-5$	0.0046	$5.00E-4$	0.0014
Stress	0.0015	0.0121	0.0046	0.0099	$5.98E-6$	0.0127	0.0031	0.0112
Strain energy	$3.28E-4$	0.0069	$9.5E-4$	0.0065	$1.3E-5$	0.0061	$6.70E-4$	0.0049

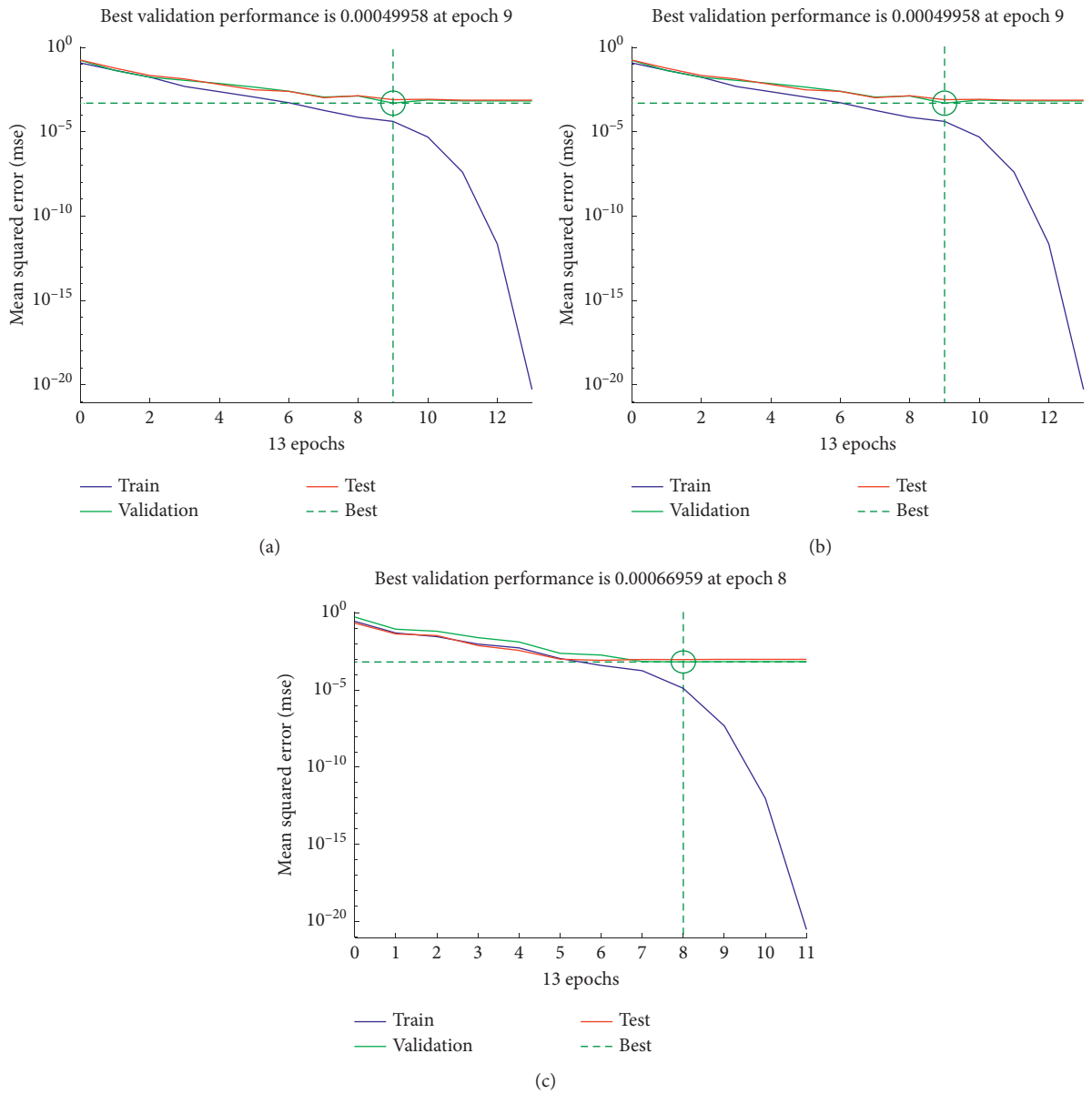


Figure 13: The best validation performance: (a) for deformation, (b) for stress, and (c) for strain energy.

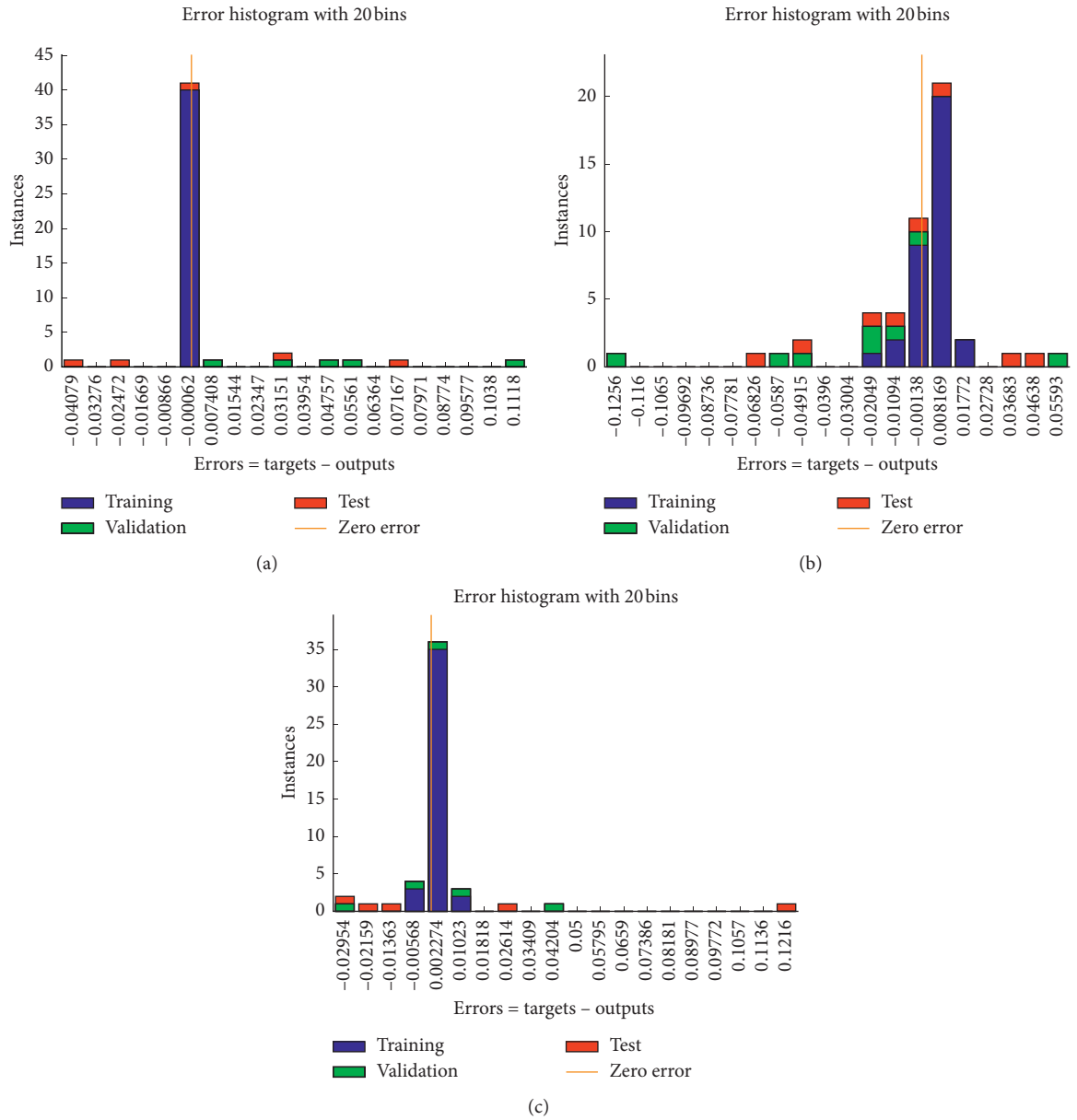
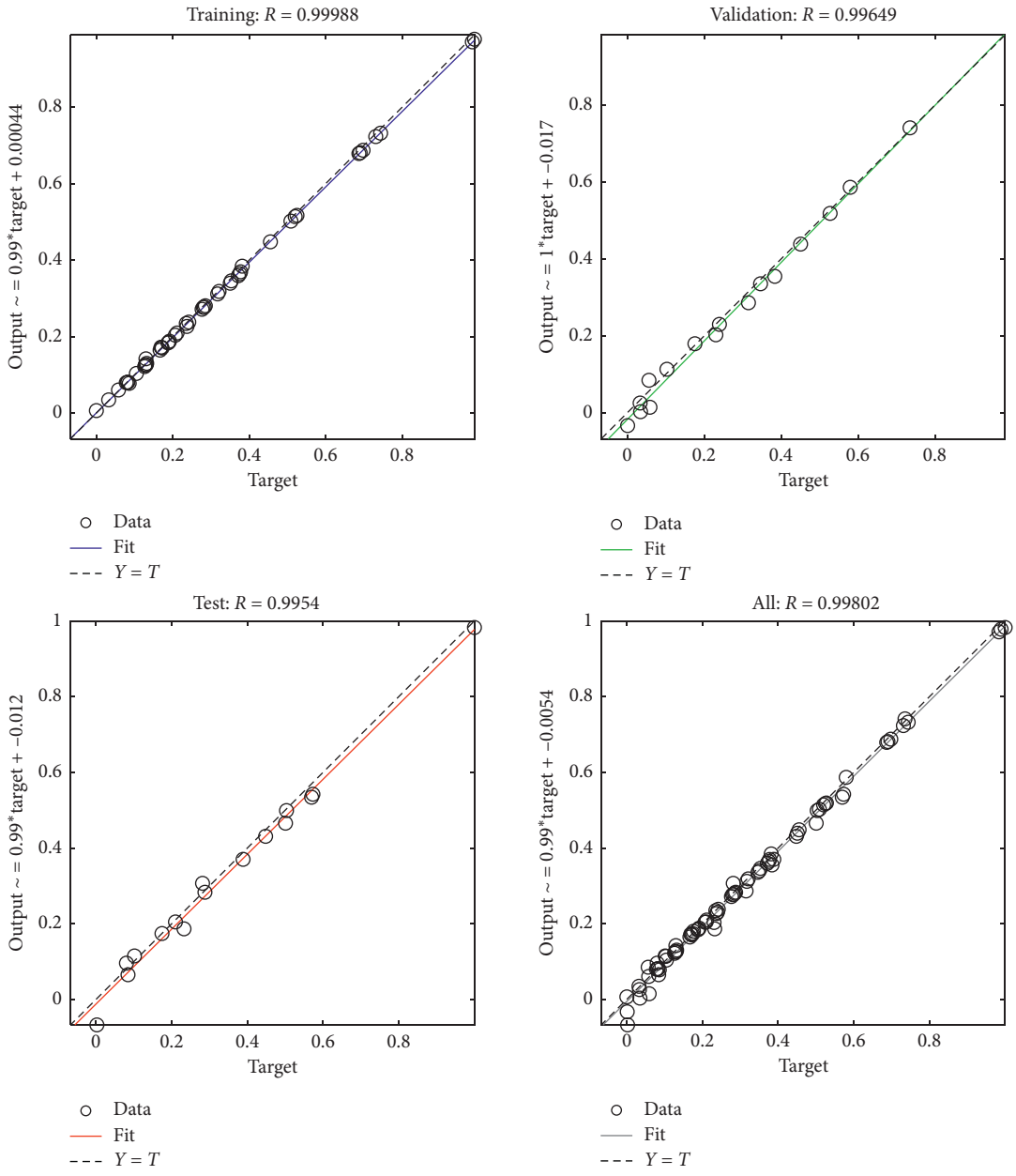
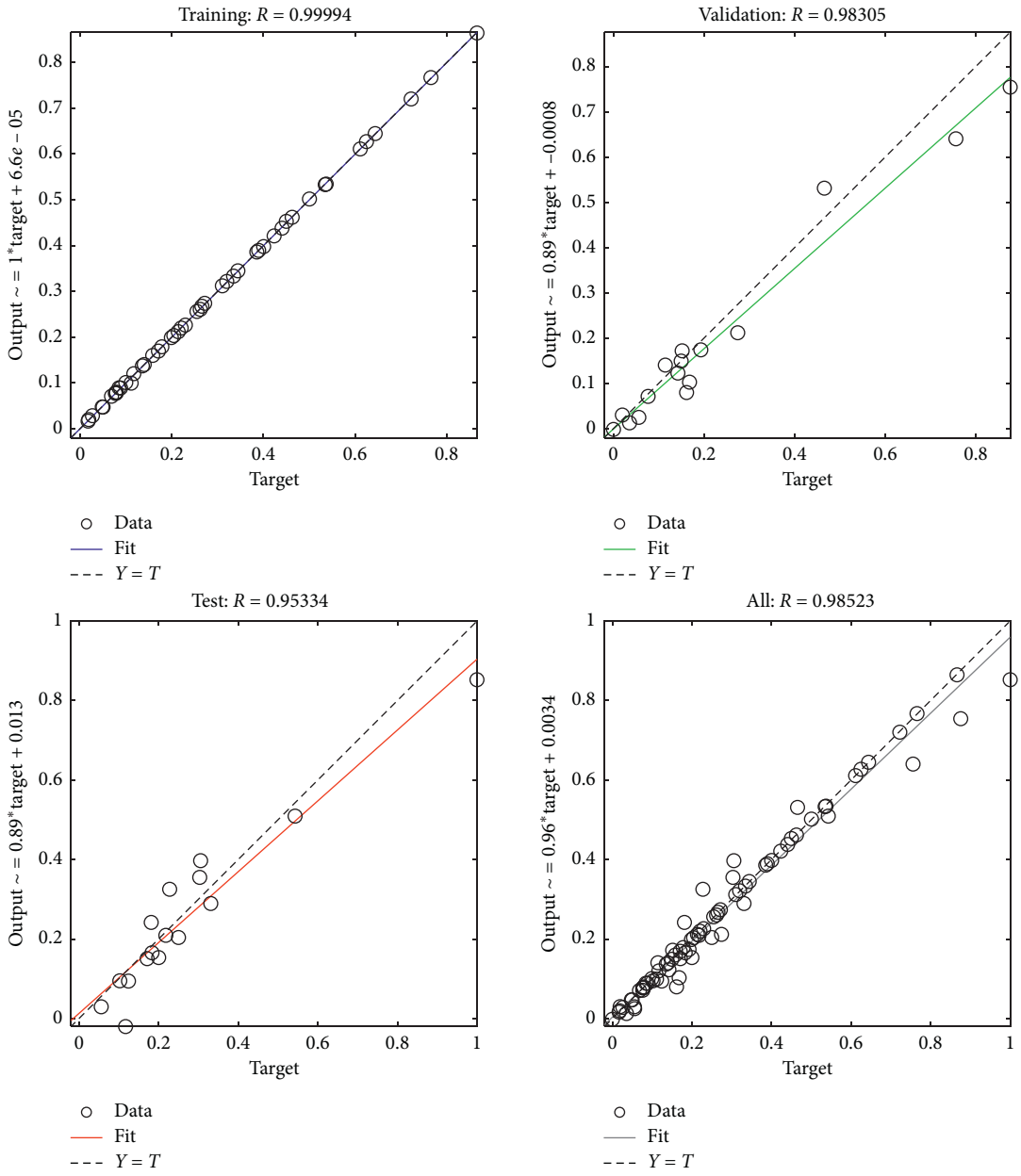


Figure 14: Error histogram (a) for deformation, (b) for stress, and (c) for strain energy.



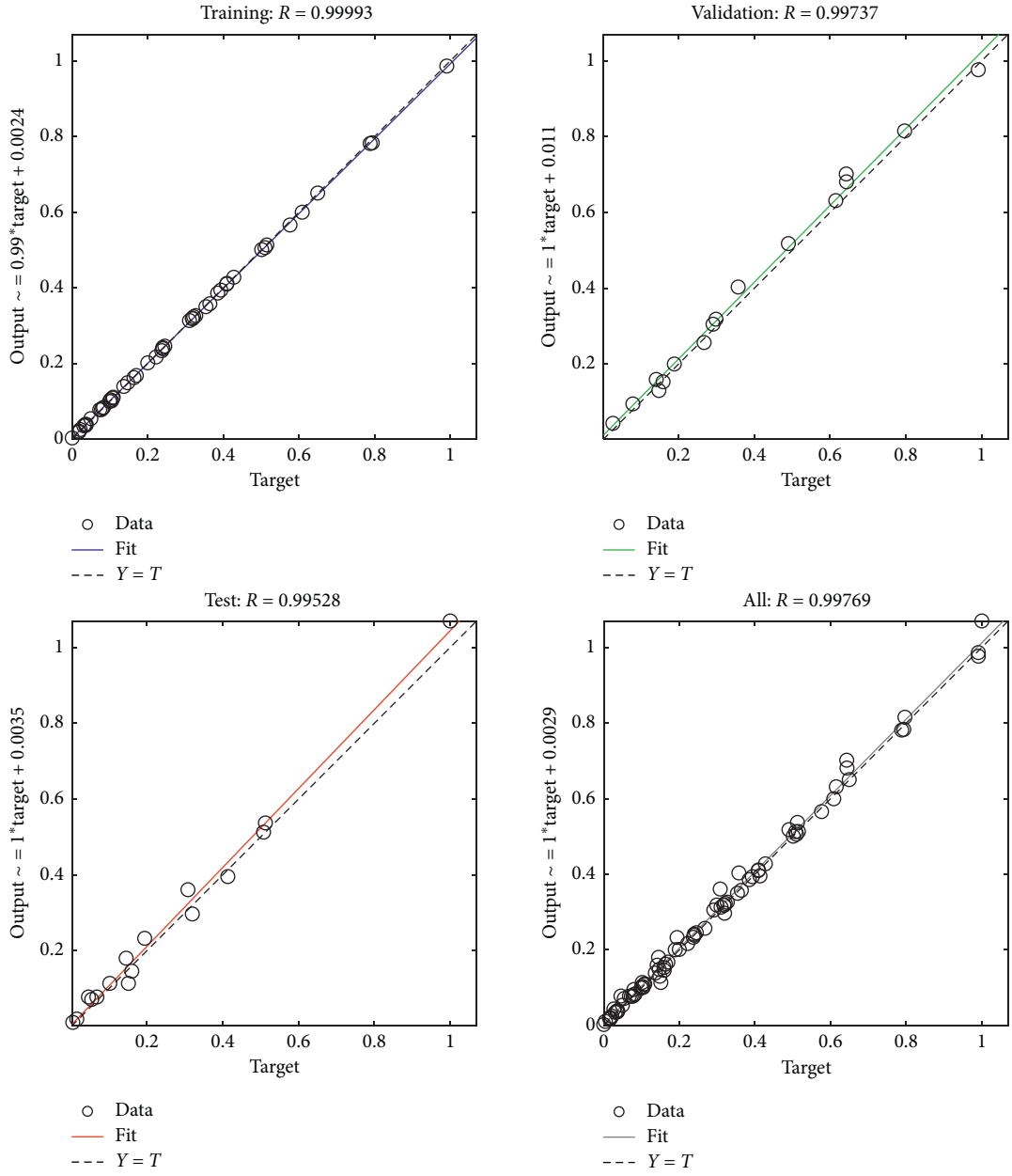
(a)

Figure 15: Continued.



(b)

Figure 15: Continued.



(c)

Figure 15: The fitness of model (a) for deformation, (b) for stress, and (c) for strain energy.

Table 16: Optimal results.

No.	WCA		ER-WCA	
	Strain energy (mJ)	Time (s)	Strain energy (mJ)	Time (s)
1	0.011509	2.62E + 03	0.011509	3.05E + 03
2	0.011509	2.59E + 03	0.011509	3.08E + 03
3	0.010654	2.63E + 03	0.010654	2.84E + 03
4	0.010654	2.63E + 03	0.010654	2.90E + 03
5	0.011127	2.62E + 03	0.011127	2.96E + 03
6	0.011314	2.63E + 03	0.011314	3.61E + 03
7	0.011364	2.63E + 03	0.011364	3.56E + 03
8	0.011509	2.62E + 03	0.011509	3.57E + 03
9	0.010717	2.63E + 03	0.010717	3.17E + 03
10	0.011509	2.63E + 03	0.011509	3.07E + 03
11	0.010675	2.65E + 03	0.010675	3.03E + 03
12	0.011509	2.33E + 03	0.011509	2.90E + 03
13	0.011509	2.67E + 03	0.011509	2.98E + 03
14	0.011509	2.67E + 03	0.011509	3.05E + 03
15	0.011509	2.90E + 03	0.011509	3.02E + 03
16	0.010716	2.68E + 03	0.010716	3.05E + 03
17	0.011509	2.68E + 03	0.011509	2.89E + 03
18	0.011509	2.72E + 03	0.011509	2.85E + 03
19	0.011509	2.85E + 03	0.011509	2.94E + 03
20	0.011509	2.80E + 03	0.011509	2.85E + 03
21	0.011257	2.95E + 03	0.011257	2.86E + 03
22	0.010708	2.69E + 03	0.010708	2.93E + 03
23	0.010656	2.94E + 03	0.010656	3.02E + 03
24	0.011509	2.94E + 03	0.011509	3.06E + 03
25	0.011509	2.96E + 03	0.011509	3.09E + 03
26	0.010675	2.63E + 03	0.010675	3.19E + 03
27	0.011509	2.65E + 03	0.011509	3.05E + 03
28	0.011509	2.68E + 03	0.011509	3.05E + 03
29	0.011509	2.95E + 03	0.011509	3.05E + 03
30	0.010716	2.96E + 03	0.010716	2.84E + 03
Average	0.01123	2717.755	0.01119	3057.928
Standard deviation	0.000373	149.2229	0.000568	202.8228

Table 17: Results of evaluation.

	Strain energy (mJ)	Deformation (mm)	Stress (MPa)	Life (10^6 cycle)
Prediction	0.01123	33.666	79.050	
FEM	0.01102	34.236	81.475	299
Error (%)	1.87	1.69	3.06	

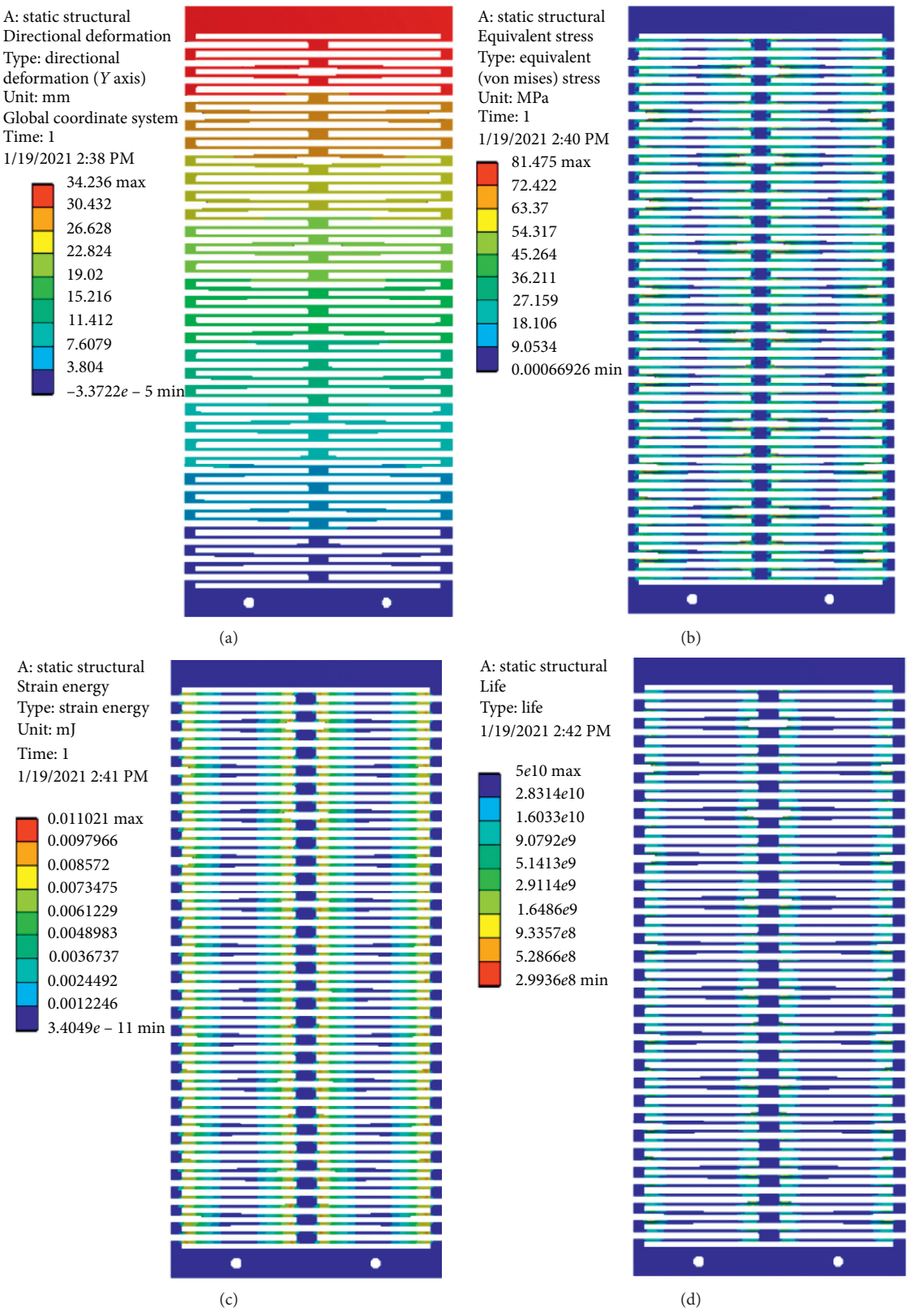


Figure 16: FEM results for (a) deformation, (b) stress, (c) strain energy, and (d) life.

small error values, it shows that the proposed optimal method has high reliability.

7. Conclusions

This paper proposes a method for the development and optimization of a planar spring used for a gravity balance mechanism. The proposed method is a combination of finite FEM, DFNN, and WCA. First, the FEM was used to collect data. Next, the approximated model was constructed using the DFNN, and finally, the geometry of the planar spring was optimized by the WCA. The effectiveness of the proposed method was tested by comparison with the optimal results of the ER-WCA. The comparison results show that the convergence speed and search stability of the proposed algorithm are better than the ER-WCA. The optimal parameters of PS are used to build 3D models. This model is used to be FEM. Comparing the FEM results with the optimal prediction results shows that the errors of energy strain, deformation, and stress are 1.87%, 1.69%, and 3.06%, respectively. This error shows that the proposed solution is highly robust. In addition, the life of the PS was also predicted with 299 million cycles. For future research, prototypes will be manufactured and measured to verify the numerical results. The optimization method will be utilized for other compliant mechanisms.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

The authors are thankful for the financial support from the HCMC University of Technology and Education, Vietnam, under Grant No. T2021-01NCS.

References

[1] S. K. Agrawal and A. Fattah, "Design of an orthotic device for full or partial gravity-balancing of a human upper arm during motion," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No. 03CH37453)*, pp. 2841–2846, Las Vegas, NV, USA, October 2003.

[2] V. Arakelian, "Gravity compensation in robotics," *Advanced Robotics*, vol. 30, no. 2, pp. 79–96, 2016.

[3] M. Uemura, Y. Mitabe, and S. Kawamura, "Simultaneous gravity and gripping force compensation mechanism for lightweight hand-arm robot with low-reduction reducer," *Robotica*, vol. 37, no. 6, pp. 1090–1103, 2019.

[4] S. K. Agrawal, G. Gardner, and S. Pledgie, "Design and fabrication of an active gravity balanced planar mechanism using auxiliary parallelograms," *J. Mech. Des.Journal of Mechanical Design*, vol. 123, no. 4, pp. 525–528, 2001.

[5] Y.-L. Chu and C.-H. Kuo, "A single-degree-of-freedom self-regulated gravity balancer for adjustable payload," *Journal of Mechanisms and Robotics*, vol. 9, 2017.

[6] W. D.W. Van Dorsser, R. Barents, B. M.B. Wisse, M. Schenk, and J. L.J. Herder, "Energy-free adjustment of gravity equilibrators by adjusting the spring stiffness," *Proceedings of the institution of Mmechanical Eengineers, Part C: Journal of Mechanical Engineering Science*, vol. 222, no. 9, pp. 1839–1846, 2008.

[7] B. M. Wisse, W. D. Van Dorsser, R. Barents, and J. L. Herder, "Energy-free adjustment of gravity equilibrators using the virtual spring concept," in *Proceedings of the 2007 IEEE 10th International Conference on Rehabilitation Robotics*, pp. 742–750, Noordwijk, The Netherlands, June 2007.

[8] T. Rahman, R. Ramanathan, R. Seliktar, and W. Harwin, "A ssimple ttechnique to ppassively ggravity-bbalance aarticulated mmechanisms," *Journal of Mechanical Design*, vol. 117, no. 4, p. 655, 1995.

[9] N. Le Chau, H.H. G. LeGiang Le, and T.-P. Dao, "A gGravity bBalance mMechanism uUsing cCompliant mMechanism," *Advances in Intelligent Systems and Computing*, in *Proceedings of the International Conference on Green Technology and Sustainable Development*, pp. 431–439, Nang City, Vietnam, November 2020.

[10] N. L. Chau, H. G. Le, T.-P. Dao, M. P. Dang, and V. A. Dang, "Efficient hybrid method of FEA-based RSM and PSO algorithm for multi-objective optimization design for a compliant rotary joint for upper limb assistive device," *Mathematical Problems in Engineering*, vol. 2019, Article ID 2587373, 14 pages, 2019.

[11] N. L.N. Le ChauChau, H. G. Le, T.-P. Dao, and V. A. Dang, "Design and optimization for a new compliant planar spring of upper limb assistive device using hybrid approach of RSM–FEM and MOGADesign and oOptimization for a nNew cCompliant pPlanar sSpring of uUpper lLimb aAssistive dDevice uUsing hHybrid aApproach of RSM-FEM and MOGA," *Arabian Journal for Science and Engineering*, vol. 44, no. 9, pp. 7441–7456, 2019.

[12] L. L. Howell, *Compliant Mmechanisms*, John Wiley & Sons, Hoboken, NJ, USA, 2001.

[13] K. J. Bathe, *Finite Eelement Mmethod*, Wiley Encyclopedia of Computer Science and Engineering, Hoboken, NJ, USA, 2007.

[14] O. C. Zienkiewicz, R. L. Taylor, P. Nithiarasu, and J. Zhu, *The Finite Element Method*, Vol. 3, McGraw-Hill, London, UK, 1977.

[15] N. Le Chau, T.-P. Dao, and V. A. Dang, "An efficient hybrid approach of improved adaptive neural fuzzy inference system and teaching learning-based optimization for design optimization of a jet pump-based thermoacoustic-Stirling heat engine," *Neural Computing and Applications*, vol. 32, no. 11, pp. 7259–7273, 2020.

[16] N. Le Chau, N. T. Tran, and T.-P. Dao, "An oOptimal dDesign mMethod for cCompliant mMechanisms," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5599624, 18 pages, 2021.

[17] F.F. S. Gómez, R.R. L. Lorza, M.M. C. Bobadilla, and R.R. E. García, "Improving the process of adjusting the parameters of finite element models of healthy human intervertebral discs by the multi-response surface method," *Materials*, vol. 10, no. 10, p. 1116, 2017.

[18] R. Lostado, R. F. Martinez, and B. J. Mac Donald, "Determination of the contact stresses in double-row tapered roller bearings using the finite element method, experimental

analysis and analytical models," *Journal of Mechanical Science and Technology*, vol. 29, no. 11, pp. 4645–4656, 2015.

[19] M. F. Alghifari, T. S. Gunawan, and M. Kartiwi, "Speech emotion recognition using deep feedforward neural network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 2, pp. 554–561, 2018.

[20] J. Seo, J. Lee, and K. Kim, "Decoding of polar code by using deep feed-forward neural networks," in *Proceedings of the 2018 International Conference on Computing, Networking and Communications (ICNC)*, pp. 238–242, Maui, HI, USA, March 2018.

[21] H. Eskandar, A. Sadollah, A. Bahreininejad, and M. Hamdi, "Water cycle algorithm–A novel metaheuristic optimization method for solving constrained engineering optimization problemsWater cycle algorithm—aA novel metaheuristic optimization method for solving constrained engineering optimization problems," *Computers & StructuresComputers & Structures*, vol. 110110-111, pp. 151–166, 2012.

[22] O. B. Haddad, M. Moravej, and H. A. Loáiciga, "Application of the water cycle algorithm to the optimal operation of reservoir systems," *Journal of Irrigation and Drainage Engineering*, vol. 141, no. 5, Article ID 04014064, 2015.

[23] W.W.-C. Chen, A.A. H. Lee, W.W.-J. Deng, and K.K.-Y. Liu, "The implementation of neural network for semiconductor PECVD process," *Expert Systems with Applications*, vol. 32, no. 4, pp. 1148–1153, 2007.

[24] S. Athreya and Y. Venkatesh, "Application of Taguchi method for optimization of process parameters in improving the surface roughness of lathe facing operation," *International Refereed Journal of Engineering and Science*, vol. 1, pp. 13–19, 2012.

[25] G. Sun, J. Fang, X. Tian, G. Li, and Q. Li, "Discrete robust optimization algorithm based on Taguchi method for structural crashworthiness design," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4482–4492, 2015.

[26] W. Lin and Z. Ma, "Using Taguchi-Fibonacci search method to optimize phase change materials enhanced buildings with integrated solar photovoltaic thermal collectors," *Energy*, vol. 106, pp. 23–37, 2016.

[27] A. Sadollah, H. Eskandar, and J. H. Kim, "Water cycle algorithm for solving constrained multi-objective optimization problems," *Applied Soft Computing*, vol. 27, pp. 279–298, 2015.

[28] A. A. Heidari, R.R. A. AbbaspourAli Abbaspour, and A.A. R. JordehiRezaee Jordehi, "An efficient chaotic water cycle algorithm for optimization tasks," *Neural Computing and Applications*, vol. 28, no. 1, pp. 57–85, 2017.

[29] A. Sadollah, H. Eskandar, A. Bahreininejad, and J. H. Kim, "Water cycle algorithm with evaporation rate for solving constrained and unconstrained optimization problems," *Applied Soft Computing*, vol. 30, pp. 58–71, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9974219, 11 pages
https://doi.org/10.1155/2021/9974219

Research Article

Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models

Vahid Najafi Moghaddam Gilani,[1] **Seyed Mohsen Hosseinian**,[1] **Meisam Ghasedi**,[2] **and Mohammad Nikookar**[2]

[1]*School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran*
[2]*Department of Civil Engineering, University of Guilan, Rasht, Iran*

Correspondence should be addressed to Meisam Ghasedi; meisam_ghasedi@msc.guilan.ac.ir

Received 15 March 2021; Revised 21 April 2021; Accepted 17 May 2021; Published 27 May 2021

Academic Editor: Adrian Neagu

Modeling the severity of accidents based on the most effective variables accounts for developing a high-precision model presenting the possibility of occurrence of each category of future accidents, and it could be utilized to prioritize the corrective measures for authorities. The purpose of this study is to identify the variables affecting the severity of the injury, fatal, and property damage only (PDO) accidents in Rasht city by collecting information on urban accidents from March 2019 to March 2020. In this regard, the multiple logistic regression and the pattern recognition type of artificial neural network (ANN) as a machine learning solution are used to recognize the most influential variables on the severity of accidents and the superior approach for accident prediction. Results show that the multiple logistic regression in the forward stepwise method has R^2 of 0.854 and an accuracy prediction power of 89.17%. It turns out that the accidents occurred between 18 and 24 and KIA Pride vehicle has the highest effect on increasing the severity of accidents, respectively. The most important result of the logit model accentuates the role of environmental variables, including poor lighting conditions alongside unfavorable weather and the dominant role of unsafe and poor quality of vehicles on increasing the severity of accidents. In addition, the machine learning model performs significantly better and has higher prediction accuracy (98.9%) than the logit model. In addition, the ANN model's greater power to predict and estimate future accidents is confirmed through performance and sensitivity analysis.

1. Introduction

Transportation, like any other industry and phenomenon, along with its advantages, has its disadvantages and limitations for road users [1, 2]. Traffic and its related predicaments have been on the rise all over the world and have had a detrimental effect on the lives and property of the people of the community. Urban pollution, rising fuel and energy consumption, wasting millions of hours a day in traffic congestion, wasting community service facilities and national assets, and, ultimately, the occurrence of accidents resulting in injury, death, and property damage are the result of poor traffic facilities and conditions [3–5].

Road traffic accidents now represent the eighth leading cause of death globally. In addition, road traffic fatalities have increased to 1.35 million a year and caused up to 50 million injuries in 2016; that is, nearly 3700 people die on the world's roads every day. In Iran, the reported number of road traffic deaths is 15932 in 2016 [6]. The statistics also reveals the high number of injury and property damage only (PDO) accidents in urban and suburban roads in the country. Therefore, the effect of factors affecting the severity of accidents should be investigated meticulously to provide practical solutions for improving safety and reducing the high number of accidents.

Various studies have been carried out on the subject area of traffic safety in recent years using multiple logistic regression. Sherafati et al. explored road traffic fatalities after receiving emergency services using multiple logistic regression in Langarud. Results showed that males,

motorbikes, and pedestrians had a positive and significant relationship with fatal accidents [7]. Intini et al. conducted a study to investigate the relationships between road familiarity/unfamiliarity and the occurrence of accidents using multiple logistic regression. Factor analysis is a very vital step in many applications. The factors of minor intersections/driveways, autumn/winter, and speed limits less than 80 km/h were more related to familiar driver crashes, but the factors of head-on and rear-end accidents, summer, and heavy vehicles were more related to unfamiliar drivers involved in accidents [8]. Also, Casado-Sanz et al. applied a logistic regression model to investigate the impact of age on accident severity in the rural crosstown road and concluded that female drivers and motorbikes had a negative impact on the likelihood of accidents [9].

Generally, metaheuristic algorithms and machine learning techniques have been widely used in different engineering studies, especially in transportation problems, which they are in desperate need of complex and accurate solutions to provide more accurate prediction models than statistical methods due to their capability of handlig more complex functions and classification problems [10–25]. Pattern recognition tools and their accurate analysis using optimized prediction tasks are a trendy topic in the two recent years [26–30]. Supplementary to this, various prediction methods have been used in different engineering problems by the emergence of various datasets [31–35]. Some studies have applied traditional methods such as regression models alongside machine learning approaches to validate the ANN-based solutions as an effective alternative method to predict the severity of accidents with higher accuracy [36–44]. Because of the close connection of accidents to human life, machine learning approaches would be applied in important functions, like predicting the type and severity of accidents, due to their higher precision [45, 46]. In order to understand the effects of the determining factors in an activity and predict incidents in the future, nonlinear links between variables with different forms of ANN may be modeled [47–49].

Chimba and Sando investigated the severity of traffic accidents using ANN and compared this method's accuracy with the ordered probit model. Results indicated that the prediction accuracy of ANN was higher [50]. Khair et al. tried to present an accident prediction model using a new framework of an artificial neural network (ANN). They affirmed that the predicted collisions were similar to the number of real incidents and thus found the proposed model [51]. Omrani presented multiple logistic regression and ANN models to predict individuals' travel mode in Luxembourg. Results showed that ANN models had better performance [52]. In order to provide a decent prediction for traffic collisions in metropolitan areas of Nuevo León, Contreras et al. utilized an innovative ANN model. The programming feature of Scilab software has been used in this analysis to verify the highest sensitivity on the expected neural network [53]. Amin used the backpropagation ANN approach to investigate gender characteristics of older driver accidents and model the variables of their accidents and finally illustrated that journey purpose was the highest contributor factor of accident risk for older drivers and lighting condition was the second most important factor [54]. Ghasedi et al. used the logit model, factor analysis, and machine learning approaches to recognize the most effective variables on suburban accidents and generate the most accurate model for predicting future accidents in the busiest suburban highway of Guilan Province located in the north of Iran. Their studies showed the outstanding role of environmental factors, such as rainy weather and inadequate lighting condition, on increasing the severity of pedestrian accidents [55].

2. Study Area and Methodology

Rasht is one of the most congested cities in Iran and also the most populous city in the north of Iran, with high traffic volume in most of the days during the year. According to the official census in 2016, its density was announced to be 414 people per square kilometer [56]. Besides, more than 60% of the accidents between 2019 and 2020 in the city of Rasht occurred in the inner ring road of this city. The effective factors on increasing the probability of accidents vary from city to city, and due to the congested traffic flows even in ring roads and high traffic interferences with other flows in several parts of this city, a separate study is needed to be conducted in Rasht metropolis. Due to the above-mentioned information and very dense urban texture and numerous traffic interferences between nonmotorized users and passing vehicles, the need to identify the most influential variables on the severity of accidents and provide the most accurate prediction model for future accidents are of great importance. As it is mentioned earlier, different approaches have been used worldwide for analyzing the number and severity of accidents considering the specific condition of crashes, including crashes with unknown severity, high-severity crashes, or the severity of crashes that may be expected to occur sometime in the future. However, in this study, due to a balanced number of datasets in terms of three categories of fatal, injury, and PDO accidents, we decided to use logit and pattern recognition type of ANN as a machine learning approach to analyze and provide a prediction model. In order to conduct research on road safety improvements, accident information that includes factors such as the time of the accident, human characteristics, environmental characteristics, and accident types should be collected meticulously. The results of consecutive visits to the Rasht Traffic Police Statistic Center led to the collection of information on the 12-month accidents from March 2019 to March 2020. In total, the statistical population includes 965 accidents that occurred in Rasht, 738 of which are related to urban accidents and the rest are related to suburban accidents. The dependent variable in this study is the severity of accidents, which are classified into three categories: fatal, injury, and PDO accidents. Since the amount of fatal accidents is small in comparison to total accidents, and the independent variables significance and goodness of fit of a model cannot be satisfied by considering the three types of dependent variables, fatal accidents have been merged with injury accidents, and the

dependent variable has been split into two categories: PDO and injury/fatal accidents [55].

Table 1 classifies the independent variables influencing the occurrence of accidents in Rasht city and the appropriate coding for each of them for modeling purposes. Moreover, the remainder of this article is structured as indicated in Figure 1.

3. Statistical and Machine Learning Forecasting Approaches

3.1. Multiple Logistic Regression. To connect a set of variables X and a dependent variable such as Y, one has a multivariate problem. In analyzing such a problem, different types of mathematical models have been used to consider the relationship's complexity. The logistics regression method is a mathematical method used to describe the relationship between several variables x and a binary dependent variable. The function used in this method is an S-shaped function called a logistics function. However, the application of logistics analysis is not just limited to the above-mentioned issue. Expanding the logistics function can also be used to solve multifaceted problems. Therefore, the logistic regression method can be used to define the variable Y in a multifaceted way. In the simplest case, $P(Y = i)$ can be considered a linear function of x_i ($P_i = x_i\beta$), where β is the vector of regression coefficients. However, one of the considerations is that the probability of P_i mentioned in the left part of the equation must be between zero and one, but the linear multiplication of $x_i\beta$ in the right part includes all real numbers. A simple way to solve this problem is to use the probability transfer function to remove the distance constraints and model the transfer function as a linear function of the parameters. This conversion takes place in two stages [57]. First, the probability of P_i becomes a chance of success from

$$\text{odds} = \frac{P_i}{1 - P_i}. \tag{1}$$

In the second step, the logarithm of equation (1) is performed so that logit or logarithm of the chance of success is obtained from

$$\text{Logit}(P_i) = \text{Log}\frac{P_i}{1 - P_i}. \tag{2}$$

The reverse transfer function, called antilogistics, is used to calculate probability in terms of logistics from

$$\text{Logit}^{-1}(z_i) = \frac{e^{z_i}}{1 + e^{z_i}}. \tag{3}$$

Logistics is a transfer function that extends the probabilities of the range (0, 1) to all real numbers. Negative logistics indicate probabilities of less than 50% and positive logistics indicate probabilities of more than 50%. Therefore, the logistics model is a general linear model that has a logistics transfer function. In other words, the probability logistics of P_i instead of the probability follow the linear model [58].

3.2. Neural Network. Neural networks have significant abilities to detect complicated data relations and could be used to extract patterns and classify methods that humans and other computer methods find extremely difficult to comprehend [45]. ANNs have been interpreted as a nonlinear system based on human's brain activity [59, 60]. The researchers could take advantage of these networks' precious ability in recognizing the unknown relationship of natural and complex systems. Due to the close relationship between accidents and the well-being of society and their direct and indirect influences on human lives, ANNs have been considered a powerful and highly precise approach to deal with accidents dilemma [61]. Therefore, in this study, to find the most accurate prediction model, urban accidents of Rasht city have been modeled by pattern recognition type of ANN, which can result in various hands-on outputs to reduce the severity and number of accident.

4. Result Analysis and Discussion

4.1. Modeling Using Multiple Logistic Regression. In this section, the multiple logistic regression is used to determine the effect of each independent variable on the dependent variables (severity of accidents). To model the severity of accidents in Rasht, initially, 63 independent variables and 3 dependent variables are defined. Due to the small number of fatal accidents (7 cases), the dependent variables are summarized into two categories, including fatal/injury and PDO accidents. To build the logit model, the entering, the forward stepwise, and the backward stepwise methods can be used. Now, it should be examined which of the three methods has the most suitable output, or in other words, which of the methods can provide a better model for accidents in Rasht. To determine this, the criteria of higher prediction accuracy and the goodness of fit index of the model are considered to identify the best model. The coefficient of determination parameter (R^2) indicates the goodness of fit of the model. To be more specific, the closer R^2 to 1, the greater the model's goodness of fit, and the higher the correct percentage for the model, the higher the strength of the model to predict accidents. Table 2 presents a summary of the three methods, including two criteria of the prediction accuracy and the goodness of fit (R^2) of the model. The forward stepwise method with the correct percentage of 89.17% and the R^2 value of 0.854 is selected as the best method to build the logit model of the severity of accidents in Rasht.

Table 3 indicates the chi-square, degree of freedom (df), and significance (sig) of the forward method in the modeling process. The chi-square statistic index is used to determine the effect of independent variables on the dependent variable and, in general, the fit of the model and it is comparable to the F-statistic in normal regression analysis. The chi-square of the model in Step 23 is equal to 61.038 with a significant value of less than 5%. Since the significance of the forward model used to predict accidents is less than 5 percent, the capability of the model to predict accidents is confirmed. Thus, independent variables affect the dependent variable and indicate a good fit.

Table 1: Description of variables used in the study.

Variable	Variable levels
Accident severity	1. PDO 2. Injury/fatal
Accident time	1. 00 : 00 to 06 : 00 2. 06 : 00 to 12 : 00 3. 12 : 00 to 18 : 00 4. 18 : 00 to 24 : 00
Accident day	1. Saturday to Tuesday (major workdays in Iran) 2. Wednesday to Friday and holidays
Road category	1. Arterial road 2. Access road
Road surface condition	1. Dry 2. Humid 3. Wet
Geometry of accident location	1. Alignment 2. Roundabout 3. U-turn/J-turn 4. Uphill 5. Downhill 6. Alley 7. Intersection
Daylight condition	1. Day 2. Night 3. Sunset and sunrise
Accident type	1. Light vehicle-light vehicle 2. Light vehicle-motorcycle 3. Light vehicle-heavy vehicle/pickup truck-truck/bus
Type of collision	1. Head-on collision 2. Rear-end collision 3. Side-impact collision
Driver gender	1. Male 2. Female
Driver age	1. Less than 18 2. 18 to 30 3. 30 to 45 4. 45 to 60 5. 60 and over
Weather condition	1. Clear 2. Cloudy 3. Rainy
Reason of accident	1. Lack of attention 2. Backover movement 3. Inability to control the vehicle 4. Exceeding lawful speed 5. Crossing a forbidden place 6. Failure to observe longitudinal spacing 7. Failure to observe lateral spacing 8. Unsafe lane changes 9. Hasty-caused accident 10. Unsafe lane changes 11. Failure to yield the right-of-way 12. Improper turns 13. Wrong way movements 14. Yaw motion of the vehicle to the left 15. Technical defect in the vehicle

Table 1: Continued.

Variable	Variable levels
Vehicle	1. KIA Pride 2. Paykan 3. Renault 4. Peugeot 5. Taxi 6. Pickup truck 7. Minibus 8. Bus 9. Truck 10. Trailer

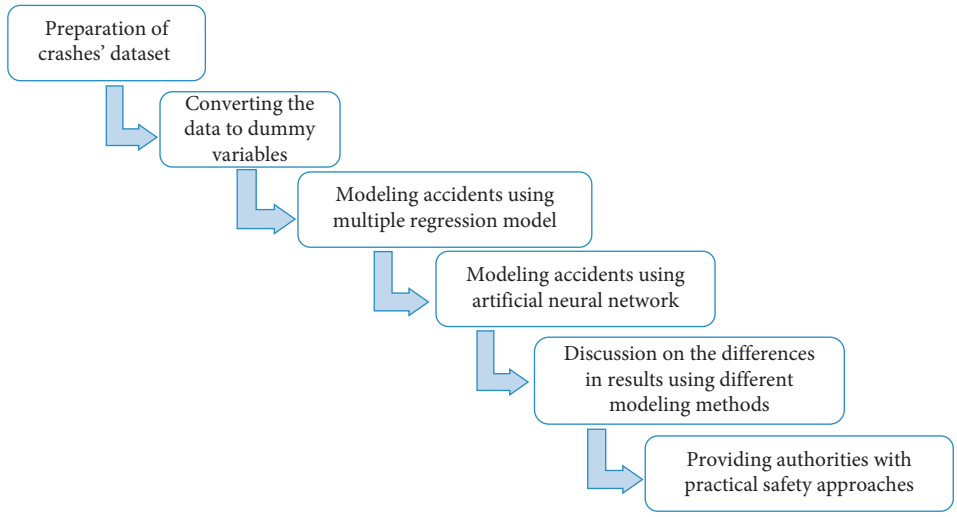


Figure 1: The overview of the proposed modeling process.

Table 2: Summary of the entering, the forward stepwise, and the backward stepwise methods.

Logistic regression type	Prediction accuracy	Goodness of fit (R^2)
Entering	82.52	0.686
Forward stepwise	89.17	0.854
Backward stepwise	86.93	0.738

Table 3: The result of the forward stepwise method.

		Chi-square	*df*	Sig.
Step 23	Step	−1.992	1	0.158
	Block	61.038	35	0.000
	Model	61.038	35	0.000

By selecting the forward stepwise method and entering all 63 selected variables in the modeling process, this method's final model is obtained in 23 steps. In this model, 11 variables are identified as the most effective variables in the severity of accidents leading to injury/fatal and PDO accidents in Rasht. Table 4 shows the variables entered in the model and their statistical indexes. According to Wald and sig, the Wald test examines the significance of the regression equation's variables and is comparable to the *t*-statistic in normal regression.

According to Table 4, the most effective variables increasing the severity of vehicle accidents are accident time, 18 : 00–24 : 00 and 12 : 00–18 : 00, KIA Pride vehicle, and rainy weather, respectively. Conversely, the most affecting variables with negative coefficients decreasing the severity of accidents include day time, accident time (6 : 00–12 : 00), and sunny weather. The most important result of the logit model underlines the role of environmental factors, including poor lighting conditions alongside unfavorable weather and the dominant role of unsafe and poor quality of vehicles on increasing the severity of accidents.

4.2. Modeling Using Artificial Neural Network. An ANN prediction model can be created using a variety of neural networks. Given the qualitative data used in this analysis, the prediction model is created using a neural network with pattern recognition capabilities. Using either supervised or unsupervised grouping, pattern recognition divides input data into objects or classes based on main characteristics [55].

The machine learning method uses the same input attributes and output labels as the variables described in

Table 4: Variables of the severity of accident logit model.

Variables	β	Standard deviation	Wald statistic	Significance	Odds ratio
6:00 to 12:00	−3.345	1.326	6.364	0.012	0.035
12:00 to 18:00	2.939	1.286	5.201	0.023	18.897
18:00 to 24:00	7.373	2.488	8.783	0.003	1592.404
Autumn	−1.918	0.725	7.010	0.008	0.147
Sunny weather	−2.446	0.806	9.205	0.002	0.087
Rainy weather	2.751	1.157	5.656	0.017	15.658
Lack of attention	1.119	0.644	3.020	0.082	3.062
Female driver	−2.063	0.783	6.946	0.008	0.127
Head-on collision	2.204	0.771	8.176	0.004	9.061
Day time	−4.343	1.422	9.323	0.002	0.013
KIA Pride vehicle	3.795	1.287	8.696	0.003	44.478
Constant	6.734	2.914	5.338	0.021	840.503

Table 5: Details of data entry.

Number of inputs	Number of hidden layers	Number of outputs	Number of training-validation-testing samples			
			—	Samples	MSE	%E
63	11	2	Training	516	$1.07429e^{-1}$	$12.81204e^{-0}$
			Validation	111	$1.36493e^{-1}$	$24.35571e^{-0}$
			Testing	111	$1.74612e^{-1}$	$24.72239e^{-0}$

Table 1. It is worth noting that the dependent variable (output class) is the different levels of accident severity, as mentioned in the "study route and methodology" section. For vehicle accidents, it has been split into two categories: fatal/injury and PDO. The ANN used in this analysis is the utilization of an existing algorithm in the software. The neural network's input data is divided into three categories:

(1) Training: these are presented to the network during training for the learning process, and the network is adjusted according to its error.

(2) Validation: these are used to measure network generalization and to halt training when generalization stops improving

(3) Testing: these have no effect on training and so provide an independent measure of network performance during and after training. In other words, it is the main criterion to realize how much the neural network's findings are similar to the actual result.

Table 5 shows the details of the accident data entry into the software and the Mean Squared Error and Percent Error. Since the number of accidents is adequate for the network training process, 70% of the data are used for network training and 15% for validation, and the remaining 15% are considered as a test of the developed network.

4.2.1. The Results of the Confusion Matrix. As can be seen in Figure 2, according to the all confusion matrix, which represents the result of the three processes of training, validation, and testing of the network, out of 632 property damage accidents, all 632 cases, and out of 106 injury/fatal accidents, 98 cases are predicted correctly by the model. The prediction accuracy of property damage accidents in the model is 100% and the prediction accuracy of injury/fatal accidents is 92.5%. Therefore, the accuracy of the model in the classification and separation of injury/fatal accidents and property damage accidents from each other is at a high level. Also, the accuracy of the whole model in determining the severity of accidents is 98.9%. With confidence of 98.9%, the model can predict the severity of accidents in terms of effective parameters.

4.2.2. The Results of the Performance of Neural Network. Figure 3 shows that the training process stopped after 92 repetitions. The point marked on the diagram shows that the answers will no longer improve from this point on, and this point with the mean squared error (MSE) 0.0231 indicates the best point for the end of the calculations and creation of the ANN for the information given.

4.2.3. Sensitivity and Specificity Analysis of the Neural Network for the Given Accident Data. Figure 3 shows the sensitivity analysis of the true positive rate of the generated ANN model compared to the false positive rate using a receiver operating characteristic (ROC) graph for accidents in Rasht. The ROC graph is a technique for visualizing, organizing, and selecting classifiers based on their performance [62]. Its popularity comes from several well-studied characteristics, such as intuitive visual interpretation of the curve and easy comparisons of multiple models [63]. When it is needed to check or visualize the multiclass classification problem's performance, the ROC curve is used. In this model, 70% of the data are considered for training, 15% for testing, and 15% for validation. As shown in Figure 4, Class 1 indicates the accuracy of the network prediction for existing

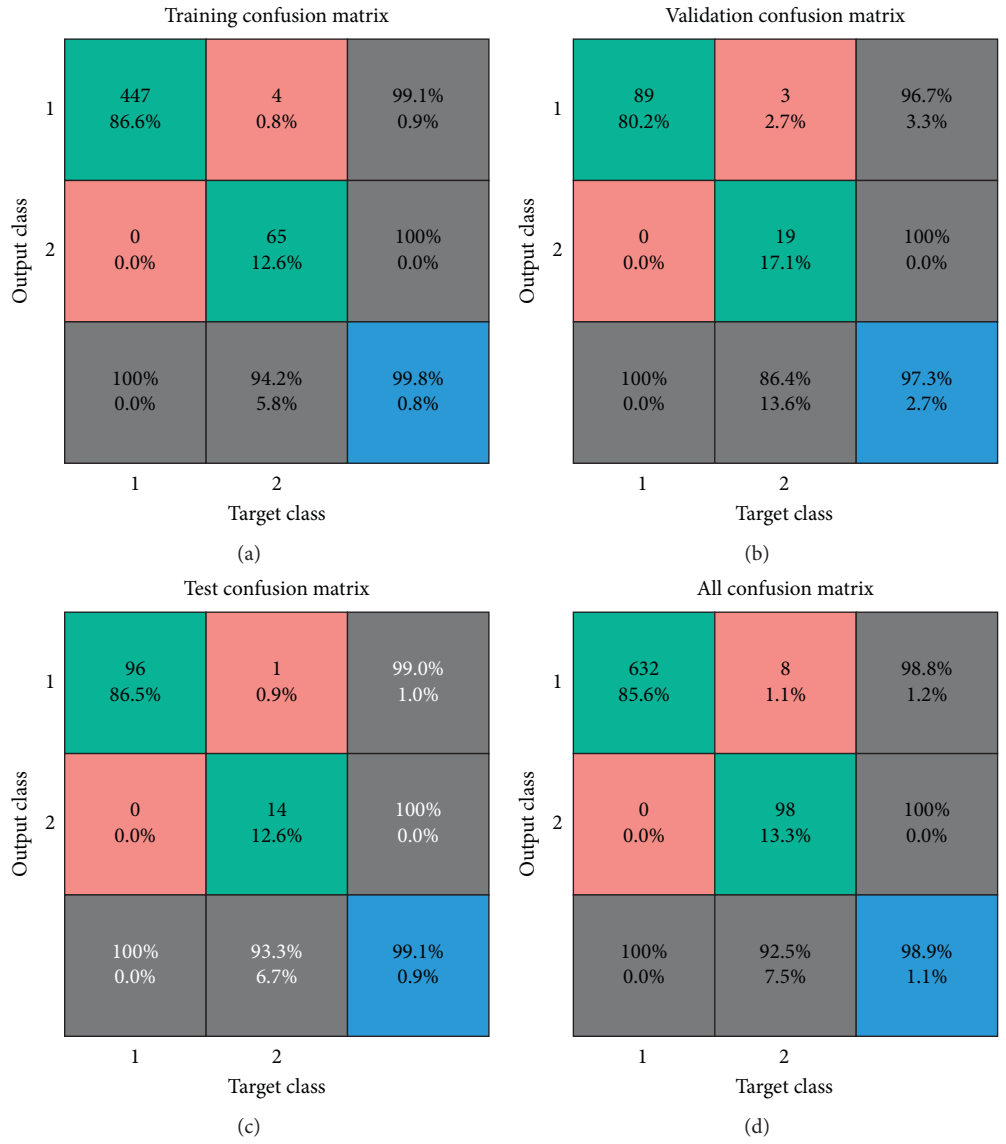


Figure 2: The confusing matrix of accidents.

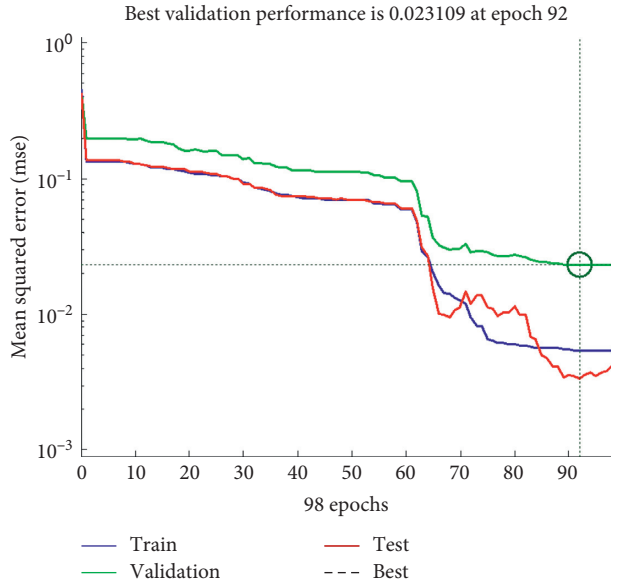


Figure 3: Performance of the neural network training process of accidents.

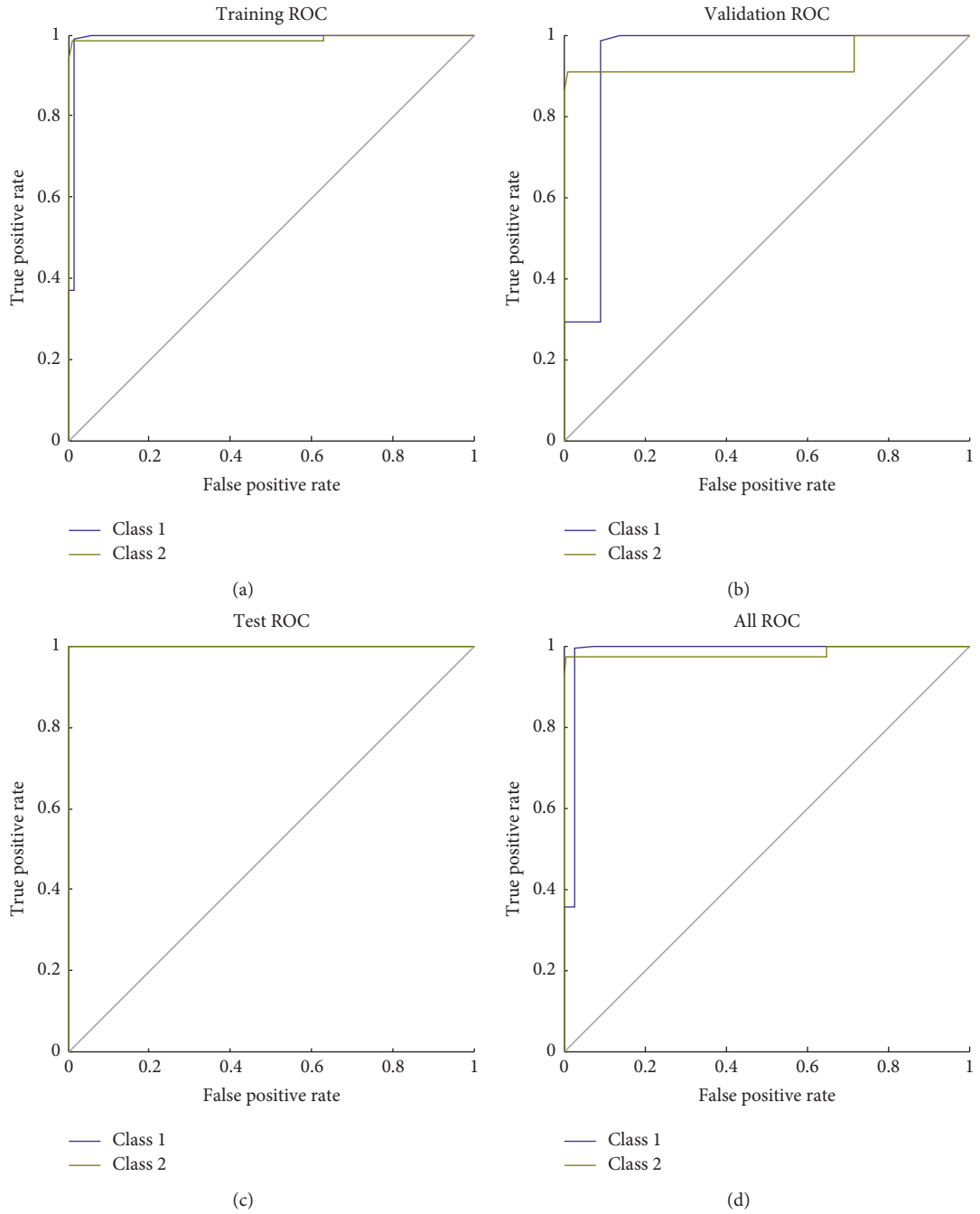


Figure 4: Sensitivity and specificity diagrams.

accidents, and Class 2 indicates the network's accuracy for future accidents. The more inclined the top and left curves, the more powerful the network in estimating and predicting more accurately [64].

5. Conclusion and Safety Approaches

In this study, aiming to investigate the main causes and severity of urban accidents in Rasht city, two models of accident analysis approaches are adopted and compared to provide practical solutions to increase overall safety and reduce the number of accidents within the city. Therefore, using the multiple logistic regression and pattern recognition type of ANN, the variables affecting the severity of accidents and the most powerful approach to predict the accidents in Rasht have been presented; the most important results and prevention strategies of accidents are as follows:

(1) Comparing the correct percentage of prediction in the multiple logistic regression and the machine learning model, the results showed that the ANN model performed better and has a higher prediction power than the logit model. To be more specific, the

prediction accuracy of the utilized ANN is 98.9%, while the logit model's prediction accuracy is 89.17%. In other words, the prediction error rate of the ANN model is 1.1%, while the logit model prediction error rate is 10.83%, which justifies the utilization of the ANN model. On the other hand, the network sensitivity analyzer diagram of ANN approach proves the high power of this model in the prediction of urban accidents.

(2) The multiple logistic regression results show that the forward stepwise method is the best method for making the logit model of accident severity in Rasht city, considering the two criteria of the goodness of fit (R^2) of 0.854 and the prediction accuracy of 89.17%.

(3) According to the logit model results, the variables of accident time (12–18) and (18–24), rainy weather, the accident reason (lack of attention), head-on collision, and KIA Pride vehicle increased the severity of accidents with a positive coefficient. In other words, the significant role of KIA Pride vehicle in the occurrence of accidents, especially at (18–24) accident time (night peak hours' traffic), is noticeable. Therefore, it is expected that the officials should conduct major improvements and corrections in lighting facilities and urban routes infrastructures and improve the quality of mass vehicle production in collaboration with car manufacturing companies.

(4) The logit model also shows that the interactive effect of the darkness of the air and head-on collision increased the likelihood of accidents, which may be due to poor visibility of drivers and cognitive distraction. A greater presence of police and speed control strategies, especially at night (between the hours of 18 and 24) and rainy weather days, is one of the best ways to reduce traffic accidents. Last but not least, increasing penalties for mobile phone users to decrease accidents due to lack of attention and providing warning signs or other pavement-based warning techniques, including pavement markers and rumble strips, may help reduce accident risk.

Since statistical analysis and programming models are usually not able to consider all the required details of the problem, for future studies, it is suggested that to utilize Geographic Information System (GIS) roadway profile data along with more deep learning and optimization techniques to have an in-depth analysis and find the most desirable solutions [65–68]. Moreover, due to the significant importance of pedestrian accidents and permanent interference of nonmotorized users and vehicles flow within a city, it is recommended to use more analytical methods and pattern recognition type of machine learning approach to provide better decision making approaches and present the most accurate prediction model for pedestrian accidents separately occurring in urban environments [55, 69]. Last but not least, to have a more in-depth analysis about pedestrian accidents and find the most dangerous conflicts, it is highly suggested to use AI-based object detection and image processing approaches.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Y. Liu, C. Yang, and Q. Sun, "Thresholds based image extraction schemes in big data environment in intelligent traffic management," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–9, 2020.

[2] Q. Zhu, "Research on road traffic situation awareness system based on image big data," *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 18–26, 2019.

[3] A. Abdi, O. Nassimi, R. Salehfard, and V. N. Moghaddam, "Analysing the influence of encroachment angle and median parameters on safety of rural highways using vehicle dynamics performance," *IOP Conference Series: Materials Science and Engineering*, vol. 471, p. 62043, 2019.

[4] I. Bargegol, V. N. M. Gilani, M. Ghasedi, and M. Ghorbanzadeh, "Delay modeling of un-signalized roundabouts using neural network and regression," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 1, pp. 28–34, 2016.

[5] S. M. Hosseinian and V. Najafi Moghaddam Gilani, "Analysis of factors affecting urban road accidents in Rasht metropolis," *ENG Transactions*, vol. 1, no. 1, pp. 1–4, 2020.

[6] World Health Organization, *Global Status Report on Road Safety 2018*, World Health Organization, Geneva, Switzerland, 2018.

[7] F. Sherafati, E. Homaie-Rad, A. Afkar, R. Gholampoor-Sigaroodi, and S. Sirusbakht, "Risk factors of road traffic accidents associated mortality in northern Iran; a single center experience utilizing oaxaca blinder decomposition," *Bulletin of Emergency & Trauma*, vol. 5, no. 2, pp. 116–121, 2017.

[8] P. Intini, N. Berloco, P. Colonna, V. Ranieri, and E. Ryeng, "Exploring the relationships between drivers' familiarity and two-lane rural road accidents. a multi-level study," *Accident Analysis & Prevention*, vol. 111, pp. 280–296, 2018.

[9] N. Casado-Sanz, B. Guirao, and D. Gálvez-Pérez, "Population ageing and rural road accidents: analysis of accident severity in traffic crashes with older pedestrians on Spanish crosstown roads," *Research in Transportation Business & Management*, vol. 30, p. 100377, 2019.

[10] M. Yazdi, N. A. Golilarz, A. Nedjati, and K. A. Adesina, "An improved lasso regression model for evaluating the efficiency of intervention actions in a system reliability analysis," *Neural Computing and Applications*, 2021.

[11] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing*, vol. 88, p. 105946, 2020.

[12] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.

[13] M. Yazdi, "A perceptual computing–based method to prioritize intervention actions in the probabilistic risk

assessment techniques," *Quality and Reliability Engineering International*, vol. 36, no. 1, pp. 187–213, 2019.

[14] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[15] M. Mirmozaffaria, "Filtering in image processing," *ENG Transactions*, vol. 1, pp. 1–5, 2020.

[16] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[17] N. A. Golilarz, H. Gao, S. Pirasteh, M. Yazdi, J. Zhou, and Y. Fu, "Satellite multispectral and hyperspectral image denoising with enhanced adaptive generalized Gaussian distribution threshold in the wavelet domain," *Remote Sensing*, vol. 13, no. 1, p. 101, 2021.

[18] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[19] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[20] N. A. Golilarz, M. Mirmozaffari, T. A. Gashteroodkhani et al., "Optimized wavelet-based satellite image de-noising with multi-population differential evolution-assisted harris hawks optimization algorithm," *IEEE Access*, vol. 8, pp. 133076–133085, 2020.

[21] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Transactions*, vol. 79, pp. 202–216, 2018.

[22] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, p. 106684, 2021.

[23] A. Anaei and A. A. Kalteh, "A new method for dental caries diagnosis using convolutional neural networks and bees algorithm," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, no. 2, pp. 52–57, 2019.

[24] I. Bargegol, M. Ghorbanzadeh, M. Ghasedi, and M. Rastbod, "Evaluation of effective factors on travel time in optimization of bus stops placement using genetic algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 245, 2017.

[25] M. Ghasedi, M. Ghorbanzadeh, and I. Bargegol, "Robust optimization of bus stop placement based on dynamic demand using meta heuristic approaches: a case study in a developing country," *Transport and Telecommunication Journal*, vol. 22, no. 1, pp. 39–52, 2021.

[26] W. Zhu, C. Ma, X. Zhao et al., "Evaluation of sino foreign cooperative education project using orthogonal sine cosine optimized kernel extreme learning machine," *IEEE Access*, vol. 8, pp. 61107–61123, 2020.

[27] G. Liu, W. Jia, M. Wang et al., "Predicting cervical hyperextension injury: a covariance guided sine cosine support vector machine," *IEEE Access*, vol. 8, pp. 46895–46908, 2020.

[28] Y. Wei, H. Lv, M. Chen et al., "Predicting entrepreneurial intention of students: an extreme learning machine with Gaussian barebone harris hawks optimizer," *IEEE Access*, vol. 8, pp. 76841–76855, 2020.

[29] H. Tang, Y. Xu, A. Lin et al., "Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted *K*-nearest neighbor classifiers," *IEEE Access*, vol. 8, pp. 35546–35562, 2020.

[30] A. Lin, Q. Wu, A. A. Heidari et al., "Predicting intentions of students for master programs using a chaos-induced sine cosine-based fuzzy *K*-nearest neighbor classifier," *IEEE Access*, vol. 7, pp. 67235–67248, 2019.

[31] B. Wang, L. Zhang, H. Ma, H. Wang, and S. Wan, "Parallel LSTM-based regional integrated energy system multienergy source-load information interactive energy prediction," *Complexity*, vol. 2019, Article ID 7414318, 13 pages, 2019.

[32] K. Qu, L. Wei, and Q. Zou, "A review of DNA-binding proteins prediction methods," *Current Bioinformatics*, vol. 14, no. 3, pp. 246–254, 2019.

[33] Z. Zhang, C. Luo, and Z. Zhao, "Application of probabilistic method in maximum tsunami height prediction considering stochastic seabed topography," *Natural Hazards*, vol. 104, no. 3, pp. 2511–2530, 2020.

[34] S. Wang, K. Zhang, L. P. H. van Beek, X. Tian, and T. A. Bogaard, "Physically-based landslide prediction over a large region: scaling low-resolution hydrological model results for high-resolution slope stability assessment," *Environmental Modelling & Software*, vol. 124, p. 104607, 2020.

[35] S. Zheng, Z. Lyu, and L. K. Foong, "Early prediction of cooling load in energy-efficient buildings through novel optimizer of shuffled complex evolution," *Engineering with Computers*, 2020.

[36] L.-Y. Chang, "Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network," *Safety Science*, vol. 43, no. 8, pp. 541–557, 2005.

[37] A. Abdulhafedh, "Crash frequency analysis," *Journal of Transportation Technologies*, vol. 6, no. 4, pp. 169–180, 2016.

[38] S. Mokhtarimousavi, J. C. Anderson, A. Azizinamini, and M. Hadi, "Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and artificial neural networks," *International Journal of Transportation Science and Technology*, vol. 9, no. 2, pp. 100–115, 2020.

[39] J. Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *ENG Transactions*, vol. 1, pp. 1–8, 2020.

[40] A. Mahani, P. Bazoobandi, S. M. Hosseinian, and H. Ziari, "Experimental investigation and multi-objective optimization of fracture properties of asphalt mixtures containing nanocalcium carbonate," *Construction and Building Materials*, vol. 285, p. 122876, 2021.

[41] W. Zhu, J. Miao, J. Hu, and L. Qing, "Vehicle detection in driving simulation using extreme learning machine," *Neurocomputing*, vol. 128, pp. 160–165, 2014.

[42] H. Behbahani, A. M. Amiri, R. Imaninasab, and M. Alizamir, "Forecasting accident frequency of an urban road network: a comparison of four artificial neural network techniques," *Journal of Forecasting*, vol. 37, no. 7, pp. 767–780, 2018.

[43] M. Yazdi, "Ignorance-aware safety and reliability analysis: a heuristic approach," *Quality and Reliability Engineering*, vol. 36, no. 2, pp. 652–674, 2020.

[44] Y. Zhang, R. Liu, A. A. Heidari et al., "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, pp. 185–212, 2020.

[45] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[46] H. Behbahani, H. Ziari, A. Amini, V. N. M. Gilani, and R. Salehfard, "Investigation of un-signalized roundabouts delay with adaptive network-based fuzzy inference system and fuzzy logic," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, pp. 140–149, 2016.

[47] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Analysis & Prevention*, vol. 43, no. 1, pp. 402–411, 2011.

[48] A. Pashaei, M. Ghatee, and H. Sajedi, "Convolution neural network joint with mixture of extreme learning machines for feature extraction and classification of accident images," *Journal of Real-Time Image Processing*, vol. 17, no. 4, pp. 1051–1066, 2019.

[49] B. Yu, Y. Chen, S. Bao, and D. Xu, "Quantifying drivers' visual perception to analyze accident-prone locations on two-lane mountain highways," *Accident Analysis & Prevention*, vol. 119, pp. 122–130, 2018.

[50] D. Chimba and T. Sando, "The prediction of highway traffic accident injury severity with neuromorphic techniques," *Advances in Transportation Studies*, vol. 19, pp. 17–26, 2009.

[51] S. J. Khair, A. F. Muaath, and F. G. Hala, "Prediction of road traffic accidents in Jordan using artificial neural network (ANN)," *Journal of Traffic and Logistics Engineering*, vol. 2, pp. 92–94, 2014.

[52] H. Omrani, "Predicting travel mode of individuals by machine learning," *Transportation Research Procedia*, vol. 10, pp. 840–849, 2015.

[53] I. E. Contreras, L. Torres-Treviño, and F. Torres, "Prediction of car accidents using a maximum sensitivity neural network," in *Smart Technology Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, G. F. Torres, J. Lozoya-Santos, E. Gonzalez Mendivil, L. Neira-Tovar, P. Ramírez Flores, and J. Martin-Gutierrez, Eds., Springer, Cham, Switzerland, 2018.

[54] S. Amin, "Backpropagation-artificial neural network (BP-ANN): understanding gender characteristics of older driver accidents in West Midlands of United Kingdom," *Safety Science*, vol. 122, p. 104539, 2020.

[55] M. Ghasedi, M. Sarfjoo, and I. Bargegol, "Prediction and analysis of the severity and number of suburban accidents using logit model, factor analysis and machine learning: a case study in a developing country," *SN Applied Sciences*, vol. 3, p. 13, 2021.

[56] Iran Statistical Center, *Official Results of Census of Populations and Houses of Iran*, Iran Statistical Center, Tehran, Iran, 2016.

[57] C. Kosun and S. Ozdemir, "The modified Brandeis dice for the traffic states dependency: the I-dice problem," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 2, pp. 76–83, 2020.

[58] K. Rosaiah, N. S. Kolli, and N. S. Sanjeeva Rao, "An application of multiple logistic regression for identifying lipid profile changes towards assessing maternal and fetal outcomes," *International Journal of Community Medicine and Public Health*, vol. 7, no. 8, p. 3014, 2020.

[59] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "Bayesian hierarchical model-based information fusion for degradation analysis considering non-competing relationship," *IEEE Access*, vol. 7, pp. 175222–175227, 2019.

[60] J. Guo, H. Zheng, B. Li, and G.-Z. Fu, "A Bayesian approach for degradation analysis with individual differences," *IEEE Access*, vol. 7, pp. 175033–175040, 2019.

[61] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2020.

[62] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[63] T. Saito and M. Rehmsmeier, "Precrec: fast and accurate precision-recall and ROC curve calculations in *R*," *Bioinformatics*, vol. 33, no. 1, pp. 145–147, 2017.

[64] A. Addeh and B. M. Maghsoudi, "Control chart patterns detection using COA based trained MLP neural network and shape features," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 1, pp. 5–8, 2016.

[65] F. Jiang, K. K. R. Yuen, and E. W. M. Lee, "Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology," *Journal of Safety Research*, vol. 75, pp. 292–309, 2020.

[66] R. U. Khan, X. Zhang, R. Kumar, A. Sharif, N. A. Golilarz, and M. Alazab, "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[67] A. Addeh and M. Iri, "Brain tumor type classification using deep features of MRI images and optimized RBFNN," *ENG Transactions*, vol. 2, no. 1, pp. 1–7, 2021.

[68] N. Amiri Golilarz and H. Demirel, "Thresholding neural network (TNN) based noise reduction with a new improved thresholding function," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 3, no. 2, pp. 81–84, 2017.

[69] S. M. Hosseinian, V. N. M. Gilani, H. T. Amoli, M. Nikookar, and A. Oroue, "Presentation of analytical methods for better decision making about the most important factor influencing rural accidents," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5564269, 16 pages, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9984840, 12 pages
https://doi.org/10.1155/2021/9984840

Research Article

Optimization of Solar CCHP Systems with Collector Enhanced by Porous Media and Nanofluid

Navid Tonekaboni,[1] **Mahdi Feizbahr**,[2] **Nima Tonekaboni**,[1] **Guang-Jun Jiang**,[3,4] **and Hong-Xia Chen**[3,4]

[1]*Mechanical-Energy Conversion, Islamic Azad University, Nour Branch, Nour, Iran*
[2]*School of Civil Engineering, Department of Engineering, University Sains Malaysia, Pulau Penang, Malaysia*
[3]*School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot, Inner Mongolia 010051, China*
[4]*Inner Mongolia Key Laboratory of Advanced Manufacturing Technology, Hohhot, Inner Mongolia 010051, China*

Correspondence should be addressed to Mahdi Feizbahr; m.feizbahr@gmail.com and Guang-Jun Jiang; jianggj_2003@163.com

Received 5 March 2021; Revised 20 March 2021; Accepted 1 April 2021; Published 16 April 2021

Academic Editor: Noorbakhsh Amiri Golilarz

The low efficiency of solar collectors can be mentioned as one of the problems in solar combined cooling, heating, and power (CCHP) cycles. For improving solar systems, nanofluid and porous media are used in solar collectors. One of the advantages of using porous media and nanoparticles is to absorb more energy under the same conditions. In this research, a solar combined cooling, heating, and power (SCCHP) system has been optimized by porous media and nanofluid for generating electricity, cooling, and heating of a 600 m^2 building in a warm and dry region with average solar radiation of Ib = 820 w/m^2 in Iran. In this paper, the optimal amount of nanofluid in porous materials has been calculated to the extent that no sediment is formed. In this study, solar collectors were enhanced with copper porous media (95% porosity) and CuO and Al_2O_3 nanofluids. 0.1%–0.6% of the nanofluids were added to water as working fluids; it is found that 0.5% of the nanofluids lead to the highest energy and exergy efficiency enhancement in solar collectors and SCCHP systems. Maximum energy and exergy efficiency of parabolic thermal collector (PTC) riches in this study are 74.19% and 32.6%, respectively. Figure 1 can be mentioned as a graphical abstract for accurately describing the cycle of solar CCHP.

1. Introduction

Due to the increase in energy consumption, the use of clean energy is one of the important goals of human societies. In the last four decades, the use of cogeneration cycles has increased significantly due to high efficiency. Among clean energy, the use of solar energy has become more popular due to its greater availability [1]. Low efficiency of energy production, transmission, and distribution system makes a new system to generate simultaneously electricity, heating, and cooling as an essential solution to be widely used. The low efficiency of the electricity generation, transmission, and distribution system makes the CCHP system a basic solution to eliminate waste of energy. CCHP system consists of a prime mover (PM), a power generator, a heat recovery system (produce extra heating/cooling/power), and thermal energy storage (TES) [2]. Solar combined cooling, heating, and power (SCCHP) has been started three decades ago. SCCHP is a system that receives its propulsive force from solar energy; in this cycle, solar collectors play the role of propulsive for generating power in this system [3].

Increasing the rate of energy consumption in the whole world because of the low efficiency of energy production, transmission, and distribution system causes a new co-generation system to generate electricity, heating, and cooling energy as an essential solution to be widely used. Building energy utilization fundamentally includes power required for lighting, home electrical appliances, warming and cooling of building inside, and boiling water. Domestic

usage contributes to an average of 35% of the world's total energy consumption [4].

Due to the availability of solar energy in all areas, solar collectors can be used to obtain the propulsive power required for the CCHP cycle. Solar energy is the main source of energy in renewable applications. For selecting a suitable area to use solar collectors, annual sunshine hours, the number of sunny days, minus temperature and frosty days, and the windy status of the region are essentially considered [5]. Iran, with an average of more than 300 sunny days, is one of the suitable countries to use solar energy. Due to the fact that most of the solar radiation is in the southern regions of Iran, also the concentration of cities is low in these areas, and transmission lines are far apart, one of the best options is to use CCHP cycles based on solar collectors [6]. One of the major problems of solar collectors is their low efficiency [7]. Low efficiency increases the area of collectors, which increases the initial cost of solar systems and of course increases the initial payback period. To increase the efficiency of solar collectors and improve their performance, porous materials and nanofluids are used to increase their workability.

There are two ways to increase the efficiency of solar collectors and mechanical and fluid improvement. In the first method, using porous materials or helical filaments inside the collector pipes causes turbulence of the flow and increases heat transfer. In the second method, using nanofluids or salt and other materials increases the heat transfer of water. The use of porous materials has grown up immensely over the past twenty years. Porous materials, especially copper porous foam, are widely used in solar collectors. Due to the high contact surface area, porous media are appropriate candidates for solar collectors [8]. A number of researchers investigated Solar System performance in accordance with energy and exergy analyses. Zhai et al. [9] reviewed the performance of a small solar-powered system in which the energy efficiency was 44.7% and the electrical efficiency was 16.9%.

Abbasi et al. [10] proposed an innovative multiobjective optimization to optimize the design of a cogeneration system. Results showed the CCHP system based on an internal diesel combustion engine was the applicable alternative at all regions with different climates. The diesel engine can supply the electrical requirement of 31.0% and heating demand of 3.8% for building.

Jiang et al. [11] combined the experiment and simulation together to analyze the performance of a cogeneration system. Moreover, some research focused on CCHP systems using solar energy. It integrated sustainable and renewable technologies in the CCHP, like PV, Stirling engine, and parabolic trough collector (PTC) [2, 12–15].

Wang et al. [16] optimized a cogeneration solar cooling system with a Rankine cycle and ejector to reach the maximum total system efficiency of 55.9%. Jing et al. analyzed a big-scale building with the SCCHP system and auxiliary heaters to produced electrical, cooling, and heating power. The maximum energy efficiency reported in their work is 46.6% [17]. Various optimization methods have been used to improve the cogeneration system, minimum system size, and performance, such as genetic algorithm [18, 19].

Hirasawa et al. [20] investigated the effect of using porous media to reduce thermal waste in solar systems. They used the high-porosity metal foam on top of the flat plate solar collector and observed that thermal waste decreased by 7% due to natural heat transfer. Many researchers study the efficiency improvement of the solar collector by changing the collector's shapes or working fluids. However, the most effective method is the use of nanofluids in the solar collector as working fluid [21]. In the experimental study done by Jouybari et al. [22], the efficiency enhancement up to 8.1% was achieved by adding nanofluid in a flat plate collector. In this research, by adding porous materials to the solar collector, collector efficiency increased up to 92% in a low flow regime. Subramani et al. [23] analyzed the thermal performance of the parabolic solar collector with Al_2O_3 nanofluid. They conducted their experiments with Reynolds number range 2401 to 7202 and mass flow rate 0.0083 to 0.05 kg/s. The maximum efficiency improvement in this experiment was 56% at 0.05 kg/s mass flow rate.

Shojaeizadeh et al. [24] investigated the analysis of the second law of thermodynamic on the flat plate solar collector using Al_2O_3/water nanofluid. Their research showed that energy efficiency rose up to 1.9% and the exergy efficiency increased by a maximum of 0.72% compared to pure water. Tiwari et al. [25] researched on the thermal performance of solar flat plate collectors for working fluid water with different nanofluids. The result showed that using 1.5% (optimum) particle volume fraction of Al_2O_3 nanofluid as an absorbing medium causes the thermal efficiency to enhance up to 31.64%.

The effect of porous media and nanofluids on solar collectors has already been investigated in the literature but the SCCHP system with a collector embedded by both porous media and nanofluid for enhancing the ratio of nanoparticle in nanofluid for preventing sedimentation was not discussed. In this research, the amount of energy and exergy of the solar CCHP cycles with parabolic solar collectors in both base and improved modes with a porous material (copper foam with 95% porosity) and nanofluid with different ratios of nanoparticles was calculated. In the first step, it is planned to design a CCHP system based on the required load, and, in the next step, it will analyze the energy and exergy of the system in a basic and optimize mode. In the optimize mode, enhanced solar collectors with porous material and nanofluid in different ratios (0.1%–0.7%) were used to optimize the ratio of nanofluids to prevent sedimentation.

2. Cycle Description

CCHP is one of the methods to enhance energy efficiency and reduce energy loss and costs. The SCCHP system used a solar collector as a prime mover of the cogeneration system and assisted the boiler to generate vapor for the turbine. Hot water flows from the expander to the absorption chiller in summer or to the radiator or fan coil in winter. Finally,

before the hot water wants to flow back to the storage tank, it flows inside a heat exchanger for generating domestic hot water [26].

For designing of solar cogeneration system and its analysis, it is necessary to calculate the electrical, heating (heating load is the load required for the production of warm water and space heating), and cooling load required for the case study considered in a residential building with an area of 600 m^2 in the warm region of Iran (Zahedan). In Table 1, the average of the required loads is shown for the different months of a year (average of electrical, heating, and cooling load calculated with CARRIER software).

According to Table 1, the maximum magnitude of heating, cooling, and electrical loads is used to calculate the cogeneration system. The maximum electric load is 96 kW, the maximum amount of heating load is 62 kW, and the maximum cooling load is 118 kW. Since the calculated loads are average, all loads increased up to 10% for the confidence coefficient. With the obtained values, the solar collector area and other cogeneration system components are calculated. The cogeneration cycle is capable of producing 105 kW electric power, 140 kW cooling capacity, and 100 kW heating power.

2.1. System Analysis Equations. An analysis is done by considering the following assumptions:

(1) The system operates under steady-state conditions

(2) The system is designed for the warm region of Iran (Zahedan) with average solar radiation Ib = 820 w/m^2

(3) The pressure drops in heat exchangers, separators, storage tanks, and pipes are ignored

(4) The pressure drop is negligible in all processes and no expectable chemical reactions occurred in the processes

(5) Potential, kinetic, and chemical exergy are not considered due to their insignificance

(6) Pumps have been discontinued due to insignificance throughout the process

(7) All components are assumed adiabatic

Schematic shape of the cogeneration cycle is shown in Figure 1 and all data are given in Table 2.

Based on the first law of thermodynamic, energy analysis is based on the following steps.

First of all, the estimated solar radiation energy on collector has been calculated:

$$\dot{Q}_s = A \times I_b \times \alpha, \tag{1}$$

where α is the heat transfer enhancement coefficient based on porous materials added to the collector's pipes. The coefficient α is increased by the porosity percentage, the type of porous material (in this case, copper with a porosity percentage of 95), and the flow of fluid to the collector equation.

Collector efficiency is going to be calculated by the following equation [9]:

Table 1: The average amount of electric charges, heating load, and cooling load used in the different months of the year in the city of Zahedan for a residential building with 600 m^2.

Month	Electrical load (kW)	Heating load (kW)	Cooling load (kW)
April	73	51	16
May	78	28	38
June	89	13	56
July	92	8	99
August	96	7	118
September	78	9	96
October	70	16	51
November	58	28	35
December	52	36	11
January	54	53	0
February	60	62	0
March	67	59	13

$$\eta_c = 0.7 - 0.41 \frac{(T_1 + T_6)/2 - T_a}{I_b}. \tag{2}$$

Total energy received by the collector is given by [9]

$$\dot{Q}_c = AI_b\eta_c. \tag{3}$$

Also, the auxiliary boiler heat load is [2]

$$\dot{Q}_b = [(h_1 - h_6)\dot{m}_{st} - AI_b\eta_c] \times \alpha. \tag{4}$$

Energy consumed from vapor to expander is calculated by [2]

$$\dot{Q}_{exp} = \dot{m}_{st}\eta_{exp}[h_1 - (h_2'(1 - x_{2s}) + h_2'' x_{2s})]. \tag{5}$$

The power output form by the screw expander [9]:

$$\dot{w}_{exp} = \eta_{exp}\dot{Q}_{exp}. \tag{6}$$

The efficiency of the expander is 80% in this case [11].

In this step, cooling and heating loads were calculated and then, the required heating load to reach sanitary hot water will be calculated as follows:

First step: calculating the cooling load with the following equation [9]:

$$\dot{Q}_{cool} = -5.45358 + 0.17373T_7. \tag{7}$$

Second step: calculating heating loads [9]:

$$\dot{Q}_{Heat} = \eta_{HE}(T_2 - T_8)\dot{m}_{heat}C_{p.W}. \tag{8}$$

Then, calculating the required loud for sanitary hot water will be [9]

$$\dot{Q}_{hotwater} = \eta_{HE}(T_4 - T_{10})\dot{m}_{hotwater}C_{pW}. \tag{9}$$

According to the above-mentioned equations, efficiency is [9]

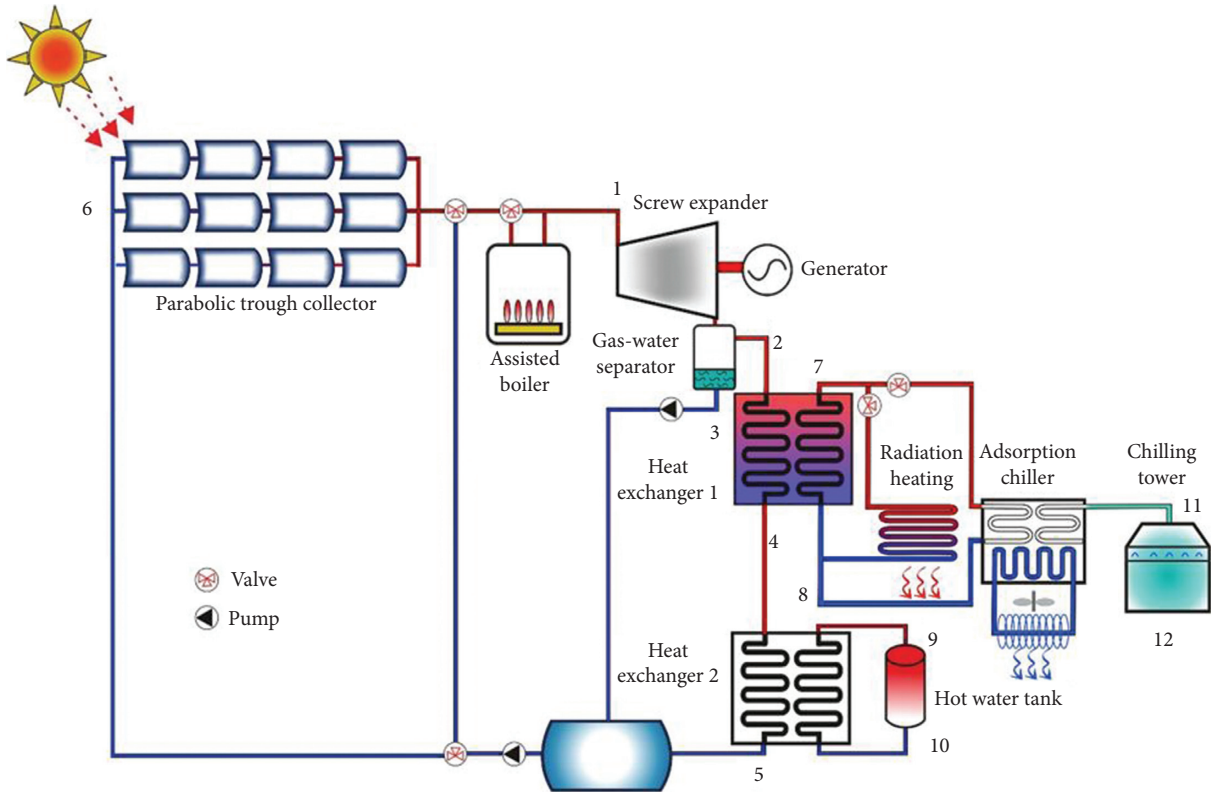


Figure 1: Schematic shape of the cogeneration cycle.

Table 2: Temperature and humidity of different points of system.

State	Temperature summer (c°)	Temperature winter (c°)	Humidity (%)	Enthalpy-summer (kj/kg)	Enthalpy-winter (kj/kg)	Entropy-summer (kj/kg)	Entropy-winter (kj/kg)
1	110	110	0.1	946.7	946.7	2.58	2.58
2	92	92	0.23	913	913	2.63	2.63
3	85	85	—	367.9	367.9	1.17	1.17
4	90	90	0.04	526	526	1.62	1.62
5	85	85	—	367.9	367.9	1.17	1.17
6	85	85	—	367.9	367.9	1.17	1.17
7	88.8	45	—	372	188.4	1.18	0.64
8	82.1	83.3	—	343.7	348.6	1.10	1.11
9	—	10	—	42	42	0.15	0.15
10	—	78	—	326.5	326.5	1.05	1.05

$$\eta_e = \begin{cases} \dfrac{\dot{W}_E + \dot{Q}_{\text{cool}}}{\dot{Q}_S + \dot{Q}_G}, & \text{summer}, \\ \dfrac{\dot{W}_E + \dot{Q}_{\text{heat}} + \dot{Q}_{\text{hotwater}}}{\dot{Q}_S + \dot{Q}_G}, & \text{winter}. \end{cases} \tag{10}$$

In the third step, calculated exergy analysis as follows.

First, the received exergy collector from the sun is calculated [9]:

$$\dot{\text{Ex}}_s = \left[1 - \frac{4}{3}\frac{\text{Ta}}{\text{Ts}}(1 - 0.28\ \ln\ f)\right]\dot{Q}_S, \quad f = 1.3 \times 10^{-5}. \tag{11}$$

In the previous equation, f is the constant of air dilution. The received exergy from the collector is [9]

$$\dot{\text{Ex}}_c = (h_1 - h_6)\dot{m}_{\text{St}} - (\text{Ta} + 273.15)(S_1 - S_6)\dot{m}_{\text{St}}. \tag{12}$$

In the case of using natural gas in an auxiliary heater, the gas exergy is calculated from the following equation [12]:

$$\dot{\text{Ex}}_G = 0.95\dot{Q}_g. \tag{13}$$

Delivering exergy from vapor to expander is calculated with the following equation [9]:

$$\dot{\text{Ex}}_{\text{exp}} = (h_1 - h_2)\dot{m}_{\text{St}} - (Ta + 273.15)(S_1 - S_2)\dot{m}_{\text{St}}. \tag{14}$$

In the fourth step, the exergy in cooling and heating is calculated by the following equation:

Cooling exergy in summer is calculated [9]:

$$\dot{E}_{\text{xcool}} = \dot{Q}_{\text{cool}}\left(\frac{T_a + 273.15}{T_{\text{cool}} + 273.15} - 1\right). \quad (15)$$

Heating exergy in winter is calculated [9]:

$$\dot{E}_{\text{xheat}} = \dot{Q}_{\text{heat}}\left(1 - \frac{T_a + 273.15}{T_{\text{heat}} + 273.15}\right). \quad (16)$$

In the last step based on thermodynamic second law, exergy efficiency has been calculated from the following equation and the above-mentioned calculated loads [9]:

$$\eta_{\text{ex}} = \begin{cases} \dfrac{\dot{W}_E + \dot{\text{Ex}}_{\text{cool}}}{\dot{\text{Ex}}_S + \dot{\text{Ex}}_G}, & \text{summer}, \\ \dfrac{\dot{W}_E + \dot{\text{Ex}}_{\text{heat}} + \dot{\text{Ex}}_{\text{hotwater}}}{\dot{\text{Ex}}_S + \dot{\text{Ex}}_G}, & \text{winter}. \end{cases} \quad (17)$$

3. Porous Media

The porous medium that filled the test section is copper foam with a porosity of 95%. The foams are determined in Figure 2 and also detailed thermophysical parameters and dimensions are shown in Table 3.

In solar collectors, copper porous materials are suitable for use at low temperatures and have an easier and faster manufacturing process than ceramic porous materials. Due to the high coefficient conductivity of copper, the use of copper metallic foam to increase heat transfer is certainly more efficient in solar collectors.

Porous media and nanofluid in solar collector's pipes were simulated in FLOW-3D software using the finite-difference method [27]. Nanoparticles Al_2O_3 and CUO are mostly used in solar collector enhancement. In this research, different concentrations of nanofluid are added to the parabolic solar collectors with porous materials (copper foam with porosity of 95%) to achieve maximum heat transfer in the porous materials before sedimentation. After analyzing PTC pipes with the nanofluid flow in FLOW-3D software, for energy and exergy efficiency analysis, Carrier software results were used as EES software input. Simulation PTC with porous media inside collector pipe and nanofluids sedimentation is shown in Figure 3.

3.1. Nano Fluid. In this research, copper and silver nanofluids (Al_2O_3, CuO) have been added with percentages of 0.1%–0.7% as the working fluids. The nanoparticle properties are given in Table 4. Also, system constant parameters are presented in Table 4, which are available as default input in the EES software.

System constant parameters for input in the software are shown in Table 5.

The thermal properties of the nanofluid can be obtained from equations (18)–(21). The basic fluid properties are indicated by the index (*bf*) and the properties of the nanoparticle silver with the index (*np*).

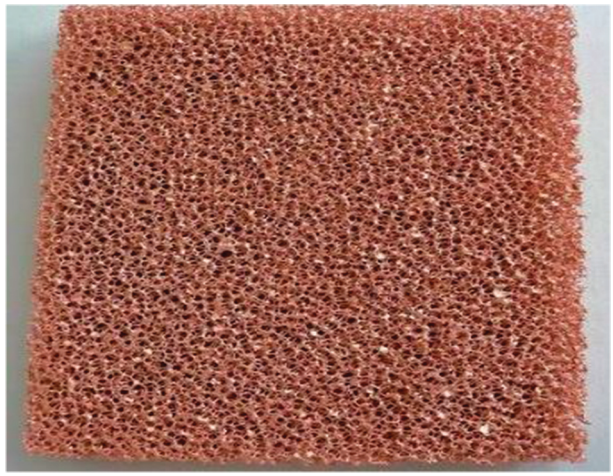


Figure 2: Copper foam with a porosity of 95%.

Table 3: Thermophysical parameters and dimensions of copper foam.

Material	Copper
Porosity	0.95
Permeability, K (m^2)	1.37×10^{-11}
Diameter of porous media, dp (mm)	31
Thermal conductivity (W/(m.k))	398

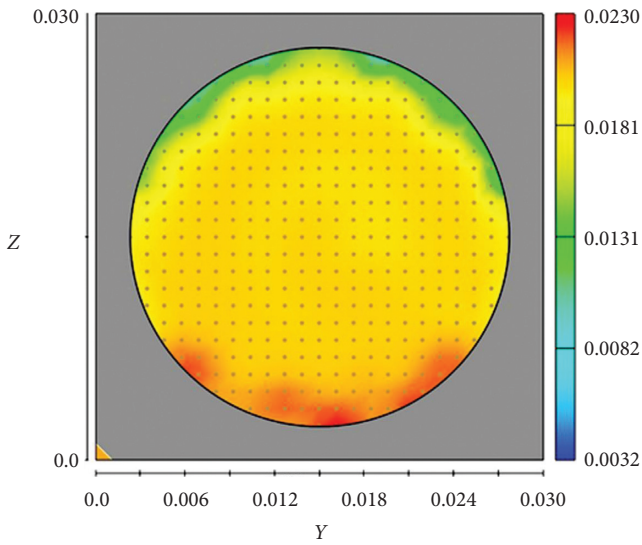


Figure 3: Simulation PTC pipes enhanced with copper foam and nanoparticles in FLOW-3D software.

Table 4: Properties of the nanoparticles [9].

Particle	ρ (kg/m^3)	k (W/mk)	cp (kj/kgk)	Density (kg/m^3)
Al_2O_3	3970	40	0.765	3950
CuO	6320	77	0.532	6310

Table 5: System constant parameters.

Parameters	Values
Average ambient temperature (T_{am})	26°C
Solar beam irradiation (I_b)	0.82 kw/m^2
Temperature difference in recuperator (ΔT_{rc})	22°C
Electromechanical efficiency of the generator (η_{mg})	95%
Turbine isentropic efficiency (η_{is})	80%
Heat exchanger efficiency (η_{he})	90%
Cover emittance (ε_c)	0.92
Average wind speed (V_{wind})	1.5 m/s
Room temperature in summer (T_{rs})	20°C
Room temperature in winter (T_{rw})	24°C
Sun temperature (T_{sun})	5770 K

The density of the mixture is shown in the following equation [28]:

$$\rho_{nf} = \rho_{bf} \cdot (1 - \varphi) + \rho_{np} \cdot \varphi, \tag{18}$$

where ρ is density and ϕ is the nanoparticles volume fraction.

The specific heat capacity is calculated from the following equation [29]:

$$c_{p.nf} = \frac{\rho_{bf \cdot (1-\phi)}}{\rho_{bf}} \cdot c_{p.bf} + \frac{\rho_{np \cdot \phi}}{\rho_{nf}} \cdot c_{p.np}. \tag{19}$$

The thermal conductivity of the nanofluid is calculated from the following equation [29]:

$$k_{nf} = k_{bf} \cdot \frac{k_{np} + 2 \cdot k_{bf} + 2 \cdot \left(k_{np} - k_{bf}\right) \cdot (1 + \beta)^3 \cdot \phi}{k_{np} + 2 \cdot k_{bf} - \left(k_{np} - k_{bf}\right) \cdot (1 + \beta)^3 \cdot \phi}. \tag{20}$$

The parameter β is the ratio of the nanolayer thickness to the original particle radius and, usually, this parameter is taken equal to 0.1 for the calculated thermal conductivity of the nanofluids.

The mixture viscosity is calculated as follows [30]:

$$\mu_{nf} = \mu_{bf} \cdot \left(1 + 2.5 \cdot \phi + 6.5 \cdot \phi^2\right). \tag{21}$$

In all equations, instead of water properties, working fluids with nanofluid are used. All of the above equations and parameters are entered in the EES software for calculating the energy and exergy of solar collectors and the SCCHP cycle. All calculation repeats for both nanofluids with different concentrations of nanofluid in the solar collector's pipe.

4. Results and Discussion

In the present study, relations were written according to Wang et al. [16] and the system analysis was performed to ensure the correctness of the code. The energy and exergy charts are plotted based on the main values of the paper and are shown in Figures 4 and 5. The error rate in this simulation is 1.07%.

We may also investigate the application of machine learning paradigms [31–41] and various hybrid, advanced optimization approaches that are enhanced in terms of exploration and intensification [42–55], and intelligent model studies [56–61] as well, for example, methods such as particle swarm optimizer (PSO) [60, 62], differential search (DS) [63], ant colony optimizer (ACO) [61, 64, 65], Harris hawks optimizer (HHO) [66], grey wolf optimizer (GWO) [53, 67], differential evolution (DE) [68, 69], and other fusion and boosted systems [41, 46, 48, 50, 54, 55, 70, 71].

At the first step, the collector is modified with porous copper foam material. 14 cases have been considered for the analysis of the SCCHP system (Table 6). It should be noted that the adding of porous media causes an additional pressure drop inside the collector [9, 22–26, 30, 72]. All fourteen cases use copper foam with a porosity of 95 percent. To simulate the effect of porous materials and nanofluids, the first solar PTC pipes have been simulated in the FLOW-3D software and then porous media (copper foam with porosity of 95%) and fluid flow with nanoparticles (AL_2O_3 and CUO) are generated in the software. After analyzing PTC pipes in FLOW-3D software, for analyzing energy and exergy efficiency, software outputs were used as EES software input for optimization ratio of sedimentation and calculating energy and exergy analyses.

In this research, an enhanced solar collector with both porous media and Nanofluid is investigated. In the present study, 0.1–0.5% CuO and Al_2O_3 concentration were added to the collector fully filled by porous media to achieve maximum energy and exergy efficiencies of solar CCHP systems. All steps of the investigation are shown in Table 6.

Energy and exergy analyses of parabolic solar collectors and SCCHP systems are shown in Figures 6 and 7.

Results show that the highest energy and exergy efficiencies are 74.19% and 32.6%, respectively, that is achieved in Step 12 (parabolic collectors with filled porous media and 0.5% Al_2O_3). In the second step, the maximum energy efficiency of SCCHP systems with fourteen steps of simulation are shown in Figure 7.

In the second step, where 0.1, −0.6% of the nanofluids were added, it is found that 0.5% leads to the highest energy and exergy efficiency enhancement in solar collectors and SCCHP systems. Using concentrations more than 0.5% leads to sediment in the solar collector's pipe and a decrease of porosity in the pipe [73]. According to Figure 7, maximum energy and exergy efficiencies of SCCHP are achieved in Step 12. In this step energy efficiency is 54.49% and exergy efficiency is 18.29%. In steps 13 and 14, with increasing concentration of CUO and Al_2O_3 nanofluid solution in porous materials, decreasing of energy and exergy efficiency of PTC and SCCHP system at the same time happened. This decrease in efficiency is due to the formation of sediment in the porous material. Calculations and simulations have shown that porous materials more than 0.5% nanofluids inside the collector pipe cause sediment and disturb the porosity of porous materials and pressure drop and reduce the coefficient of performance of the cogeneration system. Most experience showed that CUO and AL_2O_3 nanofluids with less than 0.6% percent solution are used in the investigation on the solar collectors at low temperatures and discharges [74]. One of the important points of this research is that the best ratio of nanofluids in the solar collector with a

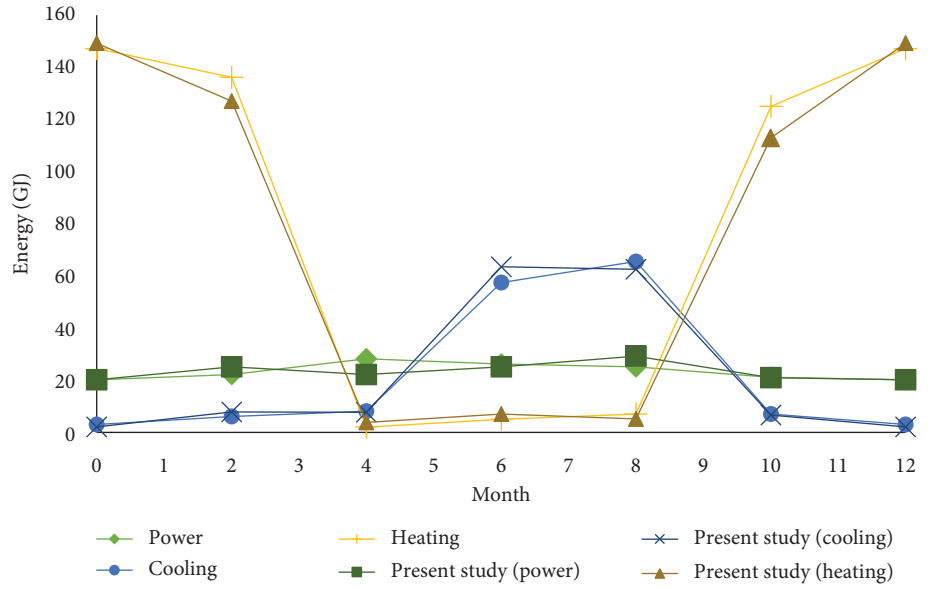


Figure 4: Verification charts of energy analysis results.

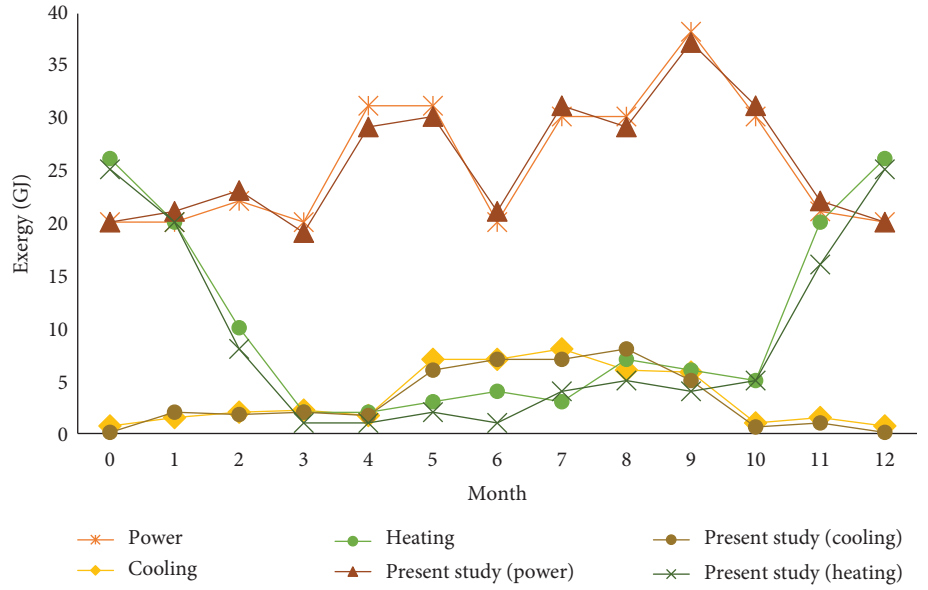


Figure 5: Verification charts of exergy analysis results.

Table 6: Collectors with different percentages of nanofluids and porous media.

1	Simple parabolic solar collectors
2	Parabolic solar collectors with fully filled porous media
3	Parabolic solar collectors with fully filled porous media and 0.1 percent CuO
4	Parabolic solar collectors with fully filled porous media and 0.1 percent Al_2O_3
5	Parabolic solar collectors with fully filled porous media and 0.2 percent CuO
6	Parabolic solar collectors with fully filled porous media and 0.2 percent Al_2O_3
7	Parabolic solar collectors with fully filled porous media and 0.3 percent CuO
8	Parabolic solar collectors with fully filled porous media and 0.3 percent Al_2O_3
9	Parabolic solar collectors with fully filled porous media and 0.4 percent CuO
10	Parabolic solar collectors with fully filled porous media and 0.4 percent Al_2O_3
11	Parabolic solar collectors with fully filled porous media and 0.5 percent CuO
12	Parabolic solar collectors with fully filled porous media and 0.5 percent Al_2O_3
13	Parabolic solar collectors with fully filled porous media and 0.6 percent CuO
14	Parabolic solar collectors with fully filled porous media and 0.56 percent Al_2O_3

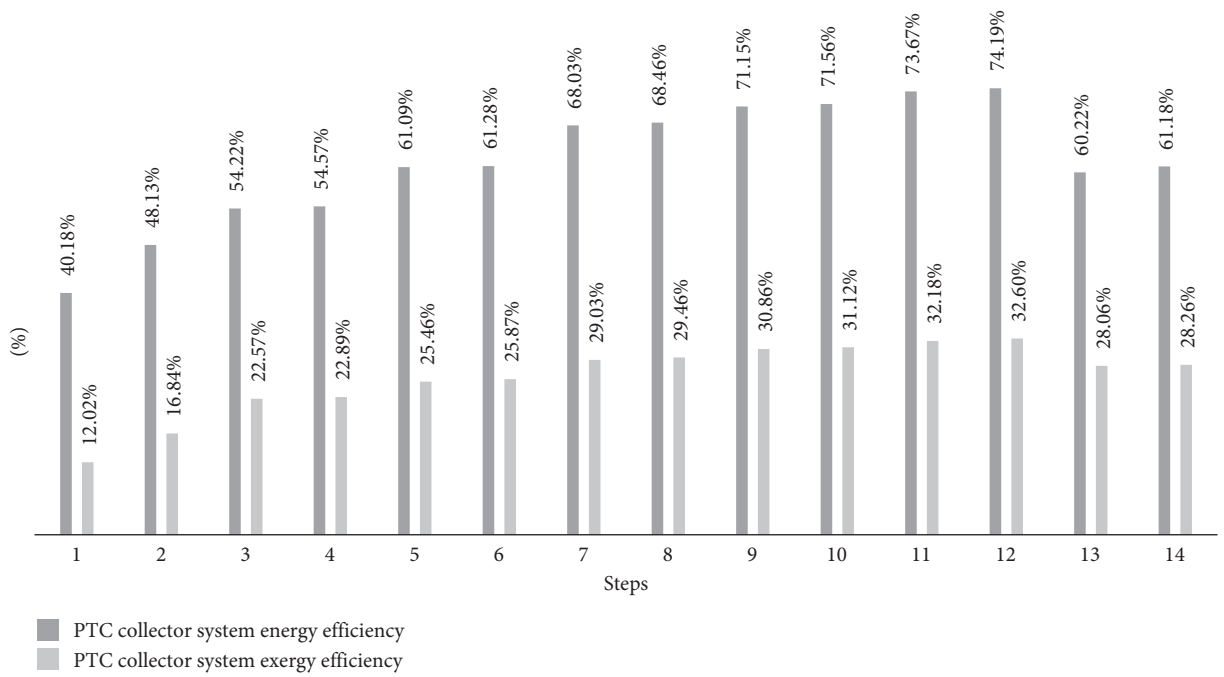


Figure 6: Energy and exergy efficiencies of the PTC with porous media and nanofluid.

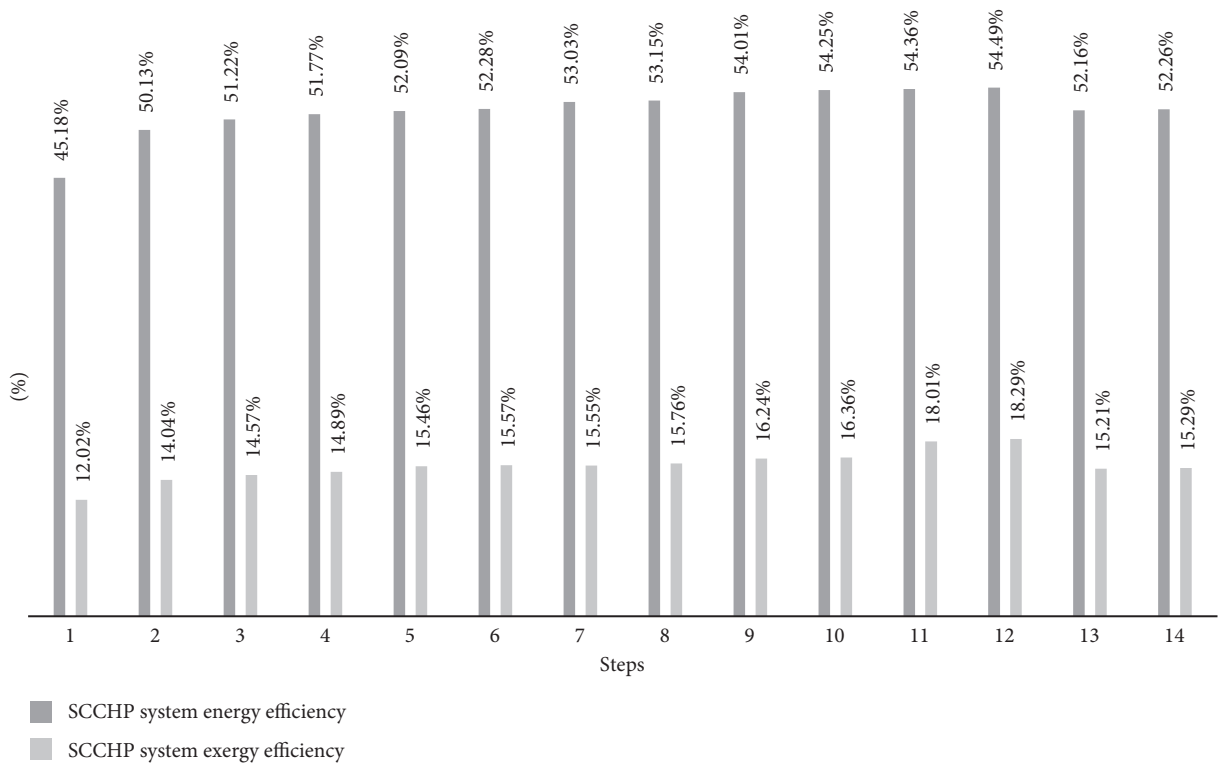


Figure 7: Energy and exergy efficiency of the SCCHP.

low temperature is 0.5% (AL_2O_3 and CUO); with this replacement, the cost of solar collectors and SCCHP cycle is reduced.

5. Conclusion and Future Directions

In the present study, ways for increasing the efficiency of solar collectors in order to enhance the efficiency of the SCCHP cycle are examined. The research is aimed at adding both porous materials and nanofluids for estimating the best ratio of nanofluid for enhanced solar collector and protecting sedimentation in porous media. By adding porous materials (copper foam with porosity of 95%) and 0.5% nanofluids together, high efficiency in solar parabolic collectors can be achieved. The novelty in this research is the addition of both nanofluids and porous materials and calculating the best ratio for preventing sedimentation and pressure drop in solar collector's pipe. In this study, it was observed that, by adding 0.5% of AL_2O_3 nanofluid in working fluids, the energy efficiency of PTC rises to 74.19% and exergy efficiency is grown up to 32.6%. In SCCHP cycle, energy efficiency is 54.49% and exergy efficiency is 18.29%.

In this research, parabolic solar collectors fully filled by porous media (copper foam with a porosity of 95) are investigated. In the next step, parabolic solar collectors in the SCCHP cycle were simultaneously filled by porous media and different percentages of Al_2O_3 and CuO nanofluid. At this step, values of 0.1% to 0.6% of each nanofluid were added to the working fluid, and the efficiency of the energy and exergy of the collectors and the SCCHP cycle were determined. In this case, nanofluid and the porous media were used together in the solar collector and maximum efficiency achieved. 0.5% of both nanofluids were used to achieve the biggest efficiency enhancement.

In the present study, as expected, the highest efficiency is for the parabolic solar collector fully filled by porous material (copper foam with a porosity of 95%) and 0.5% Al_2O_3. Results of the present study are as follows:

(1) The average enhancement of collectors' efficiency using porous media and nanofluids is 28%.

(2) Solutions with 0.1 to 0.5% of nanofluids (CuO and Al_2O_3) are used to prevent collectors from sediment occurrence in porous media.

(3) Collector of solar cogeneration cycles that is enhanced by both porous media and nanofluid has higher efficiency, and the stability of output temperature is more as well.

(4) By using 0.6% of the nanofluids in the enhanced parabolic solar collectors with copper porous materials, sedimentation occurs and makes a high-pressure drop in the solar collector's pipe which causes decrease in energy efficiency.

(5) Average enhancement of SCCHP cycle efficiency is enhanced by both porous media and nanofluid 13%.

Nomenclature

I_b: Solar radiation
a: Heat transfer augmentation coefficient
A: Solar collector area
Bf: Basic fluid
$c_{p.nf}$: Specific heat capacity of the nanofluid
F: Constant of air dilution
k_{nf}: Thermal conductivity of the nanofluid
k_{bf}: Thermal conductivity of the basic fluid
μ_{nf}: Viscosity of the nanofluid
μ_{bf}: Viscosity of the basic fluid
η_c: Collector efficiency
Q_c: Collector energy receives
$\dot{Q}_b$: Auxiliary boiler heat
$\dot{Q}_{\exp}$: Expander energy
$\dot{Q}_g$: Gas energy
$\dot{w}_{\exp}$: Screw expander work
$\dot{Q}_{\text{cool}}$: Cooling load, in kilowatts
$\dot{Q}_{\text{Heat}}$: Heating load, in kilowatts
$\dot{Q}_s$: Solar radiation energy on collector, in Joule
$\dot{Q}_{\text{hotwater}}$: Sanitary hot water load
Np: Nanoparticle
η_e: Energy efficiency
η_{HE}: Heat exchanger efficiency
$\text{x}\dot{E}_s$: Sun exergy
$\text{x}\dot{E}_c$: Collector exergy
$\text{x}\dot{E}_G$: Natural gas exergy
$\text{x}\dot{E}_{\exp}$: Expander exergy
$\dot{E}_{\text{xcool}}$: Cooling exergy
$\dot{E}_{\text{xheat}}$: Heating exergy
η_{ex}: Exergy efficiency
$\dot{m}_{\text{st}}$: Steam mass flow rate
$\dot{m}_{\text{hotwater}}$: Hot water mass flow rate
C_{pW}: Specific heat capacity of water
$\dot{w}_{\exp}$: Power output form by the screw expander
T_{am}: Average ambient temperature
ρ_{nf}: Density of the mixture.

Greek symbols

ρ: Density
ϕ: Nanoparticles volume fraction
β: Ratio of the nanolayer thickness.

Abbreviations

CCHP: Combined cooling, heating, and power
EES: Engineering equation solver.

Data Availability

For this study, data were generated by CARRIER software for the average electrical, heating, and cooling load of a residential building with 600 m^2 in the city of Zahedan, Iran.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Contract no. 71761030 and Natural Science Foundation of Inner Mongolia under Contract no. 2019LH07003.

References

[1] A. Fudholi and K. Sopian, "Review on solar collector for agricultural produce," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 9, no. 1, p. 414, 2018.

[2] G. Yang and X. Zhai, "Optimization and performance analysis of solar hybrid CCHP systems under different operation strategies," *Applied Thermal Engineering*, vol. 133, pp. 327–340, 2018.

[3] J. Wang, Z. Han, and Z. Guan, "Hybrid solar-assisted combined cooling, heating, and power systems: a review," *Renewable and Sustainable Energy Reviews*, vol. 133, p. 110256, 2020.

[4] Y. Tian and C. Y. Zhao, "A review of solar collectors and thermal energy storage in solar thermal applications," *Applied Energy*, vol. 104, pp. 538–553, 2013.

[5] J. M. Hassan, Q. J. Abdul-Ghafour, and M. F. Mohammed, "CFD simulation of enhancement techniques in flat plate solar water collectors," *Al-Nahrain Journal for Engineering Sciences*, vol. 20, no. 3, pp. 751–761, 2017.

[6] M. Jahangiri, O. Nematollahi, A. Haghani, H. A. Raiesi, and A. Alidadi Shamsabadi, "An optimization of energy cost of clean hybrid solar-wind power plants in Iran," *International Journal of Green Energy*, vol. 16, no. 15, pp. 1422–1435, 2019.

[7] I. H. Yılmaz and A. Mwesigye, "Modeling, simulation and performance analysis of parabolic trough solar collectors: a comprehensive review," *Applied Energy*, vol. 225, pp. 135–174, 2018.

[8] F. Wang, J. Tan, and Z. Wang, "Heat transfer analysis of porous media receiver with different transport and thermophysical models using mixture as feeding gas," *Energy Conversion and Management*, vol. 83, pp. 159–166, 2014.

[9] H. Zhai, Y. J. Dai, J. Y. Wu, and R. Z. Wang, "Energy and exergy analyses on a novel hybrid solar heating, cooling and power generation system for remote areas," *Applied Energy*, vol. 86, no. 9, pp. 1395–1404, 2009.

[10] M. H. Abbasi, H. Sayyaadi, and M. Tahmasbzadebaie, "A methodology to obtain the foremost type and optimal size of the prime mover of a CCHP system for a large-scale residential application," *Applied Thermal Engineering*, vol. 135, pp. 389–405, 2018.

[11] R. Jiang, F. G. F. Qin, X. Yang, S. Huang, and B. Chen, "Performance analysis of a liquid absorption dehumidifier driven by jacket-cooling water of a diesel engine in a CCHP system," *Energy and Buildings*, vol. 163, pp. 70–78, 2018.

[12] F. A. Boyaghchi and M. Chavoshi, "Monthly assessments of exergetic, economic and environmental criteria and optimization of a solar micro-CCHP based on DORC," *Solar Energy*, vol. 166, pp. 351–370, 2018.

[13] F. A. Boyaghchi and M. Chavoshi, "Multi-criteria optimization of a micro solar-geothermal CCHP system applying water/CuO nanofluid based on exergy, exergoeconomic and exergoenvironmental concepts," *Applied Thermal Engineering*, vol. 112, pp. 660–675, 2017.

[14] B. Su, W. Han, Y. Chen, Z. Wang, W. Qu, and H. Jin, "Performance optimization of a solar assisted CCHP based on biogas reforming," *Energy Conversion and Management*, vol. 171, pp. 604–617, 2018.

[15] F. A. Al-Sulaiman, F. Hamdullahpur, and I. Dincer, "Performance assessment of a novel system using parabolic trough solar collectors for combined cooling, heating, and power production," *Renewable Energy*, vol. 48, pp. 161–172, 2012.

[16] J. Wang, Y. Dai, L. Gao, and S. Ma, "A new combined cooling, heating and power system driven by solar energy," *Renewable Energy*, vol. 34, no. 12, pp. 2780–2788, 2009.

[17] Y.-Y. Jing, H. Bai, J.-J. Wang, and L. Liu, "Life cycle assessment of a solar combined cooling heating and power system in different operation strategies," *Applied Energy*, vol. 92, pp. 843–853, 2012.

[18] J.-J. Wang, Y.-Y. Jing, and C.-F. Zhang, "Optimization of capacity and operation for CCHP system by genetic algorithm," *Applied Energy*, vol. 87, no. 4, pp. 1325–1335, 2010.

[19] L. Ali, "LDA–GA–SVM: improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine," *Neural Computing and Applications*, vol. 87, pp. 1–10, 2020.

[20] S. Hirasawa, R. Tsubota, T. Kawanami, and K. Shirai, "Reduction of heat loss from solar thermal collector by diminishing natural convection with high-porosity porous medium," *Solar Energy*, vol. 97, pp. 305–313, 2013.

[21] E. Bellos, C. Tzivanidis, and Z. Said, "A systematic parametric thermal analysis of nanofluid-based parabolic trough solar collectors," *Sustainable Energy Technologies and Assessments*, vol. 39, p. 100714, 2020.

[22] H. J. Jouybari, S. Saedodin, A. Zamzamian, M. E. Nimvari, and S. Wongwises, "Effects of porous material and nanoparticles on the thermal performance of a flat plate solar collector: an experimental study," *Renewable Energy*, vol. 114, pp. 1407–1418, 2017.

[23] J. Subramani, P. K. Nagarajan, S. Wongwises, S. A. El-Agouz, and R. Sathyamurthy, "Experimental study on the thermal performance and heat transfer characteristics of solar parabolic trough collector using Al_2O_3 nanofluids," *Environmental Progress & Sustainable Energy*, vol. 37, no. 3, pp. 1149–1159, 2018.

[24] E. Shojaeizadeh, F. Veysi, and A. Kamandi, "Exergy efficiency investigation and optimization of an Al_2O_3-water nanofluid based Flat-plate solar collector," *Energy and Buildings*, vol. 101, pp. 12–23, 2015.

[25] A. K. Tiwari, P. Ghosh, and J. Sarkar, "Solar water heating using nanofluids–a comprehensive overview and environmental impact analysis," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 3, pp. 221–224, 2013.

[26] D. R. Rajendran, E. Ganapathy Sundaram, P. Jawahar, V. Sivakumar, O. Mahian, and E. Bellos, "Review on influencing parameters in the performance of concentrated solar power collector based on materials, heat transfer fluids and design," *Journal of Thermal Analysis and Calorimetry*, vol. 140, no. 1, pp. 33–51, 2020.

[27] M. Feizbahr, C. Kok Keong, F. Rostami, and M. Shahrokhi, "Wave energy dissipation using perforated and non perforated piles," *International Journal of Engineering*, vol. 31, no. 2, pp. 212–219, 2018.

[28] K. Khanafer and K. Vafai, "A critical synthesis of thermophysical characteristics of nanofluids," *International Journal of Heat and Mass Transfer*, vol. 54, no. 19-20, pp. 4410–4428, 2011.

[29] K. Farhana, K. Kadirgama, M. M. Rahman et al., "Improvement in the performance of solar collectors with

nanofluids - a state-of-the-art review," *Nano-Structures & Nano-Objects*, vol. 18, p. 100276, 2019.

[30] M. Turkyilmazoglu, "Condensation of laminar film over curved vertical walls using single and two-phase nanofluid models," *European Journal of Mechanics-B/Fluids*, vol. 65, pp. 184–191, 2017.

[31] X. Zhang, J. Wang, T. Wang, R. Jiang, J. Xu, and L. Zhao, "Robust feature learning for adversarial defense via hierarchical feature alignment," *Information Sciences*, vol. 2020, 2020.

[32] X. Zhang, T. Wang, W. Luo, and P. Huang, "Multi-level fusion and attention-guided CNN for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, 2020.

[33] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 1, pp. 1–15, 2020.

[34] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 238–255, 2019.

[35] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, 2020.

[36] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Computer Vision and Image Understanding*, vol. 1, 2020.

[37] M. Mirmozaffari, "Machine learning algorithms based on an optimization model," 2020.

[38] M. Mirmozaffari, M. Yazdani, A. Boskabadi, H. Ahady Dolatsara, K. Kabirifar, and N. Amiri Golilarz, "A novel machine learning approach combined with optimization models for eco-efficiency evaluation," *Applied Sciences*, vol. 10, no. 15, p. 5210, 2020.

[39] M. Vosoogha and A. Addeh, "An intelligent power prediction method for wind energy generation based on optimized fuzzy system," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, pp. 34–43, 2019.

[40] A. Javadi, N. Mikaeilvand, and H. Hosseinzdeh, "Presenting a new method to solve partial differential equations using a group search optimizer method (GSO)," *Computational Research Progress in Applied Science and Engineering*, vol. 4, no. 1, pp. 22–26, 2018.

[41] F. J. Golrokh, Gohar Azeem, and A. Hasan, "Eco-efficiency evaluation in cement industries: DEA malmquist productivity index using optimization models," *ENG Transactions*, vol. 1, pp. 1–8, 2020.

[42] H. Yu, "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: method and analysis," *Engineering with Computers*, vol. 1, pp. 1–29, 2020.

[43] C. Yu, "SGOA: annealing-behaved grasshopper optimizer for global tasks," *Engineering with Computers*, vol. 1, pp. 1–28, 2021.

[44] W. Shan, Z. Qiao, A. A. Heidari, H. Chen, H. Turabieh, and Y. Teng, "Double adaptive weights for stabilization of moth flame optimizer: balance analysis, engineering cases, and medical diagnosis," *Knowledge-Based Systems*, vol. 1, p. 106728, 2020.

[45] J. Tu, H. Chen, J. Liu et al., "Evolutionary biogeography-based whale optimization methods with communication structure: towards measuring the balance," *Knowledge-Based Systems*, vol. 212, p. 106642, 2021.

[46] Y. Zhang, "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 1, 2020.

[47] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection," *Engineering with Computers*, vol. 1, pp. 1–30, 2020.

[48] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[49] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[50] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

[51] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid Nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[52] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[53] X. Zhao, X. Zhang, Z. Cai et al., "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Computational Biology and Chemistry*, vol. 78, pp. 481–490, 2019.

[54] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing Journal*, vol. 88, 2020.

[55] X. Xu and H.-L. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Computing*, vol. 18, no. 4, pp. 797–807, 2014.

[56] R. U. Khan, X. Zhang, R. Kumar, A. Sharif, N. A. Golilarz, and M. Alazab, "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[57] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Transactions*, vol. 79, pp. 202–216, 2018.

[58] N. Amiri Golilarz, H. Gao, R. Kumar, L. Ali, Y. Fu, and C. Li, "Adaptive wavelet based MRI brain image de-noising," *Frontiers in Neuroscience*, vol. 14, p. 728, 2020.

[59] N. A. Golilarz, H. Gao, and H. Demirel, "Satellite image de-noising with Harris hawks meta heuristic optimization algorithm and improved adaptive generalized Gaussian distribution threshold function," *IEEE Access*, vol. 7, pp. 57459–57468, 2019.

[60] M. Eisazadeh and J. Rezapour, "Multi-objective optimization of the composite sheets using PSO algorithm," 2017.

[61] I. Bargegol, M. Nikookar, R. V. Nezafat, E. J. Lashkami, and A. M. Roshandeh, "Timing optimization of signalized intersections using shockwave theory by genetic algorithm," *Computational Research Progress in Applied Science & Engineering*, vol. 1, pp. 160–167, 2015.

[62] B. Bai, Z. Guo, C. Zhou, W. Zhang, and J. Zhang, "Application of adaptive reliability importance sampling-based extended domain PSO on single mode failure in reliability engineering," *Information Sciences*, vol. 546, pp. 42–59, 2021.

[63] J. Liu, C. Wu, G. Wu, and X. Wang, "A novel differential search algorithm and applications for structure design,"

Applied Mathematics and Computation, vol. 268, pp. 246–269, 2015.

[64] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing*, vol. 24, pp. 585–596, 2014.

[65] D. Zhao, "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowledge-Based Systems*, vol. 24, p. 106510, 2020.

[66] H. Chen, A. A. Heidari, H. Chen, M. Wang, Z. Pan, and A. H. Gandomi, "Multi-population differential evolution-assisted Harris hawks optimization: framework and case studies," *Future Generation Computer Systems*, vol. 111, pp. 175–198, 2020.

[67] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, p. 106684, 2021.

[68] G. Sun, B. Yang, Z. Yang, and G. Xu, "An adaptive differential evolution with combined strategy for global numerical optimization," *Soft Computing*, vol. 24, pp. 1–20, 2019.

[69] G. Sun, C. Li, and L. Deng, "An adaptive regeneration framework based on search space adjustment for differential evolution," *Neural Computing and Applications*, vol. 24, pp. 1–17, 2021.

[70] A. Addeh and M. Iri, "Brain tumor type classification using deep features of MRI images and optimized RBFNN," *ENG Transactions*, vol. 2, pp. 1–7, 2021.

[71] F. J. Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *Soft Computing*, vol. 1, pp. 1–8, 2020.

[72] H. Tyagi, P. Phelan, and R. Prasher, "Predicted efficiency of a low-temperature nanofluid-based direct absorption solar collector," *Journal of Solar Energy Engineering*, vol. 131, no. 4, 2009.

[73] S. Rashidi, M. Bovand, and J. A. Esfahani, "Heat transfer enhancement and pressure drop penalty in porous solar heat exchangers: a sensitivity analysis," *Energy Conversion and Management*, vol. 103, pp. 726–738, 2015.

[74] N. Akram, R. Sadri, S. N. Kazi et al., "A comprehensive review on nanofluid operated solar flat plate collectors," *Journal of Thermal Analysis and Calorimetry*, vol. 139, no. 2, pp. 1309–1343, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 5567638, 21 pages
https://doi.org/10.1155/2021/5567638

Research Article

Statistical Analysis for Study of the Effect of Dark Clothing Color of Female Pedestrians on the Severity of Accident Using Machine Learning Methods

Seyed Mohsen Hosseinian,[1] **Vahid Najafi Moghaddam Gilani**,[1] **Babak Mirbaha**,[2] **and Ali Abdi Kordani**[2]

[1]*School of Civil Engineering, Iran University of Science and Technology (IUST), Tehran, Iran*
[2]*School of Civil Engineering, Imam Khomeini International University (IKIU), Qazvin, Iran*

Correspondence should be addressed to Seyed Mohsen Hosseinian; sm_hoseinian@civileng.iust.ac.ir

Received 3 March 2021; Revised 20 March 2021; Accepted 24 March 2021; Published 14 April 2021

Academic Editor: Noorbakhsh Amiri Golilarz

The color and brightness of pedestrian clothing are among the factors that could increase the severity of their accidents due to the lack of visibility, especially at night. Today, as most Iranian females tend to wear hijab or dark clothing, the necessity of investigating female pedestrian accidents influenced by clothing color is important. Many studies have been performed to analyze the severity of pedestrian accidents, but a study has not yet been conducted to determine the effect of the dark clothing color of female pedestrians on the severity of accidents. Therefore, in this study, 12 independent variables affecting the severity of female pedestrian accidents such as clothing color, age, accident time, day, weather condition, education, pedestrian action, crossing facilities, crossing permit, job, road classification, and fault status were studied. Frequency analysis, Friedman test (FT), and Factor Analysis (FA) methods, as well as modeling methods of Multiple Logistic Regression (MLR) and Artificial Neural Networks (ANNs) using Multilayer Perceptron (MLP) and Radius Basis Function (RBF), were used. Results indicated that clothing color had a significant influence on pedestrian accidents, and chador and dark clothing color increased the probability of accidents, especially at night. The MLP model had a better prediction percentage than the rest, the prediction accuracy of which was 94.6%. Finally, safety solutions were presented according to the results to reduce pedestrian accidents and increase road safety.

1. Introduction

Walking is one of the forms of transportation that can connect a private environment with the community in the absence of other means of transportation [1]. Each person is a pedestrian during a day at different times depending on his/her activity, mode selection, or travel plan. So it is important to consider pedestrian safety along with other road users [2, 3]. Traffic management is an important issue in developed countries, and the government should invest substantially in this case for an effective transportation network [4, 5]. If no measure is taken to reduce accidents and increase road safety, the rate of accidents and casualties will always increase and result in the loss of financial and human resources [6–7].

Thousands of people are disabled or injured with long-lasting effects annually, which imposes considerable expenses on families and societies. Road Traffic Accidents (RTAs) today represent the eighth leading cause of death globally. Deaths from RTAs have increased to 1.35 million a year and caused up to 50 million injuries in 2016. In Iran, the reported number of road traffic deaths was 15,932 in 2016. Also, the estimated traffic fatality rate per 100,000 population was 20.5 [8]. So, deaths caused by Road Traffic Injuries (RTIs) should be of particular concern in Iran [9]. According to police data, the number of fatal pedestrian

accidents in Tehran from March 2017 to March 2018 was 8,582 and from March 2018 to March 2019 was 9,217. A comparison of pedestrian accidents in 2018 with 2017 shows an accident rate growth of 7.7% [10].

Pedestrians are the most vulnerable users of the highway transportation system. Various factors contribute to traffic accidents. These causes could be analyzed and useful measurements should be taken in order to reduce accidents [11–13]. Therefore, recognition of pedestrian behavior is useful in controlling and managing the transportation network [14].

In Iran, as females are required by law to wear hijabs, such as chador or other Islamic clothing, and because of their tendency to wear chador and dark clothing, a study is needed to examine the impact of their clothing on the severity of pedestrian accidents. Therefore, in this study, the effect of the dark clothing color of female pedestrians on the rate of accidents was investigated, and some strategies were presented to improve their safety.

Many studies indicated the importance of clothing color in the safety and severity of pedestrian accidents. A study was conducted by Turner et al. [15] to specify the most visible color of safety clothing in daylight for protecting the personnel in the construction of a highway. Eleven different colors were examined, including one semifluorescent color (yellow), two nonfluorescent colors (orange and yellow), and eight fluorescent colors (yellow, green, yellow-green, red-orange, yellow-orange, red mesh over white background, a combination of yellow-green with red-orange, and pink). They found the highest mean detection distance by the fluorescent red-orange colored vests. Clifton et al. [16] examined the effect of environmental and personal features on the severity of injuries in vehicle-pedestrian accidents. They found that pedestrians with dark clothing color had a lower rate of accidents compared to those wearing mixed or light clothing. Moreover, compared to those wearing lighter clothing, pedestrians with dark clothing color were also less likely to be killed or injured. These findings may be because the majority of accidents, almost 70%, happened during the day. Tyrrell et al. [17] examined the effect of extraneous points of light on the capability of 12 older and 12 younger drivers in seeing pedestrians while driving on a closed road at night. Pedestrians had black clothing alone or clothing with retro-reflective markings in 4 various configurations. They found that pedestrian motion and clothing configuration had an effect on the conspicuity.

A high number of pedestrian collisions and fatalities was reported by Zegeer and Bushell [18], occurring at night or under low-lighting circumstances. A study was performed by Wood et al. [19] on quantifying the impacts of driver's age on the conspicuity of the nighttime pedestrian and the capability of drivers in recognizing the pedestrians at night. They found that driver's age and pedestrian's clothing had a considerable effect on the distance at which the drivers initially reacted to the presence of pedestrians. When pedestrians wore a biological motion reflective clothing, they were recognized more often than when they wore a reflective vest. Borzendowski et al. [20] explored the accuracy of drivers' judgment from pedestrian conspicuity in each of the two clothing configurations while facing varying degrees of headlight glare. Their research found that by wearing a retro-reflective vest by the pedestrian, the overestimated conspicuity of participants was higher.

A research was performed by Hefny et al. [21] to analyze the severity and consequence of hospitalized trauma of pedestrian patients involved in an accident. They found that pedestrian collisions at low-lightning circumstances or nighttime had the greatest percentage and stated that the use of visibility aids, including the use of light-colored clothing or inexpensive reflective vests over the night, needs to be encouraged via public campaigns. Pour-Rouholamin and Zhou [22] studied the risk factors related to different injury severity levels experienced by pedestrians. The results showed that pedestrians increased the severity of injuries without contrasting and dark clothing. They also found that these pedestrians were more susceptible to severe injuries in comparison with those wearing contrasting clothing.

Fekety et al. [23] assessed the nighttime conspicuity advantages of the addition of electroluminescent (EL) panels in pedestrian clothing containing retro-reflective components. They indicated that the garment, including retro-reflective substances and EL, obtained longer response distances compared to the retroreflective-only garment. Besharati and Tavakoli Kashani [24] investigated Iranian crash data provided by Iranian traffic police. They showed that by both pedestrian dark clothing and illumination, the crashes that occurred on an urban highway were 9 times more likely to be fatal. Baireddy et al. [25] assessed the factors affecting pedestrian accidents in Illinois. They found that pedestrians not wearing contrasting clothing were involved in 66.29% of accidents. In addition, they were less visible and more susceptible to crashes.

Wickramasinghe et al. [26] provided a study to determine the visibility distance from the pedestrian. Detection distances were calculated for 5 pedestrian clothing colors (yellow, red, black, white, and retro-reflective). Based on the results, drivers' detection distance was remarkably influenced by the clothing color of the crossing pedestrian and headlamp light conditions. Retro-reflective color was identified to have the longest detection distance, followed by white color. Also, black color had the lowest detection distance. Abdous and Mahmoudabadi [27] explored the female pedestrian fatalities in Iran in terms of their death location and other parameters such as visibility (clothing color), day time, and age. The findings revealed that the vulnerability of female pedestrians to accidents had no considerable difference in different times of the day; however, wearing dark clothing (often chador as a prevalent robe in Iran) remarkably increased their vulnerability. Moreover, recently, well-established optimization algorithms with flexible exploratory and exploitative tendencies have been applied to various designs and engineering problems [28–45].

Modeling the severity of accidents in terms of their effective parameters makes it possible to predict the occurrence of accidents requiring relief equipment. Various feature selection and learning methods [40–58] can also be applied as a preprocessing step to these prediction cases. In

addition, using this model, the effect of each factor in intensifying accidents can be investigated. Such recognition will lead to the possibility of formulating traffic safety plans for traffic engineers, and they may also have a better understanding of the factors that have a positive or negative impact on the severity of accidents. Therefore, the aims of this study were as follows:

(1) Investigating the effect of clothing color of female pedestrians on the severity of their accidents
(2) Frequency analysis of variables affecting the accidents of the female pedestrian with dark clothing
(3) Modeling the independent variables affecting the severity of their accidents
(4) Comparison of accident models and analyses with each other and other related studies

2. Methodology

2.1. Study Area. This research was performed in the Tehran metropolis, which is the capital of the Tehran province and Iran. Tehran is the most populous metropolis in the country and western Asia, and includes 22 districts spread across 730 square kilometers and is located at 51°17′ to 51°33′ E and 35°36′ to 35°44′ N and also is 900 to 1800 meters above sea level. According to the official census in 2016, its population was 8,693,706, of which 4,369,551 were female, and the female population density was announced as 5,986 per square [59].

2.2. Data Collection. Information on pedestrian accidents in this study was collected from March 2017 to March 2019 from the Tehran traffic police. Among the recorded variables, female accidents were selected to examine the variables of clothing color, age, accident time, day, weather condition, education, pedestrian action, crossing facilities, crossing permit, job, road classification, and fault status. The total number of pedestrian accidents was 18251, of which 6128 were female pedestrian accidents, among which 3962 were female pedestrian accidents with dark clothing, indicating the high importance of female pedestrians with dark clothing in accidents. Finally, considering the abovementioned variables, the severity of female pedestrian accidents in Tehran was analyzed.

2.3. Description of Variables. In this study, the data of female pedestrian accidents in Tehran from March 2017 to March 2019 were used, which totally included 13 variables in 6128 accidents. The dependent variable in this research was accident severity, which was divided into two types: accidents inflicting injuries and fatal accidents. These variables were coded to begin modeling, as shown in Table 1.

2.4. Methods

2.4.1. Kolmogorov–Simonov (K–S) Test. The K–S test is a nonparametric test for data distribution and is generally applied to explore data normality. In the approximate significance test, by comparing the significance level and α, one can decide the normality of data distribution. By considering $\alpha = 0/05$ (in 95% certainty), if p value >0/05 can be assumed, data distribution will be normal [60].

2.4.2. Friedman Test (FT). The FT test is an extensively used nonparametric technique to compare the classifiers over multiple datasets. It is commonly applied to analyze different associated samples in various fields. For instance, FT is utilized for comparing the performance consequences of a group of classifiers over multiple datasets that cover benchmark functions, case problems, or performance indicators. FT examination process is to analyze variance by ranks. This is used in cases where there is no desire to build strong distributional assumptions [61].

2.4.3. Factor Analysis (FA). FA is a statistical method to model the covariation among a set of data, which is applied to determine the most effective variables when the number of variables is high and the relationships between them are unknown. The main purpose of FA is to decrease the dimensionality of a series of variables and determine the most important variables affecting phenomena production. FA is generally divided into two classes, namely, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA was applied in this research, which explores the latent constructs when the structure of the relationships among variables is unknown [62]. In this study, because the size of the sample is a determining factor in the accuracy of element clustering by FA, it must be assured that it is adequate. So Kaiser–Mayer–Olkin (KMO) and Bartlett's test of sphericity were applied to check the suitability of data.

2.4.4. Multiple Logistic Regression. Regression analysis is a statistical method to exploit independent variables and model the relationship between dependent and independent variables [63, 64]. The logistic regression model is an improved regression applied for dichotomous dependent variables. In this situation, binary logistic regression is utilized. Consider t as a linear function of the independent variables, in which $\beta = (\beta_0, \beta_1, \ldots, \beta_m)$ is the regression coefficient for $x = (x_1, x_2, \ldots, x_m)$ independent variables and ε is the error [65]:

$$t = \beta_0 + \beta_1 x_1 + \cdots + \beta_M x_M + \varepsilon, \tag{1}$$

The impute values are from negative infinity to positive infinity, while t must be between 0 and 1. The logistic function can solve this problem that is defined as equation (2), which has values of 0 to 1 and is interpreted as a probability:

$$\sigma(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}. \tag{2}$$

The binary logistic model will be in the form of equation (3) that presents the probability of the dependent variable:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x + \cdots + \beta_M x_M)}}. \quad (3)$$

As β_i increases positively,$\sigma(t)$ approaches 1, indicating the success probability increments [66].

2.4.5. Artificial Neural Network (ANN). Neural networks (NN), or more accurately artificial neural networks, are physically cellular systems able to achieve, store, and use experimental data [67, 68]. NNs, with their extraordinary ability to derive a general solution from imprecise or complicated data, are applied for extracting patterns and detecting trends that are too complex to be observed by human, expert systems, or other computer methods [69–77].

Network learning intelligent models have advantages over other methods such as flexibility over data, free assumptions in methodology, no requirement for professional statistical analysis before modeling, and can easily model complex multiparameter systems. This method can be very useful in finding complex nonlinear relationships when the database is high dimensional [47, 78–81]. Obtaining more efficient models to predict accidents will allow researchers to achieve a better performance in road safety. Network learning intelligent models help them to identify patterns in identifying problems and influential parameters in accidents. The advantage of intelligent models over statistical methods is that in intelligent models, the weight of variables and the amount of effect can be calculated, and the resulting network can be optimized while in statistical methods, it is not possible to optimize the model and obtain the influence of parameters concurrently [82–86].

One of the various types of network learning intelligent models is ANN. The main purpose of developing ANN models is to explore the impact of each independent variable on the goal variable. The reason for emphasizing intelligent models in this research is the high efficiency compared to other methods and determining the amount and importance of the parameters influencing the pedestrian accidents [87, 88].

Multilayer Perceptron (MLP) and Radius Basis Function (RBF) are applied in this research, among different methods of ANN models, and finally, the modeling results are compared. A general distinction between these two methods is that RBF is a localist learning kind that is responsive just to a limited section of input space, while MLP is a further distributed technique [89]. The performance of MLP and RBF is completed in two stages; the first phase is the training phase, which determines the weight of NN neurons and applies a set of patterns for training the network, and the second phase is the testing phase, which represents a list of testing patterns to the trained network to find the network output produced and compare the trained network outputs with the desired output of the testing patterns. It determines the classification accuracy of the network [29, 90, 91].

MLP and RBF models consist of neurons that are defined in three layers of input, hidden, and output. The inputs as the original predictors or the hidden layers comprise a group of variables, and the output layer includes the responses. Each variable in a layer is named a node. A node takes a changed linear combination of the outputs from the nodes beneath the layer. Then, as an output, it sends a conversion of itself as one of the inputs to one or more nodes on the subsequent layer. Typically, the conversion functions are named activation functions. There are different forms of activation functions that are applied by the nature of the problem and are selected by the network architecture. The number of neurons in input and output layers depends on independent and dependent variables, respectively. The number of neurons in the hidden layers was also specified by calibration in several runs [92, 93].

3. Results

3.1. Frequency Analysis. Analysis of pedestrian accidents based on the color of pedestrian clothing, as shown in Figure 1(a), showed that female pedestrians wore darker clothing than males (64.7% of female pedestrians had dark clothing and only 31.7% of male pedestrians had dark clothing). Frequency results showed a higher proportion of female pedestrian accidents with dark clothing (such as chador clothing) than bright clothing. Figure 1(b) also shows that among female pedestrians with dark clothing, 57.1% were females with chador clothing and 42.9% of them wore other dark clothing.

Figure 2 shows that 30.8% of female pedestrians accounting for the highest percentage of accidents were aged 51 and above, that 73.6% of them had dark clothing, and 25.8% of female pedestrians were aged 30 or less, of which only 24.1% had dark clothing, indicating that most females in the older age group tended to wear darker clothing than younger females aged 30 or less. Due to the slower walking speed and old age, more accidents had resulted, while the younger female pedestrians were more inclined to wearing bright clothing, which can reduce the occurrence of accidents, especially at night.

According to Figure 3(a), 77.9% of female accidents with dark clothing occurred at night or sunset, of which 44.7% occurred between 5 pm and 9 pm, 26.5% occurred between 9 pm and 1 am, and 6.7% happened between 1 am and 7 am. On the other hand, 62.8% of female pedestrian accidents with dark clothing (as the highest accident rate) occurred in rainy weather conditions and the least accidents (10.8%) occurred in clear weather (sunny or moonlit), as shown in Figure 3(b). It shows that in rainy weather conditions, the possibility of female pedestrian accidents with dark clothing was much greater because of insufficient visibility of drivers. Also, according to the results of analysis and Figure 3(c), 43.7% of accidents occurred on arterial roads, indicating that the probability of the occurrence of accidents was more due to a higher speed of vehicles and the presence of more females on these roads and lower resolution of dark clothing.

According to Figure 4, 58.2% of female accidents with dark clothing occurred on Wednesday to Friday and holidays (on recreational trips), and only 41.8% of accidents occurred on Saturday to Wednesday, indicating that due to more work trips, 41.5% and 36.7% of accidents were on arterial roads and highways, respectively, and only 21.8% occurred on access roads.

Table 1: Types of variables in the research.

Number	Variables	Parameters
	Dependent variable	
1	Accident severity	(1) Injury (2) Fatal
	Independent variables	
1	Clothing color	(1) Bright clothing (2) Chador clothing (3) Other dark clothing
2	Age	(1) Less than or equal to 30 (2) 31 to 40 (3) 41 to 50 (4) 51 and over
3	Accident time	(1) 7 : 00 to 13 : 00 (2) 13 : 00 to 17 : 00 (3) 17 : 00 to 21 : 00 (4) 21 : 00 to 1 : 00 (5) 1 : 00 to 7 : 00
4	Day	(1) Saturday to Wednesday (2)Wednesday to Friday and holidays
5	Weather condition	(1) Clear (2) Rainy (3) Cloudy
6	Education	(1) Illiterate (2) Just reading and writing (3) Primary school (4) High school (5) Diploma (6) Associate's degree (7) Bachelor's degree (8) Master's degree (9) Doctoral degree
7	Pedestrian action	(1) Crossing from the authorized location (2) Crossing from the unauthorized direction (3) Standing alongside the road (4) Getting out of the vehicle (5) Walking in the direction of the vehicle's movement (6) Walking in the nondirection of the vehicle's movement (7) Out of the roadway (8) Working on the vehicle (9) Sudden jump in front of the vehicle (10) Sudden running on the roadway (11) Pushing the vehicle (12) Crossing the road barriers (13) Others
8	Crossing facilities	(1) Pedestrian bridge (2) Underpass (3) Crosswalk (4) Don't have
9	Crossing permit	(1) Allowed (2) Illegal
10	Job	(1) Student (2) College students (3) Self-employed (4) Clerk (5) Housewife
11	Road classification	(1) Highway (2) Arterial road (3) Access road
12	Fault status	(1) Guilty (2) Nonguilty

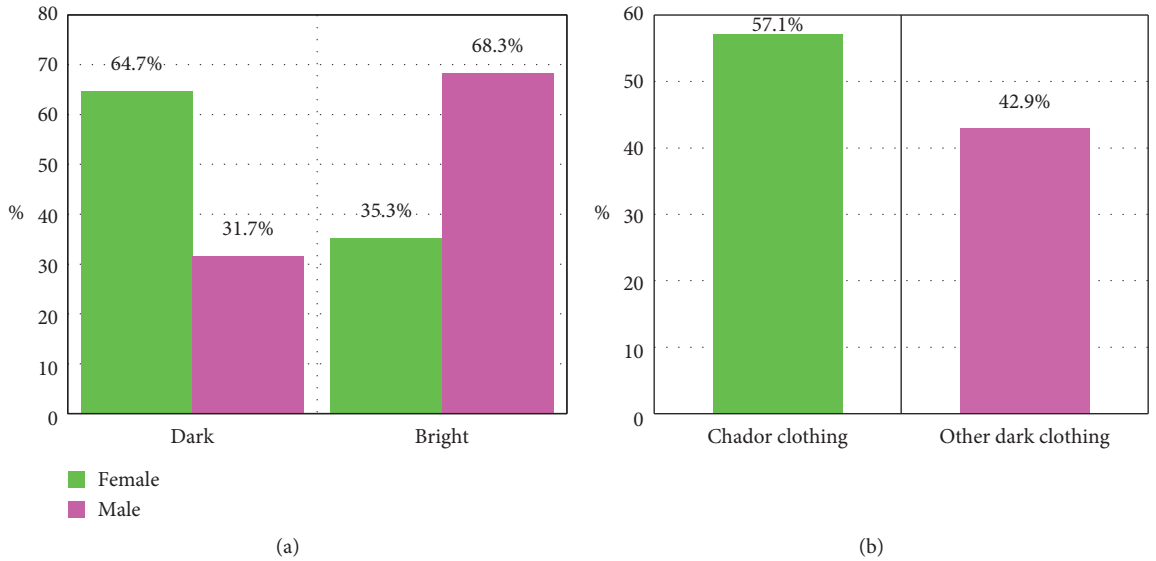


Figure 1: Pedestrian accident statistics based on: (a) pedestrians' clothing color, (b) dark clothing.

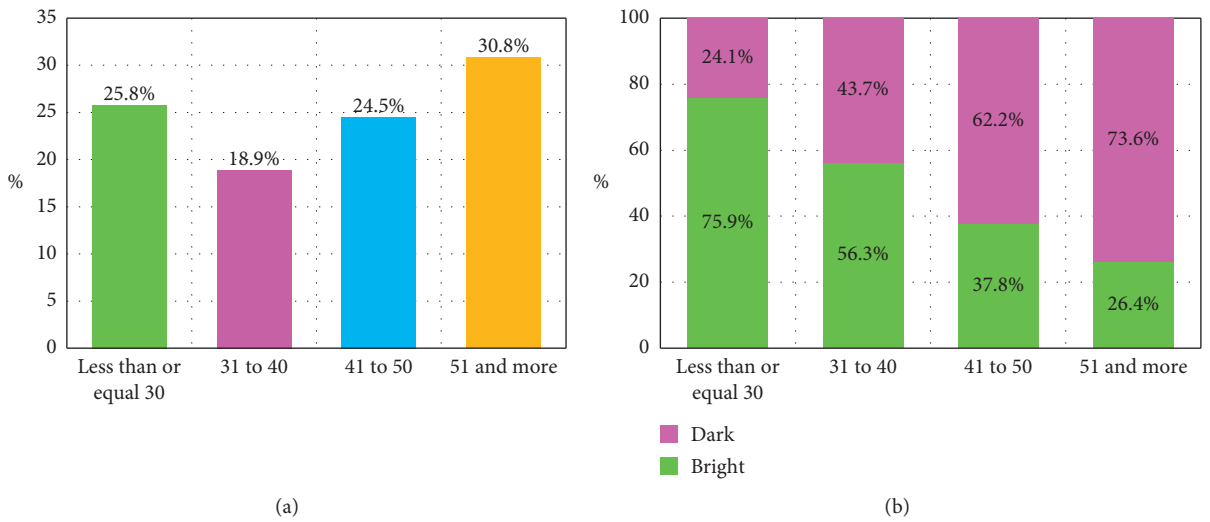


Figure 2: Female pedestrian accident statistics based on (a) age and (b) age and clothing color.

As shown in Figure 5(a), female pedestrians with dark clothing with a diploma had the highest accident rate (23.5%) and pedestrians with a doctorate had the lowest accident rate (0.3%), indicating lower awareness at the lower level of education with regard to being more cautious, and this is validated as the highest percentage of accidents occurred in the age group of 51 and above. Also, according to Figure 5(b), 34.9% of accidents involved housewives, accounting for the highest accident rate, indicating that most female pedestrians with dark clothing were housewives, and clerks had the lowest accident rates (13.8%), which indicates that they paid more attention to their surroundings.

According to Figure 6, 17.7% of female pedestrians crossed at the authorized locations, indicating that female pedestrians with dark clothing who used the allowed passages (68.2% of accidents) had more accident rates.

Also, as illustrated in Figure 7, 41.6% of female pedestrian accidents with dark clothing, the highest accident rate , occurred where there was a pedestrian bridge (Figure 7(a)), indicating that there were appropriate facilities for crossing. Also, the highest number of female pedestrian accidents occurred among those aged 51 and above, who preferred to cross the road due to difficulty in climbing the stairs of the pedestrian bridge. The lowest accident rate based on crossing facilities (6.7%) occurred where there was an underpass, indicating that the underpass had a decreasing effect on the severity of accidents. However, according to Figure 7(b), 62.9% of accidents were nonguilty pedestrians.

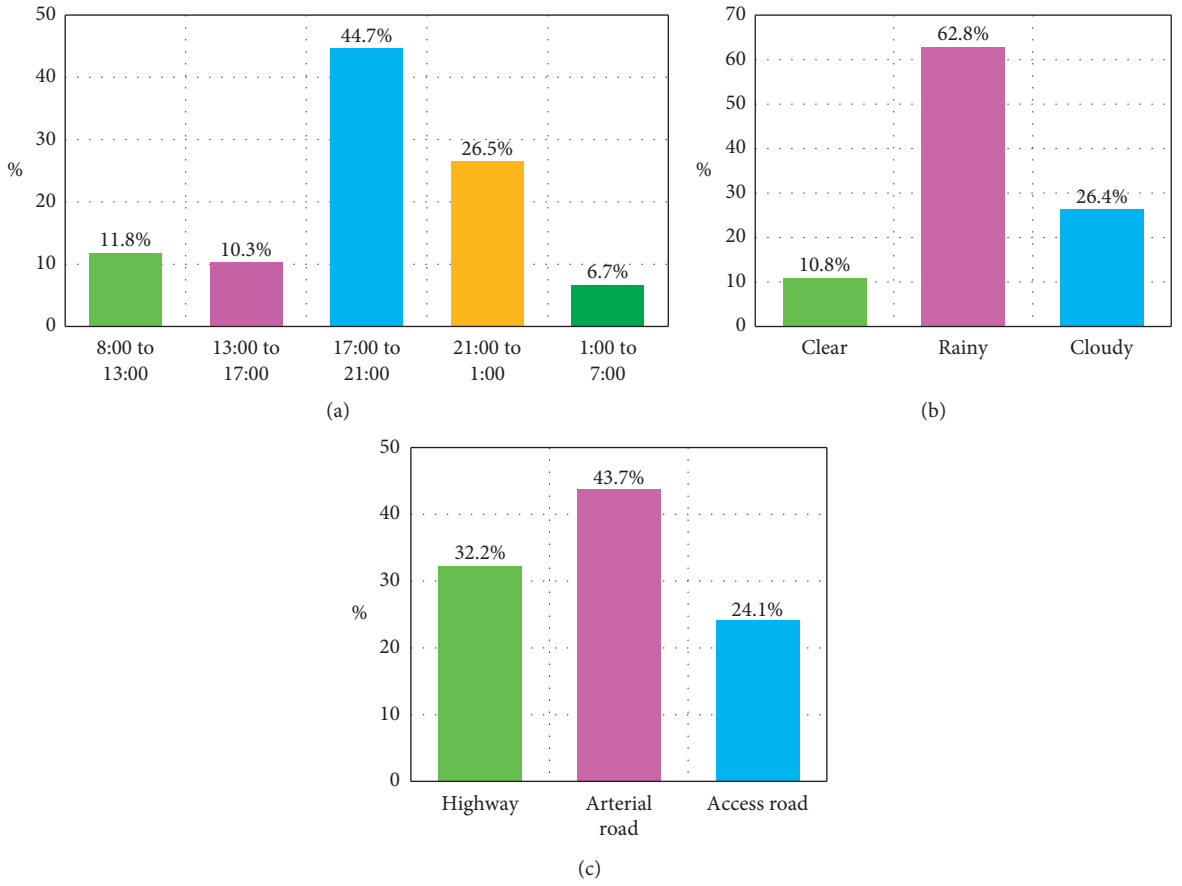


Figure 3: Female pedestrian accident statistics with dark clothing based on (a) accident time, (b) weather condition, and (c) road classification.

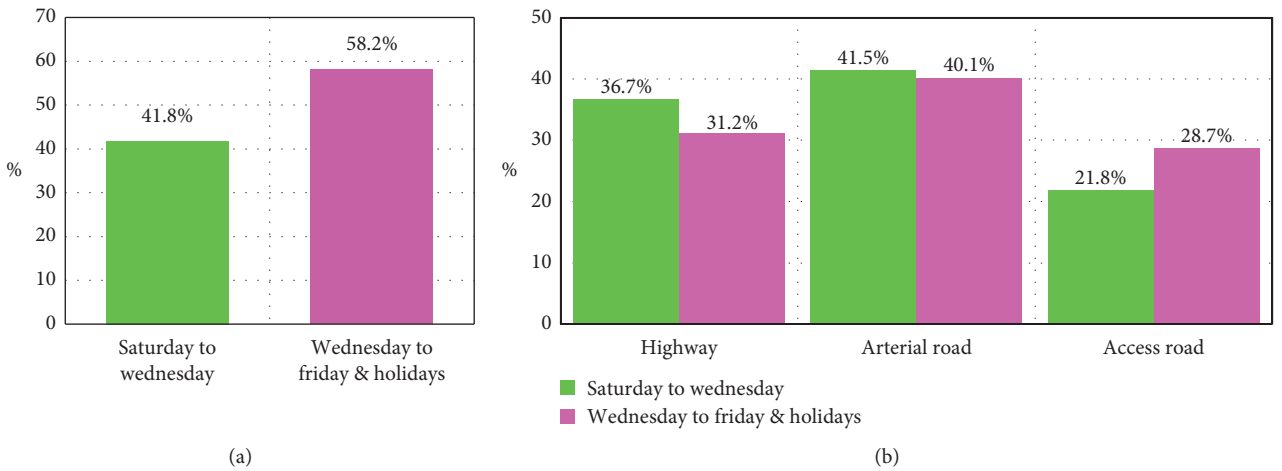


Figure 4: Female pedestrian accident statistics with dark clothing based on: (a) day, (b) road classification and day.

3.2. K–S Test Results. In order to evaluate the normality of data, the statistical distribution of data should be ensured, and the K–S test was applied to check the normality of distribution. Table 2 indicates the results of this test for dimensions.

As Table 2 indicates, the significance level in the K–S test was less than 0.05, and considering the 5% error, the null hypothesis (H_0) was rejected. So, the distribution was not normal and nonparametric tests were applied in this research.

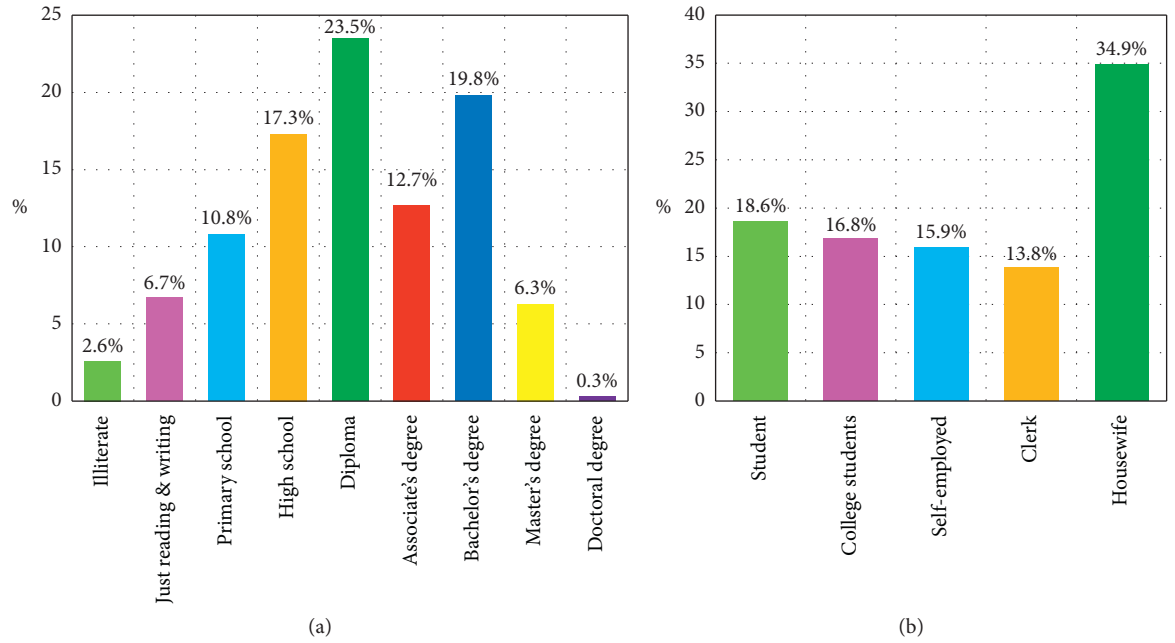


Figure 5: Female pedestrian accident statistics with dark clothing based on: (a) education, (b) job.

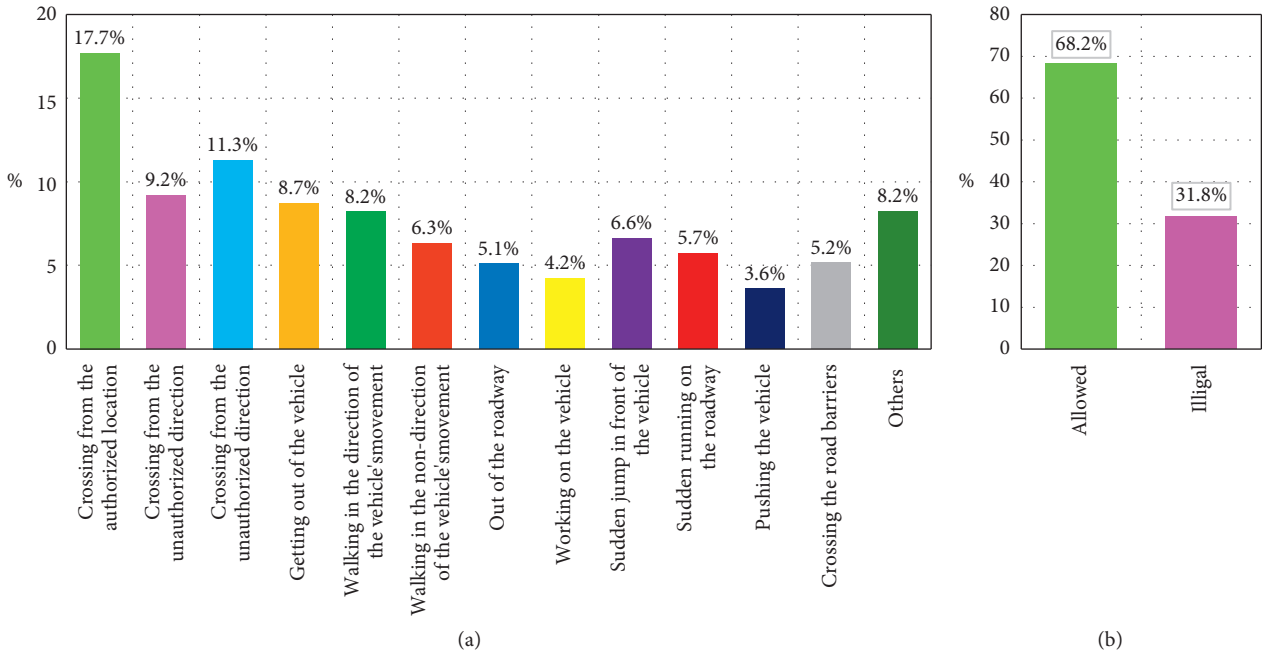


Figure 6: Female pedestrian accident statistics with dark clothing based on: (a) pedestrian action, (b) crossing permit.

3.3. FT Results. There were 12 independent variables in this study, the rank of which was examined by FT. Table 3 illustrates the statistical significance in which the sample volume, chi-square statistic, degrees of freedom, and significance level are presented.

As illustrated in Table 3, the significance level of less than 5% indicated that H_0 was rejected. Therefore, ratings were not uniform. In Table 4, the ranking of independent variables is presented. The smaller the mean rates, the higher the importance of the variable.

According to Table 4, the variables of clothing color, weather condition, and age had the highest rank, with the scores of 4.71, 5.82, and 6.05, respectively. On the other hand, the variables of job, day, and fault status were the least important factors in the occurrence of accidents. Results indicated that clothing color as a human factor had the

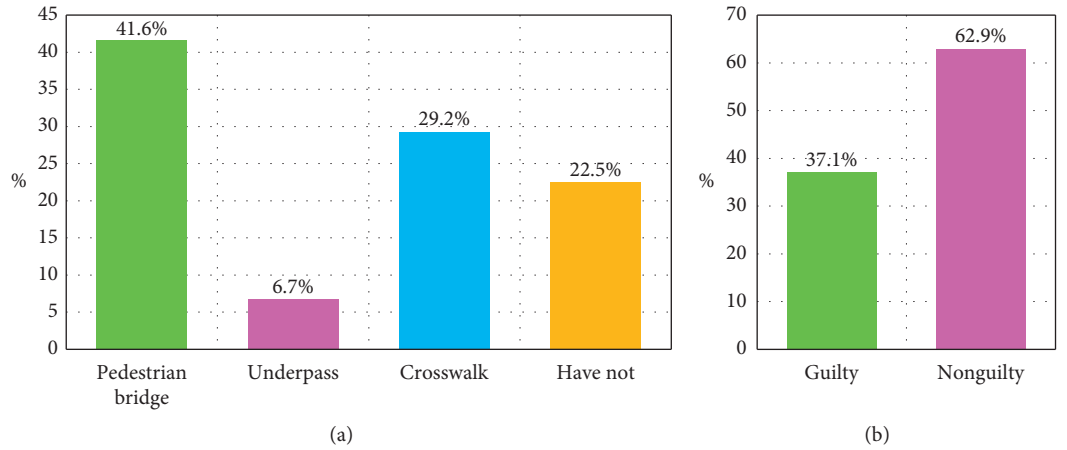


Figure 7: Female pedestrian accident statistics with dark clothing based on: (a) crossing facilities, (b) fault status.

Table 2: K-S test results.

Number	Variables	Most extreme differences			Test statistic	Asymp. Sig. (2-Tailed)
		Absolute	Positive	Negative		
1	Clothing color	0.524	0.524	−0.356	0.524	0.0
2	Age	0.472	0.472	−0.410	0.472	0.0
3	Accident time	0.177	0.143	−0.177	0.177	0.0
4	Day	0.378	0.378	−0.305	0.378	0.0
5	Weather condition	0.425	0.425	−0.357	0.425	0.0
6	Education	0.254	0.174	−0.254	0.254	0.0
7	Pedestrian action	0.387	0.387	−0.304	0.387	0.0
8	Crossing facilities	0.184	0.184	−0.172	0.184	0.0
9	Crossing permit	0.324	0.324	−0.356	0.324	0.0
10	Job	0.171	0.144	−0.171	0.171	0.0
11	Road classification	0.398	0.285	−0.398	0.398	0.0
12	Fault status	0.230	0.251	−0.230	0.230	0.0

Table 3: FT results.

Number of data	Chi-square	Degrees of freedom	Asymp. Sig.
6128	31458.703	11	0.0

greatest impact on the occurrence of female pedestrian accidents, followed by weather condition, which was illustrated as an environmental factor and the second most effective factor in increasing the occurrence of accidents. Age was identified as the third most influential factor in accidents.

3.4. FA Results. Table 5 illustrates the results of FA, indicating the KMO index and the Bartlett's test.

As indicated in previous studies, the closer the KMO index value is to one, the better the desired data for factor analysis. According to Table 4, the KMO index value was close to 0.60, indicating that it was appropriate to use FA. Moreover, similar results have been presented in previous studies [94]. Also, the significance level of the test statistic was less than 0.05, meaning that there was a significant correlation between the variables.

Table 6 indicates the eigenvalues and variance corresponding to each component. Whenever the eigenvalue of a component is higher, that component indicates more variance. In Table 6, initial eigenvalues, and eigenvalues of nonrotating and post-rotating extraction components are presented. Results indicated that the components 1 to 5 had eigenvalues greater than 1. So, they remained in the analysis.

Table 7 indicates the matrix of the remaining components. The values in Table 7 illustrate the correlation of each variable with the relevant components.

The higher the absolute value of the coefficients, the more effective the component was in the total changes of each variable [94]. FA was performed on 12 independent variables affecting female pedestrian accidents, which identified five main factors. The analysis showed that the variables of clothing color and weather condition were under the first component that the significant coefficients between the first component and each variable were 0.856 and 0.839

Table 4: Mean rank in FT.

Variables	Mean	Rank
Clothing color	4.71	1
Weather condition	5.82	2
Age	6.05	3
Accident time	6.56	4
Road classification	7.53	5
Pedestrian action	7.87	6
Education	7.99	7
Crossing permit	8.15	8
Crossing facilities	9.01	9
Job	10.15	10
Day	10.63	11
Fault status	11.07	12

Table 5: KMO and Bartlett's test.

KMO measure of sampling adequacy	Bartlett's test of sphericity		
	Approximate Chi-square	Df	Sig.
0.593	6153.749	66	0.0

and positive, respectively. Therefore, it can be said that clothing color and weather condition were the primary factors in the occurrence of female pedestrian accidents. Also, education, accident time, and age constituted the second most influential components in accidents. Accident time and age (with the coefficients of 0.728 and 0.596, respectively) positively affected the occurrence of accidents, and education (with a coefficient of −0.798) had a significant impact on reducing accidents. Similarly, road classification and pedestrian action variables constituted the third component (with coefficients of 0.737 and 0.612, respectively); day, crossing facilities, and job (with coefficients of 0.774, −0.607, and −0.528, respectively) constituted the fourth component; and crossing permit and fault status (with coefficients of −0.648 and 0.662, respectively) constituted the fifth most effective factors in female pedestrian accidents.

3.5. MLR Results. A binary model was required, as there were just two possible outcomes, namely, injury or fatal severity of accidents. For MLR modeling of the severity of female pedestrian accidents, 2 dependent variables and 55 independent sub-variables were defined and then modeled. Table 8 presents a summary of three methods of entering, forward stepwise, and backward stepwise in the MLR model.

According to Table 8, in the entering method, as all the variables entered the equation at the same time, the model did not have the opportunity to process the data properly, and therefore it was not a suitable method. So, the forward and backward stepwise methods were used. Table 8 includes two criteria for the correct percentage and goodness of fit of the model. The goodness of fit (0.648) and the correct percentage (86.3%) of the backward method were higher than the other two methods, so this method was used for the logistic model. In Table 9, the chi-square statistic, degree of freedom, and significance level of test statistic in step 20 are presented.

Results of Table 9 indicated that the amount of chi-square statistic was 1828.200. Also, the significance level of the model was less than 0.05, indicating the suitability of the fit of the MLR model. Table 10 illustrates the accuracy of the model in predicting accident severity and is used for evaluating the performance of the prediction model and, in other words, the accuracy of accident severity classification.

Table 10 indicates that out of 5268 injury accidents, 5034 cases and out of 860 fatal accidents, 252 cases were predicted correctly by the model. The prediction accuracy of the MLR model was 95.6% for injury accidents and 29.3% for fatal accidents. So, the accuracy of the MLR model in the classification of injury accidents was more than fatal accidents. Also, the accuracy of the whole model in determining the severity of accidents was 86.3%, meaning that with a confidence of 86.3%, the MLR model was able to predict the severity of accidents.

In order to model the severity of accidents, 55 independent sub-variables were included in the MLR model and finally, 12 variables were determined as the effective parameters in accidents. In Table 11, the results of the model are represented for step 20. Since the significance level was below 5%, the null hypothesis was rejected and indicated the significance of the MNR model. When β_i is positive, as variable x_i increases, the accident probability will be raised. In contrast, if β_i is negative, an increase in x_i will decrease the accident probability [94, 95].

Results of Table 11 indicated that the variables of age 51 and above, chador clothing, other dark clothing, 17 : 00 to 21 : 00 accident time, Wednesday to Friday and holidays (nonworking days), sudden jump in front of the vehicle, rainy weather, and arterial road increased the probability of female pedestrian accidents. Moreover, the variables of 1 : 00 to 7 : 00 accident hours, master's degree, walking in the nondirection of the vehicle's movement, and allowed to cross reduced the likelihood of female pedestrian accidents.

Table 6: Total variance explained.

Components	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	2.559	21.329	21.329	2.559	21.329	21.329	2.524	21.035	21.035
2	1.659	13.828	35.157	1.659	13.828	35.157	1.616	13.466	34.500
3	1.230	10.252	45.408	1.230	10.252	45.408	1.198	9.983	44.483
4	1.114	9.281	54.689	1.114	9.281	54.689	1.155	9.621	54.104
5	1.079	8.989	63.678	1.079	8.989	63.678	1.149	9.574	63.678
6	0.988	8.232	71.910						
7	0.922	7.681	79.591						
8	0.806	6.719	86.310						
9	0.784	6.534	92.844						
10	0.402	3.347	96.191						
11	0.329	2.745	98.936						
12	0.128	1.064	100.000						

Table 7: Component matrix.

Variables	Components				
	1	2	3	4	5
Clothing color	0.856	−0.055	−0.068	0.073	−0.038
Age	0.165	0.596	−0.028	0.240	0.359
Accident time	−0.045	0.728	0.061	0.036	0.006
Day	0.013	0.143	−0.041	0.774	−0.027
Weather condition	0.839	−0.006	0.017	−0.074	−0.041
Education	0.018	−0.798	0.013	−0.054	0.157
Pedestrian action	−0.036	0.078	0.612	0.048	−0.341
Crossing facilities	0.165	0.049	0.089	−0.607	0.024
Crossing permit	−0.118	−0.142	0.006	0.096	−0.648
Job	0.145	0.229	−0.073	−0.528	0.064
Road classification	0.023	−0.135	0.737	−0.387	−0.118
Fault status	0.152	0.195	−0.013	−0.103	−0.662

One of the most important results of the MNR model was that the chador clothing had the greatest effect on increasing the probability of female pedestrian accidents, followed by other dark clothing. Therefore, bright clothing color had a direct effect on reducing female pedestrian accidents, which was more important at night and in rainy weather (due to reduced driver visibility), so rainy weather as the third variable had the greatest impact on increasing the accident rate. The model also showed that the interactive effect of the darkness of the air and the darkness of the female pedestrian clothing increased the likelihood of accidents. This increase can be attributed to the reduced visibility of the pedestrian by the driver, which increases the likelihood of the pedestrian being killed in an accident. This becomes more important in arterial roads (due to the higher speed of the vehicles and the presence of pedestrians more than the highways), which had an increasing impact on the occurrence of accidents. Another important result was that in rainy weather, especially between 5 pm and 9 pm, female pedestrians aged 51 and above were more prone to accidents. Higher education levels also had a great impact on reducing female pedestrian accidents.

3.6. ANN Results

3.6.1. The MLP Model. The MLP was applied to construct a prediction model that predicts one or more dependent variables (targets). In this research, in order to segment the data randomly, the model was constructed with 70% training data, and its accuracy was evaluated based on the remaining 30%. The number of input parameters to the model was 12. The number of input layer neurons was 59, and the number of output layer neurons was 2. The automatic selection structure selected 10 neurons in the hidden layer. One hidden layer was used because it is known to have sufficient approximating power [96]. Table 12 indicates the correct percentage of each sample. The cells in the diagonal line in each case represent the correct predictions.

Table 12 shows that 3586 of 3703 accidents were correctly classified as injury accidents and 478 of 594 accidents were correctly predicted as fatal accidents. Moreover, in 94.6% of cases, the training sample was correctly classified. So, the accuracy of the MLP model was 94.6%. Also, cross-entropy error in training and testing samples was 619.025 and 328.277, respectively.

Table 8: Summary of the three methods of entering, forward stepwise, and backward stepwise of the severity of accidents.

Number	Logistic regression methods	Correct percentage	Goodness of fit (R^2)
1	Enter	81.5	0.626
2	Forward stepwise	83.4	0.517
3	Backward stepwise	86.3	0.648

Table 9: Backward stepwise model results.

		Chi-square	Df	Sig.
Step 20	Step	173.028	1	0.000
	Block	1828.200	29	0.000
	Model	1828.200	29	0.000

Table 10: Classification table.

Observed		Predicted		
		Accident severity		Correct percentage
		Injury	Fatal	
Accident severity	Injury	5034	234	95.6
	Fatal	608	252	29.3
Overall percentage		86.3		

Figure 8 shows an intuitive representation of the sensitivity to specificity for all levels of classification and illustrates two curves for injury and fatal accidents. Since there were only two categories, the curves were symmetric to the 45° line. The greater the curve inclined to up and left, the network power was more accurate in prediction [97].

In binary classification problems, the area under the curve (AUC) is commonly applied to examine the performance of a prediction model [98]. As the AUC is a portion of the unit square area, the AUC value is always between zero and one. The area under the diagonal line in the ROC curve is always 0.5. So, the AUC should be more than 0.5 [99]. According to Table 13, the probability of injury or fatal accidents was measured as 0.948, and because $0.948 > 0.5$, the response from the model was positive.

Figure 9 shows the extent to which the predicted amounts of the network change with the change in the amounts of independent variables. Normalization of this significance was obtained by dividing the importance values by their largest value and is expressed as a percentage. In Figure 9, the variables of clothing color, weather condition, education, and accident time had the greatest effect on the severity of female pedestrian accidents, respectively. This model showed that clothing color had the greatest impact on female pedestrian accidents.

3.6.2. The RBF Model. In order to model the RBF method, 70% of data were used to train the network and 30% of data were used for testing. The number of input layer neurons was 60, and the number of output layer neurons was 2. The automatic selection structure placed 9 neurons in the hidden layer. Table 14 shows the accuracy of predicting injury and fatal accidents.

Results of Table 14 indicated that the RBF network worked correctly in 86% of the cases. This model was successful in recognizing accidents leading to injuries. Also, the sum of squared error in training and testing samples of this method were 371.608 and 175.846, respectively. Moreover, Figure 10 shows the ROC curve for injury and fatal accidents. The closer the curve to the northwest, the more accurate the prediction of injury or fatal accidents [99].

According to the results of the AUC values presented in Table 15, the probability of an accident resulting in injury or fatalilty was calculated as 0.866, and since $0.866 > 0.5$, the response was positive by the model.

In Figure 11, the relative importance of each of the independent variables on the dependent variable is shown, which illustrates the extent to which these variables have an impact in predicting accidents. Figure 11 shows that the most influential variable on the severity of female pedestrian accidents in the RBF model was clothing color. Then, weather condition and accident time had the most effect on the prediction, respectively.

3.7. Comparison of Analysis and Modeling Methods Results. In this research, various methods were used to determine more accurately the variables affecting the severity of female pedestrian accidents. Each method investigated the effective variables by its specific algorithm. In FT, the variables of clothing color, weather condition, and age were ranked from the first to third rankings that caused accidents, respectively, while in FA, the variables of clothing color and weather condition were identified as the first effective components and education, accident time, and age were recognized as the second effective components in female pedestrian accidents. This indicates that in FT and FA, clothing color and weather condition were identified as the most important parameters affecting accidents.

Comparing the correct prediction percentage in the MLR and the ANN models, the results indicated that the

Table 11: MLR results in the twentieth step.

Variables	β	S.E.	Wald	Sig.	Exp (β)
Age 51 and above	3.405	0.889	14.670	0.012	30.114
Chador clothing	4.624	1.578	8.587	0.023	101.901
Other dark clothing	4.125	1.287	10.273	0.003	61.868
1 : 00 to 7 : 00	−1.658	0.861	3.708	0.001	0.191
17 : 00 to 21 : 00	3.074	1.026	8.977	0.025	21.628
Wednesday to Friday and holidays	1.493	1.070	1.947	0.003	4.450
Master's degree	−2.786	0.757	13.545	0.000	0.062
Sudden jump in front of the vehicle	3.603	1.116	10.423	0.000	36.708
Walking in the nondirection of the vehicle's movement	−2.261	0.880	6.601	0.001	0.104
Allowed to cross	−1.098	0.855	1.649	0.016	0.334
Rainy weather	4.007	1.131	12.552	0.000	54.982
Arterial road	2.867	0.779	13.545	0.000	17.584
Constant	5.260	2.361	4.963	0.000	192.481

Table 12: Classification in the MLP model.

Sample	Observed	Predicted		
		Injury	Fatal	Correct percentage
Training	Injury	3586	117	96.8
	Fatal	116	478	80.5
	Overall percentage	86.2%	13.8%	94.6
Testing	Injury	1504	61	96.1
	Fatal	67	199	74.8
	Overall percentage	85.8%	14.2%	93.0

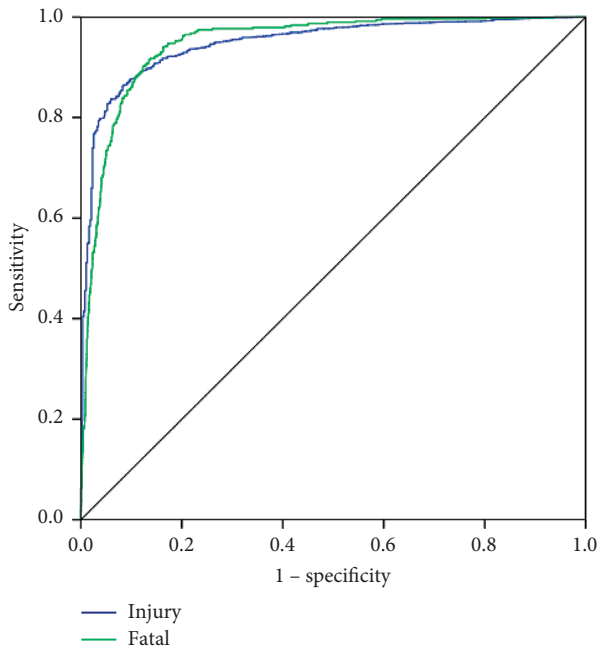


Figure 8: Receiver operating characteristic (ROC) curve of the MLP model.

Table 13: AUC for the MLP model.

		Area
Accident severity	Injury	0.948
	Fatal	0.948

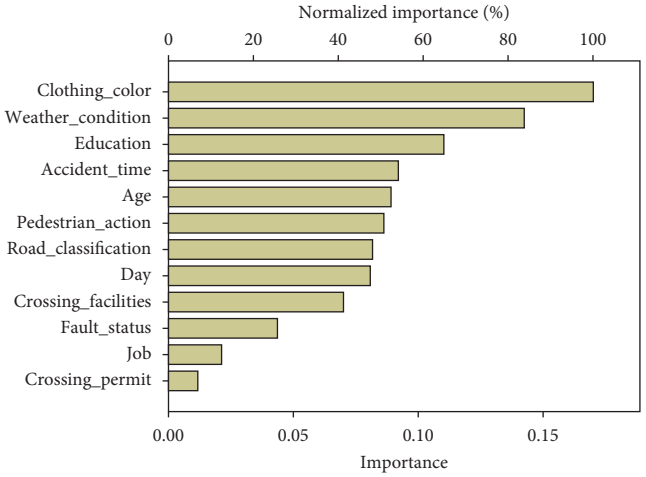


Figure 9: Independent variable importance in the MLP model.

Table 14: Classification in the RBF model.

Sample	Observed	Predicted		
		Injury	Fatal	Correct percentage (%)
Training	Injury	3631	101	97.3
	Fatal	506	103	16.9
	Overall percentage	95.3%	4.7%	86.0
Testing	Injury	1504	32	97.9
	Fatal	203	48	19.1
	Overall percentage	95.5%	4.5%	86.8

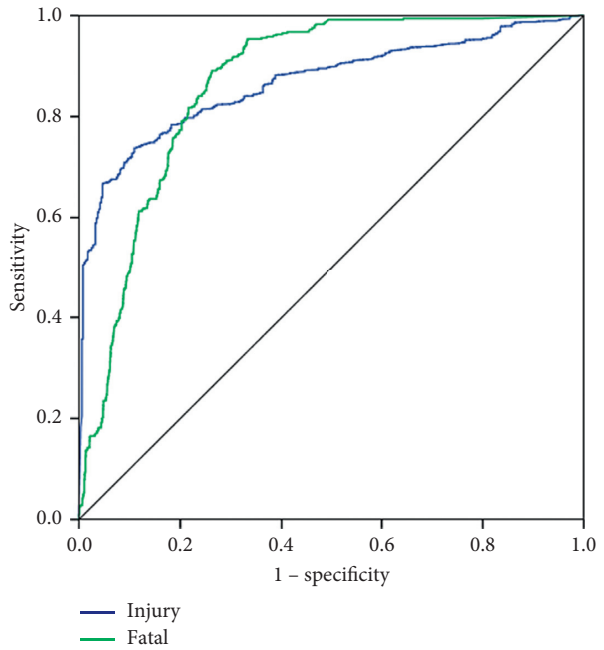


Figure 10: . TheROC curve of the RBF model.

MLP model performed better and had a higher prediction percentage than the rest. The prediction accuracy of the MLR model was 86.3%, the prediction accuracy of the MLP model was 94.6%, and the prediction accuracy of the RBF was 86.0%. On the other hand, the prediction accuracy of the MLR model was more than RBF. The MLR model showed

Table 15: The AUC for the RFB model.

		Area
Accident severity	Injury	0.866
	Fatal	0.866

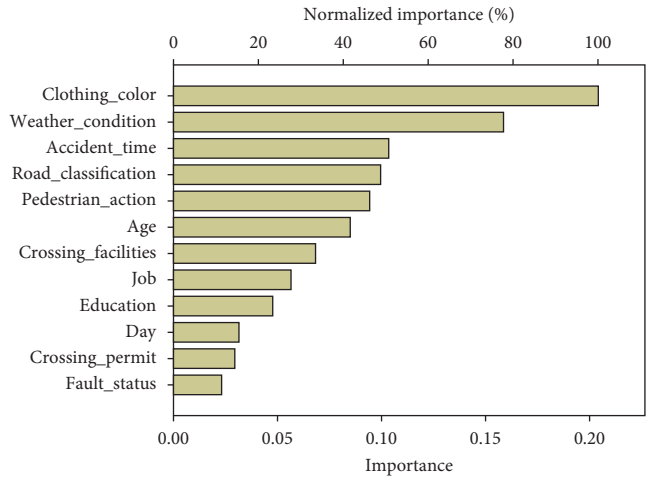


Figure 11: Independent variable importance in the RBF model.

that chador clothing and other dark clothing, respectively, had the greatest impact on increasing the probability of female pedestrian accidents, and then rainy weather had the highest effect on the occurrence of accidents. Higher education level also had a great impact on reducing accidents in the MLR model, while in the MLP model, due to better prediction, clothing color, weather condition, and education were the most effective factors in the occurrence of accidents. Also, in the RBF model, clothing color, weather condition, and accident time were the most effective factors. Consequently, clothing color and weather condition were the joint result of these three models of accident prediction as the most effective predictor.

The common point of analyses and models performed in this study was the clothing color of female pedestrians as the most important variable affecting the severity of accidents, especially at night. According to the MLR model, chador clothing and dark clothing color increased the probability of accident occurrence. In fact, the most dangerous situation for pedestrians was walking in the direction of the vehicle's movement at night, as it was very difficult for drivers to see a pedestrian at night. According to surveys conducted in this study, 64.7% of female pedestrians who were involved in an accident with vehicles at night had dark clothing.

The role of clothing color has also been highlighted in a previous study (Turner et al. [15]), which, different from this research, found that fluorescent red-orange-colored vests had the highest mean detection distance. But, in this research, bright clothing color had an important role in reducing the severity of accidents. This research showed that dark clothing color increased the severity of accidents. On the contrary, according to Clifton et al. [16], pedestrians with dark clothing had a lower rate of accidents. These findings may be due to the fact that the majority of accidents, almost 70%, happened in the daytime, while in this research, about 22.1% of accidents occurred during daylight hours. But in compliance with Pour-Rouholamin and Zhou's [22] investigations, pedestrians not wearing contrasting clothing and pedestrians with dark clothing were more susceptible to accidents. On the other hand, in a study conducted in Iran by Besharati and Tavakoli Kashani [24], low light conditions and dark clothing of pedestrians were factors that increased the likelihood of accidents, which confirms the results of this study. So, according to Baireddy et al.'s [25] study, wearing contrasting or reflective clothing can reduce the occurrence of accidents and make pedestrians less prone to accidents, especially at night.

Various previous studies investigated the conspicuity of pedestrians that the results of the study confirm them [17–22]. The decreased ability of a driver to observe a pedestrian during nighttime is a contributing factor to many pedestrian accident types. This article is consistent with Tyrrell et al. [17], in which clothing color and pedestrian action influenced conspicuity, and pedestrians who wore black clothing, especially at night, reduced the conspicuity. According to Zegeer and Bushell [18] and Hefny et al. [21], similar to the results of this study, nighttime or low lighting conditions increased the risk of accidents. So pedestrians, especially in dark clothing, need to be more aware of the fact that they may be invisible to drivers when walking in the dark; thus, it is important for pedestrians to carry a flashlight or wear retro-reflective clothing/material or light-colored clothing at night. So, according to Fekety et al.'s [23] research, adding electroluminescent (EL) panels to pedestrian clothing will increase the conspicuity of pedestrians, and this is especially important for female pedestrians, especially for

females with chador clothing at night, so chador clothing should be produced with retro-reflective materials or EL panels.

In accordance with Wood et al. [19] and Borzendowski et al.'s [20] study, pedestrian clothing had a significant impact to be recognized by drivers, especially at night, and wearing reflective clothing reduced the likelihood of accidents. Also, in accordance with previous studies of the visibility distance from the pedestrian (Wickramasinghe et al. [26]), results showed that wearing white clothing color was safer at nighttime while pedestrians wearing dark clothing color had lower visibility in the night.

The results of this study showed that 77.9% of female accidents with dark clothing occurred at night or sunset, which is in contrast to a research by Abdous and Mahmoudabadi [27] that showed that the vulnerability of female pedestrians at different times of the day is not significantly different, but the most important results of this study in common with Abdous and Mahmoudabadi's [27] results were that wearing chador clothing significantly increased their vulnerability.

4. Conclusion and Future Directions

The purpose of this study was to provide useful and practical solutions to improve the safety of female pedestrians using different statistical methods and intelligent models using police data collected from March 2017 to March 2019 to analyze the severity of female pedestrian accidents. The most important results can be mentioned:

(1) In general, female pedestrians had darker clothing than males, most of whom had chador clothing. Moreover, most of the female pedestrian accidents with dark clothing occurred between 17: :00 and 21: :00 on Wednesday to Friday and holidays, especially on the arterial roads, and in the conditions of rainy weather. Most of these accidents occurred while crossing at authorized locations, and involved female pedestrians who had a diploma and were housewives. Most accidents occurred in the presence of pedestrian bridge facilities, which indicates there were appropriate facilities for crossing the pedestrians. On the other hand, most of the female pedestrians involved were nonguilty and aged 51 and above.

(2) Based on FT, the most important factors affecting female pedestrian accidents were identified as clothing color, weather condition, and age, indicating that the most influential factor in female pedestrian accidents was clothing color as a human factor and the second most important factor was weather condition as an environmental factor affecting the occurrence of accidents.

(3) FA results showed that the variables of clothing color and weather condition constituted the first component, indicating that these two variables were the most effective factors in the occurrence of female pedestrian accidents that had a large impact (due to the coefficient value) on the occurrence of accidents. Education, accident time, and age were also known to be the second influencing component on accidents, in which accident time and age positively and increasingly affected the occurrence of accidents, and education level had a significant effect on reducing accidents. Likewise, the variables of road classification and pedestrian action constituted the third component, day of the accident, crossing facilities, and job were constituted the fourth component, and, ultimately, fault status and crossing permit constituted the fifth effective component in female pedestrian accidents. According to the results of this analysis, the most effective factor in reducing the incidence of accidents was education. The most important result of FA was that clothing color had a significant effect on increasing the incidence of accidents. Due to the high percentage of female pedestrians with dark clothing (64.7%), they had the highest accident rate, especially at night.

(4) In the MLR model, the effect of independent variables on the severity of accidents was investigated and the results showed that the accuracy of the model in predicting accidents in the backward stepwise method was 86.3%. According to the results, wearing dark clothing was an important factor in increasing accidents. Females in Iran, due to wearing a veil and the tendency to wear dark clothing (chador and other dark clothing), have a significant influence on the incidence of accidents. As illustrated in this study, it is especially critical at 17 : :00 to 21 : :00 because of less clear clothing. Traffic reached its peak between 17: :00 and 21: :00, indicating an increase in traffic accidents' rate due to the increased number of vehicles, especially on the arterial roads and highways, and the presence of more pedestrians across the city. Females aged 51 and above were more likely to be involved in accidents due to a lower walking speed and, more importantly, dark clothing (73.6% of females were 51 years and older). In rainy weather, the risk of pedestrian accidents was increased due to the reduced visibility of the pedestrian by the driver relative to the environment and the difficulty in controlling the vehicle when exposed to danger, especially at higher speeds. Whereas in clear weather conditions, the driver's view from the environment was greater and the likelihood of an accident was reduced. The results of accidents related to a sudden jump in front of the vehicle indicated that the severity of pedestrian accidents was directly related to the personality characteristics of pedestrians in addition to environmental and road conditions. In arterial roads, because of the higher speed of vehicles

and the presence of more females on these roads and lower resolution of dark clothing than other roads, pedestrians were more likely to be involved in an accident. Accidents were reduced between 1 : :00 and 7 : :00 due to less vehicle attendance. One of the most important results in this model was that the chador clothing and other dark clothing had the most impact on increasing the probability of female pedestrian accidents, and rainy weather had the most impact on increasing the accident rate. Higher level of education also had a great impact on reducing pedestrian accidents. Therefore, bright clothing color had a direct impact on reducing accidents. This made more sense at night and in rainy weather (due to reduced driver visibility).

(5) In the ANN models, the results of the methods used showed that the MLP was more accurate than the RBF, indicating that the MLP model was better at predicting the severity of accidents. Sensitivity analysis of the two methods showed that the MLP had more power to predict and estimate accidents than the RBF. In the MLP model, the variables of clothing color, weather condition, education, and accident time had the most impact on the severity of accidents, while in the RBF model, the variables of clothing color, weather condition, accident time, and road classification had the highest impact, respectively. But in general, the ANN results in both models showed that the clothing color variable had the most effect on the prediction of female pedestrian accidents, and considering the MLR model, dark clothing color had the most influence on the occurrence of the accidents. Therefore, females who wore chador and other dark clothing were more prone to accidents, which necessitate the adoption of appropriate solutions.

(6) For future work and analysis, we will deal with some state-of-the-art machine learning and optimization algorithms to be incorporated with the proposed approach to attain more accuracy. Moreover, we can implement the deep learning approaches to reach a better prediction percentage.

5. Safety Approach

Based on the results of the analyses and models of the research, the following measures are suggested to reduce the probability of injury and fatal accidents of female pedestrians:

(1) Due to the high percentage of female pedestrian accidents with dark clothing color, increasing public awareness of the use of bright clothing color for female pedestrians is one of the most important strategies to reduce the severity of accidents at night. In addition to advertising on the radio, television, and urban advertising, police, municipal, forensic, and health organizations are recommended to gift luminous stickers to females wearing dark clothing. Chain stores also could give these tags as an advertisement. On the other hand, chador and dark clothing can be produced, which are reflective or luminous at night, like worker clothing. The use of high-reflectance and luminous clothing and the provision of adequate street lighting and overhead lighting at pedestrian crossings, especially on arterial roads, are measures that can reduce the likelihood of accidents.

(2) For road markings like stop lines, crosswalks, and lane markings, retro-reflective substances should be utilized. By these substances, the light is reflected from vehicle headlights and from roadway illumination through glass beads with a particular design. Vests and other clothing for pedestrians should also be created with reflective substances. Investigations have indicated that the visibility of a pedestrian can be incremented by reflectorizing by a factor of 5. The retro-reflective material should be utilized on backpacks, jackets, shoes, and other clothing.

(3) In rainy weather, pedestrians are more likely to be involved in accidents due to insufficient visibility of drivers, so pedestrians need to be aware of the possibility of a vehicle losing control and harming them on rainy days. Installing pedestrian flashing lights and pedestrian signs beside the roads, removing visual error factors (such as removing natural and abnormal obstacles and creating a good visual field) are engineering solutions to reduce the likelihood of accidents. Also, the use of an intelligent pedestrian alert system for approaching drivers (by pressing a special button by the pedestrian and transmitting the message to the LED signs in any direction of the approaching traffic) is recommended to enhance the safety of female passengers, especially at night, in arterial roads and during rainy weather.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 3 months after publication of this article, will be considered by the corresponding author.

Disclosure

In this study, Iranian governmental organizations have not been partners and sponsors, and this study is purely for study purposes.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] I. Bargegol, V. N. M. Gilani, and F. Jamshidpour, "Relationship between pedestrians' speed, density and flow rate of crossings through urban intersections (case study: rasht metropolis)," *International Journal of Engineering-Transactions C: Aspects*, vol. 30, no. 12, pp. 1814–1821, 2017.

[2] Movahhed and M. Bagheri, "The effect of rain on pedestrians crossing speed," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 3, 2020.

[3] Hikoui, H. Farhadi, and A. Mahmodzadeh, "Estimated total stadium emergency evacuation's time based on existing population analysis," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 4, no. 2, 2018.

[4] I. Bargegol, V. Najafi Moghaddam Gilani, M. Ghasedi, and M. Ghorbanzadeh, "Delay modeling of un-signalized roundabouts using neural network and regression," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 1, 2016.

[5] I. Bargegol and M. Keyoumarsi, "An investigation on pedestrian accident severity in urban streets using path analysis," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 2, 2016.

[6] S. M. Hosseinian and V. Najafi Moghaddam Gilani, "Analysis of factors affecting urban road accidents in rasht metropolis," *ENG Transactions*, vol. 1, no. 1, pp. 1–4, 2020.

[7] S. M. Hosseinian, V. Najafi Moghaddam Gilani, P. Mehraban, and M. Arabani, "Investigation of moisture sensitivity and conductivity properties of inductive asphalt mixtures containing steel wool fiber," *Advances in Civil Engineering*, vol. 2020, Article ID 8890814, 2020.

[8] M. J. Samet, "Development of accident modification factors in two-lane highways," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 2, no. 4, 2016.

[9] N. Nadimia, S. S. Naseralavi, A. M. Amiri, and A. Z. Mirhosseini, "Proposing a model for ranking hotspots in rural roads using a multi-criteria decision-making method," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 3, 2017.

[10] Traffic police statistics and information center, Tehran, Iran, 2018 and 2019.

[11] M. Abolfazlzadeh and B. E. Gol, "Statistical analysis of the railway accidents causes in Iran," *International Journal of Engineering*, vol. 30, no. 12, pp. 1822–1830, 2017.

[12] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Investigation of effective factors in the severity of rural road accidents in guilan to determine the most effective factors and provide safety solutions," *Road*, vol. 29, no. 106, pp. 115–128, 2021.

[13] M. Nikookar, M. B. Movahhed, J. Ayoubinejad, V. Najafi Moghaddam Gilani, and S. M. Hosseinian, "Improving the moisture sensitivity of asphalt mixtures by simultaneous modification of asphalt binder and aggregates with carbon nanofiber and carbon nanotube," *Advances in Civil Engineering*, vol. 2021, Article ID 6682856, 2021.

[14] H. Behbahani and V. Najafi Moghaddam Gilani, "Analysis of crossing speed of the pedestrians in marked and unmarked crosswalks in the signalized and un-signalized intersections (case study: Rasht city)," *IOP Conference Series: Materials Science and Engineering*, vol. 245, no. 4, 2017.

[15] J. D. Turner, C. J. Simmons, and J. R. Graham, "High-visibility clothing for daytime use in work zones," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1585, no. 1, pp. 1–8, 1997.

[16] K. J. Clifton, C. V. Burnier, and G. Akar, "Severity of injury resulting from pedestrian-vehicle crashes: what can we learn from examining the built environment?" *Transportation Research Part D: Transport and Environment*, vol. 14, no. 6, pp. 425–436, 2009.

[17] R. A. Tyrrell, J. M. Wood, A. Chaparro, T. P. Carberry, B.-S. Chu, and R. P. Marszalek, "Seeing pedestrians at night: visual clutter does not mask biological motion," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. 506–512, 2009.

[18] C. V. Zegeer and M. Bushell, "Pedestrian crash trends and potential countermeasures from around the world," *Accident Analysis & Prevention*, vol. 44, no. 1, pp. 3–11, 2012.

[19] J. M. Wood, P. Lacherez, and R. A. Tyrrell, "Seeing pedestrians at night: effect of driver age and visual abilities," *Ophthalmic and Physiological Optics*, vol. 34, no. 4, pp. 452–458, 2014.

[20] S. A. W. Borzendowski, A. A. S. Sewall, P. J. Rosopa, and R. A. Tyrrell, "Drivers' judgments of the effect of headlight glare on their ability to see pedestrians at night," *Journal of Safety Research*, vol. 53, pp. 31–37, 2015.

[21] A. F. Hefny, H. O. Eid, and F. M. Abu-Zidan, "Pedestrian injuries in the United Arab Emirates," *International Journal of Injury Control and Safety Promotion*, vol. 22, no. 3, pp. 203–208, 2015.

[22] M. Pour-Rouholamin and H. Zhou, "Investigating the risk factors associated with pedestrian injury severity in Illinois," *Journal of Safety Research*, vol. 57, pp. 9–17, 2016.

[23] D. K. Fekety, D. E. Edewaard, A. A. Stafford Sewall, and R. A. Tyrrell, "Electroluminescent materials can further enhance the nighttime conspicuity of pedestrians wearing retroreflective materials," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 58, no. 7, pp. 976–985, 2016.

[24] M. M. Besharati and A. Tavakoli Kashani, "Which set of factors contribute to increase the likelihood of pedestrian fatality in road crashes?" *International Journal of Injury Control and Safety Promotion*, vol. 25, no. 3, pp. 247–256, 2018.

[25] R. Baireddy, H. Zhou, and M. Jalayer, "Multiple correspondence analysis of pedestrian crashes in rural Illinois," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2672, no. 38, pp. 116–127, 2018.

[26] V. Wickramasinghe, J. Edirisinghe, and S. Dissanayake, *PW 0803 Night Time Visibility of Crossing Pedestrians*, BMJ Publishing Group Ltd., London, UK, 2018.

[27] H. Abdous and A. Mahmoudabadi, *PW 0580 Vulnerability of Female Pedestrians in Different Crash Locations, A Case-Study for iran*, BMJ Publishing Group Ltd., London, UK, 2018.

[28] H. Yu, W. Li, and C. Chen, "Dynamic Gaussian bare-bones fruit fly optimizers with abandonment mechanism: method and analysis," *Engineering with Computers*, vol. 2020, pp. 1–29, 2020.

[29] J. Hu, H. Chen, A. A. Heidari et al., "Orthogonal learning covariance matrix for defects of grey wolf optimizer: insights, balance, diversity, and feature selection," *Knowledge-Based Systems*, vol. 213, Article ID 106684, 2020.

[30] C. Yu, M. Chen, and K. Cheng, "SGOA: annealing-behaved grasshopper optimizer for global tasks," *Engineering with Computers*, vol. 2021, pp. 1–28, 2021.

[31] W. Shan, "Double adaptive weights for stabilization of moth flame optimizer: balance analysis, engineering cases, and medical diagnosis," *Knowledge-Based Systems*, vol. 214, Article ID 106728, 2020.

[32] J. Tu, H. Chen, J. Liu et al., "Evolutionary biogeography-based whale optimization methods with communication structure: towards measuring the balance," *Knowledge-Based Systems*, vol. 212, Article ID 106642, 2021.

[33] D. Zhao, L. Liu, F. Yu et al., "Chaotic random spare ant colony optimization for multi-threshold image segmentation of 2D Kapur entropy," *Knowledge-Based Systems*, vol. 216, Article ID 106510, 2020.

[34] H. Chen, A. A. Heidari, H. Chen, M. Wang, Z. Pan, and A. H. Gandomi, "Multi-population differential evolution-assisted Harris hawks optimization: framework and case studies," *Future Generation Computer Systems*, vol. 111, pp. 175–198, 2020.

[35] X. Xu and H.-l. Chen, "Adaptive computational chemotaxis based on field in bacterial foraging optimization," *Soft Computing*, vol. 18, no. 4, pp. 797–807, 2014.

[36] Y. Xu, H. Chen, J. Luo, Q. Zhang, S. Jiao, and X. Zhang, "Enhanced Moth-flame optimizer with mutation strategy for global optimization," *Information Sciences*, vol. 492, pp. 181–203, 2019.

[37] N. A. Golilarz, H. Gao, and H. Demirel, "Satellite image de-noising with harris hawks meta heuristic optimization algorithm and improved adaptive generalized Gaussian distribution threshold function," *IEEE Access*, vol. 7, pp. 57459–57468, 2019.

[38] N. A. Golilarz et al., "ORCA optimization algorithm: a new meta-heuristic tool for complex optimization problems," in *Proceedings of the 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, Chengdu, China, December 2020.

[39] N. A. Golilarz, M. Mirmozaffari, T. A. Gashteroodkhani et al., "Optimized wavelet-based satellite image de-noising with multi-population differential evolution-assisted harris hawks optimization algorithm," *IEEE Access*, vol. 8, pp. 133076–133085, 2020.

[40] A. Addeh, A. Khormali, and N. A. Golilarz, "Control chart pattern recognition using RBF neural network with new training algorithm and practical features," *ISA Transactions*, vol. 79, pp. 202–216, 2018.

[41] N. A. Golilarz, A. Addeh, H. Gao et al., "A new automatic method for control chart patterns recognition based on ConvNet and harris hawks meta heuristic optimization algorithm," *IEEE Access*, vol. 7, pp. 149398–149405, 2019.

[42] Tarhan, B. Cem, and M. Ranjbar, "A Review on multidisciplinary design optimization of dynamic engineering systems," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 3, 2020.

[43] S. Semir and M. Ranjbar, "A review on cutting tool optimization approaches," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 6, no. 3, 2018.

[44] H. Ziari, H. Divandari, S. M. Seyed Ali Akbar, and S. M. Hosseinian, "Investigation of the effect of crumb rubber powder and warm additives on moisture resistance of SMA mixtures," *Advances in Civil Engineering*, vol. 2021, 2021, In press.

[45] A. Izadi, F. Jamshidpour, and H. Hadizadeh, "Evaluation of the performance of bus special lines according to indicators of quality and transport and its performance improvement using ultra-innovative intelligent models (Case study: Rasht city)," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, no. 4, 2019.

[46] X. Zhang et al., "Robust feature learning for adversarial defense via hierarchical feature alignment," *Information Sciences*, vol. 560, 2020.

[47] X. Zhang, T. Wang, W. Luo, and P. Huang, "Multi-level fusion and attention-guided CNN for image dehazing," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[48] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2020.

[49] X. Zhang, D. Wang, Z. Zhou, and Y. Ma, "Robust low-rank tensor recovery with rectification and alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, pp. 238–255, 2019.

[50] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[51] X. Zhang, T. Wang, J. Wang, G. Tang, and L. Zhao, "Pyramid channel-based feature attention network for image dehazing," *Computer Vision and Image Understanding*, vol. 197-198, Article ID 103003, 2020.

[52] X. Zhao, D. Li, B. Yang, C. Ma, Y. Zhu, and H. Chen, "Feature selection based on improved ant colony optimization for online detection of foreign fiber in cotton," *Applied Soft Computing*, vol. 24, pp. 585–596, 2014.

[53] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection," *Engineering with Computers*, pp. 1–30, 2020.

[54] F. Jafari Golrokh, G. Azeem, and A. Hasan, "Eco-efficiency evaluation in cement industries: DEA malmquist productivity index using optimization models," *Engineering Transactions*, vol. 1, no. 1, 2020.

[55] F. Jafari Golrokh and A. Hasan, "A comparison of machine learning clustering algorithms based on the DEA optimization approach for pharmaceutical companies in developing countries," *Engineering Transactions*, vol. 1, no. 1, 2020.

[56] Y. Razieh, "An intelligent method for down syndrome detection in fetuses using ultrasound images and deep learning neural networks," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, no. 3, 2019.

[57] V. Mahdi and A. Addeh, "An intelligent power prediction method for wind energy generation based on optimized fuzzy system," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 5, no. 2, 2019.

[58] Shourabi, R. Zamani, and M. R. Moradian, "An intelligent method based on model predictive torque control and optimized ANFIS for induction motor speed control," *Computational Research Progress in Applied Science & Engineering (CRPASE)*, vol. 4, no. 4, 2018.

[59] Iran Statistical Center, *Official Results of Census of Populations and Houses of Iran*, Iran Statistical Center, Tehran, Iran, 2016.

[60] I. Bargegol, V. N. M. Gilani, and S. Farghedayn, "Analysis of the effect of vehicles conflict on pedestrian's crossing speed in signalized and un-signalized intersection," *Advances in Environmental Biology*, vol. 8, no. 21, pp. 502–510, 2014.

[61] R. Eisinga, T. Heskes, B. Pelzer, and M. T. Grotenhuis, "Exact p values for pairwise comparison of Friedman rank sums, with application to comparing classifiers," *BMC Bioinformatics*, vol. 18, no. 1, p. 68, 2017.

[62] S. M. Hosseinian, V. Najafi Moghaddam Gilani, H. T. Amoli, M. Nikookar, and A. Orouei, "Presentation of analytical

methods for better decision making about the most important factor influencing rural accidents," *Mathematical Problems in Engineering*, vol. 2021, Article ID 5564269, 2021.

[63] T. Sanaz, A. Aslani, and H. Forotan, "Prediction of energy consumption by using regression model," *Computational Research Progress in Applied Science & Engineering*, vol. 2, no. 3, 2016.

[64] A. Mahyar, "Rough-set theory in solving road pavement management problems (case study: Ahwaz-Shush highway)," *Computational Research Progress in Applied Science & Engineering*, vol. 3, no. 2, 2017.

[65] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Statistical analysis and presentation of accident prediction model leading to injuries and deaths of pedestrians in rural roads of Gilan," *Journal of Transportation Research*, 2020.

[66] R. K. Young and J. Liesman, "Estimating the relationship between measured wind speed and overturning truck crashes using a binary logit model," *Accident Analysis & Prevention*, vol. 39, no. 3, pp. 574–580, 2007.

[67] N. Amiri Golilarz and H. Demirel, "Thresholding neural network (TNN) based noise reduction with a new improved thresholding function," *Computational Research Progress in Applied Science & Engineering*, vol. 03, no. 02, 2017.

[68] H. Behbahani, G. H. Hamedi, and V. Najafi Moghaddam Gilani, "Predictive model of modified asphalt mixtures with nano hydrated lime to increase resistance to moisture and fatigue damages by the use of deicing agents," *Construction and Building Materials*, vol. 265, Article ID 120353, 2020.

[69] Y. ZhangR. Liu et al., "Towards augmented kernel extreme learning models for bankruptcy prediction: algorithmic behavior and comprehensive analysis," *Neurocomputing*, vol. 430, 2020.

[70] H.-L. Chen, G. Wang, C. Ma, Z.-N. Cai, W.-B. Liu, and S.-J. Wang, "An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease," *Neurocomputing*, vol. 184, pp. 131–144, 2016.

[71] L. Hu, G. Hong, J. Ma, X. Wang, and H. Chen, "An efficient machine learning approach for diagnosis of paraquat-poisoned patients," *Computers in Biology and Medicine*, vol. 59, pp. 116–124, 2015.

[72] L. Shen, H. Chen, Z. Yu et al., "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, pp. 61–75, 2016.

[73] J. Xia, H. Chen, Q. Li et al., "Ultrasound-based differentiation of malignant and benign thyroid nodules: an extreme learning machine approach," *Computer Methods and Programs in Biomedicine*, vol. 147, pp. 37–49, 2017.

[74] C. Li, L. Hou, B. Y. Sharma et al., "Developing a new intelligent system for the diagnosis of tuberculous pleural effusion," *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 211–225, 2018.

[75] X. Zhao, X. Zhang, Z. Cai et al., "Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients," *Computational Biology and Chemistry*, vol. 78, pp. 481–490, 2019.

[76] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing Journal*, vol. 88, 2020.

[77] M. Wang, H. Chen, B. Yang et al., "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, vol. 267, pp. 69–84, 2017.

[78] H. Behbahani, V. Najafi Moghaddam Gilani, A. Amini, N. Kamboozia, and S. M. Hosseinian, "Fuzzy-neural analysis of pedestrian flow crossing urban intersections," in *Proceedings of the 18th International Conference on Traffic & Transportation Engineering*, Tehran, Iran, January 2020.

[79] V. Najafi Moghaddam Gilani, S. M. Hosseinian, and M. Nikookar, "Presentation of a new deicer with the least moisture and fatigue failures in asphalt mixtures," *Arabian Journal for Science and Engineering*, 2021.

[80] X. Zhang, M. Fan, D. Wang, P. Zhou, and D. Tao, "Top-k feature selection framework using robust 0-1 integer programming," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[81] X. Zhang, R. Jiang, T. Wang, and J. Wang, "Recursive neural network for video deblurring," *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[82] V. Najafi Moghaddam Gilani, S. M. Hosseinian, H. Behbahani, and G. H. Hamedi, "Prediction and pareto-based multi-objective optimization of moisture and fatigue damages of asphalt mixtures modified with nano hydrated lime," *Construction and Building Materials*, vol. 261, Article ID 120509, 2020.

[83] V. Najafi Moghaddam Gilani, S. M. Hosseinian, D. Safari, and M. B. Movahhed, "Investigation of the impact of deicer materials on thermodynamic parameters and its relationship with moisture susceptibility in modified asphalt mixtures by carbon nanotube," *Arabian Journal for Science and Engineering*, 2020.

[84] H. Behbahani, H. Ziari, A. Amini, V. Najafi Moghaddam Gilani, and R. Salehfard, "Investigation of un-signalized roundabouts delay with adaptive-network-based fuzzy inference system and fuzzy logic," *Computational Research Progress in Applied Science & Engineering*, vol. 2, no. 04, pp. 1–7, 2016.

[85] J. Qian, S. Feng, Y. Li et al., "Single-shot absolute 3D shape measurement with deep-learning-based color fringe projection profilometry," *Optics Letters*, vol. 45, no. 7, pp. 1842–1845, 2020.

[86] H. Chen, A. Chen, L. Xu et al., "A deep learning CNN architecture applied in smart near-infrared analysis of water pollution for agricultural irrigation resources," *Agricultural Water Management*, vol. 240, Article ID 106303, 2020.

[87] A. Mahani, P. Bazoobandi, S. M. Hosseinian, and H. Ziari, "Experimental investigation and multi-objective optimization of fracture properties of asphalt mixtures containing nano-calcium carbonate," *Construction and Building Materials*, vol. 285, Article ID 122876, 2021.

[88] V. Najafi Moghaddam Gilani, S. M. Hosseinian, G. H. Hamedi, and D. Safari, "Presentation of predictive models for two-objective optimization of moisture and fatigue damages caused by deicers in asphalt mixtures," *Journal of Testing and Evaluation*, vol. 49, 2020 In press.

[89] I. Yilmaz and O. Kaynar, "Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5958–5966, 2011.

[90] H. T. Abdelwahab and M. A. Abdel-Aty, "Artificial neural networks and logit models for traffic safety analysis of toll plazas," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1784, no. 1, pp. 115–125, 2002.

[91] R. U. Khan, X. Zhang, R. Kumar, A. Sharif, N. A. Golilarz, and M. Alazab, "An adaptive multi-layer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[92] M. Kashaninejad, A. A. Dehghani, and M. Kashiri, "Modeling of wheat soaking using two artificial neural networks (MLP and RBF)," *Journal of Food Engineering*, vol. 91, no. 4, pp. 602–607, 2009.

[93] A. Addeh and M. Iri, "Brain tumor type classification using deep features of MRI images and optimized RBFNN," *Engineering Transactions*, vol. 2, no. 1, 2021.

[94] N. Kamboozia, M. Ameri, and S. M. Hosseinian, "Statistical analysis and accident prediction models leading to pedestrian injuries and deaths on rural roads in Iran," *International Journal of Injury Control and Safety Promotion*, vol. 27, no. 4, pp. 493–509, 2020.

[95] G. H. Hamedi, M. R. Esmaeeli, V. Najafi Moghaddam Gilani, and S. M. Hosseinian, "The effect of aggregate-forming minerals on thermodynamic parameters using surface free energy concept and its relationship with the moisture susceptibility of asphalt mixtures," *Advances in Civil Engineering*, vol. 2021, Article ID 8818681, , 2021.

[96] J. B. Yang, K. Q Shen, C. J Ong, and X. P Li, "Feature selection for MLP neural network: the use of random permutation of probabilistic outputs," *IEEE Transactions on Neural Networks*, vol. 20, no. 12, pp. 1911–1922, 2009.

[97] T. Saito and M. Rehmsmeier, "Precrec: fast and accurate precision-recall and ROC curve calculations in R," *Bioinformatics*, vol. 33, no. 1, pp. 145–147, 2017.

[98] E. LeDell, M. Petersen, and M. van der Laan, "Computationally efficient confidence intervals for cross-validated area under the ROC curve estimates," *Electronic Journal of Statistics*, vol. 9, no. 1, p. 1583, 2015.

[99] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.