

Article

Machine Learning Modeling of Aerobic Biodegradation for Azo Dyes and Hexavalent Chromium

Zulfiqar Ahmad ¹, Hua Zhong ¹, Amir Mosavi ^{2,3,4,5,*}, Mehreen Sadiq ⁶, Hira Saleem ⁶, Azeem Khalid ⁶, Shahid Mahmood ⁶ and Narjes Nabipour ⁷

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China; zulfiqar@whu.edu.cn (Z.A.); zhonghua21cn@126.com (H.Z.)

² Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany

³ Department of Automation, Obuda University, 1034 Budapest, Hungary

⁴ Department of Mathematics, J. Selye University, 94501 Komarno, Slovakia

⁵ Department of Informatics, J. Selye University, 94501 Komarno, Slovakia

⁶ Department of Environmental Sciences, PMAS-Arid Agriculture University, Rawalpindi 46300, Pakistan; mehreenfayyaz2@gmail.com (M.S.); hiraraoenv@gmail.com (H.S.); azeem@uaar.edu.pk (A.K.); shmahmood@uaar.edu.pk (S.M.)

⁷ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; narjesnabipour@duytan.edu.vn

* Correspondence: amir.mosavi@mailbox.tu-dresden.de

Received: 9 March 2020; Accepted: 3 June 2020; Published: 4 June 2020



Abstract: The present study emphasizes the efficacy of a biosurfactant-producing bacterial strain *Klebsiella* sp. KOD36 in biodegradation of azo dyes and hexavalent chromium individually and in a simultaneous system. The bacterial strain has exhibited a considerable potential for biodegradation of chromium and azo dyes in single and combination systems (maximum 97%, 94% in an individual and combined system, respectively). Simultaneous aerobic biodegradation of azo dyes and hexavalent chromium (SBAHC) was modeled using machine learning programming, which includes gene expression programming, random forest, support vector regression, and support vector regression-fruit fly optimization algorithm. The correlation coefficient includes the dispersion index, and the Willmott agreement index was employed as statistical metrics to assess the performance of each model separately. In addition, the Taylor diagram was used to further investigate the methods used. The findings of the present study were that the support vector regression-fruitfly optimization algorithm (SVR-FOA) with correlation coefficient (CC) of 0.644, (scattered index) SI of 0.374, and (Willmott's index of agreement) WI of 0.607 performed better than the autonomous support vector regression (SVR), gene expression programming (GEP), and random forest (RF) methods. In addition, the standalone SVR model with CC of 0.146, SI of 0.473, and WI of 0.408 ranked the second best. In summary, the SBAHC can be accurately estimated using the hybrid SVR-FOA method. In other words, FOA has proven to be a powerful optimization algorithm for increasing the accuracy of the SVR method.

Keywords: biodegradation; chromium; azo dyes; prediction; SVR-FOA

1. Introduction

Many salts, in particular azo dyes and chromium sulphate, are the most commonly used chemicals in various processes in the leather tanning industry [1]. Azo dyes are regarded as xenobiotics because they are very resistant. Their chemical structure is designed to resist discoloration when exposed to

sunlight, sweat, water, microbes, or other chemical agents [2]. These dyes tend to persist for a long time, circulate, and accumulate in the food chain [3]. The improper disposal of liquid wastes containing azo dyes and their derivatives is visually unpleasant in the aquatic ecosystem, leading to a reduction in solar radiation and, thus, to a reduction in photosynthesis and dissolved oxygen (DO) concentration, which ultimately leads to serious toxic effects on marine life and serious environmental damage [4]. These dyes also have many toxic effects in the form of eye infections, tumors, allergies, cancer, and respiratory diseases in humans [5].

Wastewater from the textile industry containing organic pollutants often contains free or complex ionic metals [6]. The activity of microorganisms can be affected by a high concentration of lead, chromium, copper, zinc, arsenic, and cadmium, which are known as heavy metals [7]. The composition of the microbial may influence the decomposition properties of microorganisms [8]. Chromium exists in two biologically active forms that are stable under their oxidizing conditions. The toxicity of hexavalent chromium (Cr(VI)) is about 100 times higher than that of the trivalent chromium (Cr(III)) and is also mutagenic, carcinogenic, and teratogenic [9]. Due to its high solubility and permeability to the outer layer of the cell, Cr(VI) changed to Cr(III) and sticks more strongly to DNA to form the DNA-Cr complex and Cr protein, which tend to alter the structure and function of enzymes and cause mutations when exposed to humans, leading to skin and respiratory infections that are irritating to the eyes and even carcinogenic [10].

Co-contamination (azo dyes and chromium) is a complex problem in the biodegradation process. Physico-chemical and bioremediation technologies are currently available to reduce the hexavalent chromium of a more toxic form to less toxic Cr(III) in industrial effluents [11] and the biodegradation of azo dyes, but these processes are costly and less effective in removing combined pollutants or have secondary effects on pollution to remove these toxic substances. Several microorganisms, in particular, bacteria producing biosurfactants, are able to grow in a medium amended with hexavalent chromium and changed it to trivalent chromium. However, such a process is limited due to the small number of potential microorganisms that can be overcome by the exogenous application of biosurfactant-producing bacteria and a nutrient medium, but this also requires the optimization of the biological process to develop an effective system for biodegradation of azo dyes and Cr(VI).

Statistical methods have been used to estimate medium parameter optimization in some previous studies [11–13]. These include the traditional optimization method, where one factor is optimized and the other is kept constant [14,15]; however, this process takes more time. Taguchi design [15–17], Plackett–Burman design [18–20], response surface methodology (RSM) with central composite design (CCD), and *n*-factor design [21–24] can be considered as the conventional tools for experimental design. Sometimes, second-order polynomial cannot fully cope with nonlinear biological interactions [25,26]. This leads researchers to employ computational intelligence techniques to predict and estimate the optimal parameters of simulated bioreduction of azo dyes and Cr(VI).

Gene expression programming (GEP) is a metaheuristic approach in the presence of a decision tree (DT) structure to combine the properties of other data-driven approaches [27]. Likewise, the RF technique, which includes a number of decision trees and naturally integrates feature selection and interaction into the learning process, is a good way to estimate factors that influence bioprocesses [28]. Other methods, such as SVR, are a powerful machine learning (ML) technique developed from statistical learning theory. The SVR is popular due to its high generalization performance and its capability to manage nonlinear models in the presence of kernel techniques [29]. Fruit fly optimization algorithm (FOA) was recently introduced by Pan [30] as an evolutionary optimization algorithm. This technique has advantages, like being easy to understand and less time-consuming. As a new intelligent optimization algorithm, this technique has attracted attention and has been successfully employed in the autonomous control of surface vessels [31] and the optimization of control bioprocesses [32].

The existing study is novel in the sense that it describes simultaneous biodegradation of azo dyes and chromium (VI) using biosurfactant-producing bacteria, which have advantageous over traditional bacteria having potential of biosurfactant production, which results in micelle formation

and enhances solubility of compounds in aqueous system. Moreover, application of SVR-FOA for estimation of simultaneous biodegradation of azo dyes and chromium (VI) in the current study is a novel strategy as the technique is superior compared to previously employed algorithm programs, as SVR-FOA has fewer parameters and is easy to program. Hence, it optimizes the complex linear regression problem inspired by the fruit fly food searching phenomena by a specialized way of smell and vision. The main contribution of the study was to develop a novel modeling approach based on machine learning programming for estimating and predicting optimal parameters for SBAHC by the newly isolated bacterial strain *Klebsiella* sp. KOD36.

2. Methodology

2.1. Culture Medium, Chemicals, and Microorganism

Reactive black dyes and di-potassium chromate were used in this study to determine possible degradation by bacterial stain. All solutions were prepared in distilled water at 121 °C for 20 min. A working solution of 15, 100, and 150 mg L⁻¹ azo dyes and 2, 5, and 10 mg L⁻¹ chromium was prepared from the stock solution. The diphenyl carbide reagent was used to determine the chromium concentration in the spectrophotometer. Degradation studies using mineral salts (MSM), NaCl (1.0 g/L), CaCl₂ (0.1 g/L), KH₂PO₄ (1.0 g/L), MgSO₄ · 7H₂O (0.5 g/L), Na₂HPO₄ (1.0 g/L), and yeast extract (4.0 g/l) amended with various concentration of Cr(VI) and azo dyes were used for the purification and streaking of bacterial strains already reported to produce biosurfactants. The strain used was isolated previously [33] and identified as *Klebsiella* sp. through 16 s RNA. The sequence of identified strain was submitted with accession number KT364873 to gene bank [33].

2.2. Evaluation of *Klebsiella* sp. KOD36 Potential for Biodegradation of Azo Dyes and Chromium (VI) in Single and Combined System

The strain *Klebsiella* sp. KOD36, which produces biosurfactants, was investigated for capability to utilize as bioremediation agent in biodegradation of azo dyes and hexavalent chromium in single and combined aerobic systems. For this purpose, the MSM broth was supplemented with various concentrations of chromium (0, 5, 10 mg L⁻¹) and azo dyes (0, 100, 150 mg L⁻¹) in single or combination systems (100 mg L⁻¹ azo dyes and 5 mg L⁻¹ Cr, 100 mg L⁻¹ azo dyes and 10 mg L⁻¹ Cr, 150 mg L⁻¹ azo dyes and 5 mg L⁻¹ Cr, 150 mg L⁻¹ azo dyes and 10 mg L⁻¹ Cr) inoculated with *Klebsiella* sp. KOD36 uniform bacterial community 2.3×10^7 colony forming unit (CFU) mL⁻¹. The glass vials were placed at 150 rpm in a stirred, temperature-controlled incubator for 24 h. Aliquots were regularly taken alternately from each vial to determine their concentration of reduced species. The degradation potential was measured using the technique as described by Desai et al. [34].

2.3. Estimation of Optimal Growth Factors for Simultaneous Bioremoval of Azo Dyes and Chromium (VI) in Single and Combined System

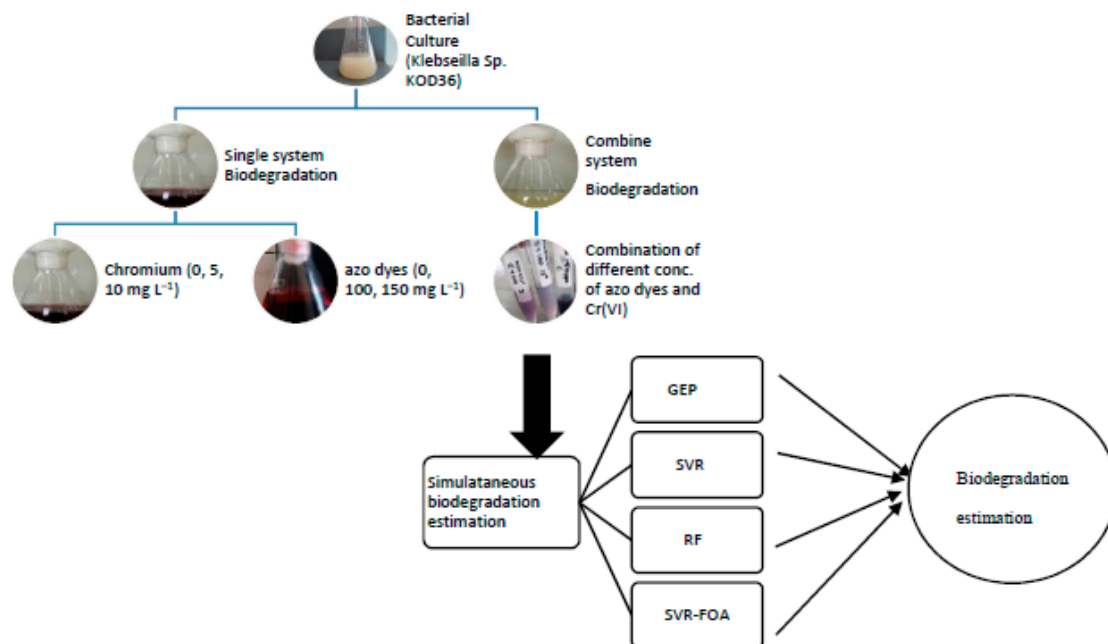
Various parameters for the simultaneous biodegradation of azo dyes and chromium were evaluated using the Taguchi model, as shown in Table 1. The measured values of simultaneous biodegradation of azo dyes and chromium are given in Table 2, while machine learning employed is depicted by Figure 1.

Table 1. Parameters and their levels.

Independent Variables	Levels			
	Unit	−1	0	1
Temperature	°C	30	35	40
pH	Unit less	3	6	9
IP	h	8	16	24
Carbon source	20 mg/L	Glucose	Sucrose	Starch
Nitrogen source	2 mg/L	Yeast	Ammonium sulfate	Urea
Shaking	rpm	150	200	250

Table 2. Measured average value of simultaneous reduction of azo dye and chromium by biosurfactant-producing bacteria with their dependent variables.

Temperature °C	pH	IP (Hours)	Source of Carbon (20 mg/L)	Source of Nitrogen (2 mg/L)	Shaking (rpm)	Simultaneous Reduction of Azo Dye and Chromium (%)
30	3	8	1	1	1	43
30	6	8	1	2	2	93
30	9	8	1	3	3	84
30	3	16	2	1	1	63
30	6	16	2	2	2	10
30	9	16	2	3	3	45
30	3	24	3	1	1	11
30	6	24	3	2	2	90
30	9	24	3	3	3	35
35	3	8	1	1	1	66
35	6	8	1	2	2	80
35	9	8	1	3	3	38
35	3	16	2	1	1	25
35	6	16	2	2	2	74
35	9	16	2	3	3	30
35	3	24	3	1	1	69
35	6	24	3	2	2	44
35	9	24	3	3	3	86
40	3	8	1	1	1	16
40	6	8	1	2	2	64
40	9	8	1	3	3	95
40	3	16	2	1	1	99
40	6	16	2	2	2	98
40	9	16	2	3	3	60
40	3	24	3	1	1	20
40	6	24	3	2	2	50
40	9	24	3	3	3	46

**Figure 1.** Schematic diagram of simultaneous biodegradation of azo dyes and Cr(VI) and their estimation through machine learning programming.

2.4. Methods of Analysis

Azo Dye and Chromium Concentration Measurement

Chromium and azo dye concentrations were measured using the technique described by Desai et al. [34]. In short, 1 mL of desired sample was drawn from a tube and subjected to centrifugation

at 10,000 rev/min for 20 min. The chromium concentration was determined by complexing the chromium with a diphenyl carbide reagent. The pink color was developed, which was determined at wavelengths of 540 nm. The concentration of the azo dye at 570 nm was also calculated for each dye with a spectrophotometer.

2.5. Machine Learning Methodologies

In the present study, our major aim was to design a novel methodology to estimate and predict the simultaneous biodegradation of azo dyes and chromium (VI) in a single and combine system by means of machine learning programming. The acquired data were subjected to use in training GEP, RF, SVR, and SVR-FOA procedures. The gene expression program was up-to-date algorithm methodology that described the relationship between given input and resulting output variables through software development [35].

In contrast to genetic algorithm (GA) and genetic programming (GP), the GEP is actually the combination of these two [35,36]. GEP is a specialized program for solving regression problems. Support vector regression (SVR) is specifically important for its capability of management and required performance in solving problems, particularly nonlinear regression problems [37]. The selection of basic parameters is an important aspect in SVR technique. The main parameters of SVR technique include c , ϵ , and kernel functions [38]. The drawback of this technique is, in certain cases, the incorrect values of SVR basic parameters may result in either under- or over-tuning. Therefore, an optimal value of each basic parameter should be chosen while in the SVR training phase. Various methodologies have been adopted to select the optimal value of these basic parameter of SVR. Fruit fly optimization algorithm (FOA) is such a technique, introduced by Pan [30], which employs the foraging strategy of a fruit fly to search the optimal position of each basic value in support vector regression. These methods are described as following.

2.5.1. Gene Expression Programming (GEP)

Genetic expression programming is a specialized genetic programming technique that is capable of solving optimal problems by developing an expression tree (ET). The baseline of this technique is a specialized tree-like structure which is first trained as living organisms by changing shape, size, composition, and other affecting factors. Genetic expression programs, similar to living organisms, are coded as fixed traits on chromosomes. Therefore, the GEP is just like a genotype-phenotype expression system that uses a simple genetic information code to exhibit the biologically inspired phenotype. The particular fixed-length chromosome has genetic information similar to actual information stored in a chromosome part. Each chromosome contains several genes that are called sub-entity types. In genetic expression programming, all the other sub-entity types are connected to a root, similar to a tree, and make a connection to each other. These sub-connections in genetic expression technique include division, multiplication, subtraction, and addition [39]. These particular genes, irrespective of their fixed length, have varying size and shape. So, the varying length of different genes allows these genes in GEP to progress adoption and evolution. Each specific area in a gene is known as open reader frame (ORF), which provides solution while exhibiting as code in expression tree [40]. Similar to other evolutionary programs, the genetic expression program is based on scattered information on chromosomes. In a general set of data (population), fitness function is used to evaluate each chromosome and designate a specific value. Different fitness functions in a genetic expression program have been used previously [35]. Suitable chromosomes are picked up in the next generation. These chromosomes are further controlled by particular gene operators after being selected. The process of selection continues until a proper optimal value is attained [35,41].

2.5.2. Support Vector Regression

SVR is normally used for solving regression problems and has been used as an estimation and prediction system in biological systems [42–44]. This is a supervised program which utilizes structural

risk minimization (SRM), in contrast to empirical risk minimization (ERM) used in conventional neural network. ERM basically reduces the error of a training dataset while SRM is helpful in reducing errors at higher extent. Therefore, SVM has the potential to reduce errors of commonly practiced neural networks [37,38]. The main background of SVR is to design the given input data into more precise dimensions and improve the efficiency of linear regression problems by adopting kernel function; yet these kernel functions are not adequate specifically in more complex linear regression problems. Each function (kernel) must have two features, symmetricity and compliance, with Cauchy-Schwarz criteria to address these issues. These features ensure that new space is capable of being defined by these functions (kernel). In general, the support vector regression performance depends on its parameters, which include ϵ (intensive zone, which is normally used to fire the training dataset), C (trade-off), and γ (determine of relative error and smoothness). To optimize the various SVR parameters, different algorithms have been developed.

2.5.3. FOA (Fruit Fly Optimization Algorithm)

The FOA is a biological program inspired by the behavior of *Drosophila* insect for food search (Pan, 2012). The *Drosophila* insect has a unique quality of superior smell and vision, which differentiate and make it superior from other insects. The insect employs a sense of vision when approaching a food source. This information is transmitted to the insect body, which determines the route leading to the food source identified [45]. The schematic diagram of SVR-FOA methodology, adapted by Nabipour et al. [46], is shown in Figure 2.

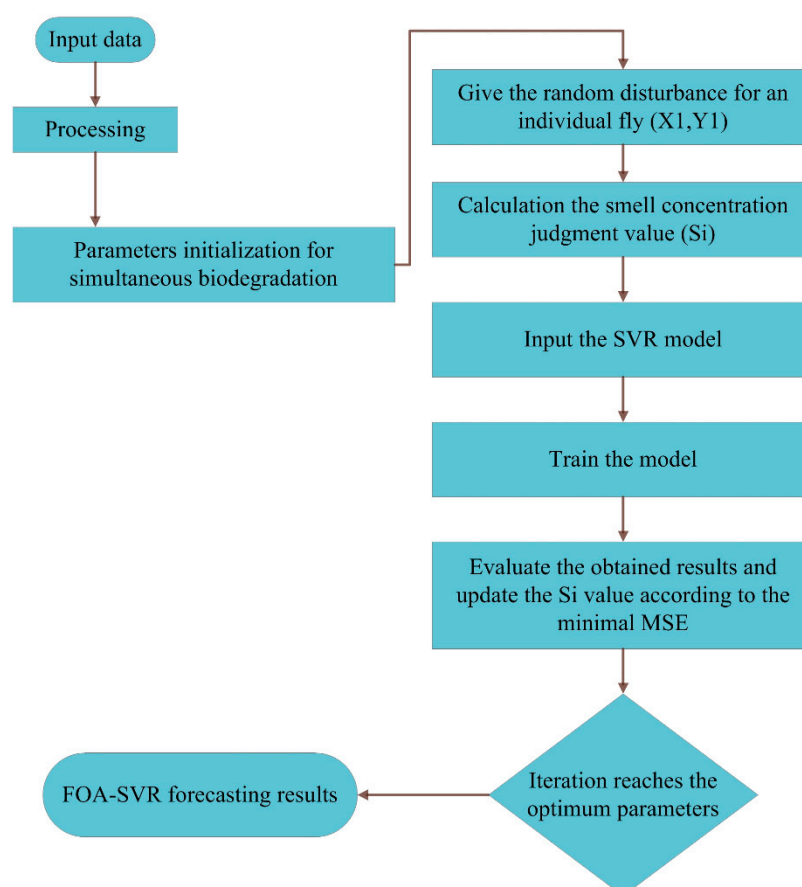


Figure 2. The schematic diagram of SVR-FOA adapted by Nabipour et al. [46].

2.6. Modeling Methodologies

The simultaneous biodegradation of azo dyes and chromium was estimated and modeled using DT techniques (e.g., GEP, SVR, RF, and SVR-FOA). DT techniques present simple interpretation of results by handling nonlinear and nonparametric variables. Evaluating the performance of the mentioned techniques was performed by comparing results with the empirical relationships presented by other researchers.

$$\begin{aligned} \text{Simultaneous degradation of azo dyes and chromium (\%)} \\ = -3.87091^{\text{Nitrogen} + \text{Incubation} - \text{Nitrogen}} \\ + \frac{1.28408 \text{Nitrogen}}{\text{Carbon} - 0.929413} + \text{pH} + \sqrt{\text{pH}^{\text{Carbon}} - \sqrt{\text{Temperature}}} \\ + \text{Nitrogen}(\text{Carbon} - \text{Nitrogen} - \text{pH} + \text{Temperature}) \end{aligned}$$

In the above formulation, set the mentioned following values for nitrogen as yeast, ammonium sulfate, and urea, while for carbon source glucose set sucrose and starch.

Parameters for Evaluation of Models' Performance

The performance (predictive) of the recommended model was evaluated as CC, SI, and WI. These statistics are presented as follows [47,48]:

I: CC expressed as:

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

II: SI follows as:

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

III: WI expressed as:

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

The targeted and predicted values are denoted as O_i and P_i , respectively.

3. Results and Discussion

In the present study, the potential of biosurfactant-producing strain *Klebsiella* sp. KOD36 was tested for its simultaneous reduction of chromium and reactive black-5 azo dyes (RB-5). Additionally, optimization of environmental and nutritional parameters during simultaneous biodegradation of chromium and azo dyes was assessed using GEP, SVR, RF, and SVR-FOA. Table 3 presents the statistical analysis parameters of the used data.

Table 3. Statistical analysis parameter values of the used data.

Variable	Mean	Minimum	Maximum	Standard Deviation	Coefficient of Variation	Skewness	Correlation with Simultaneous Degradation of Azo Dyes And chromium (%)
Shaking (rpm)	200.0	150.0	250.0	41.6	0.21	0.0	0.176
pH	6.0	3.0	9.0	2.50	0.42	0.0	0.176
IP (h)	16.0	8.0	24.0	6.66	0.42	0.0	−0.210
Temperature (C)	35.0	30.0	40.0	4.16	0.12	0.0	0.121
Simultaneous degradation of azo dyes and chromium (%)	56.8	10.0	99.0	28.2	0.50	−0.08	1

3.1. Evaluation of *Klebsiella* sp. KOD36 for Biodegradation of Azo Dyes and Chromium (VI)

3.1.1. In Single System

Results presented in Figure 3 represent that for chromium at 5 mg L^{−1}, 98% reduction was observed as compared to control after 24 h of inoculation of bacterial strains *Klebsiella* sp. KOD36, while for 10 mg L^{−1} of chromium up to 80% reduction occurred, and for chromium (5 mg L^{−1}) 80% reduction of chromium was observed compared to control (Figure 4).

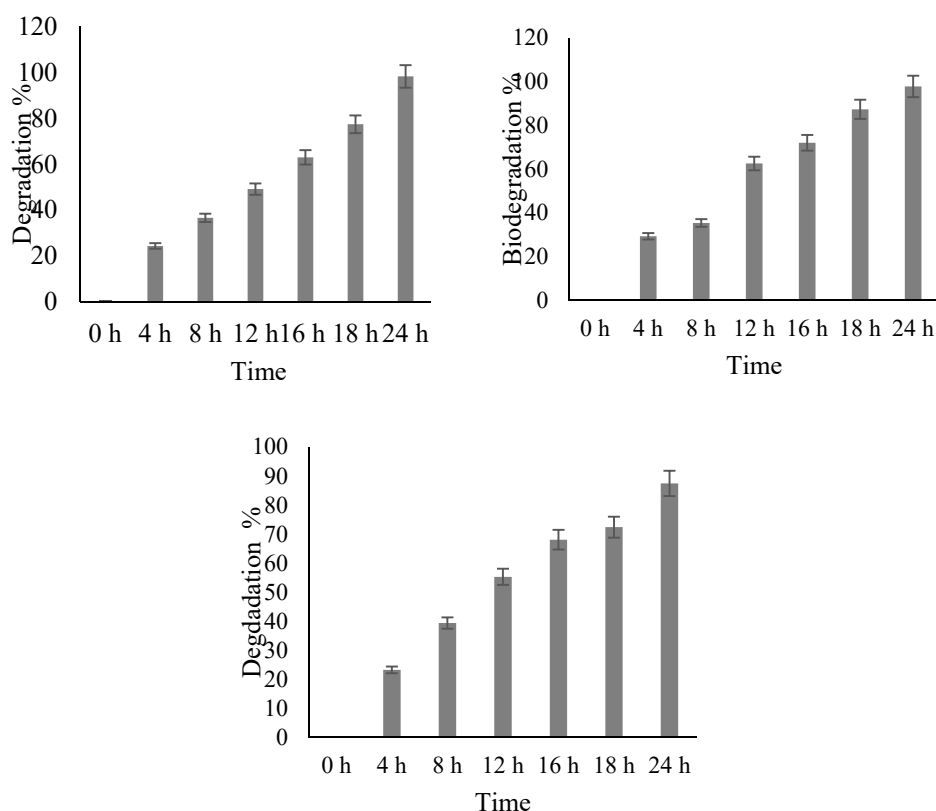


Figure 3. Biodegradation of azo dyes at various concentrations (15, 100, and 150 mg L^{−1}) by biosurfactant-producing *Klebsiella* sp. KOD36.

Regarding azo dyes at 100 mg L^{−1}, a maximum 90% degradation of azo dyes was observed, while for 150 mg L^{−1} azo dyes' concentration, a maximum 87% degradation was observed after 24 h. The inoculation of biosurfactant-producing strain *Klebsiella* sp. KOD36 significantly enhanced the biodecolorization of reactive black dyes compared with the sample lacking the bacterial strain. Based on results of the previous study, *B. circulans* BWL1061 decolorizes azo dyes [37], which lead to

achieving the biodecolorization by improving the enzymes responsible for degradation and dyes. The significant improvement in biodegradation of azo dyes and chromium (VI) by biosurfactant at critical micelle concentration (CMC) and biosurfactant-producing bacteria indicates that electrostatic attraction forces and hydrophobic part of the biosurfactant play a vital role in biodegradation of dyes. A similar finding and mechanism was described by a previous study conducted by Liu et al. [49], who described the isolated strain BWL1061 exhibited degradation potential for azo dyes, which may likely have been due to interaction of biosurfactant hydrophobic moiety of biosurfactant and dyes. Similarly, in another study conducted, Thacker and Madamwar [50] described the capability of the biosurfactant-producing bacterial strain *Ochrobactrum* sp. and *Bacillus* sp. for the reduction of hexavalent chromium in a batch study experiment.

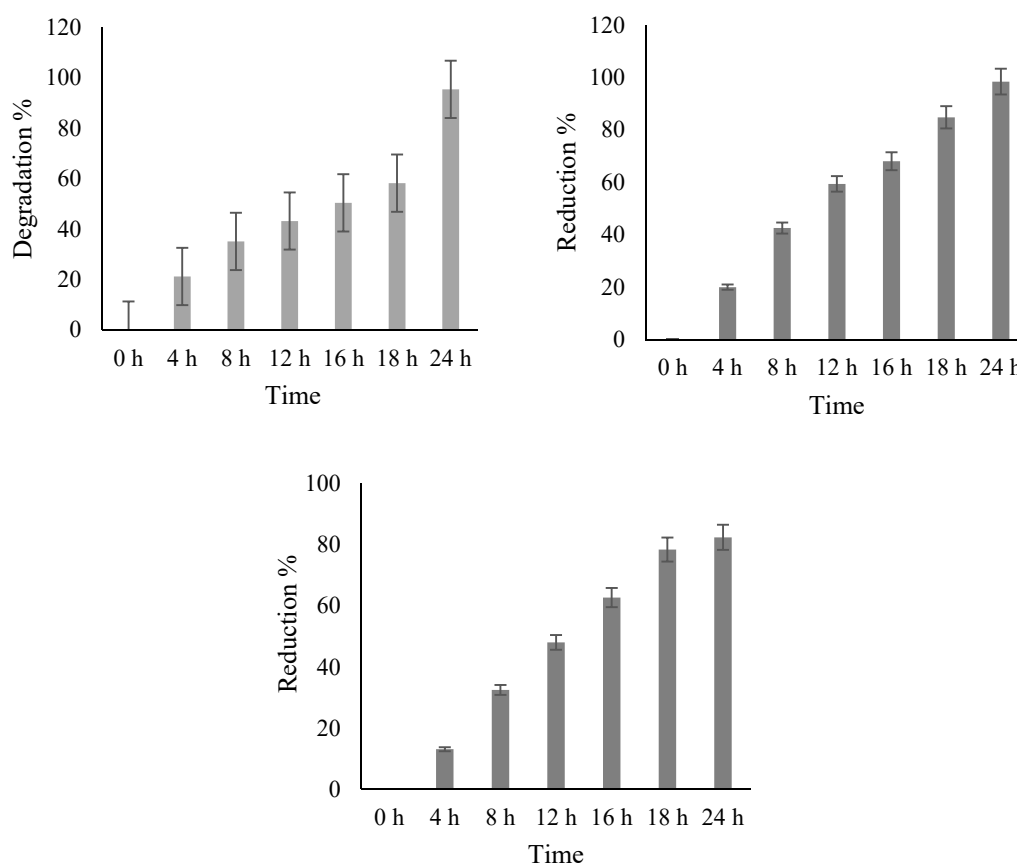


Figure 4. Bioreduction of Cr(VI) at various concentrations (2, 5, and 10 mg L⁻¹) by biosurfactant-producing *Klebsiella* sp. KOD36.

3.1.2. In Combined System

Microorganisms are important in the way that they are excellent bioremediation agents for heavy metal contamination (soil and water). Microorganisms showing a significant resistance to heavy metals have potential as remediation agent in detoxification of these heavy metals. However, in certain cases, more specifically under co-contamination, their efficiency is hindered, as heavy metals cause toxicity to microorganisms and lower their efficiency for biodegradation of azo dyes. In this scenario, the following investigation was designed to investigate the efficacy of biosurfactant and biosurfactant-producing bacteria for decolorization of azo dyes with various concentrations of chromium (VI).

Regarding simultaneous degradation results (azo dyes at 100 mg L⁻¹ and Cr at 5 mg L⁻¹ concentration), 89% degradation was observed after 24 h of inoculation of bacterial strains *Klebsiella* sp. KOD36 after 24 h, as compared to control (Figure 5). While for 100 mg L⁻¹ (azo dyes and 10 mg L⁻¹ Cr), up to 94% simultaneous degradation was observed after 24 h of inoculation of bacterial strains

Klebsiella sp. KOD36, as compared to control. The percent degradation for 150 mg L⁻¹ azo dyes and 5 mg L⁻¹ Cr, and 150 mg L⁻¹ azo dyes and 10 mg L⁻¹ Cr were 91% and 82%, respectively, after 24 h, as compared to control. Similar effects could also be observed in the study conducted by Halmi et al. [51], who isolated a novel potential stain for decolorization of four different dyes, namely amaranth dye, Biebrich scarlet, direct blue, and metanil yellow, under aerobic environment. The isolate bacterial strain exhibited decolorization a maximum of 52% of dyes (initial concentration 150 ppm potassium dichromate) in nutrient broth medium after an incubation of 24 h under shaking at 150 rpm. The enhanced biodecolorization effect could be likely have been due to the fact that biosurfactants reduce the toxicity of hexavalent chromium by entrapping it in micelles and reduce their bioavailability to microorganisms and, meanwhile, bacterial decolorization for azo dyes was enhanced.

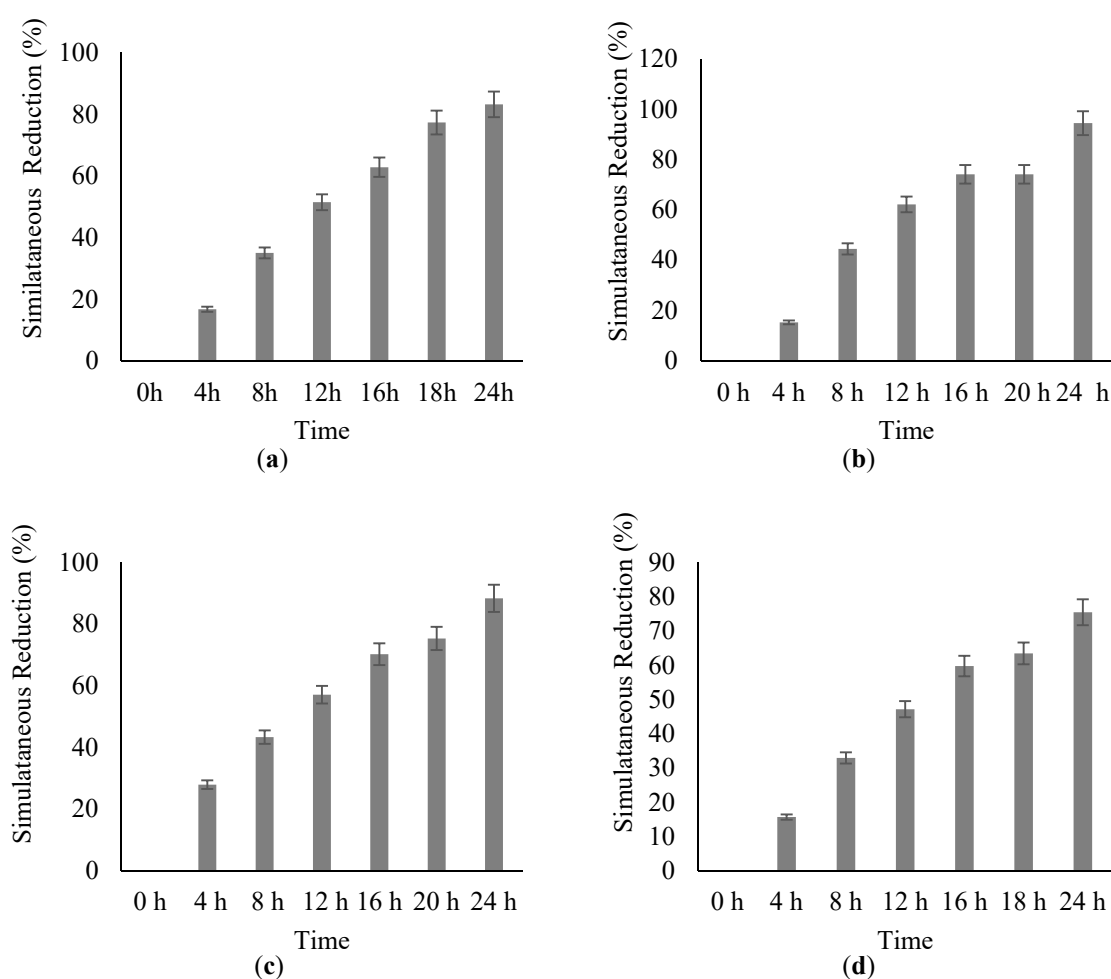


Figure 5. Time course study of simultaneous reduction (%) of azo dyes and chromium at concentration (a) (100 mg L⁻¹, 5 mg L⁻¹), (b) (100 mg L⁻¹, 10 mg L⁻¹), (c) (150 mg L⁻¹, 5 mg L⁻¹), (d) (150 mg L⁻¹, 10 mg L⁻¹) by *Klebsiella* sp. KOD36.

The presence of organic and non-organic compounds emit mixed pollution in industrial zones [38]. Conventional wastewater contains various types of organic and inorganic contaminants, which require immediate attention. Chang et al. [52] conducted a study using a high salinity-tolerant bacterial strain, A12 and L, for its biodegradability for sulfamethoxazole (SMX). They also found that under aerobic and anaerobic conditions the bacterial strain denoted as A12 and L showed a significant degradation of SMX in milkfish culture pond sediment batch experiments. Biosurfactant plays a vital role in this regard. Micelle formation in biosurfactant entrapped the heavy metal (chromium) in its core, thus reducing the bioavailability to bacterial cell by preventing the cells of chromium [30]. Subsequently, bacterial cell

efficiently decolorized azo dyes in the presence of Cr(VI). The above findings showed that *Klebsiella* sp. KOD36 is a proper choice for reclaiming azo dyes and metal (Cr(VI))-contaminated sites.

3.2. Modeling Outcomes

There is not any significant instruction for splitting training and testing data. In this study, data were divided into training (67%) and testing (33%) to develop GEP, RF, SVR, and SVR-FOA models for SBAHC estimation. Moreover, SVR-FOA optimized the default values of SVR for increasing the accuracy of predictions. So, the default and optimized values of SVR and SVR-FOA are presented in Table 4. Also, the GEP modeling functional parameters are shown in Table 5. Therefore, with default and optimized parameters, the defined scenarios for SVR, SVR-FOA, GEP, and RF models' parameters are shown in Table 6.

Table 4. Parameters of the SVR and SVR-FOA models.

Parameter	SVR	SVR-FOA
C	1.0000	1.7242
γ	0.0100	0.0517
ϵ	0.0010	0.0468

Table 5. Parameters of the GEP model.

Parameter	Value
Head size	8
Linking Function	Addition (+)
Number of Genes	3
Chromosomes	30
Mutation Rate	0.044
Inversion Rate	0.1
One-Point RR	0.3
Two-Point RR	0.3
Gene RR	0.1
Gene Transposition Rate	0.1
Used functions	+, −, ×, ÷, power

Table 6. Various model evaluation performance parameters' comparison.

Parameter	SVR	SVR-FOA	GEP	RF
CC	0.146	0.644	0.387	0.41
SI	0.473	0.374	0.647	0.519
WI	0.408	0.607	0.456	0.507

This is clear in Table 6 that SVR-FOA showed the maximum estimation performance. In other words, SVR-FOA with CC value of 0.644, SI value of 0.374, and WI value of 0.607 estimated SBAHC more accurately than other considered models and, hence, chosen as the best model among others studies, followed by SVR with CC value of 0.146, SI value of 0.473, and WI value of 0.408. Although the CC value of SVR was low, due to lower SI error, it may be more appropriate than GEP and RF models. Additionally, among GEP and RF models, GEP showed weak performance with CC value of 0.387, SI value of 0.647, and WI value of 0.456. Furthermore, it can be concluded from Table 6 that SVR-FOA increased CC values of SVR, GEP, and RF by 341.1%, 66.4%, and 57.1%, respectively. Also, it reduced SI values of the mentioned models by 20.9%, 42.2%, and 27.9%, respectively. Finally, SVR-FOA increased the WI values of the mentioned models by 48.8%, 33.1%, and 19.7%, respectively. Although the GEP

had lower accuracy in predicting simultaneous biodegradation, the model could be used for SBAHC. The mentioned GEP formulation is presented below.

$$\text{SBAHC} = -3.87091^{\text{Nitrogen}} + \text{Incubation} - \text{Nitrogen} + \frac{1.28408\text{Nitrogen}}{\text{Carbon}-0.929413} + \text{pH} + \sqrt{\text{pH}^{\text{Carbon}}} - \sqrt{\text{Temperature}} + \text{Nitrogen}(\text{Carbon} - \text{Nitrogen} - \text{pH} + \text{Temperature})$$

In the above formulation, the following values should be considered for nitrogen, yeast 1, ammonium 2, urea 3; and for carbon, glucose 1 and sucrose 2.

These performance parameters of various models used in the present study are also shown as bar chart (Figure 6). It is obvious from the chart that SVR-FOA had highest potential for predicting SBAHC in prediction and estimation.

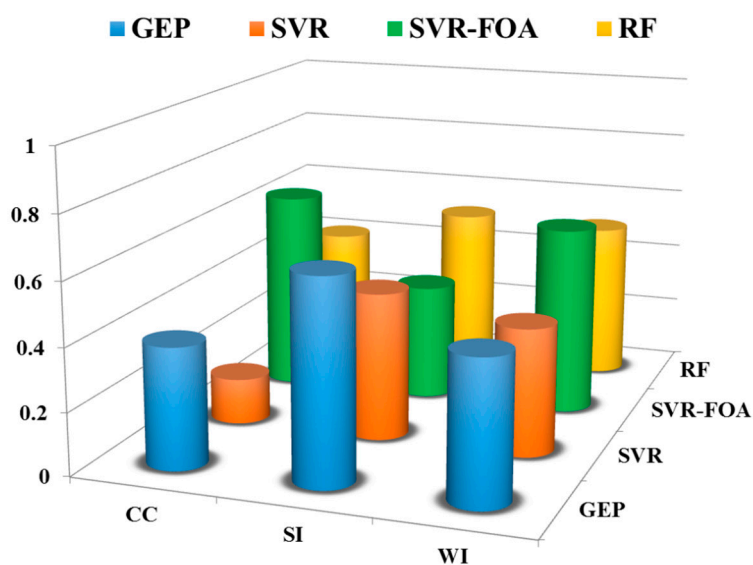


Figure 6. Three-dimensional bar graphs of the statistical parameters.

The SBAHC predictive results of various models is also shown in Figure 7. It can be observed from Figure 5 that SVR-FOA had a higher performance than other considered models. Furthermore, Figure 8 indicates scatter plots of prediction of the SBAHC values with SVR-FOA, SVR, GEP, and RF models. The less-scattered points exhibited by SVR-FOA is a clear indication that the values of SVR-FOA were more accurate than standalone SVR, GEP, and RF models.

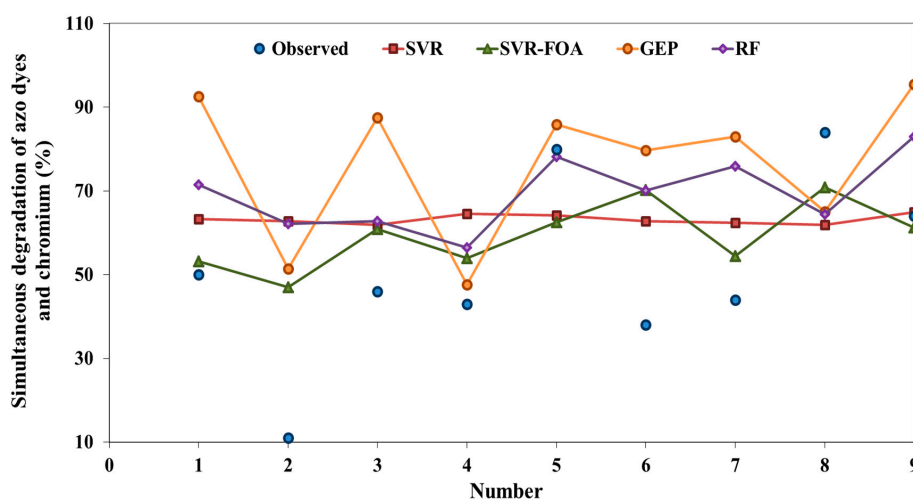


Figure 7. Estimated and observed values comparison for simultaneous degradation of azo dyes and chromium (%) of various models studied.

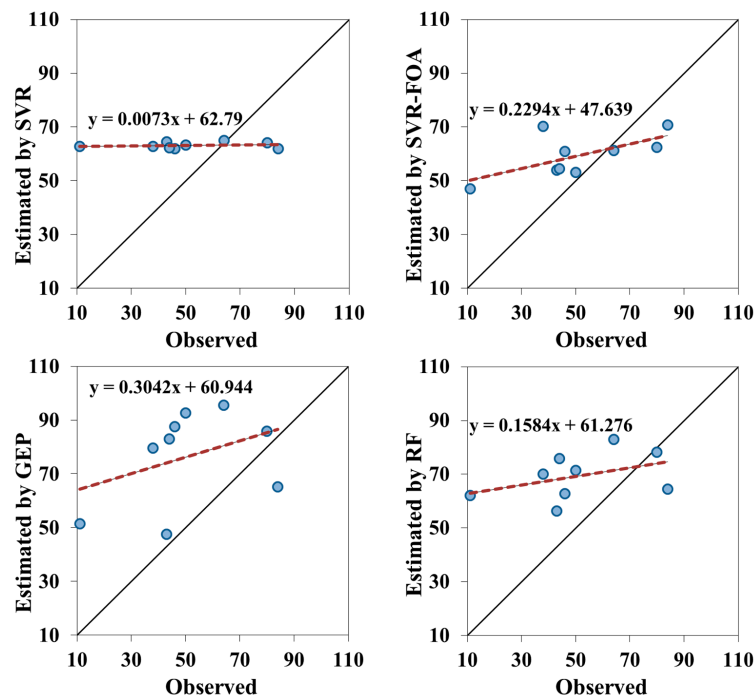


Figure 8. The estimated and observed values for simultaneous degradation of azo dyes and chromium (%) by various models used in the present study.

Furthermore, Taylor diagrams (TD) were employed to examine standard deviation (SD) and CC values for the SVR-FOA, SVR, GEP, and RF models. Figure 9 presents TD for all models. It can be understood from Figure 7 that SVR-FOA (a point with grey color), due to a shorter distance from the observed green point, provided relatively precise predictions of SBAHC values. As a conclusive remark, it can be stated that SVR-FOA with optimized values ($C = 1.7242$, $\gamma = 0.0517$ and $\varepsilon = 0.0468$) and using input parameters of temperature, pH, incubation period (IP), and shaking is more capable for accurate estimation of SBAHC in comparison to standalone SVR and GEP models, and may be recommended for further implementations [36,53,54].

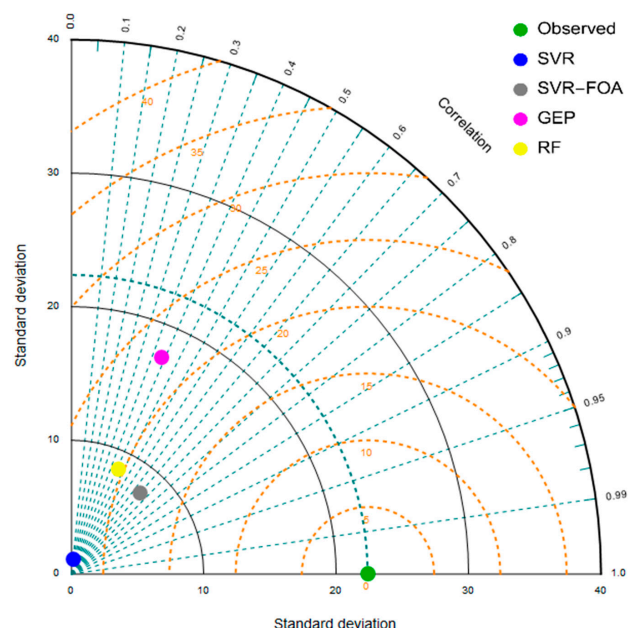


Figure 9. Taylor diagrams of estimated values of simultaneous degradation of azo dyes and chromium (%).

The need for a robust model for estimation of large number of input variables is obvious in the recent world. Another study, investigated by Amato et al. [55], illustrated that in a well-defined geographic area the application of social nets (on-line) may increase the detection efficiency (real time) and alert diffusion. They proposed a multicomplex big data system that uses clustering event detection techniques along with multimedia content and biologically inspired programming to develop alerts.

4. Conclusions

Klebsiella sp. KOD36 significantly boosted the biodegradation of azo dyes. Additionally, the capabilities of SVR, SVR-FOA, GEP, and RF models in estimation of SBAHC values were inspected. Accordingly, the enactments of studied methods were comprehensively examined using CC, SI, and WI parameters. Also, Taylor diagrams were utilized for further assessment. The obtained results indicated that SVR-FOA with CC of 0.644, SI of 0.374, and WI of 0.607 had better performance comparing to standalone SVR, GEP, and RF methods. Moreover, standalone SVR model ranked the second best with CC of 0.146, SI of 0.473, and WI of 0.408. This can be verified by the presented Taylor diagram. Because of less-scattered points exhibited by SVR-FOA, it can be concluded that the estimates of SVR-FOA were much more accurate than other studied models. Conclusively, a fruit fly optimization algorithm had remarkable impact in reducing the prediction errors of a standalone SVR method and it can be recommended for SBAHC estimation. However, the main drawback in using the SVR-based model is that, when there is a big dataset or large number of population load, the time consumed as learning/training is very high, thus determination of parameters involves mainly the researcher experience.

Author Contributions: Conceptualization, Z.A., H.Z., M.S., A.K. and S.M.; formal analysis, H.Z., M.S., H.S., A.K. and S.M.; investigation, H.Z. and A.K.; methodology, Z.A., A.M. and H.S.; project administration, H.S. and S.M.; resources, Z.A., H.Z. and M.S.; software, A.M.; supervision, A.M.; validation, N.N.; visualization, M.S. and N.N.; writing—original draft, Z.A., and A.M.; writing—review and editing, Z.A., A.M. and N.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China (51629901) and Major Science and Technology Program for Water Pollution Control and Treatment of China (2017ZX07108-001).

Acknowledgments: We acknowledge the financial support of this work by the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 project and the 2017-1.3.1-VKE-2017-00025 project. This research has been supported by the Project: “Support of research and development activities of the J. Selye University in the field of Digital Slovakia and creative industry” of the Research and Innovation Operational Programme (ITMS code: NFP313010T504) co-funded by the European Regional Development Fund.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

SBAHC	Simultaneous aerobic biodegradation of azo dyes and hexavalent chromium
GEP	Gene expression programming
RF	Random forest
SVR	Supportive vector regression
FOA	Fruit fly optimization algorithm
SRM	Structural risk minimization
ERM	Empirical risk minimization
CC	Correlation coefficient
IP	Incubation period
DO	Dissolved oxygen
DT	Decision tree
ML	Machine learning
SI	Scattered index
WI	Willmott agreement index

References

- Sandrin, T.R.; Maier, R.M. Impact of metals on the biodegradation of organic pollutants. *Environ. Health Perspect.* **2003**, *111*, 1093–1100. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vishwakarma, S.K.; Singh, M.P.; Srivastava, A.K.; Pandey, V.K. Azo dye (direct blue 14) decolorization by immobilized extracellular enzymes of *Pleurotus* species. *Cell. Mol. Boil.* **2012**, *58*, 21–25.
- Simmons, J.E.; Yang, R.S.H.; Berman, E. Evaluation of the nephrotoxicity of complex mixtures containing organics and metals: Advantages and disadvantages of the use of real-world complex mixtures. *Environ. Health Perspect.* **1995**, *103* (Suppl. 1), 67–71. [\[PubMed\]](#)
- Saratale, R.G.; Saratale, G.D.; Chang, J.S.; Govindwar, S. Bacterial decolorization and degradation of azo dyes: A review. *J. Taiwan Inst. Chem. Eng.* **2011**, *42*, 138–157. [\[CrossRef\]](#)
- Ali, N.; Hameed, A.; Siddiqui, M.; Ghumro, P.; Ahmed, S. Application of *Aspergillus niger* SA1 for the enhanced bioremoval of azo dyes in simulated textile effluent. *Afr. J. Biotechnol.* **2009**, *8*, 3839–3845.
- Moawad, W.M.; El-Rahim, W.M.A.; Khalafallah, M. Evaluation of biotoxicity of textile dyes using two bioassays. *J. Basic Microbiol.* **2003**, *43*, 218–229. [\[CrossRef\]](#)
- Baath, E. Thymidine incorporation into macromolecules of bacteria extracted from soil by homogenization centrifugation. *Soil Boil. Biochem.* **1992**, *24*, 1157–1165. [\[CrossRef\]](#)
- Bader, J.L.; González, G.; Goodell, P.C.; Pillai, S.D.; Ali, A.S.; Pillai, S.D. Chromium-resistant bacterial populations from a site heavily contaminated with hexavalent chromium. *Water Air Soil Pollut.* **1999**, *109*, 263–276. [\[CrossRef\]](#)
- Huq, S.I. Critical environmental issues relating to tanning industries in Bangladesh. In Proceedings of the ACIAR, Coimbatore, India, 31 January–4 February 1998; pp. 22–28.
- Mahmood, S.; Khalid, A.; Mahmood, T.; Arshad, M.; Ahmad, R. Potential of newly isolated bacterial strains for simultaneous removal of hexavalent chromium and reactive black-5 azo dye from tannery effluent. *J. Chem. Technol. Biotechnol.* **2012**, *88*, 1506–1513. [\[CrossRef\]](#)
- Ho, W.S.W.; Poddar, T.K. New membrane technology for removal and recovery of chromium from waste waters. *Environ. Prog.* **2001**, *20*, 44–52. [\[CrossRef\]](#)
- Onwosi, O.; Odibo, F.J.C. Use of response surface design in the optimization of starter cultures for enhanced rhamnolipid production by *Pseudomonas nitroreducens*. *Afr. J. Biotechnol.* **2013**, *12*, 2611–2617.
- Abbasi, H.; Hamed, M.M.; Lotfabad, T.B.; Zahiri, H.S.; Sharafi, H.; Masoomi, F.; Moosavi-Movahedi, A.A.; Ortiz, A.; Amanlou, M.; Noghabi, K.A. Biosurfactant-producing bacterium, *Pseudomonas aeruginosa* MA01 isolated from spoiled apples: Physicochemical and structural characteristics of isolated biosurfactant. *J. Biosci. Bioeng.* **2012**, *113*, 211–219. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lotfy, W.A.; Ghanem, K.M.; El-Helow, E.R. Citric acid production by a novel *Aspergillus niger* *Klebseilla pneumoniae*. II. Optimization of process parameters through statistical experimental designs. *Bioresour. Technol.* **2007**, *98*, 3470–3477. [\[CrossRef\]](#) [\[PubMed\]](#)
- Tanyildizi, S.; Ozer, D.; Elibol, M. Optimization of α -amylase production by *Bacillus* sp. using response surface methodology. *Process Biochem.* **2005**, *40*, 2291–2296. [\[CrossRef\]](#)
- Locner, H.; Matar, J.E. *Designing for Quality*; Productivity Press: New York, NY, USA, 1990.
- Yang, W.; Tarng, Y. Design optimization of cutting parameters for turning operations based on the Taguchi method. *J. Mater. Process. Technol.* **1998**, *84*, 122–129. [\[CrossRef\]](#)
- Kalyani, L.T.; Sireesha, G.N.; Aditya, A.K.G.; Sankar, G.G.; Prabhakar, T. Production optimization of rhamnolipid biosurfactant by *Streptomyces coelicoflavus* (NBRC 15399T) using Plackett–Burman design. *Eur. J. Biotechnol. Biosci.* **2014**, *1*, 7–13.
- Zhao, F.; Mandlaa, M.; Hao, J.; Liang, X.; Shi, R.; Han, S.; Zhang, Y. Optimization of culture medium for anaerobic production of rhamnolipid by recombinant *Pseudomonas stutzeri* Rhl for microbial enhanced oil recovery. *Lett. Appl. Microbiol.* **2014**, *59*, 231–237. [\[CrossRef\]](#)
- Mabrouk, M.E.; Youssif, E.M.; Sabry, S.A. Biosurfactant production by a newly *Klebseilla pneumoniae* soft coral-associated marine *Bacillus* sp. E34: Statistical optimization and characterization. *Life Sci. J.* **2014**, *11*, 756–768.
- Amodu, O.S.; Ntwampe, S.K.; Ojumu, T.V. Optimization of biosurfactant production by *Bacillus licheniformis* STK 01 grown exclusively on *Beta vulgaris* waste using response surface methodology. *BioResources* **2014**, *9*, 5045–5065. [\[CrossRef\]](#)

22. Chandankere, R.; Yao, J.; Masakorala, K.; Jain, A.K.; Kumar, R. Enhanced production and characterization of biosurfactant produced by a newly isolated *Bacillus amyloliquefaciens* USTBb using response surface methodology. *Int. J. Curr. Microbiol. Appl. Sci.* **2014**, *3*, 66–80.
23. Sen, R.; Swaminathan, T. Application of response-surface methodology to evaluate the optimum environmental conditions for the enhanced production of surfactin. *Appl. Microbiol. Biotechnol.* **1997**, *47*, 358–363. [[CrossRef](#)]
24. Mnif, I.; Sahnoun, R.; Ellouze-Chaabouni, S.; Ghribi, D. Evaluation of *B. subtilis* SPB1 biosurfactants' potency for diesel-contaminated soil washing: Optimization of oil desorption using Taguchi design. *Environ. Sci. Pollut. Res.* **2014**, *21*, 851–861. [[CrossRef](#)] [[PubMed](#)]
25. Mutalik, S.R.; Vaidya, B.K.; Joshi, R.M.; Desai, K.M.; Nene, S.N. Use of response surface optimization for the production of biosurfactant from *Rhodococcus* spp. MTCC 2574. *Bioresour. Technol.* **2008**, *99*, 7875–7880. [[CrossRef](#)] [[PubMed](#)]
26. Chen, H.C. Optimizing the concentrations of carbon, nitrogen and phosphorous in a citric acid fermentation with response surface method. *Food Biotechnol.* **1996**, *10*, 13–27. [[CrossRef](#)]
27. Zounemat-Kermani, M.; Rajaei, T.; Ramezani-Charmahineh, A.; Adamowski, J.F. Estimating the aeration coefficient and air demand in bottom outlet conduits of dams using GEP and decision tree methods. *Flow Meas. Instrum.* **2017**, *54*, 9–19. [[CrossRef](#)]
28. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
29. Cheng, K.; Lu, Z. Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression. *Comput. Struct.* **2018**, *194*, 86–96. [[CrossRef](#)]
30. Pan, W.T. A new fruit fly optimization algorithm: Taking the financial distress model as an example. *Knowl. Based Syst.* **2012**, *26*, 69–74. [[CrossRef](#)]
31. Abidin, Z.; Hamzah, M.; Arshad, M.; Ngah, U. A calibration framework for swarming ASVs' system design. *Indian J. Mar. Sci.* **2012**, *41*, 581–588.
32. Liu, Y.; Wang, X.; Li, Y. A modified fruit-fly optimization algorithm aided PID controller designing. In Proceedings of the Intelligent Control and Automation (WCICA), 10th World Congress on IEEE, Beijing, China, 6–8 July 2012; pp. 233–238.
33. Ahmad, Z.; Arshad, M.; Asghar, H.N.; Sheikh, M.A.; Crowley, D.E. Isolation, screening and functional characterization of biosurfactant producing bacteria isolated from crude oil contaminated site. *Int. J. Agric. Boil.* **2016**, *18*, 542–548.
34. Desai, C.; Jain, K.; Patel, B.; Madamwar, D. Efficacy of bacterial consortium-AIE2 for contemporaneous Cr (VI) and azo dye bioremediation in batch and continuous bioreactor systems, monitoring steady-state bacterial dynamics using qPCR assays. *Biodegradation* **2009**, *20*, 813. [[CrossRef](#)] [[PubMed](#)]
35. Ferreira, C. *Gene Expression Programming in Problem Solving*, in *Soft Computing and Industry*; Springer: New York, NY, USA, 2002; pp. 635–653.
36. Baghban, A.; Mosavi, A. Insight into the antiviral activity of synthesized schizonepetin derivatives: A theoretical investigation. *Sci. Rep.* **2020**, *25*, 1–5. [[CrossRef](#)] [[PubMed](#)]
37. Suykens, J.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
38. Karballaezadeh, N.; Mohammadzadeh, S.D.; Shamshirband, S.; Hajikhodaverdikhan, P.; Mosavi, A.; Chau, K.W. Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road). *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 188–198. [[CrossRef](#)]
39. Ebtehaj, I.; Bonakdari, H.; Zaji, A.H.; Azimi, H.; Sharifi, A. Gene expression programming to predict the discharge coefficient in rectangular side weirs. *Appl. Soft Comput.* **2015**, *35*, 618–628. [[CrossRef](#)]
40. Lopes, H.S.; Weinert, W.R. EGIPSYS: An enhanced gene expression programming approach for symbolic regression problems. *Int. J. Appl. Math. Comput. Sci.* **2004**, *14*, 375–384.
41. Ferreira, C. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*; Springer: New York, NY, USA, 2006; Volume 21.
42. Yadav, B.; Ch, S.; Mathur, S.; Adamowski, J. Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach. *J. Hydrol.* **2016**, *543*, 373–385. [[CrossRef](#)]
43. Cao, Q.; Leung, K.M. Prediction of chemical biodegradability using support vector classifier optimized with differential evolution. *J. Chem. Inf. Model.* **2014**, *54*, 2515–2523. [[CrossRef](#)]

44. Tang, L.; Zhang, M.; Wen, L. Support vector machine classification of seismic events in the Tianshan orogenic belt. *J. Geophys. Res. Solid Earth* **2020**, *125*, e2019JB018132. [\[CrossRef\]](#)
45. Xiao, C.; Hao, K.; Ding, Y. An improved fruit fly optimization algorithm inspired from cell communication mechanism. *Math. Probl. Eng.* **2015**, *2015*, 492195. [\[CrossRef\]](#)
46. Nabipour, N.; Daneshfar, R.; Rezvanjou, O.; Mohammadi-Khanaposhtani, M.; Baghban, A.; Xiong, Q.; Li, L.K.; Habibzadeh, S.; Doranehgard, M.H. Estimating biofuel density via a soft computing approach based on intermolecular interactions. *Renew. Energy* **2020**, *152*, 1086–1098. [\[CrossRef\]](#)
47. Qasem, S.N.; Samadianfard, S.; Kheshtgar, S.; Jarhan, S.; Kisi, O.; Shamshirband, S.; Chau, K.W. Modeling monthly pan evaporation using wavelet support vector regression and wavelet artificial neural networks in arid and humid climates. *Eng. Appl. Comput. Fluid* **2019**, *13*, 177–187. [\[CrossRef\]](#)
48. Samadianfard, S.; Majnooni-Heris, A.; Qasem, S.N.; Kisi, O.; Shamshirband, S.; Chau, K.W. Daily global solar radiation modeling using data-driven techniques and empirical equations in a semi-arid climate. *Eng. Appl. Comput. Fluid Mech.* **2019**, *13*, 142–157. [\[CrossRef\]](#)
49. Liu, W.; Liu, C.; Liu, L.; You, Y.; Jiang, J.; Zhou, Z.; Dong, Z. Simultaneous decolorization of sulfonated azo dyes and reduction of hexavalent chromium under high salt condition by a newly isolated salt-tolerant strain *Bacillus circulans* BWL1061. *Ecotoxicol. Environ. Saf.* **2017**, *141*, 9–16. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Thacker, U.; Madamwar, D. Reduction of toxic chromium and partial localization of chromium reductase activity in bacterial isolate DM1. *World J. Microbiol. Biotechnol.* **2005**, *21*, 891–899. [\[CrossRef\]](#)
51. Halmi, M.I.E.; Abdullah, S.R.S.; Shukor, M.Y. Characterization of Chromate Reducing *Pseudomonas Aeruginosa* Strain Mie3 Isolated from Juru River Sludge and its Potential on Azo Dye Decolorization. *J. Chem. Pharmac. Sci.* **2017**, *10*, 522–526.
52. Chang, B.V.; Chao, W.L.; Yeh, S.L.; Kuo, D.L.; Yang, C.W. Biodegradation of Sulfamethoxazole in Milkfish (*Chanos chanos*) Pond Sediments. *Appl. Sci.* **2019**, *9*, 4000. [\[CrossRef\]](#)
53. Shamshirband, S.; Hadipoor, M.; Baghban, A.; Mosavi, A.; Bukor, J.; Várkonyi-Kóczy, A.R. Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases. *Mathematics* **2019**, *7*, 965. [\[CrossRef\]](#)
54. Nabipour, N.; Mosavi, A.; Baghban, A.; Shamshirband, S.; Felde, I. Extreme learning machine-based model for Solubility estimation of hydrocarbon gases in electrolyte solutions. *Processes* **2020**, *8*, 92. [\[CrossRef\]](#)
55. Amato, F.; Moscato, V.; Picariello, A.; Sperli'i, G. Extreme events management using multimedia social networks. *Future Gener. Comput. Syst.* **2019**, *94*, 444–452. [\[CrossRef\]](#)



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).