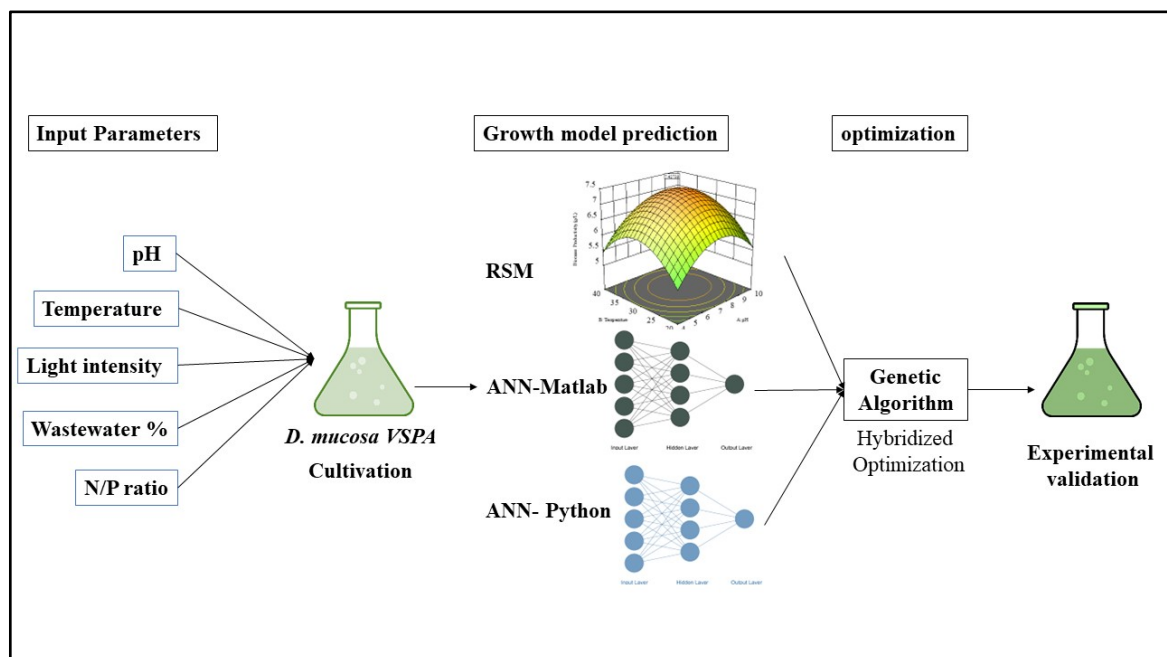


Chapter 4

A multiple input optimisation of *Diplosphaera mucosa* VSPA Biomass Production and Bioremediation Using artificial neural network- genetic algorithms



A multiple input optimisation of *Diplosphaera mucosa* VSPA biomass production and bioremediation using artificial neural network- genetic algorithms

4.1. Introduction

The RSM-GA and ANN-GA hybrid approaches are computational techniques that combine RSM or ANN models with GA to solve optimisation and modelling problems. In the RSM-GA hybrid, RSM is used to create a mathematical approximation of the response surface, which is then optimised using GA for improved efficiency. In the ANN-GA hybrid, ANN acts as the underlying model, capturing complex relationships, and GA optimises the ANN's parameters or architecture. Both approaches benefit from the global optimisation capabilities of GA and the modelling flexibility of RSM or ANN. However, they may suffer from increased computational complexity, require careful parameter tuning, and lack interpretability. Despite these challenges, RSM-GA and ANN-GA hybrid approaches offer robust solutions for challenging optimisation and modelling tasks[297].

Therefore, based on the discussed hypothesis, the current study aimed to optimise different input parameters affecting the growth of *Diplosphaera mucosa* VSPA in textile and carpet effluent. Five critical parameters were optimised: temperature, pH, light intensity, wastewater composition and N/P ratio for increasing biomass production and nutrient removal efficiency. The optimisation process started with generating different models using RSM and ANN. After that, models were solved using GA to determine the input parameters' optimised values, leading to high biomass production and nutrient removal efficiency.

4.2. Materials and Methods

4.2.1. Strain isolation and characterisation

D. mucosa VSPA strain was isolated from the inlet of the Sewage Treatment Plant located at Bhagwanpur, Varanasi, India (25° 16' 21" N, 83° 0' 16.92" E). The isolation,

sequencing, and characterisation procedures were that used by [298]. The sequence of the strain was analysed using the BLAST program. The strain showed close similarity with *D. mucosa* (89.60%), therefore named *D. mucosa* VSPA. To the best of the author's knowledge, no available literature has optimised input parameters for the growth of *D. mucosa* VSPA in carpet and textile industry wastewaters.

4.2.2. Wastewater Collection and Characterisation

The collection and characterisation of wastewater were performed according to standard and internally accepted methods [299]. Carpet effluent was collected from local carpet industries in Bhadohi, India (25° 22' 48" N, 82° 35' 24" E). While textile effluent was sampled from local textile industries in Varanasi, India (25° 18' 0" N, 82° 55' 48" E). The temperature of the samples was recorded at the collection site using a digital thermometer. After that, sample bottles were immediately sealed and packed with dry ice. Bottles were directly opened in the lab, and samples were characterised for various parameters. Initial concentration of ammonium nitrogen ($\text{NH}_4^+\text{-N}$), nitrate-nitrogen ($\text{NO}_3^-\text{-N}$), phosphate-phosphorus ($\text{PO}_4^{3-}\text{-P}$), COD and BOD was found to be 54.6 ± 3.26 mg/L, 2.12 ± 0.03 mg/L, 6.3 ± 0.16 mg/L, 372 ± 16.32 mg/L and 87.2 ± 3.26 mg/L in textile effluent, respectively. While, in carpet effluent, they were 82.8 ± 1.63 mg/L, 3.5 ± 0.24 mg/L, 9.3 ± 2.44 mg/L, 252 ± 9.79 mg/L and 1570 ± 16.32 mg/L.

4.2.3. Experimental set-up

The strain was cultivated in textile and carpet effluent under batch modes in a 500 ml Erlenmeyer flask with a working volume of 200 ml. For inoculum preparation, the stock culture of the strain was cultivated in 20 ml BBM medium (10% of the working volume) in a 100 ml flask. Growth was monitored until the mid-log phase, and actively growing cells were used to inoculate each experimental run. This step ensured that a uniform concentration of inoculum was supplied at the start of each experimental run. The input

parameters pH, temperature, light intensity, wastewater composition and N/P ratio were varied according to the RSM and ANN experimental design, individually for carpet and textile effluents as shown in **Table 4.1** and **4.2**. Flasks were kept in a shaking incubator (Remi make, India) at 120 rpm shaking speed, which was used for varying the temperature. pH was varied manually through stock solutions of 1 N HCl and 1 N NaOH and measured using a digital pH meter (Eutech pH Tutor make, US). The light was supplied using 30 W cool fluorescent tubes (Philips make, India), and intensity was regulated using the electric regulator (Anchor make, India). The light source's intensity was measured using a lux meter (Sigma make, China). Photoperiod (light/dark cycle) duration was maintained at 16 hours. Various studies have reported that a photoperiod of 16 hours (16:8 light/dark cycles) positively impacts microalgae growth instead of natural photoperiod, i.e. 12 hours [300], [301]. Wastewater was diluted with distilled water to vary its composition in terms of BOD, COD, N and P content. The N/P ratio of the wastewater samples was varied using standard stocks of NH_4Cl and KH_2PO_4 . Cultivation continued until the stationary phase was achieved, approximately ten days during each trial. The stationary phase was determined by measuring the optical density (OD) at 680nm of the samples withdrawn from the flasks at regular intervals. All experiments were performed in triplicate and average values were used for further optimisation. The same experimental set-up was used for experimentally verifying the optimal values of process parameters obtained after optimisation.

Table 4.1. Experimental design and predicted values of biomass concentration by RSM and ANN during cultivation in carpet effluent.

S.No.	A	B	C	D	E	Biomass Concentration Prediction (g/L)						
						Exp.	RSM	SE	ANN (MATLAB)	SE	ANN (Python)	SE.
1	10	40	60	40	5	3.10	3.07	0.00	3.10	0.00	3.09	0.00
2	4	40	180	40	5	3.40	3.41	0.00	3.46	0.00	3.38	0.00
3	10	20	180	40	5	2.70	2.85	0.02	2.70	0.00	2.70	0.00
4	7	30	0	60	15	2.90	3.37	0.22	2.90	0.00	2.90	0.00
5	4	20	180	40	5	2.90	2.86	0.00	3.34	0.19	2.94	0.00
6	14	30	120	60	15	2.50	2.63	0.02	2.50	0.00	2.50	0.00
7	7	30	120	60	15	6.80	7.16	0.13	7.00	0.04	7.00	0.04
8	7	30	120	60	15	6.90	7.16	0.07	7.00	0.01	7.00	0.01
9	10	40	60	80	5	3.80	3.84	0.00	3.80	0.00	3.80	0.00
10	0	30	120	60	15	1.60	1.95	0.12	1.68	0.01	1.68	0.01
11	7	30	120	60	15	6.80	7.26	0.21	7.00	0.04	6.80	0.00
12	4	40	180	80	25	4.80	4.61	0.04	4.80	0.00	4.80	0.00
13	4	20	60	40	5	2.90	2.86	0.00	2.90	0.00	2.90	0.00
14	10	40	180	40	25	5.20	5.10	0.01	5.47	0.08	5.27	0.01
15	4	40	60	40	5	3.30	3.23	0.00	3.30	0.00	3.33	0.00
16	10	40	180	80	5	4.40	4.37	0.00	4.40	0.00	4.40	0.00
17	7	6	120	60	15	1.30	1.40	0.01	1.30	0.00	1.31	0.00
18	4	20	60	40	25	4.10	4.01	0.01	4.14	0.00	4.14	0.00
19	10	40	180	80	25	5.60	5.77	0.03	6.47	0.77	6.07	0.23
20	7	30	120	60	15	7.10	7.16	0.00	7.00	0.01	7.10	0.00
21	10	20	180	80	5	3.80	3.71	0.01	3.80	0.00	3.43	0.14
22	10	20	60	80	25	5.40	5.31	0.01	5.40	0.00	5.30	0.01
23	4	40	60	80	25	4.60	4.57	0.00	4.60	0.00	4.50	0.01
24	10	40	60	80	25	5.60	5.41	0.04	5.60	0.00	5.80	0.04
25	4	40	60	80	5	3.90	3.92	0.00	3.90	0.00	4.00	0.01
26	10	20	180	80	25	5.50	5.49	0.00	5.05	0.20	5.25	0.06
27	10	40	60	40	25	4.90	4.77	0.02	4.90	0.00	4.90	0.00
28	4	40	180	80	5	4.10	4.13	0.00	4.10	0.00	4.30	0.04
29	10	20	60	40	25	4.60	4.59	0.00	4.60	0.00	4.40	0.04
30	7	53	120	60	15	2.10	2.18	0.01	2.01	0.01	2.21	0.01
31	7	30	120	60	38	5.80	6.06	0.07	5.80	0.00	5.60	0.04
32	4	20	180	40	25	3.90	3.84	0.00	3.90	0.00	3.94	0.00
33	7	30	120	60	15	7.10	7.16	0.00	7.00	0.01	6.90	0.04
34	4	20	180	80	25	4.50	4.50	0.00	4.50	0.00	4.60	0.01
35	7	30	120	60	15	7.20	7.16	0.00	7.00	0.04	7.24	0.00
36	7	30	120	12	15	5.50	5.64	0.02	5.40	0.01	5.50	0.00
37	10	20	180	40	25	4.90	4.75	0.02	4.90	0.00	4.90	0.00
38	4	20	180	80	5	3.50	3.65	0.02	3.50	0.00	3.40	0.01
39	10	40	180	40	5	3.60	3.58	0.00	3.60	0.00	3.50	0.01

40	7	30	262	60	15	3.70	3.60	0.01	3.70	0.00	3.60	0.01	
41	4	40	60	40	25	4.10	4.01	0.01	4.29	0.04	4.19	0.01	
42	7	30	120	60	15	7.00	7.16	0.03	7.00	0.00	7.00	0.00	
43	7	30	120	100	15	7.30	7.34	0.00	7.30	0.00	7.10	0.04	
44	4	40	180	40	25	3.90	4.02	0.01	3.90	0.00	4.00	0.01	
45	10	20	60	80	5	3.40	3.36	0.00	3.40	0.00	3.45	0.00	
46	4	20	60	80	5	3.70	3.62	0.01	3.32	0.15	3.32	0.15	
47	7	30	120	60	15	8.30	7.16	1.30	8.11	0.04	8.41	0.01	
48	7	30	120	60	0	3.10	3.02	0.01	3.21	0.01	3.11	0.00	
49	4	20	60	80	25	4.70	4.65	0.00	4.70	0.00	4.80	0.01	
50	10	20	60	40	5	2.60	2.52	0.01	2.20	0.16	2.40	0.04	
							MSE	0.05				0.04	0.02
							RMSE	0.22				0.19	0.14

A: pH; B: Temperature; C: Light Intensity; D: Wastewater percentage; E: N/P Ratio; SE: Standard Error; MSE: Mean Square Error; RMSE: Root Mean Square Error;

Table 4.2. Experimental design and predicted values of biomass concentration by RSM and ANN during cultivation in textile effluent.

S.No.	A	B	C	D	E	Biomass Concentration Prediction (g/L)						
						Exp.	RSM	SE	ANN (MATLAB)	SE	ANN (Python)	
											SE	SE.
10	40	60	40	5	2.6	2.49	0.01	1.90	0.49	2.56	0.00	10
4	40	180	40	5	2.8	2.74	0.00	2.78	0.00	2.76	0.00	4
10	20	180	40	5	2.1	2.32	0.05	2.10	0.00	2.13	0.00	10
7	30	0	60	15	2.3	2.86	0.31	2.35	0.00	2.32	0.00	7
4	20	180	40	5	2.3	2.26	0.00	2.31	0.00	2.34	0.00	4
14	30	120	60	15	2.1	2.36	0.07	2.12	0.00	2.10	0.00	14
7	30	120	60	15	6.2	6.05	0.02	6.01	0.04	6.18	0.00	7
7	30	120	60	15	6.3	6.05	0.06	6.01	0.08	6.04	0.07	7
10	40	60	80	5	3.3	3.26	0.00	3.33	0.00	3.31	0.00	10
0	30	120	60	15	0.88	1.35	0.22	0.89	0.00	0.89	0.00	0
7	30	120	60	15	6.2	6.05	0.02	6.01	0.04	6.21	0.00	7
4	40	180	80	25	4.2	3.77	0.18	4.14	0.00	4.20	0.00	4
4	20	60	40	5	2.3	2.21	0.01	2.27	0.00	2.29	0.00	4
10	40	180	40	25	4.6	4.33	0.07	4.50	0.01	4.60	0.00	10
4	40	60	40	5	2.8	2.63	0.03	2.68	0.01	2.78	0.00	4
10	40	180	80	5	3.8	3.70	0.01	3.81	0.00	3.81	0.00	10
7	6	120	60	15	0.7	0.82	0.01	0.69	0.00	0.69	0.00	7
4	20	60	40	25	3.5	3.24	0.07	3.67	0.03	3.67	0.03	4
10	40	180	80	25	4.9	4.97	0.00	4.94	0.00	4.94	0.00	10
7	30	120	60	15	6.5	6.05	0.20	6.61	0.01	6.61	0.01	7
10	20	180	80	5	3.1	3.04	0.00	2.82	0.08	2.82	0.08	10
10	20	60	80	25	4.7	4.57	0.02	4.53	0.03	4.53	0.03	10
4	40	60	80	25	4.2	3.89	0.10	4.20	0.00	4.20	0.00	4
10	40	60	80	25	5.1	4.67	0.18	5.07	0.00	5.07	0.00	10

4	40	60	80	5	3.4	3.39	0.00	3.12	0.08	3.12	0.08	4
10	20	180	80	25	5	4.80	0.04	5.36	0.13	5.36	0.13	10
10	40	60	40	25	4.2	3.94	0.07	3.53	0.45	3.53	0.45	10
4	40	180	80	5	3.5	3.41	0.01	2.86	0.41	3.46	0.00	4
10	20	60	40	25	3.9	4.80	0.81	3.91	0.00	3.91	0.00	10
7	53	120	60	15	0.9	1.52	0.38	0.99	0.01	0.97	0.01	7
7	30	120	60	38	4.1	5.04	0.88	4.17	0.01	4.07	0.00	7
4	20	180	40	25	3.3	4.15	0.72	3.19	0.01	3.29	0.00	4
7	30	120	60	15	5.5	6.05	0.30	6.01	0.26	6.01	0.26	7
4	20	180	80	25	3.9	3.82	0.01	3.52	0.14	3.92	0.00	4
7	30	120	60	15	5.6	6.05	0.20	5.71	0.01	5.91	0.10	7
7	30	120	12	15	4.2	4.56	0.13	4.27	0.00	4.37	0.03	7
10	20	180	40	25	4.4	4.12	0.08	4.17	0.05	4.37	0.00	10
4	20	180	80	5	2.9	3.40	0.25	2.86	0.00	2.96	0.00	4
10	40	180	40	5	3.1	3.02	0.01	3.33	0.05	3.13	0.00	10
7	30	262	60	15	3.1	3.28	0.03	3.32	0.05	3.12	0.00	7
4	40	60	40	25	3.4	3.17	0.05	3.12	0.08	3.42	0.00	4
7	30	120	60	15	5.4	6.05	0.42	5.91	0.26	5.61	0.04	7
7	30	120	100	15	5.9	6.27	0.14	5.89	0.00	5.93	0.00	7
4	40	180	40	25	3.3	3.14	0.03	3.19	0.01	3.29	0.00	4
10	20	60	80	5	2.7	2.67	0.00	2.69	0.00	2.72	0.00	10
4	20	60	80	5	3.1	3.40	0.09	3.42	0.10	3.12	0.00	4
7	30	120	60	15	6.2	6.05	0.02	6.01	0.04	6.21	0.00	7
7	30	120	60	0	2.5	2.10	0.16	2.57	0.00	2.47	0.00	7
4	20	60	80	25	4.2	3.99	0.04	4.11	0.01	4.21	0.00	4
10	20	60	40	5	1.9	1.86	0.00	1.79	0.01	1.89	0.00	10
									MSE	0.13	0.06	0.03
									RMSE	0.36	0.25	0.16

A: pH; B: Temperature; C: Light Intensity; D: Wastewater percentage; E: N/P Ratio; SE: Standard Error; MSE: Mean Square Error; RMSE: Root Mean Square Error;

4.2.4. Experimental Design

4.2.4.1. Response Surface Methodology (RSM)

Experiments were designed using the CCD (Central Composite Design) approach, and the input parameters' effect on output parameters were analysed using RSM. The CCD-RSM combined approach was performed in the Design Expert software. Five independent input parameters (temperature, pH, light intensity, wastewater composition and N/P ratio) were selected to optimise one output parameter (biomass concentration). The CCD design consists of 50 experimental runs at three coded levels, as indicated in [Table 4.3](#). The input parameter range was decided according to a search in the previous literature [302].

Table 4.3. Experimental range and levels of five input parameters for microalgae cultivation.

Factor	Input Parameter	Coded Levels		
		Low (-1)	Mean (0)	High (+1)
A	pH	4	7	10
B	Temperature (°C)	20	30	40
C	Light Intensity ($\mu\text{mol m}^{-2}$ s^{-1})	2500	5000	7500
D	Wastewater (%)	40	60	80
	(Textile Effluent: mg/L)	($\text{NH}_4^+\text{-N}$: 21.84; $\text{PO}_4^{3-}\text{-P}$: 2.52; BOD: 34.88; COD: 148.8)	($\text{NH}_4^+\text{-N}$: 32.76; $\text{PO}_4^{3-}\text{-P}$: 3.78; BOD: 52.32; COD: 223.2)	($\text{NH}_4^+\text{-N}$: 43.68; $\text{PO}_4^{3-}\text{-P}$: 5.04; BOD: 69.76; COD: 297.6)
E	(Carpet Effluent: mg/L)	($\text{NH}_4^+\text{-N}$: 33.12; $\text{PO}_4^{3-}\text{-P}$: 3.72; BOD: 41.60; COD: 628)	($\text{NH}_4^+\text{-N}$: 49.68; $\text{PO}_4^{3-}\text{-P}$: 5.58; BOD: 62.4; COD: 942)	($\text{NH}_4^+\text{-N}$: 66.24; $\text{PO}_4^{3-}\text{-P}$: 7.44; BOD: 83.2; COD: 1256)
	N/P Ratio	5	15	25

After that, RSM was used for analysing the effects of input parameters on biomass production and determining their optimum level leading to high biomass concentration. A full quadratic model was used for the analysis purpose as indicated in Eq. (4.1):

$$Y_0 = b_0 + \sum_i b_i X_i + \sum_{ii} b_{ii} X_i^2 + \sum_{ij} X_i X_j \quad (4.1)$$

where Y_0 is the biochemical response, b_0 is the offset term, X_i and X_j are the independent variables, and b_i , b_{ii} and b_{ij} are the i^{th} linear coefficient, i^{th} quadratic coefficient and ij^{th} interaction coefficient, respectively.

4.2.4.2. Artificial Neural Network (ANN)

The current study used the feed-forward ANN (FFANN) modelling technique to optimise process parameters. MATLAB (Version 2017b) and Python, both programming languages, were used for constructing ANN models. In MATLAB, ANN models were constructed using the Neural Network tool (NN). In Python, three libraries: (i) NumPy (for working with arrays), (ii) TensorFlow (for building and training machine learning models), and (iii) Pandas (for data manipulation and analysis), were used to construct ANN models. Feed-forward Artificial Neural Networks (ANNs) are a class of neural networks that consist of a series of interconnected layers of artificial neurons. The data obtained after RSM-mediated experimental runs were used for ANN modelling, which included all five input parameters and biomass concentration values (output parameter). The ANN structure consists of 3 essential layers: (i) input layer, (ii) hidden layer, and (iii) output layer, as shown in [Figure 4.1](#). All these layers are connected through “nodes” or “neurons”, through which the transfer of information occurs from one to the next layer [303]. ANN model construction starts with feeding the five input parameters into the input layer. After that, they were transferred to the hidden layer, where processing occurs (training, validation and testing). Based on the complexity of the problems, there can be more than one hidden layer and multiple neurons in the hidden layer. The number of neurons in hidden layers was decided by the trial-and-error method, where different numbers are tested until the error is close to zero. The following equation describes the connection between layers [304]:

$$Z_j = f \sum_{i=1}^N |xw + b|_i \quad (4.2)$$

where Z and x are the i^{th} input and j^{th} output, respectively, f indicates the operating function, w indicates the weight and b indicate the bias. MATLAB carried out the processing using single hidden layers with 10 neurons. While Python carried out the operation using two hidden layers having 32 neurons and 16 neurons each. Hidden layers

predicted the output based on the input, and the results were then transferred to the output layer. Here, the comparison of the experimental and predicted results is carried out, and error is predicted.

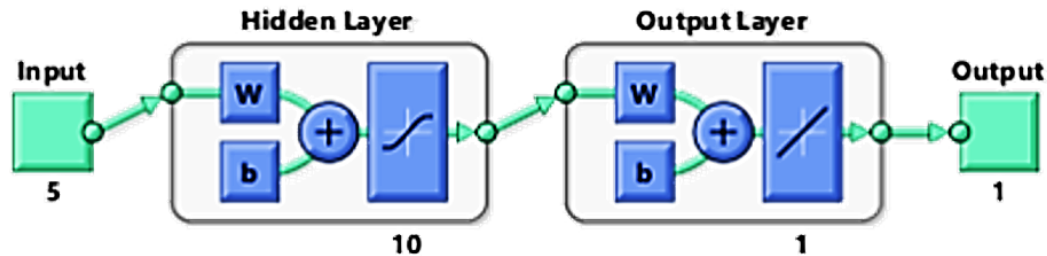


Figure 4.1. The architecture of the artificial neural network with hidden layer and output layer.

4.2.4.2.1. Training, validation, and testing

The experimental data were divided randomly into three sets: (i) Training (34 data points), (ii) Testing (8 data points), and (iii) Validation (8 data points). The division of the data was performed using the “dividerand” function. In the training step, data from both the training set was used to train the network using input and output data. The training process typically involves adjusting the model’s parameters using an optimisation algorithm. Levenberg-Marquardt algorithm, belonging to the Quasi Newton class, was used to train the network. Training can be carried out in many cycles/iterations; each iteration in ANN modelling is called an epoch. The goal is to teach the model to recognise patterns in the data and make accurate predictions. The data in the validation set was then used to fine-tune the model by testing the model’s performance after each epoch. After each validation step, the model’s hyperparameters are adjusted to improve the performance of the model. After the training step was completed, the final performance of the model was evaluated in the testing subset. Only input data were fed to the machine in the testing subset, and trained models were made to predict the output. Error is computed based on the difference between predicted and experimental results. The predicted output by the ANN models and the corresponding error

was reported for carpet and textile effluent (see supplementary material). It should be noted that data in the validation and testing set remained hidden from the machine during the training step.

4.2.4.3. Genetic Algorithm based optimisation

After constructing RSM and ANN models, they were integrated with GA to generate optimum values of all five input parameters leading to high biomass production (RSM-GA and ANN-GA). GA is a heuristic search method that draws on Darwin's theory of evolution, used to solve optimisation puzzles. When competing for scarce resources, the fittest always triumph over the weakest, and the best individuals move towards the global solution [36]. GA process first starts with the random generation of the initial population. Each individual represents a chromosome comprising five genes, representing the 5 input parameters. These chromosomes were then evaluated based on their fitness value generated using the RSM and ANN models. The chromosomes/individuals with the highest fitness values were selected and subjected to selection, crossover and mutation to generate offspring of the next generation. The hyperparameters affecting the performance of GA were selected according to the trial-and-error method till minimum error was achieved. The process was continued for 100 generations to obtain a global solution and the best fitness value. The results of the GA were then experimentally verified to determine the final accuracy of the models.

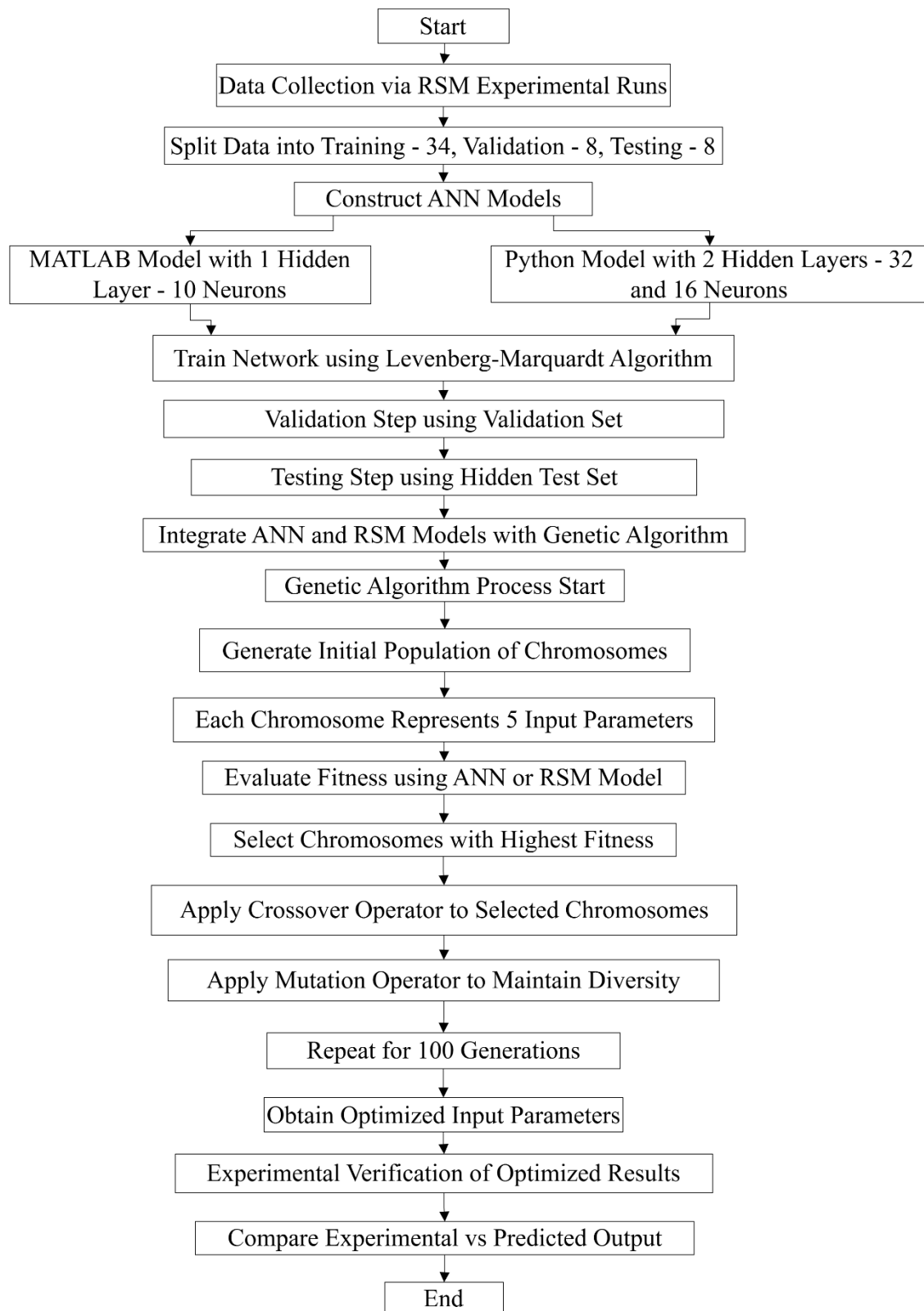


Figure 4.2. Workflow of ANN-GA-based optimization, showing model construction, GA processing, and experimental validation.

4.2.5. Analytical Techniques

The biomass concentration after each experimental run was estimated by measuring the absorbance (Abs.) of the sample at 680 nm and calculating the biomass concentration (Conc.) using Eq. (4.3):

$$Abs. = 0.1324 \times Conc. \quad (R^2: 0.98) \quad (4.3)$$

This relationship between absorbance and biomass concentration was determined experimentally. Compared to the gravimetric method, the absorbance method of biomass measurement is rapid and can provide reproducible data by correctly calibrating spectrophotometer devices [305]. After that, the remaining pollutant concentration was measured in terms of ammonium nitrogen ($\text{NH}_4^+\text{-N}$) and phosphate-phosphorus ($\text{PO}_4^{3-}\text{-P}$). $\text{NH}_4^+\text{-N}$ concentration was estimated by the Phenate method and $\text{PO}_4^{3-}\text{-P}$ by vanadomolyddophosric acid [299]. The final pollutant removal efficiency (RE) was determined by Eq.(4):

$$RE = \frac{P_o - P_f}{P_o} \times 100 \quad (4.4)$$

where RE indicates the percent removal efficiency and P_f and P_o indicate the final and initial pollutant concentration (mg L^{-1}).

4.2.6. Statistical Analysis

Each experiment was conducted in triplicate to ensure accuracy, and the average results were utilised for the further optimisation process. ANOVA was conducted to determine each input parameter's significance and interaction using the F-value and p-value (see supplementary material). The test was considered significant when p value was less than 0.05. The predictive performance of RSM and ANN models was determined via coefficient of regression (R^2), mean square error (MSE) and root mean square error (RMSE).

4.3. Results and Discussion

4.3.1. RSM Modelling

The statistical significance of the generated model, effects of individual input parameters, their squares and interactions were evaluated using p-values and F-values. The p-values of pH (A), temperature (B), light intensity (C), wastewater percentage (D), N/P ratio (E), two-way interaction of pH and N/P ratio (AE) and the squares of all input parameters (A, B, C, D, and E) were less than 0.05 at 95% confidence interval, in both the effluents. This indicates a statistically significant effect of all parameters on biomass production. Among the individual parameters, the N/P ratio was the most critical input parameter, followed by wastewater percentage. In contrast, light intensity has the least impact on biomass production while cultivating *D. mucosa VSPA* in both effluents.

The multiple regression method was applied to fit the experimental data of biomass production to the generated RSM quadratic equation. The coded quadratic equation predicting the biomass concentration during cultivation in carpet and textile effluents is given in Eq. (5) and (6), respectively:

$$X_C = 7.16 + 0.21A + 0.16B + 0.09C + 0.35D + 0.64E + 0.23AE + 0.04BC - 0.02BD - 0.09BE + 0.01CD - 0.04CE - 0.03DE - 0.88A^2 - 0.95B^2 - 0.66C^2 - 0.11D^2 - 0.46E^2 \quad (R^2 = 0.98) \quad (4.5)$$

$$X_T = -11.99 + 0.98A + 0.54B + 0.04C + 0.05D + 0.18E + 0.002AB + 0.005AC + 0.00005AD + 0.007AE + 0.0002BC - 0.0004BD - 0.001BE - 0.00001CD - 0.00005CE - 0.00004DE - 0.08A^2 - 0.008B^2 - 0.0001C^2 - 0.0002D^2 - 0.004E^2 \quad (R^2 = 0.95) \quad (4.6)$$

Here, X_C and X_T indicate biomass concentration generated during cultivation in carpet and textile effluents, respectively. These equations were used for predicting the

experimental data. Both equations can be further simplified by removing insignificant terms, as shown in Eq. (7) and Eq. (8):

$$X_C = 7.16 + 0.21A + 0.16B + 0.09C + 0.35D + 0.64E + 0.23AE - 0.88A^2 - 0.95B^2 - 0.66C^2 - 0.11D^2 - 0.46E^2 \quad (Ad.R^2 = 0.97 \& Pred.R^2 = 0.96) \quad (4.7)$$

$$X_T = -11.99 + 0.98A + 0.54B + 0.04C + 0.05D + 0.18E + 0.007AE - 0.08A^2 - 0.008B^2 - 0.0001C^2 - 0.0002D^2 - 0.004E^2 \quad (Ad.R^2 = 0.92 \& Pred.R^2 = 0.83) \quad (4.8)$$

Individual parameters and the interaction of pH and N/P ratio positively affect biomass production, while square terms have a negative impact. A high F-value and low p-value (<0.001) indicated that constructed model was significant at a 95% confidence level. There is only a 0.01% chance that this large F-value could occur due to noise during experimental trials. Both models adjusted R² values (0.97 and 0.92) indicated a strong correlation between biomass production and input parameters. The predicted (Pred.) R² of 0.96 and 0.83 for both models reasonably agree with the Adjusted (Ad.) R² of 0.97 and 0.92; i.e., the difference is less than 0.2. Thus, models can be used for the mathematical prediction of biomass production with variations in the input parameters. In addition, the lack of fit F-value of 0.87 implies the lack of fit is insignificant relative to the pure error, indicating that the model can be used to predict biomass production.

4.3.2. Response plots

2D and 3D surface plots generated after RSM model construction were analysed to better understand input parameters' effects on biomass production. 2D plots represent the individual effect of parameters on biomass production during the cultivation of *D. mucosa* *VSPA* in carpet and textile effluents, respectively (see supplementary material).

The highest biomass production was achieved in the pH 7-8 range during the cultivation of *D. mucosa VSPA* in both effluents. pH generally affects the activity of enzymes responsible for various metabolic activities occurring in the microalgal cells and the solubility of various nutrients in the medium. Every enzyme has its own pH optima range, which in turn decides the optimum pH range for microalgae growth. Most microalgal species are active around the neutral pH range (6.5-7.5) [306]. pH also determines the mechanism of CO₂ uptake by the microalgal cells. Microalgal cells can directly uptake CO₂ from the atmosphere or medium through diffusion. While at high pH, atmospheric CO₂ gets dissolved in the medium in the form of bicarbonate. The microalgae consume bicarbonate through the activity of carbonic anhydrase enzyme present in the periplasmic space [68]. There are various publications available in the literature that studied the effect of pH on microalgae growth. There altered the pH of the medium either using chemical buffers or purging CO₂ at different concentrations [195]–[197], [307]. Nayak et al. (2016) cultivated *Scenedesmus* sp. at different pH in domestic wastewater. According to their findings, the highest biomass concentration and nutrient removal efficiency was achieved at pH 7 [195]. The next important conclusion from the 2D plots revealed that the temperature of around 30° C resulted in high biomass production during the cultivation of *D. mucosa VSPA* in both effluents. Various studies revealed that most microalgae species are active around 25°-30°C [182], [308], but some can tolerate high temperatures [188]. Some can even survive below 15°C [189]. As the temperature increases from the optimum level, the rate of photorespiration increases, causing a determinantal effect on growth [190]. While at low temperature, the rate of photosynthesis decreases, ultimately causing the growth to decrease [191]. With pH, temperature also affects the activity of enzymes and the solubility of gases in the medium [192]. During the cultivation of *Chlorella vulgaris* at different temperatures in aquaculture wastewater, a high growth rate was noticed at 25°C [309].

Light intensity is one of the most crucial factors affecting microalgae growth, as it supplies the energy required for microalgae growth, mainly during the autotrophic mode of cultivation. It is necessary to provide the optimum level of light intensity. High light intensity will lead to photoinhibition, and low light intensity will not be enough for photosynthesis [310]. In the present study, a supply of light intensity of around $120 \mu\text{mol m}^{-2} \text{s}^{-1}$ was enough to increase biomass production during cultivation in both effluents. However, as per F-value, the light intensity has the least impact on the growth of *D. mucosa VSPA* compared to other input parameters. One possible reason for this phenomenon may be the presence of high organic content in both effluents (CF: 252 mg/L; TF: 164 mg/L). Microalgae may shift the mode of cultivation towards the mixotrophic or heterotrophic mode if more reduced organic carbon sources, such as acetate, are present in the medium. Thus, reducing the dependency on light. This shift in cultivation mode by microalgae has also been explained on the molecular level through differential gene expression profiles [311]. So similarly, in the present case, *D. mucosa VSPA* may have shifted towards the mixotrophic mode, utilising organic carbon sources present in both effluents. The next input parameter studied in the present case was wastewater percentage. *D. mucosa VSPA* was able to tolerate 70-80% of wastewater, achieving high biomass productivity. However, as per the trend in Figure, biomass concentration became nearly stable after 80%, indicating that *D. mucosa VSPA* may tolerate a high wastewater percentage. Effluents from industries contain various toxic chemicals that may inhibit the growth if the microalgae species are directly cultivated in 100% wastewater effluent. Therefore, they are slowly acclimatised by increasing the wastewater percentage in the medium in a step-wise manner [312]. However, some algal species can survive in the medium, having a 100% wastewater percentage with a high growth rate [313].

The last input parameter optimised in the current study was the N/P ratio. At the N/P ratio of 20-25, the highest biomass production was obtained in both effluents. The average stoichiometric formula of microalgal biomass is $C_{106}H_{181}O_{45}N_{16}P$, determined via the Redfield ratio [314]. Therefore, the N/P ratio of the medium should be around 16:1 or higher if one needs to increase the growth rate and biomass production of microalgae. But its level should be optimised as a very high N/P ratio will lead to growth inhibition due to nutrient toxicity. Mayers et al. (2014) cultivated *Nannochloropsis* sp. at different N/P ratios ranging from 16:1 to 80:1. According to their findings, high biomass productivity was noticed between the N/P ratio of 16:1 to 32:1. While, as the N/P ratio was increased beyond 64:1, decrease in productivity was noticed [315]. A very low N/P ratio is also not suitable for microalgae cultivation, as concluded by Wagner et al. (2021) in the study. They noticed a decrease in the growth rate of the *Chlorella* and *Scenedesmus* sps., when the N/P ratio declined below 5:1 [316]. The interaction between pH and N/P ratio was only significant among the two-way interaction results of input parameters ($p < 0.05$). pH influences the solubility of the nutrients in the medium. Most of the nutrient gets precipitated at high pH in the medium, directly affecting the N/P ratio. Therefore, pH should be regularly monitored and regulated, especially during photosynthesis. Microalgae uptakes inorganic carbon sources during photosynthesis, increasing the medium's pH. The surface plot representing the interaction between the pH and N/P ratio in carpet and textile effluent is represented in [Figure 4.3](#) and [4.4](#) respectively.

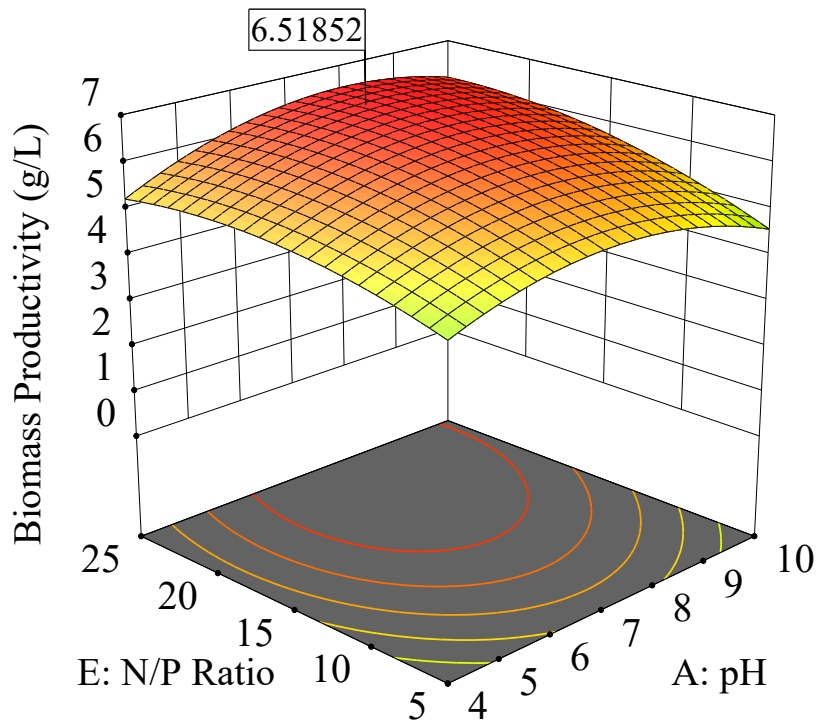


Figure 4.3. 3D surface plot representing the two-way interaction between pH and N/P ratio during cultivation of *D. mucosa VSPA* in Carpet effluent

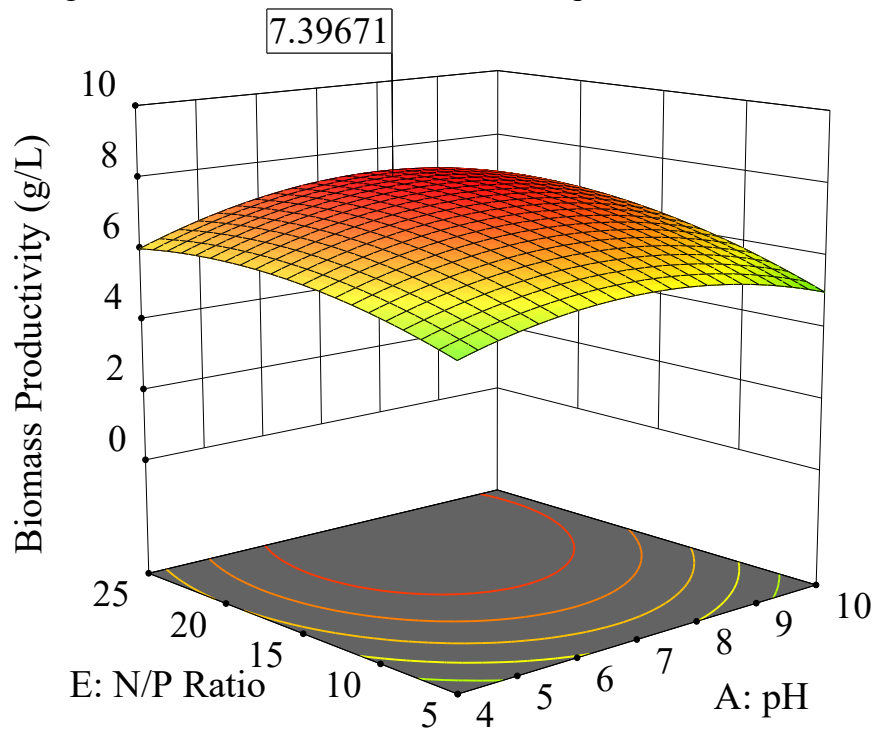


Figure 4.4. 3D surface plot representing the two-way interaction between pH and N/P ratio during cultivation of *D. mucosa VSPA* in Textile effluent.

4.3.3. ANN Modelling

ANN models for determining the correlation between input parameters and biomass production were constructed using a 5-10-1 topology (Input-Hidden-Output) in MATLAB and 5-16-1 & 5-32-1 in Python. ANN models successfully predicted the biomass concentration values during the cultivation of *D. mucosa VSPA* in carpet and textile effluent as shown in **Table 4.4**. MSE and RMSE values of prediction by ANN were less than RSM, indicating its better performance in prediction. There are various studies available in the literature that reported the better prediction performance of ANN than RSM. Hossain et al. (2022) optimised process parameters for increasing biomass productivity and nutrient removal efficiency during the cultivation of *Chlorella kessleri* in municipal wastewater. According to their findings, ANN performed better than RSM with high prediction efficiency ($R^2 = 0.99$) and minimum error (0.7) [37]. ANN also performed better than RSM during the optimisation of lipid and biodiesel productivity by microalgal species [317]. Although ANN is more accurate than RSM, these models work best together when interpreting experimental results. While RSM emphasises the statistical significance of the individual input parameters and their interactions via ANOVA, ANN is more accurate in capturing the non-linear link between the input and output parameters [38].

ANN models generated by Python performed the prediction task with less error than those generated by MATLAB. Though in the present case, Python performed better than MATLAB, it's not the same in all cases. In a basic comparison study based on numerical simulation, Python performed better than MATLAB [318]. While in another simulation study, Python presented worse results than MATLAB [319]. Depending on the task and the availability of high-level features in each system to address it, the productivity and performance results for each language differ. As a result, the optimal language to use in a given situation will be heavily influenced by the particulars of the issues that must be

resolved [39]. Therefore, more comparison studies on the optimisation problem must be conducted to conclude the best language between MATLAB and Python. The simulation curves obtained after training, validation and testing of ANN models are given in **Figure 4.5** and **4.6** for carpet and textile effluents, respectively. The ANN models were trained using 34 data points in 5 epochs (R^2 - CF:0.98 & TF:0.99). The trained models after each epoch were validated using 8 datapoints (R^2 - CF:0.98 & TF:0.97). Models were finally validated by predicting the 8 data points in the testing set (R^2 - CF:0.96 & TF:0.99).

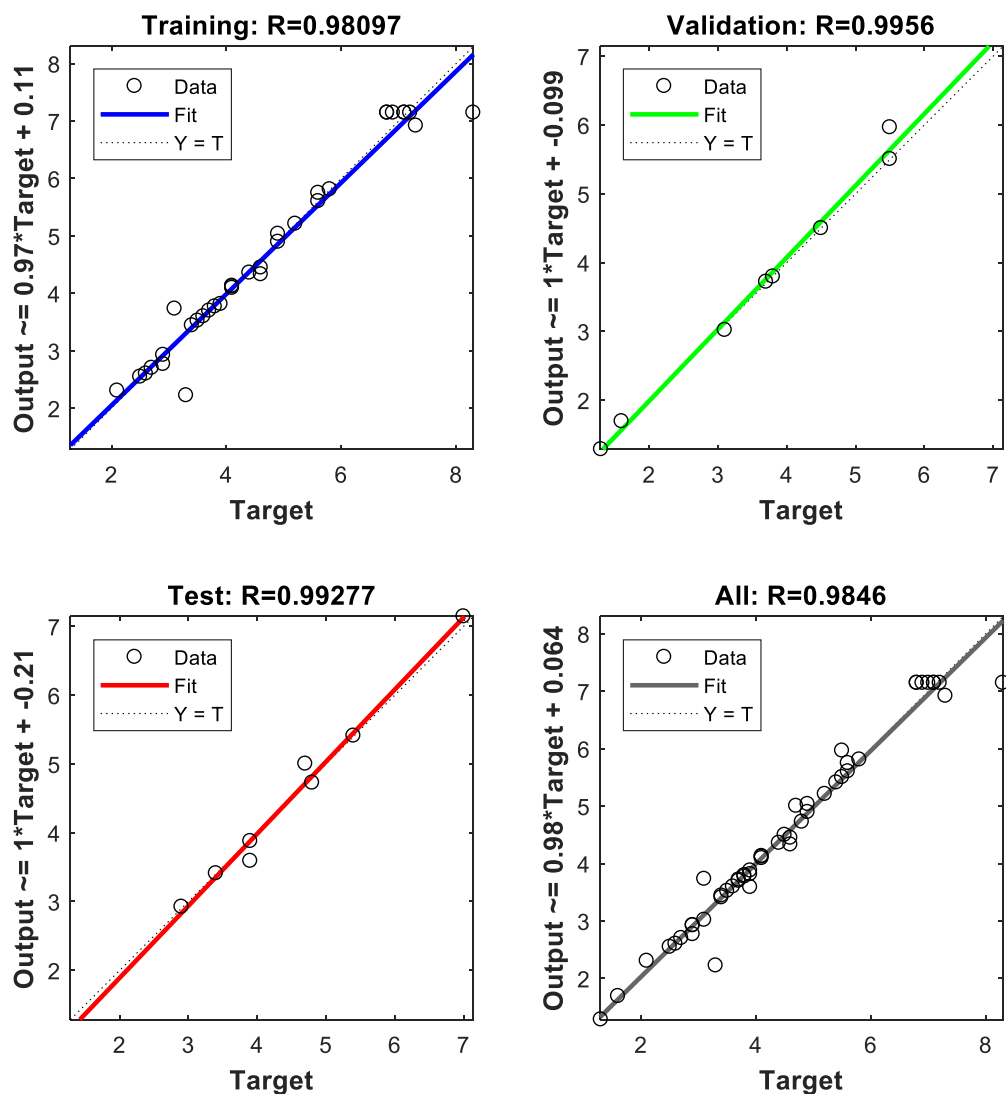


Figure 4.5. Simulation results obtained after training, validation and testing of ANN models during cultivation of *D. mucosa* VSPA in Carpet Effluent

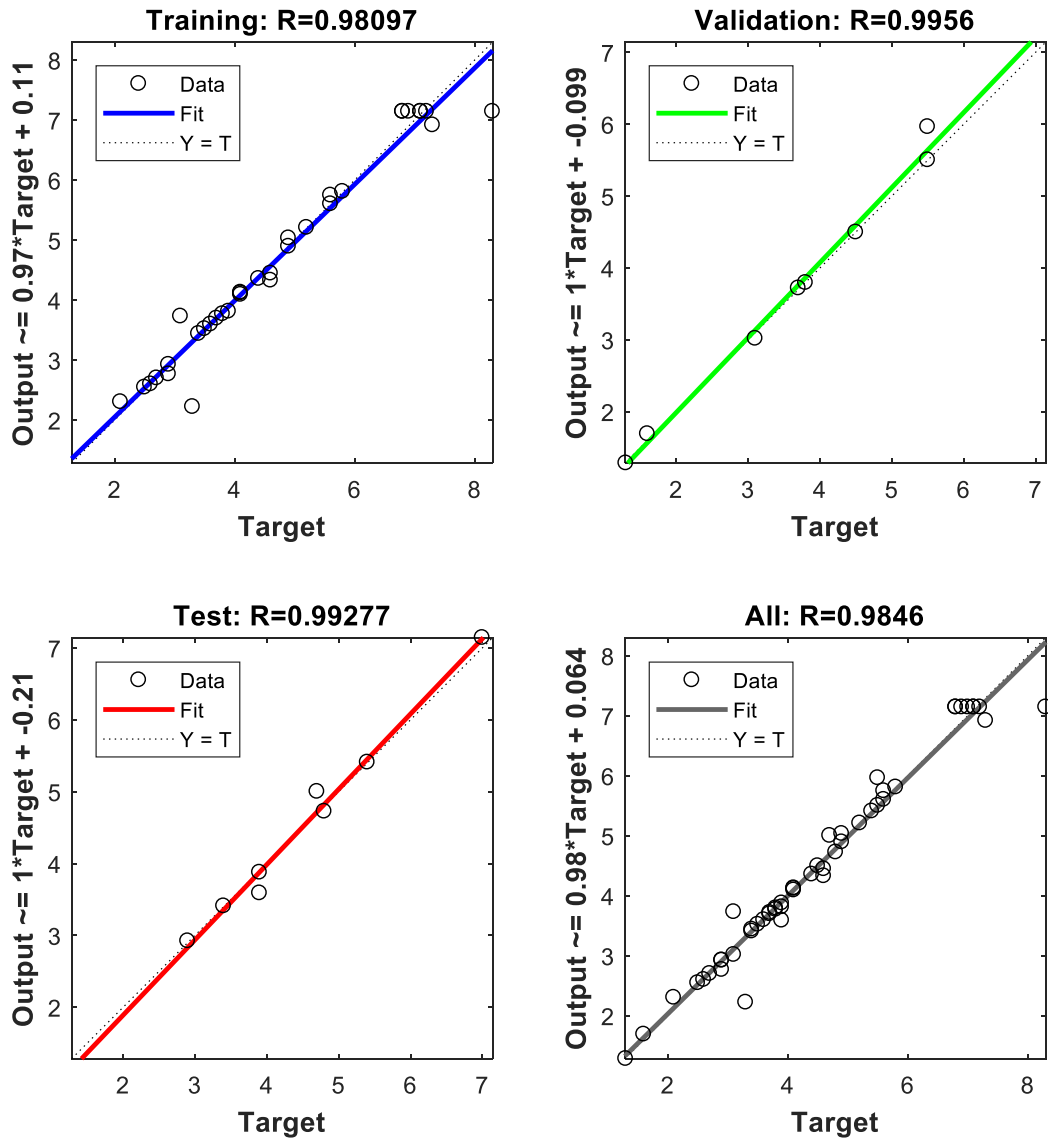


Figure 4.6. Simulation results obtained after training, validation and testing of ANN models during cultivation of *D. mucosa* VSPA in Textile effluent.

Residual histogram plots further analysed the performance of the ANN models are shown in **Figure 4.7** and **4.8** for carpet and textile effluent respectively. This plot represents the frequency of errors/residual values of training, testing and validation datasets. It is evident from the Figure that a high frequency of residuals occurred at zero for all three datasets. The frequency drastically dropped as the residual values moved away from zero in both positive and negative directions. This confirms that ANN models can predict new values with high accuracy and minimum error.

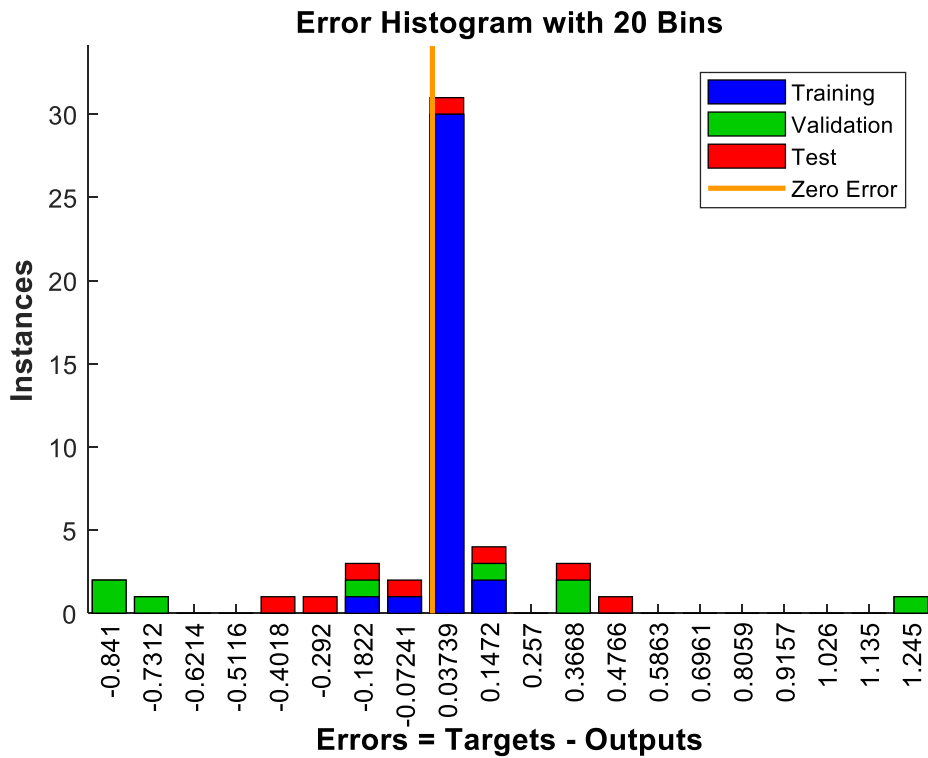


Figure 4.7. Residual histogram plots obtained after training, validation and testing of ANN models during cultivation of *D. mucosa* VSPA in Carpet Effluent

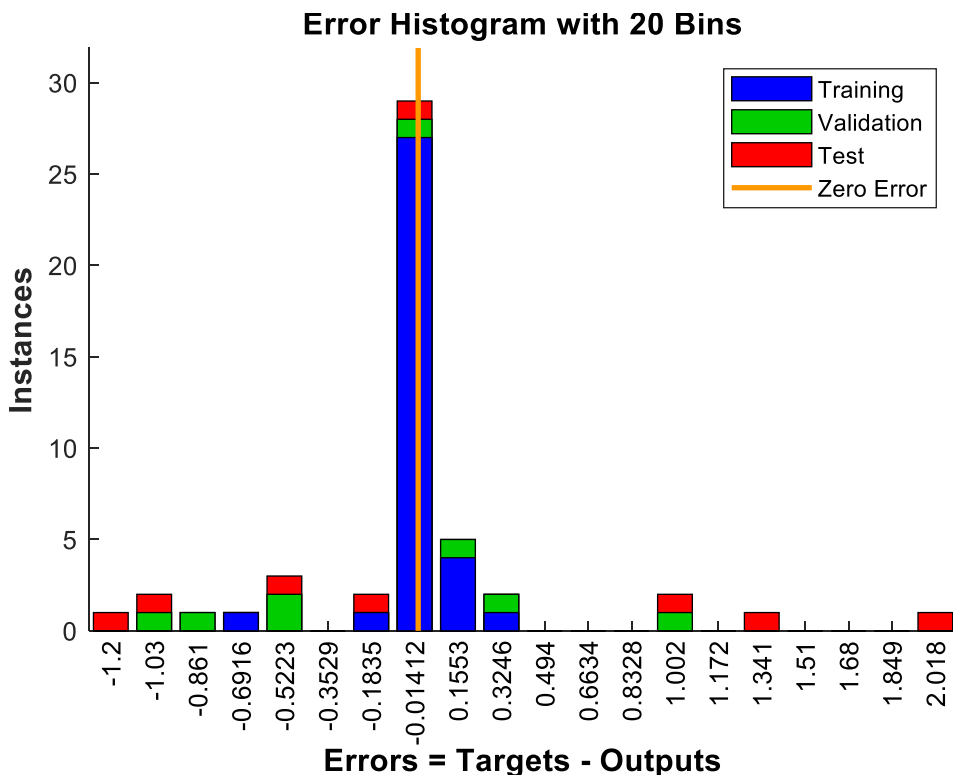


Figure 4.8. Residual histogram plots obtained after training, validation and testing of ANN models during cultivation of *D. mucosa* VSPA in Textile effluent.

Table 4.4. Comparison of the results obtained in the present study with some previously reported studies.

Species	Wastewater type	Optimization technique	R ²	Size of database	Input parameter optimized	Biomass conc. g/L	References
<i>Chlorella kessleri</i>	Municipal wastewater	RSM MLP-ANN	0.99	Not reported	Temperature (T), light-dark cycle (LD), and nitrate-to-phosphate (N:P)	0.5	S. M. Zakir Hossain et al. 2022
<i>Chlorella vulgaris</i>	BG-11 media	MLP-ANN	0.99	1925	Cultivation time and pH	2.66	Vinoj Chamilka Liyanaarachch et al. 2021
<i>Chlorella vulgaris</i>	Not reported	MLP-ANN	Not reported	Not reported	Glucose, nitrate and phosphate	3.1	M.H. Morowvat, Y. Ghasemi, et al. Clin. Res. 8 (2016)
<i>Desmodesmus sp. VV2</i>	Not reported	RSM-ANN	0.96	20	Nitrogen, Ca and MgO concentrations	1.2	Elamathi Vimali et al. 2022
<i>Chlorella kessleri</i>	Synthetic wastewater	Not reported	Not reported	Not reported	Not reported	1.2	Z. Arbib et al. 2014
<i>Scenedesmus sp.</i>	Domestic wastewater	ANN-GA	0.9947	30	Light intensity, photoperiod, temperature and initial pH	3.07	Manoranjan Nayak et al. 2018
<i>D. mucosa VSPA</i>	Textile effluent	RSM ANN-GA	0.96	50	pH, Light intensity, Temperature,	6.68	This study

4.3.4. RSM-GA and ANN-GA-based optimisation

The models generated by RSM and ANN were set as objective functions. After that, they were hybridised with GA to generate the optimised values of input parameters leading to high biomass productivity. The performance of the GA process was highly affected by values of hyperparameters, such as initial population size, elite size and mutation rate. Values of these parameters were selected by a trial-and-error method: population size (100), elite size (10) and mutation rate (0.1). The optimisation process terminated after 100 generations till the achievement of the global solution. This suggests that no more crossover or mutation exists within the individuals that could influence biomass productivity. The optimum conditions of input parameters obtained using the RSM-GA and RSM-ANN hybrid approaches have been given in [Table 4.5](#). Comparison of the results obtained in the present study with some previously reported studies have been given in [Table 4.4](#).

Table 4.5. The optimal global solution obtained after the genetic algorithm hybridisation approach and experimental verification.

Input Parameters (Coded Value)	Textile Effluent				Carpet Effluent			
	RS M	RSM -GA	ANN- GA (MATLAB)	ANN- GA (PYTHON)	RS M	RSM -GA	ANN- GA (MATLAB)	ANN- GA (PYTHON)
A	7.4 30.	7.80	7.70	7.90	7.5 32.	7.30	7.70	7.80
B	90 53	31.50	31.40	32.50	70 45	34.7	28.40	28.8
C	91 79.	4826	5304	5261	26 73.	4391	4570	4574
D	30	89.30	94.20	74.30	09	81.60	82.80	93.10

E	21.				22.			
	70	17.20	19.30	17.10	90	22.10	20.90	23.50
Output								
Parameter								
(Prediction)								
	6.5				7.3			
Biomass Conc.	1	6.52	6.61	6.68	0	7.30	7.51	7.64
Experimental								
Verification								
	6.3	6.35	6.5	6.61	7.2	7.29	7.49	7.63
Biomass Conc.	1				5			
	3.0	1.69	0.76	0.45	2.0	1.23	0.27	0.13
Error	7				5			
	95.				97.			
NRE	14	96.21	96.81	97.23	13	97.87	98.91	99.20
	76.				78.			
PRE	14	77.21	77.81	78.23	13	78.87	79.91	79.80
	86.				87.			
CRE	2	87.2	87.4	88.13	8	90.81	94.23	96.21

A: pH; B: Temperature (°C); C: Light Intensity (lux); D: Wastewater percentage; E: N/P Ratio; NRE: Nitrogen Removal Efficiency; PRE: Phosphorus Removal Efficiency; CRE: COD Removal Efficiency; Conc.: Concentration;

After obtaining the optimum values of input parameters, they were subjected to experimental verification. After each experiment, biomass concentration and pollutant removal efficiency in terms of $\text{NH}_4^+\text{-N}$ and $\text{PO}_4^{3-}\text{-P}$ was measured. As shown in [Figure 4.9](#), in all cases, the error was less than 5%, with minimum error in ANN-GA (Python) experiments. This indicates that the ANN-GA-based approach provides a robust and reliable approach for optimisation. The pollutant removal efficiency was also high in all experimental set-ups, with more than 95% removal of $\text{NH}_4^+\text{-N}$ and 75% removal of $\text{PO}_4^{3-}\text{-P}$ in both effluents. The optimisation process increased biomass production and pollutant

removal efficiency compared to that reported in the previous study. Singh et al. (2023) cultivated *D. mucosa* VSPA in textile effluent without an optimisation process. According to their findings, 6.14 g/L of biomass concentration was obtained with 90% removal of $\text{NH}_4^+\text{-N}$ and 70% of $\text{PO}_4^{3-}\text{-P}$ [298]. ANN-GA hybrid approach for the optimisation process can also be extended to optimise other process parameters such as air flow rate, CO_2 content, photoperiod and other important parameters affecting microalgae growth in wastewater.

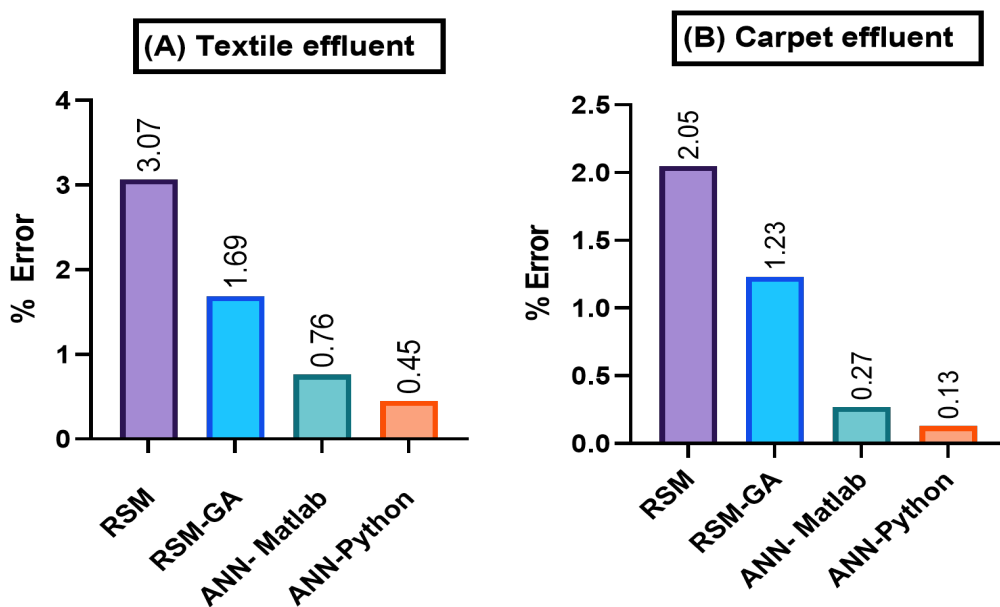


Figure 4.9. Error obtained during RSM-GA and ANN-GA based hybridization approach.

After obtaining the optimum values of the input parameters, the performance of the system was evaluated based on biomass concentration and pollutant removal efficiencies, as summarized in **Table 4.6**. The optimized conditions led to a significant improvement across all measured parameters in both textile and carpet effluents. Specifically, biomass concentration in textile effluent increased from 3.98 g/L to 6.68 g/L (a 67.83% increase), while in carpet effluent it rose from 4.26 g/L to 7.63 g/L (79.1% increase). Similarly, nitrogen removal efficiency (NRE) improved from 90.4% to 96.81% in textile effluent and 91.2% to 99.20% in carpet effluent, reflecting an enhancement of 7.09% and 8.77%,

respectively. Phosphorus removal efficiency (PRE) also showed notable increments, improving by 10.99% in textile effluent and 11.4% in carpet effluent. Additionally, chemical oxygen demand (COD) removal efficiency (CRE), which was not previously included in the summary table, was calculated and revealed improvements of 11.5% and 4.68% in textile and carpet effluents, respectively. These enhancements underscore the effectiveness of the ANN-GA-based optimisation approach in improving biomass productivity and pollutant removal. The outcomes align with the hypothesis that intelligent optimisation frameworks can significantly boost system efficiency by fine-tuning operational parameters under complex wastewater conditions.

Table 4.6. Percentage increment in biomass productivity, nitrogen removal efficiency (NRE), phosphorus removal efficiency (PRE), and chemical oxygen demand (COD) removal efficiency (CRE) after optimization compared to control conditions.

Parameter	Textile Effluent			Carpet Effluent		
	Control Value (without optimization)	Optimized Value (ANN-GA)	% Increment	Control Value (without optimization)	Optimized Value (ANN-GA)	% Increment
Biomass(g/L)	3.98	6.68	67.83	4.26	7.63	79.1
NRE	90.4	96.81	7.09	91.2	99.20	8.77
PRE	70.1	77.81	10.99	71.6	79.80	11.4
CRE	79.0	88.13	11.5	91.9	96.21	4.68

4.3.5. Future Implications

The practical use of RSM-GA and ANN-GA hybrid models will assist in designing and automating the operation of pilot-scale and large-scale photobioreactors. These models can accurately predict the behaviour of the photobioreactor and provide insights into optimal operating conditions for maximum microalgal growth and pollutant removal. Further, this hybrid approach can be extended to optimise other important parameters, including CO₂ content, aeration rate, and photoperiod. By incorporating pollution control parameters

specified by regulatory bodies like WHO and CPCB (Central Pollution Control Board), the modelling process results can ensure that the bioreactor's performance meets the required pollution control regulations by tuning the hyperparameters required for model construction. These regulations may include limits on parameters such as pollutant concentrations, effluent quality, and environmental impact. By adhering to these guidelines, the RSM-ANN-GA hybrid models facilitate environmentally friendly and compliant bioreactor operations that contribute to effective pollution control and sustainable practices.

4.4. Conclusion

This study optimised the treatment process for *D. mucosa* VSPA-based textile and carpet effluent, resulting in high biomass production. Two computing techniques, RSM and ANN, were employed to create predictive models. The optimisation process utilised RSM-GA and ANN-GA hybrid approaches. ANOVA analysis revealed that the pH-N/P ratio and all input parameters significantly influenced biomass productivity. Python-based ANN models showed superior prediction performance (CE: MSE-0.02, RMSE-0.14; TE: MSE-0.03, RMSE-0.16). Hybridisation with GA produced a global optimal solution, with the ANN-GA approach performing best (error: 0.14). These techniques are promising for optimising multi-input/multi-objective problems, including microalgae-based wastewater treatment.