

Chapter 6

**Computational and Experimental Validation of generate
ML models**

Computational and Experimental Validation of generate ML models

6.1. Computational Validation

Computational validation of the ML models was done by testing the rules extracted from them on some recently reported experimental results collected from references cited in Table 6.1. It should be noted that the data tabulated in Table 6.1 was not used during any training and testing step of decision tree construction.

Table 6.1. Computational validation of the machine learning models.

S.No.	Experimental Data Class			Prediction by Model			Reference
	NRE	PRE	Biomass Conc.	NRE	PRE	Biomass Conc.	
1	High	Low	High	High	High	High	[234]
2	High	Low	High	High	Low	High	[234]
3	High	Low	Low	Low	Low	High	[420]
4	Low	High	Low	Low	High	Low	[421]
5	High	Low	High	High	Low	High	[422]
6	Low	Low	High	Low	Low	Low	[422]
7	High	High	High	High	High	High	[423]
8	High	High	High	High	High	Low	[423]
9	High	High		High	Low		[555]
10	Low	High		High	High		[555]
11	High	High		Low	High		[555]
Classification Accuracy →				72.7%	81%	75%	

As summarized in Table 6.1, the model was able to correctly classify 8 data points out of 11 for NRE, 9 data points out of 11 for PRE, and 6 data points out of 8 for biomass production, corresponding to 72.7%, 81% and 75% accuracy each for NRE, PRE and biomass production, respectively. This result indicates the reliability of rules extracted from the constructed decision tree models and they can be further used to design new experimental procedures.

6.2. Experimental validation

In this study, the rules generated by the ML models for biomass production optimization, were verified experimentally by isolating two new microalgae strains.

6.2.1. Material and Methods

6.2.1.1. Strain Isolation

The combinations of input parameters obtained from decision tree analysis were experimentally verified by applying them to two newly isolated strains. Strains were named as VSVM1 and VSVM2. VSVM1 was isolated from the water of Ganga River near Ravidas Ghat, Varanasi (25° 17' 2.36" N, 83° 0' 33.13" E) in March 2022. VSVM2 was isolated from the water flowing in inlet of the sewage treatment plant (STP) situated at Bhagwanpur, Varanasi (25° 16' 21" N, 83° 0' 16.92" E) in April 2022. Both samples were serially diluted ten times and spread over agar plated containing Bold Basal Media (BBM). Plates were artificially illuminated by LED tube lights at 5000 lux intensity at a duration of 14:08 hrs light/dark cycle, kept at 25°C in an incubator. After seven days, individual colonies were picked up, identified under a microscope and streaked on fresh plates. Repeated streaking was performed until a pure colony was obtained (Figure 6.1).



Figure 6.1. Isolated colonies cultured on agar plates: VSVM1 and VSVM2.

6.2.1.2. Strain Identification

Both isolated colonies were identified by 18s rRNA sequencing. A spin column kit (High Media Pvt. Ltd., Make Mumbai, India) was used for the extraction of chromosomal DNA. The amplification of the 18S rRNA gene was carried out by polymerase chain reaction in a thermal cycler. The amplified products were purified by Exonuclease I-Shrimp Alkaline Phosphatase (Exo-SAP). Using the Sanger Method, purified amplicons were sequenced in an ABI 3500xl genetic analyser (Life Technologies Pvt. Ltd., Make USA). Sequencing files (.ab1) edited with CHROMASLITE (version 1.5) were examined by using the Basic Local Alignment Search Tool (BLAST), which located regions of local similarity between sequences, using the closet culture sequence retrieved from the National Centre for Biotechnology Information (NCBI) database.

6.2.1.3. Wastewater Preparation

As per the constructed dataset, three types of wastewaters: (i) municipal (ii) industrial and (iii) livestock, were mostly used for microalgae cultivation. In the present study decision tree models were verified by cultivating the isolated strains in synthetically simulated municipal, livestock and sugar industry wastewater. To minimise the risk of batch-to-batch variation that may affect validation results, synthetic simulated wastewaters were used in the study. The motive behind selecting sugar industry is that, it is one of the most prominent industries in Asia including India, China, Thailand and Pakistan. It is one of the highly water intensive industries as the production of one ton of sugar requires 70 m³ of water, producing effluent which has large amount of organic load [586]. Compositions of all synthetic simulated wastewater are represented in Table 6.2. N/P ratio of the prepared wastewater was adjusted by using ammonium chloride and potassium dihydrogen phosphate as per requirement. Composition of trace element solution was 0.275 mg/L MnCl₂.7H₂O, 0.44 mg/L ZnSO₄.7H₂O; 1.45 mg/L FeCl₃; 0.391 mg/L CuSO₄.5H₂O; 0.42 mg/L CoCl₂.6H₂O.

Table 6.2. Composition of synthetically simulated wastewater sources used in experiments.

S.No.	Wastewater	Composition	Reference
1	Municipal	5.4 mg/L NH ₄ Cl, 20 mg/L KH ₂ PO ₄ , 0.4 mg/L MgSO ₄ .7H ₂ O, 32 mg/L CaCl ₂ , 0.025 g/L yeast extract, 0.1 g/L meat extract and soya oil and 1 ml of trace element solution	[587]
2	Livestock	0.2g/L NO ₃ ⁻ -N (stock solution was aqueous solution of NaNO ₃ containing NO ₃ ⁻ -N 10 g/L), 2 g/L glucose, 0.5 g/L NaHCO ₃ , 11.6 g/L K ₂ HPO ₄ , 9 g/L	[588]

		KH_2PO_4 and 1 ml of trace element solution	
3	Sugar Industry	2.5 g/L Sucrose, 0.125 g/L $(\text{NH}_4)_2\text{SO}_4$, 0.065 g/L K_2HPO_4 , 0.050 g/L KH_2PO_4 , 0.50 g/L $(\text{NH}_4)\text{HCO}_3$, 4 g/L NaHCO_3 , 4 g/L KHCO_3 , 0.025 g/L yeast extract and 1 ml of trace element solution	[589]

6.2.1.4. Experimental design and conduction

The experiment set was designed according to the combinations of input parameters obtained through decision tree as summarised in Table 5. Experimental setup used during verification process is represented in Figure 6.2.

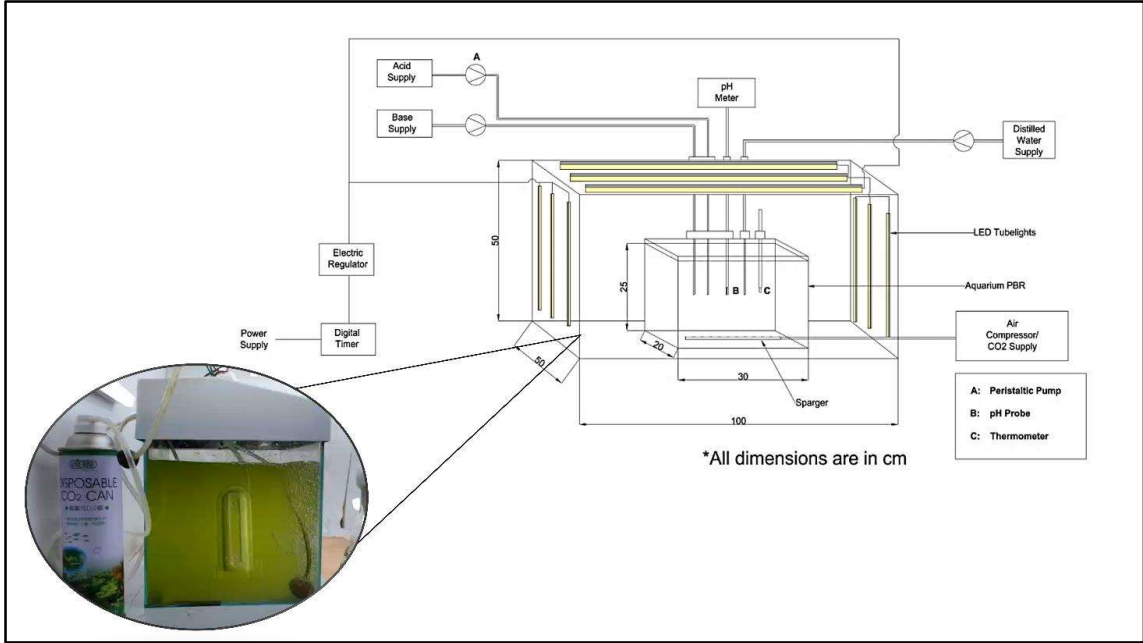


Figure 6.2. Experimental setup used for verification of decision tree model (Designed in AUTOCAD 2023 Student Version).

The experiments were conducted in a 15 L (30 cm x 20 cm x 25 cm) aquarium photobioreactor, with working volume 10.5 L. Wastewater was supplied through wastewater supply port of the

reactor. The reactor was kept in the rectangular cardboard chamber (100 cm x 50 cm x 50 cm), having LED tube lights on the upper, back and side walls as represented in Figure 6.2. Intensity of the LED lights were controlled by an electric regulator and timing of photoperiod was controlled by a digital timer. Intensity of the LED lights was measured by a luxmeter. pH of the reactor was continuously monitored and manually maintained by supplying 1 N NaOH and HCl reservoirs using peristaltic pump. Temperature was monitored by thermometer and controlled by keeping the chamber in an air-conditioned room. Air in the reactor was sparged at the rate of 3 L/minute through an air compressor. CO₂ was supplemented with the air as and when required from the cans through diffusor. The distilled water was supplied for maintaining the desired volume, in order to recover the evaporative loss from the reactor. The strains were cultivated in 1 L BBM media (~10% of 10.5 L working volume of reactor) and growth was monitored till mid log phase by monitoring the culture at 680 nm. At the attainment of the mid log phase, inoculation culture was transferred to reactor through inoculum port. This step was necessary for maintaining the uniform biomass concentration at the start of every experiment. After inoculation, each set of experiment was conducted for 10 days till stationary phase was attained. In order to determine the contamination with other species, samples were withdrawn randomly during the reactor operation for microscopic analysis.

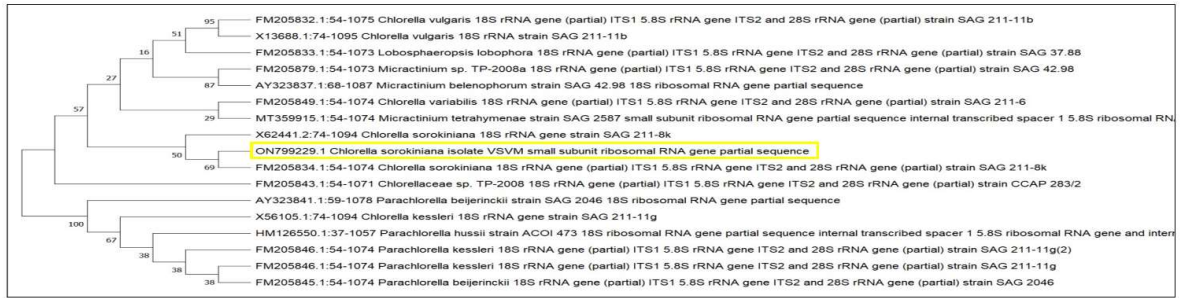
6.2.1.5. Biomass Measurement

100 ml of the sample was withdrawn after 10 days of cultivation period for the final measurement of biomass concentration. Sample was centrifuged at 6000 rpm for 15 minutes. Pellet was resuspended in buffer solution and centrifuged. This step was necessary to remove adsorbed materials from the surface of microalgal cells. Pellet was transferred in cup shaped aluminium foil. Initial weight of cup without pellet was also noted earlier. Cup was dried in furnace at 80°C for 2 hours and final weight of cup was measured. Drying process and measurement continued at interval of every 15 minutes, till a constant weight was achieved.

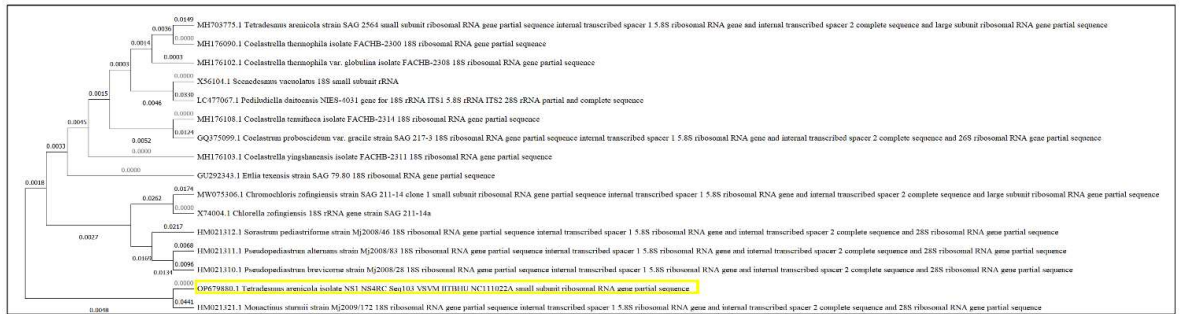
6.3. Results and Discussion

6.3.1.1. Strain Identification

Both strains were identified via 18S rRNA gene sequencing. A partial sequence of length 1018 bp was obtained for the VSVM1 sequence and a 1035 bp sequence was obtained for the VSVM2 sequence. Sequences were analysed by the BLAST program, by selecting the option ‘Standard database’ and from type material. The VSVM1 and VSVM2 strains showed high similarity with *Chlorella sorkiniana* (98.54%) (class *Trebouxiophyceae*) and *Tetradesmus arenicola* (99.32%) (class *Chlorophyceae*), respectively. Hence, VSVM1 and VSVM2 strains were designated as *Chlorella sorkiniana* VSVM1 and *Tetradesmus arenicola* VSVM2. Both strains were submitted to NCBI under the accession numbers ON799229 and OP679880, respectively. The phylogenetic tree for the strains were constructed using MEGA 11 (Figure 6.3). Notably, as per a search from the Web of Science and Scopus database, none of the publications has reported the cultivation of *Tetradesmus arenicola* in the wastewater.



(A)



(B)

Figure 6.3. Phylogenetic tree representing both strains: (A) *Chlorella sorokiniana* VSVM1 and (B) *Tetrademus arenicola* VSVM2.

6.3.1.2. Experimental Verification

The paths in decision tree analysis having more than five nodes and supported by more than 10 data points were taken into consideration. Values of input parameters not present in the path were decided according association rule mining. Total 38 experiments were conducted and the results of the theoretical class predicted from ML and experimental class were compared as in Table 6.3. The high and low classes represent the two classes of biomass output variable used in the Decision Tree and Association Rule Mining analysis. They were formed based on the median value (1.22 g/L) indicated in the previous chapters.

Table 6.3. Comparison between theoretical values obtained from the decision tree and experimental values.

S	Path	Strain	WW	C	PT	RT	CO ₂	Tem	III	p	LI	PP	N/P	Th.	Exp.	Exp.
.	No.			T	M			p.		H				Clas	Conc.	Clas
N														ss	*(g/L)	ss
o																
.																
1	7	VSV	M	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	1.1±0.0	L
		M1								5			9		9	
2	7	VSV	Li	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	2.51±0.	H
		M1								5			9		03	
3	7	VSV	I	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	0.81±0.	L
		M1								5			9		07	
4	7	VSV	M	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	1.47±0.	H
		M2								5			9		13	
5	7	VSV	Li	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	2.93±0.	H
		M2								5			9		04	
6	7	VSV	I	A	P	PBR	0.04	25	0.1	7.	200	16	8.6	H	0.97±0.	L
		M2								5			9		19	
7	8	VSV	M	M	P	PBR	0.04	25	0.1	7.	80	16	16	H	1.73±0.	H
		M1								5					01	
8	8	VSV	Li	M	P	PBR	0.04	25	0.1	7.	80	16	16	H	1.17±0.	L
		M1								5					08	
9	8	VSV	M	M	P	PBR	0.04	25	0.1	7.	80	16	16	H	0.71±0.	L
		M2								5					09	
1	8	VSV	L	M	P	PBR	0.04	25	0.1	7.	80	16	16	H	1.19±0.	L
0		M2								5					15	
1	11	VSV	M	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	0.87±0.	L
1		M1								5					04	
1	11	VSV	Li	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	1.97±0.	H
2		M1								5					01	

1	11	VSV	I	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	0.51±0.	L
3		M1								5					02	
1	11	VSV	M	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	1.67±0.	H
4		M2								5					09	
1	11	VSV	Li	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	2.25±0.	H
5		M2								5					15	
1	11	VSV	I	A	P	PBR	0.04	25	0.1	7.	200	22	36	H	1.57±0.	H
6		M2								5					07	
1	12	VSV	M	A	P	PBR	2.5	25	0.2	7	150	16	16	H	2.11±0.	H
7		M1													02	
1	12	VSV	Li	A	P	PBR	2.5	25	0.2	7	150	16	16	H	3.71±0.	H
8		M1													08	
1	12	VSV	I	A	P	PBR	2.5	25	0.2	7	150	16	16	H	1.42±0.	H
9		M1													02	
2	12	VSV	M	A	P	PBR	2.5	25	0.2	7	150	16	16	H	2.78±0.	H
0		M2													05	
2	12	VSV	Li	A	P	PBR	2.5	25	0.2	7	150	16	16	H	4.12±0.	H
1		M2													11	
2	12	VSV	I	A	P	PBR	2.5	25	0.2	7	150	16	16	H	1.59±0.	H
2		M2													09	
2	13	VSV	M	A	P	PBR	2.5	25	0.2	7	200	16	16	H	2.39±0.	H
3		M1													06	
2	13	VSV	Li	A	P	PBR	2.5	25	0.2	7	200	16	16	H	3.97±0.	H
4		M1													21	
2	13	VSV	I	A	P	PBR	2.5	25	0.2	7	200	16	16	H	1.79±0.	H
5		M1													09	
2	13	VSV	M	A	P	PBR	2.5	25	0.2	7	200	16	16	H	2.99±0.	H
6		M2													05	
2	13	VSV	Li	A	P	PBR	2.5	25	0.2	7	200	16	16	H	4.45±0.	H
7		M2													13	

2	13	VSV	I	A	P	PBR	2.5	25	0.2	7	200	16	16	H	1.81±0.03	H
8		M2														
2	14	VSV	Li	A	P	PBR	0.04	25	0.2	7.	200	16	16	H	0.87±0.09	L
9		M1								5						
3	14	VSV	I	A	P	PBR	0.04	25	0.2	7.	200	16	16	H	0.63±0.08	L
0		M1								5						
3	14	VSV	Li	A	P	PBR	0.04	25	0.2	7.	200	16	16	H	0.91±0.01	L
1		M2								5						
3	14	VSV	I	A	P	PBR	0.04	25	0.2	7.	200	16	16	H	0.71±0.05	L
2		M2								5						
3	17	VSV	M	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	1.67±0.05	H
3		M1														
3	17	VSV	Li	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	1.91±0.09	H
4		M1														
3	17	VSV	I	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	1.29±0.04	H
5		M1														
3	17	VSV	M	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	1.81±0.05	H
6		M2														
3	17	VSV	Li	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	2.21±0.11	H
7		M2														
3	17	VSV	I	A	AC	PBR	2.5	25	0.2	7	200	16	8	H	1.38±0.13	H
8		M2														

VSVM1: *Chlorella sorkiniana* VSVM1; VSVM2: *Tetrademus arenicola* VSVM2; WW Source: Wastewater; M: Municipal; Li: Livestock; I: Industrial; CT: Cultivation Type; A: Autotrophic; M: Mixotrophic; P: Primary; AC: Autocalve; PBR: Photobioreactor; Temp.: Temperature (°C); IIL: Initial Inoculum Level (g/L); LI: Light Intensity ($\mu\text{mol m}^{-2} \text{s}^{-1}$); PP: Photoperiod (hours); N/P: Nitrogen/Phosphorus Ratio; Th.: Theoretical; Exp.: Experimental; RT: Reactor Type;

The results of 27 experiments resulted in a high-class level out of 38 (Table 6.3). The accuracy of the decision tree after experimental verification was found to be 71.05%. Compared to the computational accuracy of 81.25, difference of 10.20% was noticed in the experimental data. Around 10% error seemed to be fair enough as biological data suffer from variations due to the

environmental conditions. The error can be minimised by constructing a more intensive dataset and conducting in depth research on the most critical input parameters. Experimental verification also provided some other important inferences during microalgae cultivation in wastewater. Among the two strains, VSVM2 belonging to class *Chlorophyceae*, achieved higher biomass production in all types of wastewaters as compared to the VSVM1 strain, belonging to class *Trebouxiophyceae*. A more comparative study is needed between these two classes for their performance in the wastewater. Some publications reported that species belonging to class *Chlorophyceae* are better than species belonging to class *Trebouxiophyceae*. However, while some of the studies reported reverse results [560], [590]–[592]. As an alternative, species belonging to these classes can be co-cultivated with each other, increasing wastewater treatment efficiency and simultaneously generating biomass in high concentration [557], [593]. Another important inference was that livestock wastewater supported higher biomass productivity than other wastewater. Livestock wastewater is known to support high microalgae biomass production in comparison to other wastewater sources due to its high nutrient load and N/P ratio around 16:1 [594].

6.4. Conclusions

Computational and experimental validation indicated the reliability of rules extracted from the constructed decision tree models and they can be further used to design new experimental procedures. Compared to the computational accuracy of 81.25, difference of 10.20% was noticed in the experimental data. Around 10% error seemed to be fair enough as biological data suffer from variations due to the environmental conditions. As per experimental verification, the newly isolated strain *Tetradesmus arenicola* VSVM2 provided better removal efficiency than *Chlorella sorokiana* VSVM2.

