

Chapter 5

Application of association rule mining for finding the specific condition of predictor variables for increasing wastewater treatment capability and biomass production

Application of association rule mining for finding the specific condition of predictor variables for increasing wastewater treatment capability and biomass production

5.1. Introduction

In chapters 3rd and 4th, ML tools have been applied to determine the best combinations of cultivation parameters for increasing the efficiency of microalgae-based wastewater treatment (MBWT) by Decision Tree algorithms. However, it is also necessary to analyse the effect of individual parameters on MBWT and to determine their range which will favour microalgae growth in wastewater. This study can be performed by using association rule mining, one of the data mining tools. ARM algorithms can be utilized to mine a large dataset of MBWT and to find the relationships or rules between cultivation parameters/predictor variables and output variables (nutrient removal efficiency and biomass production) that can lead to high performance of microalgae-based wastewater treatment. Therefore, in the present investigation, a dataset was constructed representing predictor variables and output variables including nitrogen removal efficiency (NRE), phosphorus removal efficiency (PRE) and biomass production. After database construction, one of the ARM algorithms, the apriori algorithm was applied to determine the relationship between input and output variables leading to high biomass productivity and high nutrient removal efficiency.

5.2. Materials and Methods

5.2.1. Data collection and pre-processing

Dataset with nearly 500 data entries representing biomass productivity and pollutant removal capability of microalgae during wastewater treatment was constructed by collecting data from publications [24], [25], [30], [31], [136], [141], [145], [146], [149], [153], [156], [159], [167], [170], [172]–[174], [178], [190], [192], [205], [210], [214], [219], [220], [226], [236], [248], [250]–[256], [258]–[260], [264], [266], [267], [269], [273]–[277], [280], [286], [289], [294],

[297], [300], [302], [303], [305], [307]–[310], [312]–[314], [316], [317], [324], [327], [329], [331], [333], [336], [338], [339], [343], [344], [346]–[349], [385], [409], [416], [430], [433], [435], [439], [441], [445], [448]–[450], [454], [456], [460]–[473], [476]–[478], [480], [481], [483], [484], [487]–[489], [491]–[513], [515], [516], [523]–[557]. These publications were extracted from authorized sources (Elsevier, Springer, ACS and Wiley) based on keywords. The final dataset was composed of eleven input/predictor variables and three output variables namely biomass production, NRE and PRE. Predictor variables involved both categorical and continuous variables as represented in Table 5.1 and

Table 5.2 respectively. Categorical variables were used as it is for association rules mining and continuous variables were discretized into different levels depending upon their range. The units of all variables were also standardized. The output variables were divided into two class (high and low level) depending on the median of the class [417]. The median value of NRE, PRE and biomass production was 83.93%, 88.70% and 1.22 g/L, respectively. Values greater or equal to median were grouped in high-level class while values less than the median were grouped in low-level class.

Table 5.1. Categorical variables used in the analysis and their identities.

S.No.	Categorical Variable	Discrete Identities
1	Microalgae Class	<i>Chlorodendrophyceae</i> , <i>Chlorophyceae</i> , <i>Consortium</i> , <i>Cyanophycean</i> , <i>Eustigmatophyceae</i> , <i>Trebouxiophyceae</i> <i>Xanthophyceae</i>
2	Wastewater source	Domestic, Municipal, Agriculture, Industrial, Synthetic, Livestock
3	Pre-treatment method	Primary, Aerobic, Anaerobic, Autoclave

4	Reactor Type	Flasks, PBR
---	--------------	-------------

Table 5.2. Continuous variables and their discretized classes.

S.No	Continuous Variables	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
1	CO ₂ Content (v/v)	[0-0.53]	(0.53-2.53]	(2.53-5.53]	(5.53-9.53]	(9.53-20.53]	(20.53-45.53]
2	Temperature (°C)	[4-15.2]	(15.2-20.8]	(20.8-23.6]	(23.6-25]	(25-30.6]	(30.6-45]
3	Initial Inoculum Level (g/L)	[0.005-0.105]	(0.105-0.205]	(0.205-0.305]	(0.305-0.405]	(0.405-0.505]	(0.505-11.905]
4	pH	(4.86-6.75]	(6.75-7.02]	(7.02-7.56]	(7.56-8.10]	(8.10-9.72]	
5	Light Intensity ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	[0]	(0.1-50]	(50-100]	(100-200]	(200-1500]	(1500-4050]
6	Photoperiod (h)	[0]	(10.8-13.5]	(13.5-16.2]	(16.2-24]		
7	Nitrogen/Phosphorus Ratio (N/P)	(0.26-5.26]	(5.26-10.26]	(10.26-15.26]	(15.26-50.26]	(50.26-495.26]	

Round bracket ('()') indicates that value has been included in the range and square bracket ('[]') shows that value has not been included in the range.

5.2.2. Computational details

The apriori algorithm was employed using MATLAB (Version 9.3; R2017b) in order to construct association rules. The generated rules were accepted and extrapolated based on support, confidence and lift values. Support indicates the fraction of a particular level of input variable that leads to a specific class of output variables in the total dataset (Equation no. (5.1)) [558]:

$$S(A \rightarrow B) = P(A \cap B) = \frac{freq(A,B)}{N} \quad (5.1)$$

where, $freq(A, B)$ represents the number of datapoints of level A of predictor variables leading to output class B and N is the total number of entries in the database. The rule is reliable if the support value is greater than 0.01. Next, the accuracy of the model was determined by the confidence parameter. Confidence is the ratio of datapoints of a particular input variable level leading to a specific output class to the total number datapoints of that particular level (Equation no. (5.2)) [558]:

$$C(A \rightarrow B) = \frac{P(A \cap B)}{P(A)} = \frac{freq(A,B)}{freq(A)} \quad (5.2)$$

where, $freq(A, B)$ represents the total datapoints of level A of input variable leading to output variable B and $freq(B)$ represents the total datapoints of level A present in the dataset. The rule is more accurate if its confidence value is more than 0.5.

At last, the significance of the accepted rules was decided based on lift values. Lift values are the main indicator that provide positive or negative impacts of a particular level of input on the output variable. It is calculated by dividing the confidence by the ratio of total number of datapoints of the specific output class to the total number of entries (Equation no. 5.3) [558]:

$$L(A \rightarrow B) = \frac{C(A \rightarrow B)}{P(B)} = \frac{C(A \rightarrow B)}{freq(B)/N} \quad (5.3)$$

where, $\text{freq}(B)$ represents the total no. datapoints of output class B. Lift value less than 1 indicates a negative relation between input and output variable and vice-versa. The plot of the lift values was made by Origin Lab software (Version 2017).

5.3. Results and Discussion

ARM technique was used to determine the range or level of predictor variables leading to high biomass production and pollutant removal efficiency during microalgae-based wastewater treatment process. The results were analysed on the basis of support, confidence and lift values. As the main indicator was lift, therefore the Tables and Figures represent the obtained lift values. In the figures, blue and green circles represent high class and low output variables, respectively. The higher lift value of the blue circle indicates that a particular level of input variable will have positive impact on the output variable. In contrast, a higher value of the green circle shows that the level will have negative sway on the output variable. The circles' size is proportional to the no. of cases or data points of a particular level. The results of association rule mining are concluded based on two parameters, lift and count values. Count values indicate the number of data points that support the particular rule. The rules with high positive lift and count values are selected. Results are summarized based on lift and count values as shown in Table 5.14.

5.3.1. Microalgae Class

Microalgae species belonging to various classes such as *Chlorophyceae*, *Trebouxiophyceae*, *Cyanophyceae* and many more have been applied to treat wastewater. Table 5.3 and Figure 5.1 represent the lift and count values for rules for microalgae class each for NRE, PRE and biomass production.

Table 5.3. Lift values for the association rules derived for Microalgae Class.

Level	Microalgae Class	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	<i>Chlorodendrophyceae</i>	1.25	0.75	0.33	1.67	1.50	0.50
2	<i>Chlorophyceae</i>	1.08	0.92	1.17	0.83	0.89	1.11
3	Consortium	0.94	1.06	1.15	0.85	1.08	0.92
4	<i>Cyanophyceae</i>	1.60	0.40	0.67	1.33	1.20	0.80
5	<i>Eustigmatophyceae</i>	1.60	0.40	0.00	2.00	1.60	0.40
6	<i>Trebouxiophyceae</i>	0.91	1.10	0.84	1.16	1.05	0.95
7	<i>Xanthophyceae</i>	1.33	0.67	0.80	1.20	1.50	0.50

ARM rules indicate that the species belonging to class *Chlorophyceae* have high nutrient removal efficiency while species belonging to class *Trebouxiophyceae* have efficient biomass productivity than other classes. The lift and count value of nitrogen removal efficiency by *Chlorophyceae* was 1.08 and 109, and for phosphorus removal efficiency, it was 1.17 and 120. Similarly, high biomass productivity by *Trebouxiophyceae* was supported by a lift and count value of 1.05 and 112. So, based on both lift and count values, these rules were accepted. While, as per Figure 5.1, only 4 data points supports high nitrogen removal by *Xanthophyceae* and high biomass productivity by *Eustigmatophyceae*. These were not enough to conclude the rules generated by association rule mining.

It has been suggested that class *Chlorophyceae* has higher nutrient efficiency than other classes including class *Trebouxiophyceae* [516]. Singh et al. (2022) cultivated two strains, *Chlorella vulgaris* belonging to class *Trebouxiophyceae* and *Scenedesmus vacuolatus* belonging to class *Chlorophyceae* at different concentrations of municipal wastewater. Performance of *S. vacuolatus* was efficient than that of *C. vulgaris*, it was able to recover 59.11-

92.28% of nitrogen and phosphorus from municipal wastewater [559]. Gupta et al. (2016) concluded that *Chlorophyceae* has high nutrient removal efficiency and *Trebouxiophyceae* has capability of high biomass productivity. Authors cultivated two strains *C. sorokiniana* (*Trebouxiophyceae*) and *S. obliquus* (*Chlorophyceae*) in 2 L flasks containing raw sewage at LI of $80 \mu\text{mol m}^{-2} \text{s}^{-1}$ with 16:8 h light: dark cycle under continuous aeration. *S. obliquus* removed 98.54% of nitrogen and 97.99% of phosphorus whereas *C. sorokiniana* was able to remediate 86.93% of nitrogen and 68.24% of phosphorus in 100% wastewater concentration [560].

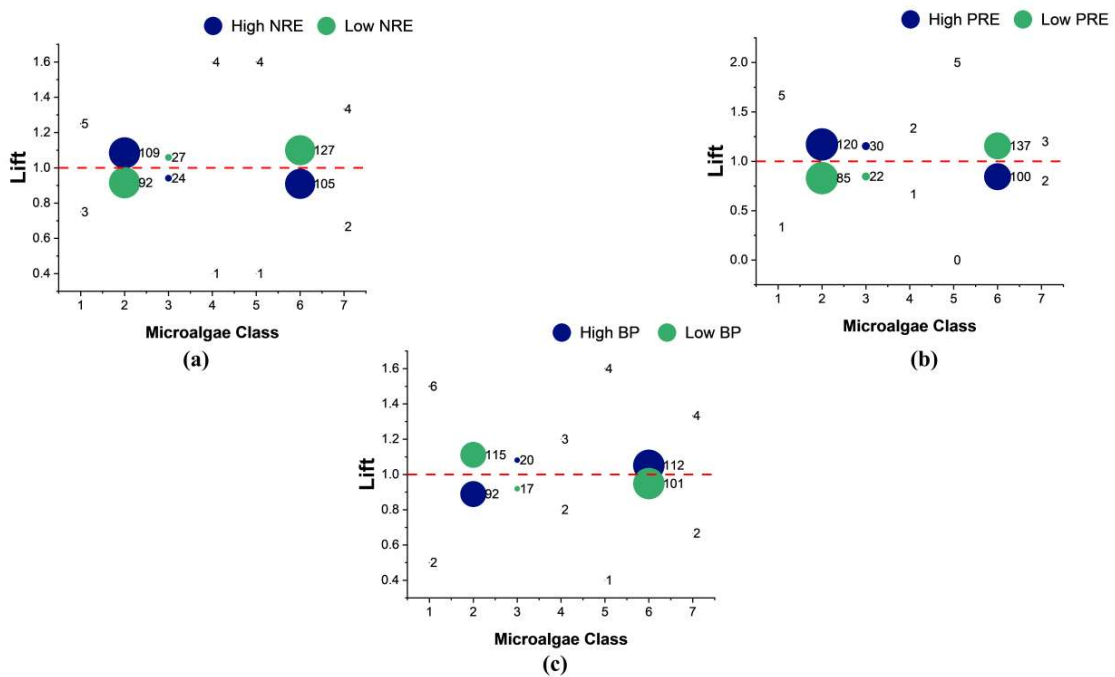


Figure 5.1. Lift and count values for ARM derived for microalgae class: (a) NRE, (b) PRE and (c) Biomass production.

5.3.2. Wastewater Source

Microalgae species have been used to treat various sources of wastewater, including agricultural field run-off, domestic, industrial, livestock and municipal wastewater. Table 5.4 and Figure 5.2 represents the lift and count values for ARM rules with reference to wastewater source for NRE, PRE and biomass production each.

Table 5.4. Lift values for the association rules derived aimed at wastewater source.

Level	Wastewater Source	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	Agriculture	1.40	0.60	0.80	1.20	1.00	1.00
2	Domestic	0.60	1.40	0.52	1.48	0.42	1.58
3	Industrial	0.91	1.09	0.84	1.16	1.19	0.81
4	Livestock	0.76	1.24	0.84	1.16	1.14	0.86
5	Municipal	1.26	0.74	1.19	0.81	0.89	1.11
6	Synthetic	1.35	0.65	1.59	0.41	1.02	0.98

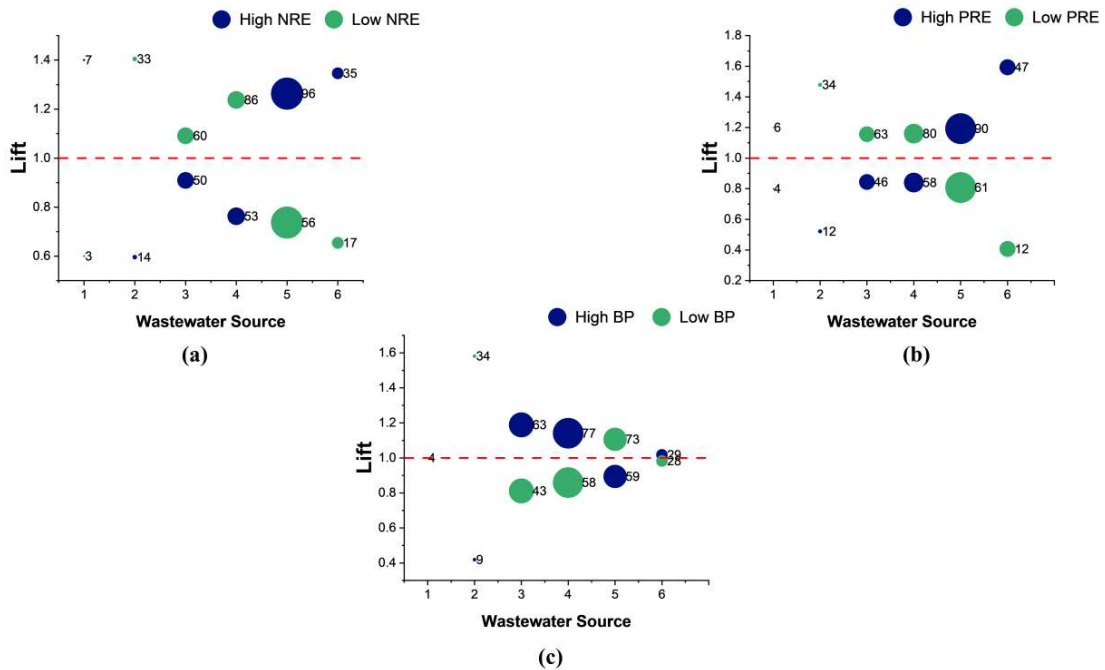


Figure 5.2. Lift and count values for ARM derived for wastewater source: (a) NRE, (b) PRE and (c) Biomass production.

According to the lift and count values of ARM for wastewater sources, industrial and livestock wastewater supports high biomass productivity than other wastewater sources. In contrast, the removal efficiency was higher in municipal and synthetic wastewater sources as compared to

other sources. Industrial wastewater is composed of effluents discharged from various industries such as tannery [28], brewery [561], while livestock wastewater includes effluents originating from units related to livestock such as dairy and slaughter houses [550], [562]. Industrial and livestock wastewater sources have very high nutrient concentration, which supports high biomass productivity. However, microalgae may not utilise high levels of nutrients completely due to culture inhibition such as low light penetration, thus resulting in low nutrient removal efficiency than other sources.

Domestic wastewater comprises of effluents generated from daily activities of residential locations and municipal wastewater is composed of 80-95% domestic wastewater with 5-20% industrial wastewater [563], [564]. Nevertheless, domestic wastewater has low nutrient concentration. Therefore, both biomass productivity and nutrient removal efficiency are low, as predicted by association rules. Agriculture wastewater is constituted by excess nutrients transferred from agricultural fields due to the excessive usage of pesticides and insecticides [168]. However, there were no significant association rules derived for agricultural wastewater as its lift values was one and very few data points were available for it in the literature. Therefore, agricultural wastewater can be a promising source for future research as it has a high nutrient concentration to support microalgae growth. Arora et al. (2016) cultivated *Chlamydomonas debaryana* IITRIND3 in domestic, municipal, industrial and livestock wastewater in 1 L flasks [494]. According to their results, the highest biomass production of 3.66 g/L was obtained in dairy wastewater, while the highest total NRE of 83.86% was obtained in municipal wastewater [494]. These results were in accordance with the ARM rules. Choudhary et al. (2016) cultivated microalgae consortia in domestic wastewater, livestock wastewater and synthetic media (BBM) in 250 ml flasks. The highest removal efficiency and biomass productivity were obtained in livestock wastewater with 99% removal of ammonium

nitrogen and 83% removal of phosphate phosphorus. A biomass concentration of 1.9 g/L was obtained which was 51% more than the synthetic media [258].

5.3.3. Pre-treatment Method

Wastewater comprises of particulate matters, toxins, high nutrient concentration and microbial consortia. Therefore, wastewater needs pre-treatment before it can be used for microalgae cultivation [565]. The lift and count values for ARM rules derived for pre-treatment, each for NRE, PRE and biomass production have been represented in Table 5.5 and Figure 5.3.

Table 5.5. Lift values for the association rules derived for pre-treatment.

Level	Pre-treatment Method	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	Aerobic	0.67	1.33	0.93	1.07	0.85	1.15
2	Anaerobic	0.66	1.34	0.66	1.34	0.96	1.04
3	Autoclave	1.04	0.96	1.12	0.88	1.15	0.85
4	Primary	1.22	0.78	1.08	0.92	0.98	1.02

Lift and count values derived from ARM indicate that autoclave pre-treatment is better suited for biomass productivity and nutrient removal. The reason being, autoclave pre-treatment decreases microbial load from wastewater which causes inhibition of microalgae growth.

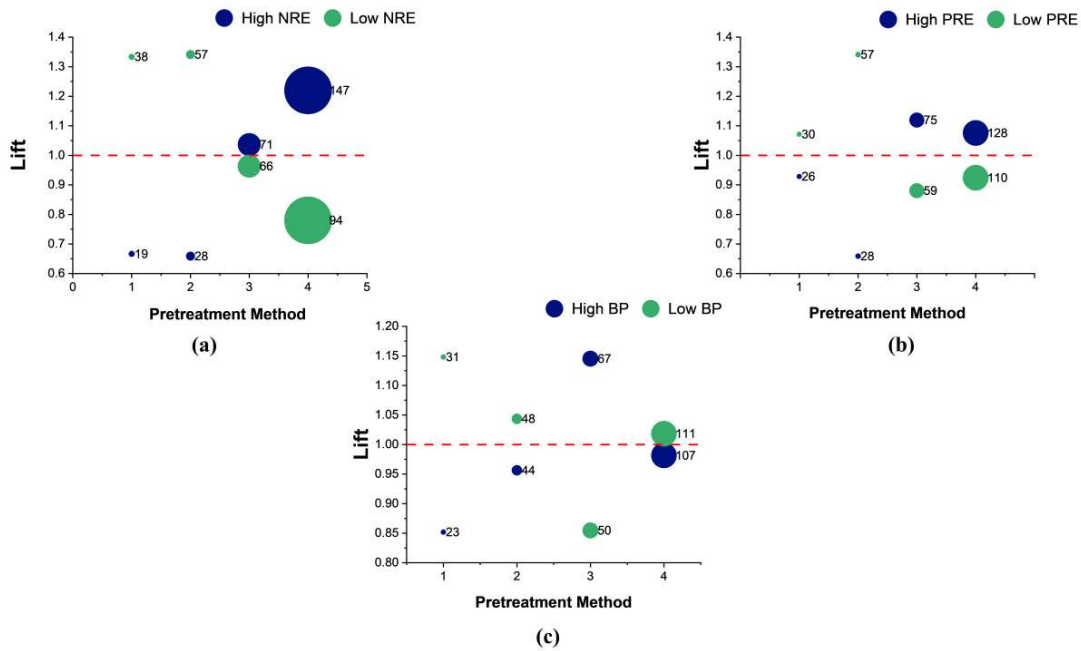


Figure 5.3. Lift and count values for ARM derived for pre-treatment: (a) NRE, (b) PRE and (c) Biomass production.

Autoclaving also breakdowns complex nutrients into simpler compounds that are easily assimilable by microalgae. However, autoclave pre-treatment is an energy expensive one as compared to other pretreatments [148]. Next to autoclave, the primary pre-treatment provides promising result for high performance of microalgae in wastewater. Primary pre-treatment includes conventional physical methods such as centrifugation, gravity settling and filtration [565]. Primary pre-treatment is sufficient to support microalgae growth in wastewater medium for large scale wastewater treatment. The next two pre-treatment methods, aerobic and anaerobic are applied when organic content in the wastewater is high. However, in comparison to other methods their performance is low. No positive relation was detected between these methods and biomass productivity. As per the lift value, the performance of both methods is quite similar. Anaerobic processes are generally favourable than aerobic methods as anaerobic method generates biogas as a by-product which has significant economic importance [170], [210].

These results were consistent with previously reported experimental findings [260]. AlMomani and Ormeci (2016) cultivated *Chlorella vulgaris* in primary, aerobic and anaerobic treated municipal effluent in a 500 ml flask at 20°C. According to their findings, the highest biomass concentration (2.71 g/L) was obtained in primary treated effluent. The second highest (2.37 g/L) was obtained in anaerobically treated effluent and the lowest (1.86 g/L) was obtained in aerobically treated effluent. The rationale behind this was the presence of high ammonia concentration or other toxic elements generated during the aerobic and anaerobic treatment process which inhibited microalgae growth [260]. Choix et al. (2018) cultivated *Chlorella* sp. in filtered (primary treatment) and bio-digested (anaerobic treatment) Tequila vinasses, generated during Tequila production. As predicted, the highest biomass concentration (2.30 g/L) was obtained in filtered Tequila vinasses [504].

5.3.4. Reactor type

Two reactors that were considered for evaluation in the present analysis were: (i) flasks (ii) Photobioreactor (PBR). The lift and count values for association mining rules are depicted in Table 5.6 and Figure 5.4.

Table 5.6. Lift values for the association rules derived for Reactor.

Level	Reactor	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	Flasks	0.98	1.02	0.98	1.02	1.02	1.02
2	PBR	1.02	0.98	1.02	0.98	0.98	0.98

According to the results presented in Table 5.6 and Figure 5.4, there is an apparent relationship between high biomass productivity, high nutrient removal efficiency and PBR. PBR is used for pilot and large-scale microalgae-based wastewater treatment, while flasks carry out

experimental study at a laboratory scale. PBR are closed-controlled illuminated vessels that provide better control over process parameters such as pH, temperature, LI with control over microbial contamination [205].

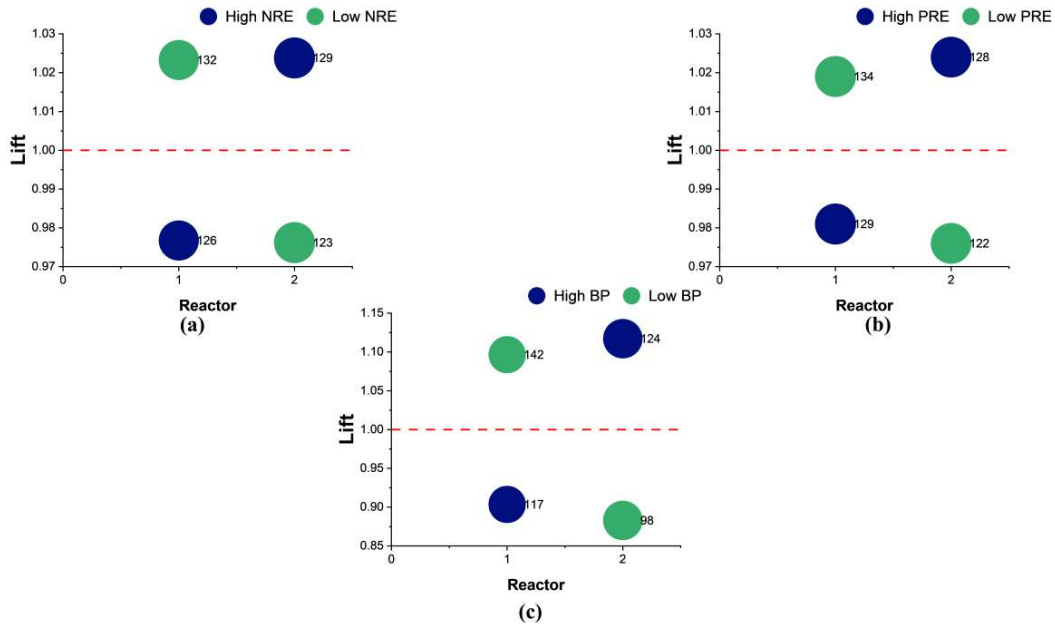


Figure 5.4. Lift and count values for ARM derived for reactor: (a) NRE, (b) PRE and (c) biomass production.

5.3.5. CO₂ Content

CO₂ content is one of the critical factors influencing microalgae growth as it provides energy for them. Microalgae fix inorganic carbon by capturing CO₂ from the atmosphere or flue gases. Microalgae can even utilize soluble carbonates present in the medium [48], [566]. Lift and count values for different association mining rules derived for CO₂ content each for NRE, PRE and biomass production have been represented in Table 5.7 and Figure 5.5.

Table 5.7. Lifts values for the association rules derived for CO₂ content.

Level	CO ₂ content Range	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	0-0.53	0.90	1.10	0.80	1.20	0.93	1.07
2	0.53-2.53	1.36	0.64	1.38	0.62	1.33	0.67
3	2.53-5.53	1.11	0.89	1.46	0.54	1.31	0.69
4	5.53-9.53	1.29	0.71	1.43	0.57	0.57	1.43
5	9.53-20.53	1.36	0.64	1.31	0.69	0.50	1.50
6	20.53-45.53	0.00	2.00	0.00	2.00	1.00	1.00

It is evident that when CO₂ content of level 2 (0.53-2.53%) was supplied to the microalgae-based wastewater treatment process, it increased both treatment efficiency and biomass productivity. Any deviation from it had negative impact on the treatment. The productivity was low at lower CO₂ concentrations. However, a higher concentration of CO₂ (> 5%) inhibited the growth. A decrease in productivity at high CO₂% was might be due to the change in the characteristics of the photosynthetic process. At high CO₂ concentration, affinity towards CO₂ decreases while photosynthetic sensitivity towards O₂ rises and the activity of carbonic anhydrase gets reduced. This is one of the main enzymes responsible for inorganic carbon assimilation [217], [218].

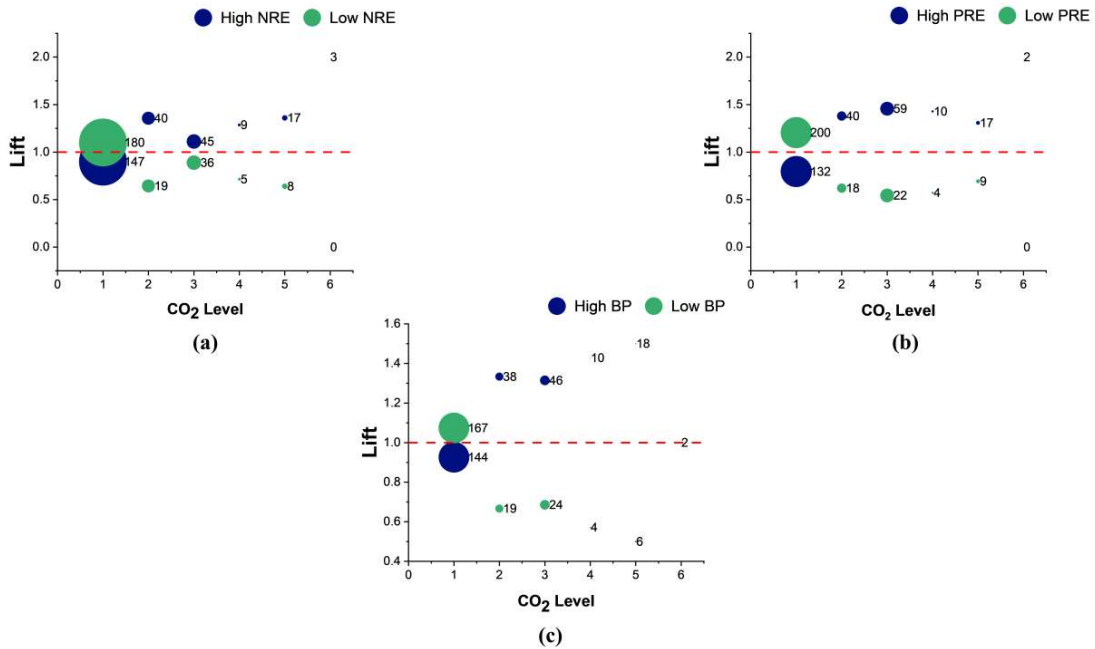


Figure 5.5. Lift and count values for ARM derived for CO₂ content: (a) NRE (b) PRE and (c) Biomass production.

Nayak et al. (2016) cultivated *Scenedesmus sp.* in domestic wastewater in airlift PBR at different CO₂ concentrations ranging from 0.03 to 10%. Both biomass production and nutrient removal efficiency increased with the rise in CO₂ concentration. However, a declining trend was observed after 5% concentration. Highest biomass production of 1.37 g/L and removal efficiency (NH₄⁺-N: 98.3%; NO₃⁻-N: 70.2%; PO₄³⁻-P: 78.9%) was obtained at CO₂ concentration of 2.5% [226]. This biomass yield was 1.54 times more than the control. This was in accordance with the lift values obtained by association rule mining.

5.3.6. Temperature

Temperature influences microalgae growth by affecting the metabolic process, enzyme activity and cytoplasmic viscosity. The photoinhibition occurs at higher temperature [557]. Lift and count values for different association mining rules was derived for temperature, each for NRE, PRE and biomass production (Table 5.8 and Figure 5.6).

Table 5.8. Lifts values for the association rules derived for temperature.

Level	Temperature Range	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	4-15.2	0.50	1.50	0.75	1.25	1.20	0.80
2	15.2-20.8	1.02	0.98	0.84	1.16	1.00	1.00
3	20.8-23.6	1.34	0.66	1.34	0.66	0.82	1.18
4	23.6-25	0.88	1.12	0.91	1.09	1.02	0.98
5	25-30.6	1.13	0.87	1.10	0.90	1.07	0.93
6	30.6-45	0.73	1.27	0.73	1.27	1.20	0.80

Level 4 (23.6-25 °C) was significant according to the lift and count values for association mining rules derived for temperature and biomass production. However, level 5 (25-30.6 °C) was more significant for the pollutant removal. Any deviation from these levels showed negative impression on microalgae performance in wastewater. Higher temperature reduces CO₂ solubility in the culture medium due to which the Rubisco enzyme starts bonding with O₂, leading to photorespiration [568]. High temperature leads to cell disruption and extracellular polymer substance secretion, resulting in cell aggregation [569]. The low temperature inhibits Rubisco activity, thus reducing the photosynthetic rate and bio-transformation of CO₂ [570]. Zhang et al. (2021) studied the effect of different temperatures during aquaculture wastewater treatment by *C. vulgaris*-bacterial consortia [571]. The best treatment efficiency was obtained at 25°C (level 4 in the present study) with 68.8% NH₄⁺-N and 94.4% COD removal [571]. Treatment efficiency was still better at 20°C than that obtained above 30°C. Authors also studied the effect of temperature on Rubisco and citrate synthase activity during photosynthesis and the TCA cycle respectively. The highest activity was 1668.0 U/g-protein for Rubisco and 1110.8 U/g-protein for citrate synthase was achieved at 30°C [571].

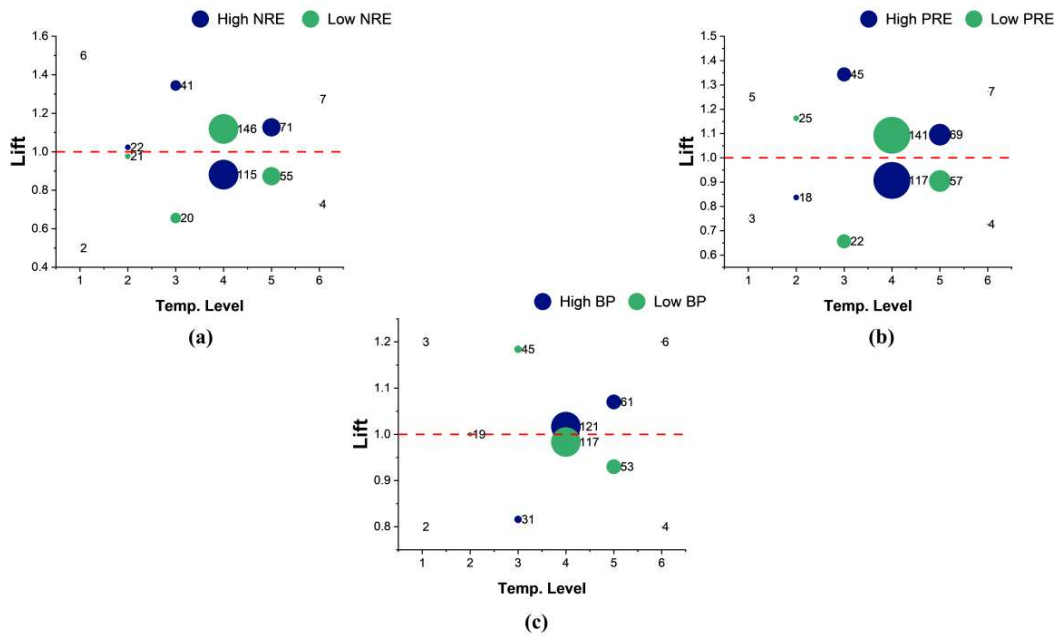


Figure 5.6. Lift and count values for ARM derived for Temperature: (a) NRE (b) PRE and (c) Biomass production.

Chu et al. (2015) demonstrated the influence of seasonal temperature variations in the cultivation of *C. pyrenoidosa*. *C. pyrenoidosa* was cultivated in starch processing wastewater in an airlift photobioreactor. Temperature during summer was found more favourable in achieving a maximum biomass concentration of 2.26 g/L and high nutrient removal efficiency as compared to other seasonal temperatures [220]. However, it has been also reported that there are scarce microalgae species present in nature that can grow below 15°C [220] or above 35°C [270].

5.3.7. Initial Inoculum Level

IIL acts as one of the vital factors during the start of the microalgae cultivation, particularly in wastewater treatment. Lift and count values for different association mining rules were derived for the IIL, each for NRE, PRE and Biomass production (Table 5.9 and Figure 5.7).

Table 5.9. Lifts values for the association rules derived for the initial inoculum level.

Level	Initial Inoculum	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	0.005-0.105	1.00	1.00	0.93	1.07	0.58	1.42
2	0.105-0.205	0.84	1.16	1.09	0.91	1.30	0.70
3	0.205-0.305	1.10	0.90	1.00	1.00	1.57	0.43
4	0.305-0.405	1.20	0.80	1.27	0.73	1.30	0.70
5	0.405-0.505	1.40	0.60	1.60	0.40	2.00	0.00
6	0.505-11.905	0.94	1.06	0.90	1.10	1.59	0.41

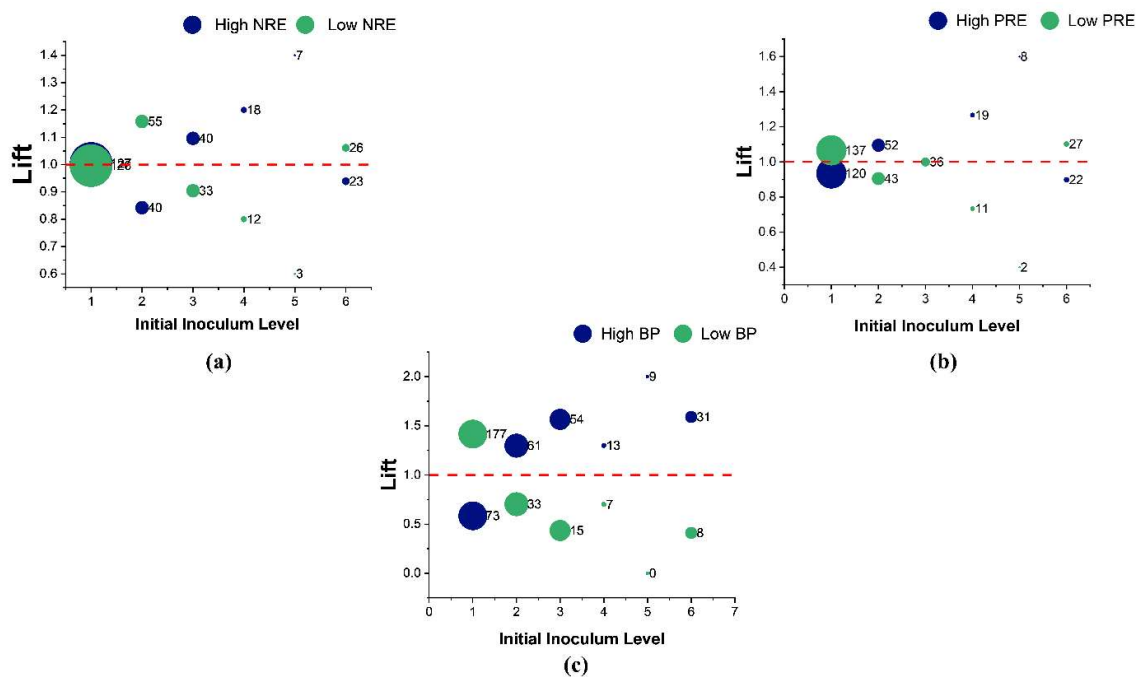


Figure 5.7. Lift and count values for ARM derived for Initial Inoculum Level: (a) NRE (b) PRE and (c) Biomass production.

Level 3 (0.205-0.305 g/L) and level 4 (0.305-0.405 g/L) proved to be beneficial for increasing biomass productivity and nutrient removal capacity according to the lift and count values for

ARM derived for IIL. Nevertheless, level 5 (0.405-0.505 g/L) also provided promising results (few data points supported it). Input of high inoculum level of microalgae at the start of the treatment helps in acclimatizing the microalgae species in the adverse environment of wastewater and thus decreases the lag phase of growth [229]. However, it should be noted that the supply of very high IIL leads to mutual shading due to which less light penetrates in the culture [520]. It has been also observed that very high IIL is responsible for auto inhibition in the treatment [521]. Kumar et al. (2014) cultivated *Chlorella* sp. (PSDK01) in shrimp-cultured effluent at different IIL ranging from 0.25 to 12 g/L [522]. Microalgae growth increased with an increase in IIL. However, after a certain IIL value, a declining trend in the growth was observed. Authors reported that 0.5 g/L IIL (Level 5 in the present case) was most suitable for both biomass productivity (0.78 g/L/d) and nutrient removal including phosphate (96%) and nitrate (69%) [522]. In contrast to this, 0.25 g/L (Level 3) was found to be most suitable for ammonia consumption (63%) [522].

IIL also affects biofilm-based microalgae production during the early stages of culture [572]. According to Li et al. (2021), IIL affected both biofilm production and microalga growth of *Chlorella vulgaris* during its cultivation in Bristol medium on sterile Petri dishes. There was a negative relation between IIL and microalgae growth due to less penetration of light during the supply of high IIL [572]. Thus, selecting the right IIL is crucial as it reduces the cost of operation, particularly in the large-scale wastewater treatment [471].

5.3.8. pH

Lift and count values for different association mining rules derived for pH, each for NRE, PRE and biomass production are shown in Table 5.10 and Figure 5.8.

Table 5.10. Lifts values for the association rules derived for pH.

Level	pH Range	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	4.86-6.71	1.09	0.91	1.06	0.94	1.03	0.97
2	6.71-7.18	1.04	0.96	1.02	0.98	1.14	0.86
3	7.18-7.64	0.94	1.06	0.95	1.05	0.92	1.08
4	7.64-8.10	0.98	1.02	1.03	0.97	0.93	1.07
5	8.10-9.5	0.78	1.22	1.00	1.00	0.96	1.04

Analysis of the influence of the pH on the growth of microalgae is difficult as every microalgal species has its own optimum pH range depending upon CO₂ solubility and enzyme activity [225]. However, based on lift and count values of the association rules, the optimum pH range for microalgae growth was of level 2 (6.71-7.18). pH affects the solubility, form and biotransformation of CO₂ and other nutrients present in the medium. In low pH range, the primary inorganic carbon source in the medium is CO₂. Microalgae directly uptake this form through the diffusion. At high pH, bicarbonate is the main form of inorganic carbon in the medium. Carbonic anhydrase assists in the uptake of this form of inorganic carbon [57]. pH also affects the form of ammonium nitrogen in the medium. As the pH is increased by a unit, the ratio of the free ammonia to total ammonium nitrogen is increased by a factor of 10. High concentration of free ammonia inhibits the growth of microalgae [573].

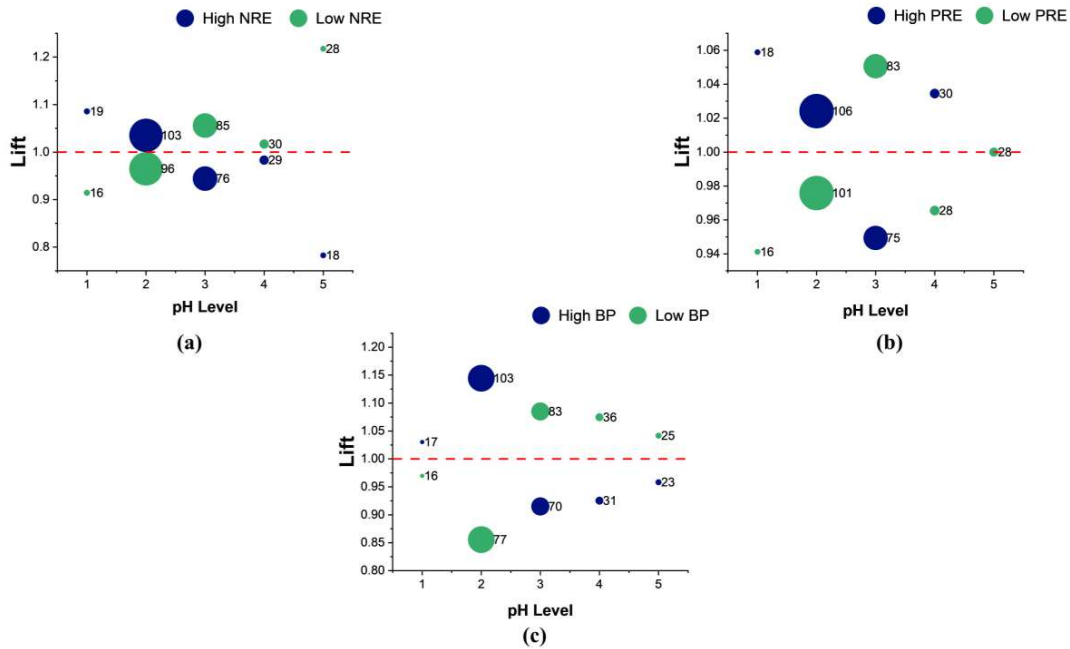


Figure 5.8. Lift and count values for ARM derived for pH: (a) NRE (b) PRE and (c) Biomass production.

Nayak et al. (2016) demonstrated that the growth of *Scenedesmus* sp. in domestic wastewater surged with the increase in pH. However, after a certain point, growth declined with further increase in pH. The highest biomass production of 1.37 g/L and more than 98% nutrient removal efficiency was obtained at pH 6.8 (level 2 in the present analysis) [226]. There are few reports which state that some of the the well-adapted microalgae species can grow at high pH. Li et al. (2019) reported a *Spirulina platensis* species that can grow at pH 9.5, proving that different microalgae species have different tolerance to pH levels [574].

5.3.9. Light Intensity and Photoperiod

Light conditions, including LI, photoperiod, wavelength is crucial for growth of microalgae as they provide the energy required for photosynthesis. Table 5.11 and Figure 5.9 represent the lift and count values for ARM derived for LI.

Table 5.11. Lifts values for the association rules derived for light intensity.

Level	LI Range	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	0	0.20	1.80	0.20	1.80	0.20	1.80
2	0.1-50	0.90	1.10	0.89	1.11	1.01	0.99
3	50-100	1.21	0.79	1.04	0.96	0.86	1.14
4	100-200	0.85	1.15	0.94	1.06	1.00	1.00
5	200-1500	1.05	0.95	1.22	0.78	1.14	0.86
6	1500-4050	1.38	0.62	1.00	1.00	1.67	0.33

As per the lift and count represented in Table 5.11 and Figure 5.9, level 5, i.e., LI in the range of 200-1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ resulted in the highest biomass productivity and nutrient removal efficiency. Level 2 and 6 also support high microalgae growth rate. However, they are supported by low lift and count values. LI provides the energy required to convert inorganic carbon into organic microalgal biomass by synthesising protein, carbohydrate and lipid, affecting the critical enzymes involved in the photosynthesis. Therefore, the supply of low LI will lead to a low growth rate and low biomass productivity, whereas high LI will decline the growth due to the formation of hydrogen peroxide. Arun et al. (2021) cultivated a mixed microalgae consortium at different light intensities and accessed their ammonium NRE during cultivation in a 5 L photo-sequencing batch bioreactor with a 2 L working volume. 100% ammonium NRE was obtained in the range of 200-500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (Level 5 in the present case). As the LI was increased beyond 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$, only 6.2% ammonium NRE was obtained [575]. Gao et al. (2022) cultivated *C. vulgaris* at different light intensities to test the uptake of various carbon sources in the BG 11 medium. The highest microalgae growth rate was obtained in the range of 8000-12000 lux intensity (around 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$) [576]. Thus,

level 5 of LI impacts positively on both microalgae biomass productivity and nutrient removal efficiency.

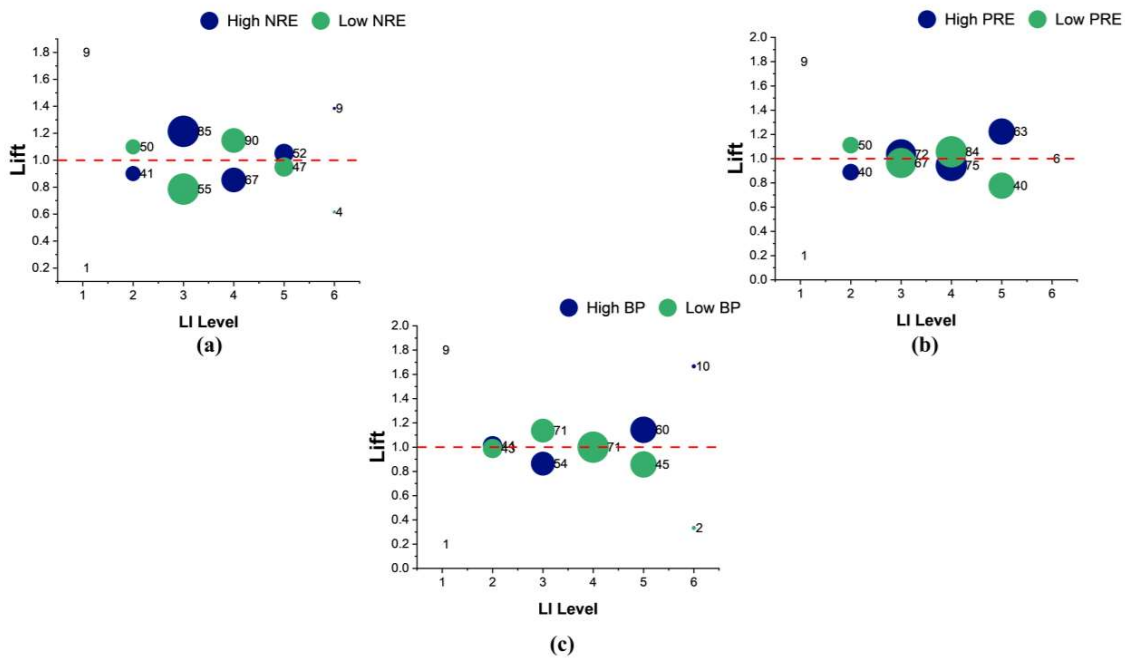


Figure 5.9. Lift and count values for ARM derived for light intensity: (a) NRE (b) PRE and (c) Biomass production.

Heterotrophic mode of cultivation, level 1 of LI and photoperiod results in higher biomass productivity than autotrophic mode of cultivation due to constant availability of carbon source [195], [196], [198]. The heterotrophic mode can be more favourable due to cost-effectiveness, simplicity and requirement of less maintenance [199]. However, results of ARM obtained in the present investigation do not support this phenomenon due to less count and lift values. Therefore, more research is needed to determine the effectiveness of the heterotrophic mode of cultivation and its scalability.

LI and photoperiod both have a synergistic effect on the growth of microalgae. Low LI with longer photoperiod will impact growth positively, whereas shorter photoperiod will have adverse influence on growth. Similarly, high LI with shorter photoperiod will have a positive

effect. However, a longer photoperiod duration will have negative effect. Table 5.12 and Figure 5.10 represent the lift and count values for ARM derived for photoperiod.

Table 5.12. Lifts values for the association rules derived for photoperiod.

Level	Photoperiod Range	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	0	0.20	1.80	0.20	1.80	0.00	1.00
2	10.8-13.5	0.92	1.08	0.97	1.03	1.02	0.49
3	13.5-16.2	1.32	0.68	1.27	0.73	1.20	0.40
4	16.2-24	0.91	1.09	0.88	1.12	0.86	0.57

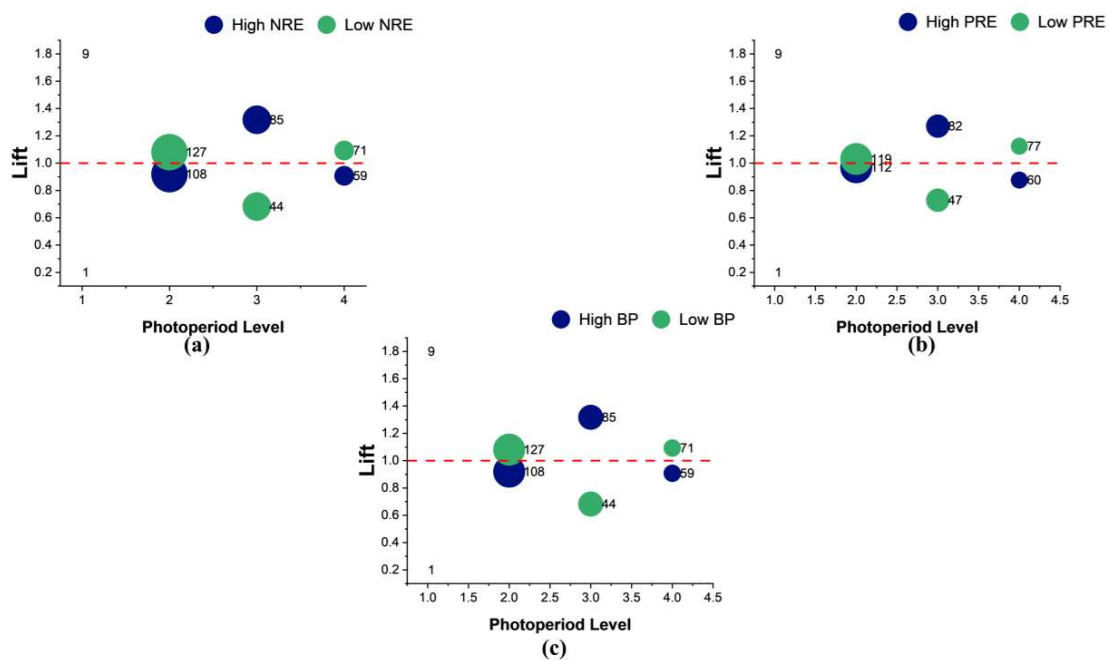


Figure 5.10. Lift and count values for ARM derived for photoperiod: (a) NRE (b) PRE and (c) Biomass production.

Level 3 (13.5-16.2 h) is beneficial for biomass productivity and nutrient removal efficiency. Any deviation from it degrades both growth and nutrient removal efficiency. Photoperiod duration of level 3 is greater than average natural sunlight duration, i.e., 12 hours. It has been

reported that increasing photoperiod duration beyond 12 hours led to high biomass productivity [577], [578]. Yan et al. (2016) demonstrated the synergistic effect of LI and photoperiod during the cultivation of *Chlorella* sp. in biogas slurry in transparent polyethylene photobioreactor [579]. Authors noted that optimized LI and photoperiod duration for high pollutant removal efficiency was: (i) low LI with long photoperiod ($300 \mu\text{mol m}^{-2} \text{s}^{-1}$; 16 h) (ii) medium LI with middle photoperiod ($600 \mu\text{mol m}^{-2} \text{s}^{-1}$; 14 h) and (iii) high LI with short photoperiod ($900 \mu\text{mol m}^{-2} \text{s}^{-1}$; 12 h) [579]. All strategies were in accordance with the association rules derived for LI (Level 5) and photoperiod (Level 3). Leong et al. (2022) cultivated *Chlorella vulgaris* in municipal wastewater and evaluated the effect of different light/dark photoperiod on growth and pollutant removal efficiency. 16:8 h light/dark duration (level 3 in the present case) yielded high biomass concentration (0.89 g/L) and more than 94% COD and ammonium nitrogen removal was obtained [580].

The third light condition, wavelength is also essential to increase lipid productivity, particularly during two-stage culture. In a two-stage culture strategy, red and blue wavelength of light are applied to increase biomass productivity. After sufficient growth, the wavelength is changed to green, acting as a stress factor. This increases lipid accumulation in the cells [581]. However, very scarce data is available in the literature that focus on the effect of wavelength on microalgae growth. Hence, in the present investigation ARM was not applied on wavelength.

5.3.10. Nitrogen/Phosphorus (N/P) Ratio

There is a synergistic effect of nitrogen and phosphorus concentration on microalgae growth [582]. In the present work, the association mining rules for nutrient concentration are presented in terms of N/P ratio (Table 5.13 and Figure 5.11).

Table 5.13. Lifts values for the association rules derived for nitrogen/phosphorus ratio.

Level	N/P Ratio Level	NRE		PRE		BP	
		High	Low	High	Low	High	Low
1	0.26:1-5.26:1	1.15	0.85	0.91	1.09	0.99	1.01
2	5.26:1-10.26:1	0.84	1.16	1.04	0.96	0.95	1.05
3	10.26:1-15.26:1	1.04	0.96	1.18	0.82	0.76	1.24
4	15.26:1-50.26:1	1.17	0.83	1.32	0.68	1.20	0.80
5	50.26:1-495.26:1	0.55	1.45	0.70	1.30	1.09	0.91

It is evident from Table 5.13 and Figure 5.11 that the highest biomass productivity was achieved in the range of 15.26:1-50.26:1 N/P (level 4) ratio with lift value 1.20 depending upon the lift and count values. Similar results were also obtained in the case of nitrogen and phosphorus removal efficiency with lift values 1.17 and 1.32, respectively. Any variation from this level led to decline in both biomass productivity and nutrient removal efficiency. The average stoichiometric formula of the elements present in the microalgal biomass is $C_{106}H_{181}O_{45}N_{16}P$ [582]. Therefore, for the high productivity, the N/P ratio of the growth medium should be around 16:1 or higher up to a specific limit, level 4 in the present case. There were 65 cases of level 4, out of which 21 cases were of livestock wastewater among all wastewater sources. Thus, livestock wastewater seemed to be one of the best mediums for achieving maximum microalgae biomass productivity [583].

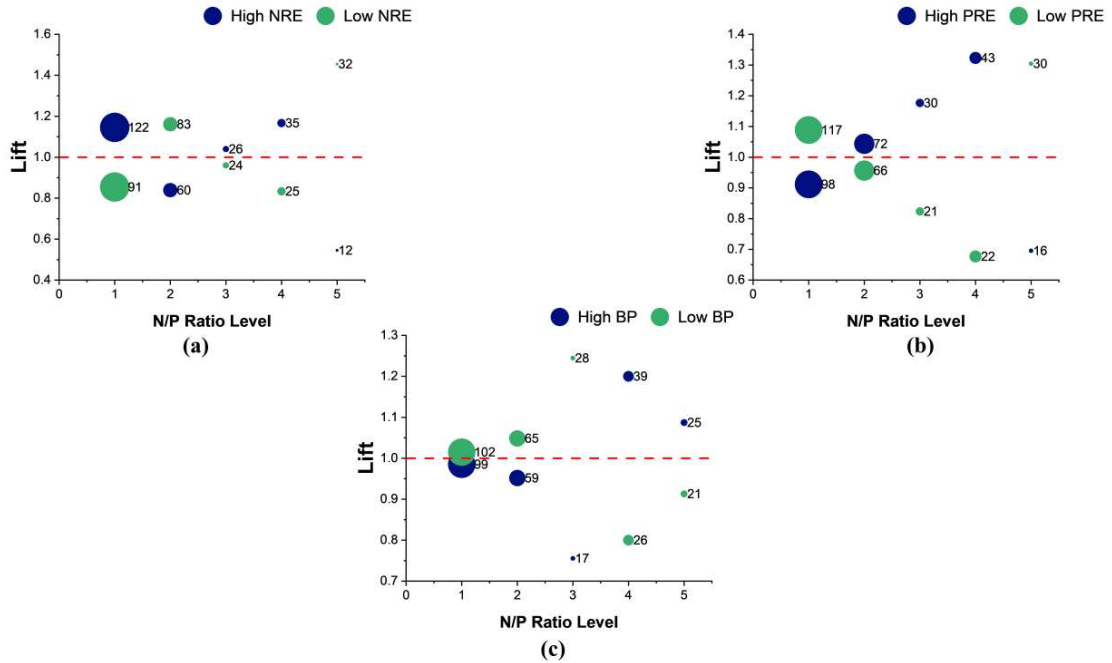


Figure 5.11. Lift and count values for ARM derived for Nitrogen/Phosphorus ratio: (a) NRE (b) PRE and (c) Biomass production.

During the cultivation of *Nannochloropsis* sp. at different N/P ratios of 16:1, 32:1, 64:1 and 80:1, the highest biomass productivity was obtained at 16:1 and 32:1 N/P ratios (level 4 in the present case). A decrease in biomass productivity was obtained at 64:1 and 80:1 N/P ratio [584]. It has been reported that as soon as the N/P ratio is decreases below 5:1, microalgae growth becomes stationary or even declines. Both biomass productivity and nutrient removal efficiency declined below the 5:1 N/P ratio during the cultivation of mixed consortia (*Chlorella sorokiniana* and *Scenedesmus* sp) and monoculture of *Chlorella* sp. [585].

5.4. Conclusions

Association rule mining technique was used to analyse the dataset for finding specific conditions/rules of cultivation parameters which lead to high biomass productivity and high nutrient removal efficiency. Various species of microalgae belonging to class *Chlorophyceae* & *Trebouxiophyceae* were found to be highly efficient in wastewater treatment. The livestock

wastewater with primary pre-treatment will be a cost-efficient choice if the sole aim is biomass production. The summary of the results has been shown in Table 5.14.

Table 5.14. Summary of the level and range/identity of critical predictor variables leading to high biomass productivity and wastewater treatment efficiency by microalgae.

S.No.	Predictor variables	Level	Range/Identity
1	Microalgae Class	2 & 6	<i>Chlorophyceae</i> & <i>Trebouxiophyceae</i>
2	Wastewater Source	4	Livestock
3	Pre-treatment Method	4	Primary
4	Reactor type	2	PBR
5	CO ₂ Content (% v/v)	2	0.53-2.53
6	Temperature (°C)	4	23.6-25
7	Initial Inoculum Level (g/L)	3 & 4	(0.205-0.305) & (0.305- 0.405)
8	pH	2	6.71-7.18
9	Light Intensity ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	5	200-1500
10	Photoperiod (h)	3	13.5-16.2
11	N/P Ratio	4	15.26:1-50.26:1

