Chapter 4

Decision tree analysis for predicting the best combination of predictor variables for increasing wastewater treatment capability and biomass production (for specific class)

Decision tree analysis for predicting the best combination of predictor variables for increasing wastewater treatment capability and biomass production (for specific class)

4.1. Introduction

In the previous chapter, analysis was done considering general strains. But, every microalga species has a different tolerance level towards cultivation parameters or predictor variables such as nutrient concentration in wastewater, LI, pH, temperature, CO_2 content and many others. Each of them has its own optimum combination of predictor variables under which they display a high potential to remediate pollutants from wastewater and increase biomass productivity. Therefore, in order to enhance the efficiency of the microalgae-based wastewater treatment, it is necessary to select the right strain and supply the right combination of cultivation parameters. Therefore, the present chapter aims at providing the right combination of predictor variables for the cultivation of microalgae belonging to two classes, i.e., *Chlorophyceae* and *Trebouxiophyceae* in wastewater. There were two main reasons for selecting these two classes: (i) The predictor importance study in Chapter 3 indicated that species belonging to classes *Chlorophyceae* and *Treboxiophyceae* highly influences nutrient removal efficiency and biomass production, (ii) According to the literature study, most microalgae species used for wastewater treatment belong to the mentioned classes [271], [426]. So, a large dataset for further Decision Tree analysis can be constructed individually for both classes.

This study started with constructing a dataset representing output variables, namely, nitrogen removal efficiency (NRE), phosphorus removal efficiency (PRE) and biomass production for both the classes, by collecting data from recent scientific articles (2010-2021). Thereafter, the decision tree function was applied to analyse the constructed dataset. The best combination of predictor variables was detected for maximising the nutrient removal capability and biomass productivity of both the classes.

4.2. Methodology

4.2.1. Dataset construction and pre-processing

The dataset referring to output variables NRE (239 data points), PRE (239 data points), biomass production (214 data points) for class *Trebouxiophyceae*, and NRE (200 data points), PRE (204 data points), biomass production (206 data points) for class *Chlorophyceae* was extracted from information available in published literature [24], [29]–[31], [136], [145], [146], [153], [156], [158], [159], [170], [172], [175], [176], [178], [180], [192], [205], [210], [214], [219], [220], [226], [236], [247]–[250], [253], [254], [256], [259], [260], [262]–[264], [266], [267], [269], [273]–[275], [277], [280], [286], [289], [294], [300], [302], [303], [305], [308]–[310], [312]–[314], [316], [317], [320], [322], [324], [330], [331], [333], [336], [338]–[340], [343], [344], [346], [348], [392], [413], [416], [427]–[516]. The cultivation parameters or predictor variables for evaluation were considered based on correlation analysis [209]. The selected predictor variables showing correlation with the output variables and their range or identities considered in studies are shown in Table 4.1. Variables such as reactor volume, type of light source, light wavelength and cultivation time were not considered, as they did not show any correlation with output variables. Predictor variables were comprised of both categorical and continuous variables, which were classified into three groups: (i) pre-cultivation stage deciding factors; (ii) growth components in wastewater; (iii) operating variables. The hot encoding technique encoded categorical variables. The output variables were clustered into two classes: (i) low-level class and (ii) high-level class. The median of output variables was used to set up the threshold limit. Values lower than the threshold were clustered in the low-level class. Values higher than the threshold were clustered into high-level class. At the last step of data-pre-processing, all the null values were removed as they affect the performance of machine learning algorithms and decrease their accuracy. Also, units of all variables were standardised.

Table 4.1. Cultivation parameters used during decision tree analysis.

S.No.	Predictor variables	Ranges for continuous variables or identities for categorical variables
1	Microalgae Class	*Chlorodendrophyceae, Chlorophyceae, Consortium, Cyanophycean, Eustigmatophyceae, Trebouxiophyceae Xanthophyceae*
2	Wastewater source	Domestic, Municipal, Agriculture, Industrial, Synthetic, Livestock
3	Cultivation type	Autotrophic, Heterotrophic, Mixotrophic
4	Pre-treatment method	Primary, Aerobic, Anaerobic, Autoclave
5	Reactor Type	Flasks, PBR
6	CO_2 Content	0.03-45% (v/v)
7	Temperature	4-43 °C
8	Initial Inoculum Level	0.0058-12 g/l
9	pH	4.86-9.5
10	Light Intensity	0-4033 µmol m^{-2} s^{-1}
11	Photoperiod	0-24 h
12	N/P	0.26-494

4.2.2. Decision Tree Analysis

The decision tree algorithm was used for analysing the dataset. The combinations of predictor variables generating low and high levels of output variables for each class were determined. Figure 4.1 shows the steps involved in decision tree analysis. MATLAB (version 9.2; R2017a) function "fitctree" was applied for constructing decision trees by default Classification and Regression Trees (CART) algorithm. "OptimizeHyperparameters", one of the options of the "fitctree" function, was used to build optimum decision tree models by optimising hyperparameters and minimising the cross-validation error. Hyperparameters indicate the parameters whose values needed to be set before performing machine learning analysis [517]. In the present analysis, all hyperparameters including "split criterion", "MaxNumSpilts", "MinLeafSize", and "NumVariablesToSample" were optimised.

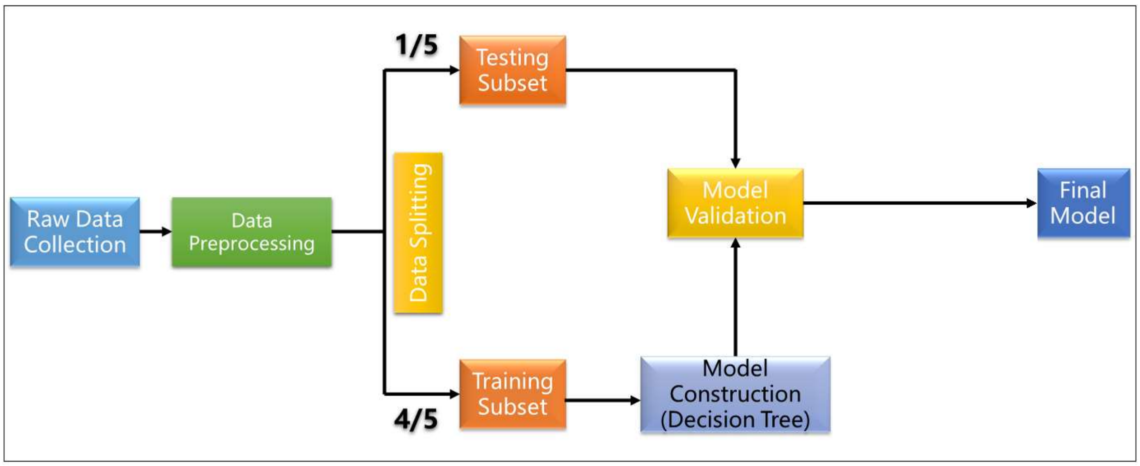

Figure 4.1. General steps involved during decision tree analysis.

The training and testing of the model were executed by dividing the whole dataset randomly into two subgroups: (i) training set consisting of 80% of the whole dataset; (ii) testing subset comprising of 20% of the whole dataset. This step is needed for building a generalised tree model that can categorise any new experimental finding with high precision. Division of the dataset into training and testing sets was performed by MATLAB function "cvpartition", and the "HoldOut" validation procedure performed validation of the trained models. Different decision tree models were constructed using the training set. The model with minimum cross validation error was further picked up for testing the dataset and predicting the final accuracy of the selected model. It should be noted that no data from the testing set was used during the training step of decision tree models. Finally, the significant predictor variables which affected the construction of the decision tree models and the microalgae-based wastewater treatment were determined by MATLAB function "predictorImportance".

4.3. Results and Discussion

Three decision tree models for both classes were constructed. Various combinations of predictor variables leading to low and high-level of output variables were determined afterwards by decision tree models.

4.3.1. Decision Tree Analysis for class *Trebouxiophyceae*

4.3.1.1. Nitrogen Removal Efficiency

The optimum tree model representing various combinations of predictor variables leading to the low and high level of NRE for class *Trebouxiophyceae* is shown in Figure 4.2. Here, anaerobic pretreatment serves as the root of the decision tree.

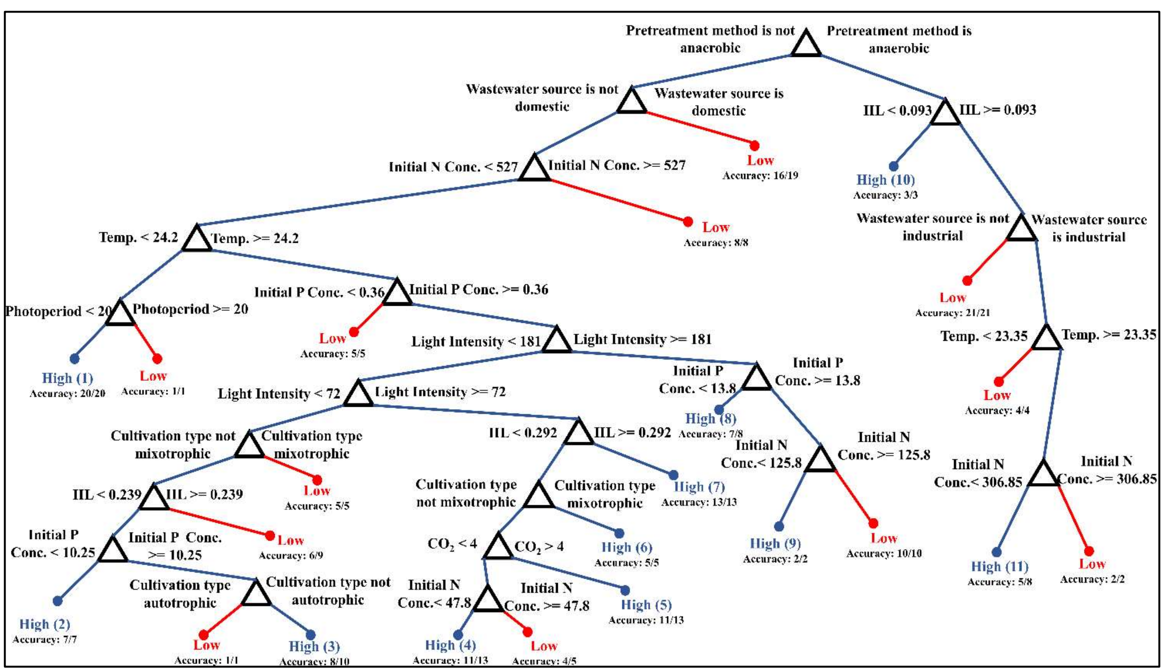

Figure 4.2. Decision tree classification model for NRE for class *Trebouxiophyceae* (refer to Table 4.1 for units)

This tree has 11 different combinations or paths of predictor variables showing high NRE. Starting from the left-hand side, when anaerobic treatment is not done, domestic wastewater source provides the next branching level. Other predictor variables that influenced the growth of the tree were initial phosphorus concentration (Initial P. concentration), initial nitrogen concentration (Initial N. concentration), temperature (Temp.), initial inoculum level (IIL), CO_2 content, LI, photoperiod, and cultivation type. There were nine different paths that led to a high level of NRE. Among them, the first and seventh path provided the most generalised rules. In the case of the first path, pretreatment was other than anaerobic, wastewater source was other than domestic, Initial N. concentration < 527 mg/L, Temp. < 24.2 ˚C and photoperiod < 20 hours. The accuracy of this path was nearly 100% and reinforced by 20 data points. Statistically, this path was the most reliable for class *Trebouxiophyceae*. Conditions for the seventh path were pre-treatment method (other than anaerobic), wastewater source (other than domestic), Initial N. concentration < 527 mg/L, temp. >=24.2 ˚C, Initial P. concentration ≥ 0.36 mg/L, LI between 72-181 µmol m^{-2} s^{-1} and IIL ≥ 0.292 g/L. The accuracy of this path was also 100%. However, this was reinforced by 13 data points only.

For the right-hand side, IIL provided the next branching level when anaerobic pretreatment was done. Other factors that influenced the tree's growth were wastewater source, Temp., and initial N. concentration. There were only two paths that led to high NRE. The most generalised combination of predictor variables on the right-hand side was provided by path no. eleven where pretreatment method was anaerobic, IIL ≥ 0.09 g/L, wastewater source was industrial, temp. ≥ 23.35 ˚C and initial N. concentration < 306.85 mg/L. The accuracy of this path was 62.5%, reinforced by eight data points. At last, the general accuracy of the model was determined by a confusion matrix, which was constructed by simulating the rules extracted from the tree on the testing set, as shown in Figure 4.3. Confusion matrix for NRE for class *Trebouxiophyceae*.. The data in the testing set was new for the models and never used during the training step. As shown in Figure 17, the model correctly classified 17 low-level data points out of 25 and 17 high-level data points out of 22 with a general accuracy of 72.34%.

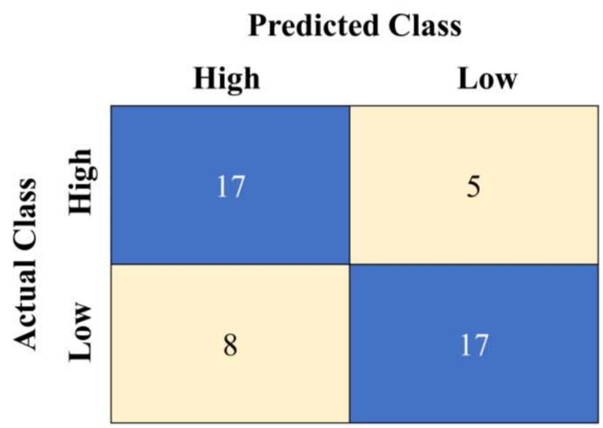

Figure 4.3. Confusion matrix for NRE for class *Trebouxiophyceae*.

Now, for determining the influence of each input variable on the growth of the decision tree models, the "predictorImportance" function was used, and the results are shown in Figure 4.4.

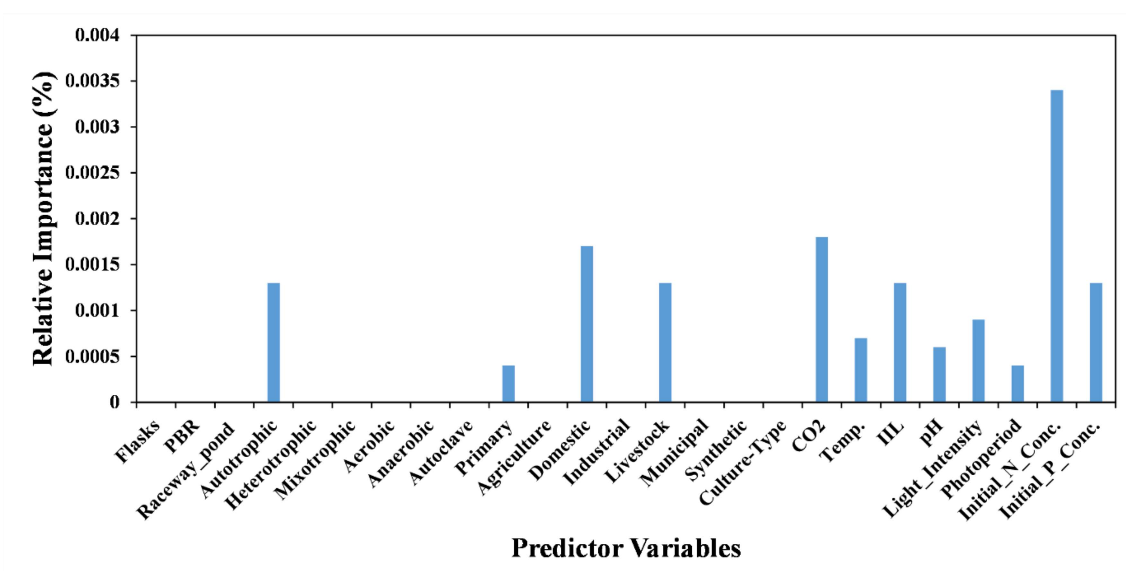

Figure 4.4. Relative importance of the predictor variables on NRE for class *Trebouxiophyceae*.

As evident from Figure 4.4, operating variables including temp., IIL, pH, LI, CO_2 content and nutrient concentration were more influential than other predictor variables during the construction of decision tree models. As expected, the initial N. concentration was more significant than the initial P. concentration. A higher concentration of nitrogen (>527 mg/L) as predicted by the first path acted as a growth reducing factor for class *Trebouxiophyceae*. Tam and Wong (1996) estimated the growth of *Chlorella* species belonging to class *Trebouxiophyceae*, at different N. concentrations. The authors observed that growth was inhibited at 700 mg/L ammonium [518]. Akerstrom et al. (2014) cultivated *Chlorella sp.* at different nitrogen concentrations ranging from 221-1210 mg/L in sludge liquor. NRE increased from 47.05% to 83.74 %, with the surge in N concentration from 221 mg/L to 363 mg/L. After that, NRE decreased with a further increase of N [448]. Therefore, pretreatment of wastewater such as dilution is needed in case of high N. concentration (> 500 mg/L) in the wastewater.

4.3.1.2. Phosphorus Removal Efficiency

Decision tree analysis determined different combinations of predictor variables leading to high PRE for class *Trebouxiophyceae* which is represented by optimum tree structure as shown in Figure 4.5.

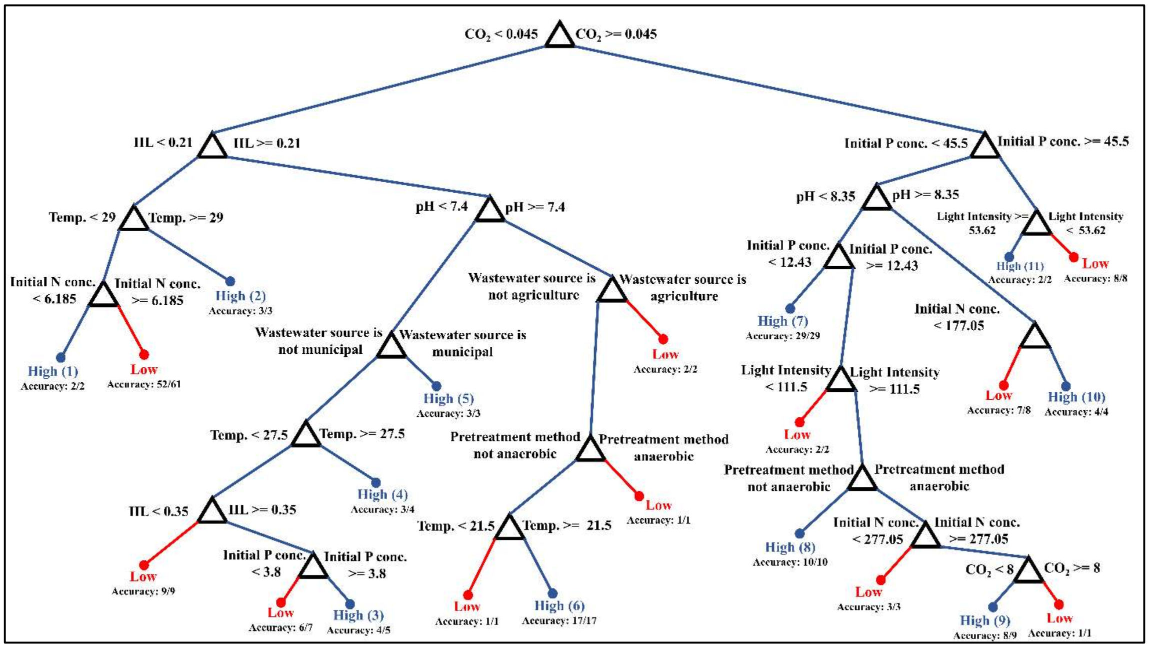

Figure 4.5. Decision tree classification model for PRE for class *Trebouxiophyceae* (refer to Table 4.1 for units).

The tree first separates the dataset according to the CO_2 content supplied to the cultivation medium. Decision tree analysis identified 11 different paths leading to high PRE. In case of CO_2 content less than 0.045 %(v/v), IIL provides the next cut. Branching continues further with operating variables, nutrient concentration, wastewater source and pretreatment method, providing six different combinations of predictor variables leading to high PRE. Among these, the sixth path provided the most generalised combination as it has higher accuracy (100%) than others and is supported by 17 data points. Rules provided by this path were $CO_2 < 0.045$ %(v/v), IIL ≥ 0.21 g/L, pH ≥ 7.4, no presence of agricultural wastewater, pre-treatment method is not anaerobic and Temp. ≥ 21.5^{o}C. For the case of CO_2 content more than 0.045% (v/v), initial P. concentration provides the next cut and branching is continued by pH, LI, initial N. concentration and pretreatment method, providing five paths or different combinations of predictor variables leading to high PRE. Out of these, the seventh path provided the most generalised rules, which were $CO_2 \geq 0.045$%(v/v), pH < 8.35 and initial P. concentration < 12.43 mg/L.

The accuracy of this path was 100% and was reinforced by 29 data points. Rules extracted from this path are the most reliable for increasing PRE for the case of class *Trebouxiophyceae*. Next, to further justify the rules extracted from the decision tree analysis, they were tested on the testing dataset and accuracy was determined by a confusion matrix.

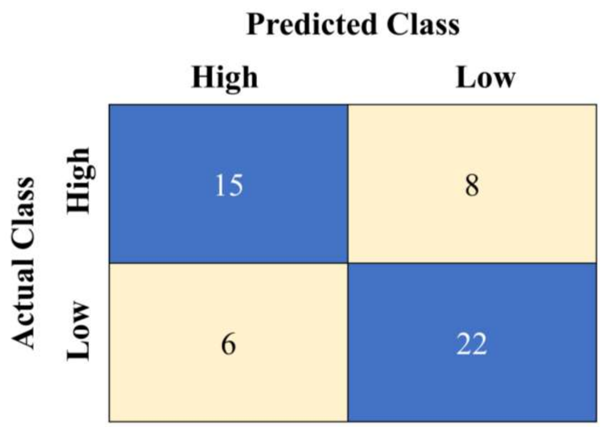

Figure 4.6. Confusion matrix for PRE for class *Trebouxiophyceae*.

The model correctly classified 22 low-level data points out of 28 and 15 high-level data points out of 23 with a general accuracy of 72.5%.

Relative importance of the predictor variables is shown in Figure 4.7.

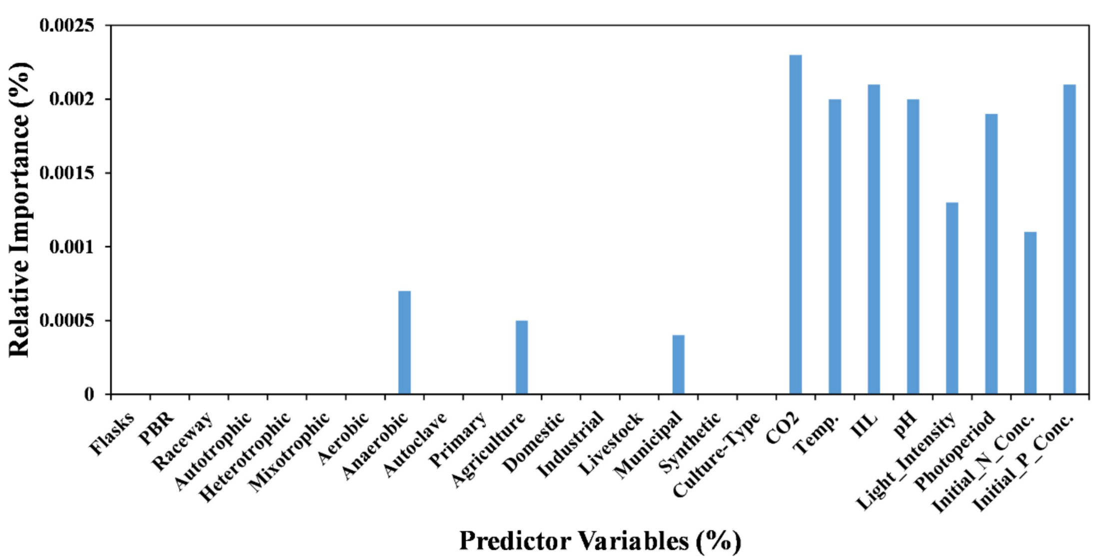

Figure 4.7. Relative importance of the predictor variables on PRE for class *Trebouxiophyceae*.

Similar to the NRE, the construction of the model was influenced significantly by nutrient concentration in wastewater and operating variables. In the case of PRE, initial P. concentration was more influential than initial N. concentration. With the increase of initial P. concentration, there was a slight decrease in PRE as compared to the trend of NRE [448]. This may be due to the fact that *Trebouxiophyceae* class has high P. uptake capacity. They can remediate and store excess P. in the form of polyphosphate granules, which can be utilised during phosphate deficient conditions [519]. Among the pretreatment methods, the anaerobic pretreatment method exceedingly influenced PRE over other methods.

4.3.1.3. Biomass Production

The optimum tree model representing various combinations of predictor variables leading to the low and high level of biomass production for class *Trebouxiophyceae* is shown in Figure 4.8.

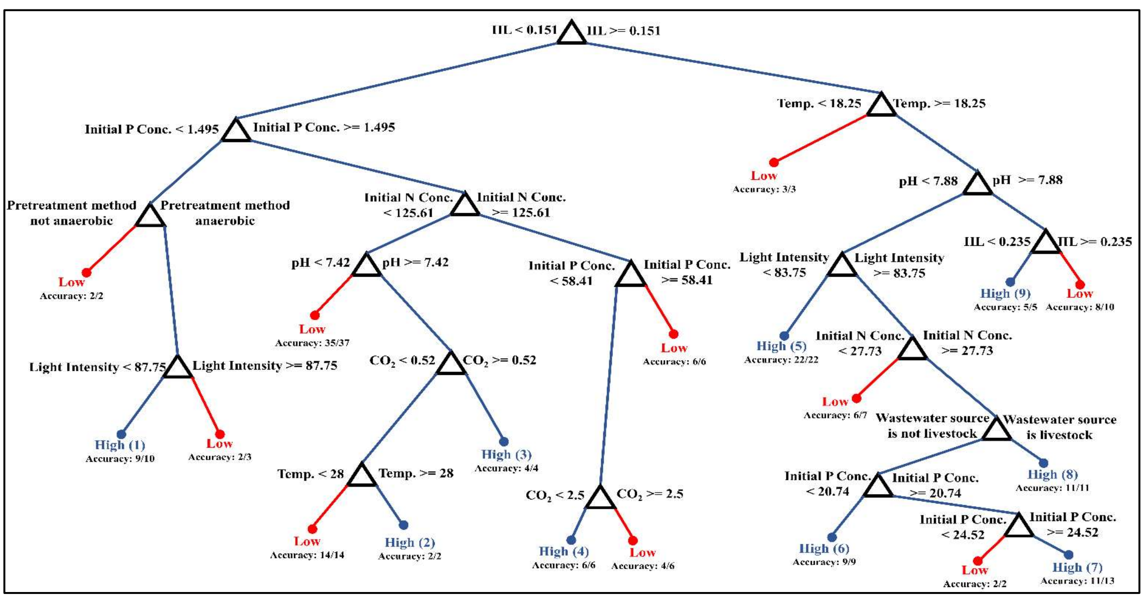

Figure 4.8. Decision tree classification model for biomass production for class *Trebouxiophyceae* (refer to Table 4.1 for units).

The tree splits the whole dataset according to IIL level. It detected nine different paths leading to high biomass production for class *Trebouxiophyceae*. For the left-hand side, where IIL < 0.151 g/L, initial P. concentration provides the next cut, and the tree's growth was influenced by pretreatment method, nutrient concentration, and operating variables, detecting four different paths leading to high biomass production. The first path provided the most generalised rule whose conditions were IIL < 0.151 g/L, initial P. conc. < 1.495 mg/L, anaerobic pretreatment and LI < 87.75 μmol m^{-2} s^{-1}, out of these. Accuracy of this path was 100% and reinforced by 10 data points. It is worth noting that initial nutrient concentration becomes one of the significant variables at a low IIL level on the left side. The reason for this phenomenon has been provided during the relative importance analysis of predictor variables. For the right-hand side, where IIL ≥ 0.151 g/L, Temp. provides the next cut and growth of the tree is influenced by operating variables and initial nutrient concentration, detecting five different paths leading to high biomass production. Out of them, path no. fifth and seventh provided the most generalised rules for increasing biomass production for class *Trebouxiophyceae*. Conditions for the fifth path were IIL ≥ 0.151 g/L, Temp. ≥ 18.25 ˚C, pH < 7.88 and LI < 83.75 μmol m^{-2} s^{-1}. The accuracy of this path was 100% and was supported by 22 data points. For path no. seven, the conditions were IIL ≥ 0.151 g/L, Temp. ≥ 18.25 ˚C, pH < 7.88 and LI ≥ 83.75 μmol m^{-2} s^{-1}, initial N. concentration ≥ 27.73 mg/L, wastewater source other than livestock, initial P. concentration ≥ 24.52 mg/L. The accuracy of this path was 84.61% and was held by 13 data points. The final accuracy of the model was determined by the confusion matrix, as shown in Figure 4.9. The model correctly predicted 17 low-level data points out of 23 and 16 high-level data points out of 19, with a general accuracy of 78.57%.

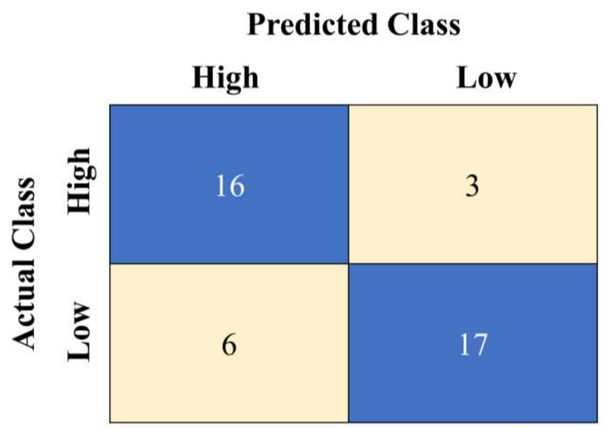

Figure 4.9. Confusion matrix for biomass production for class *Trebouxiophyceae*.

The relative importance of the predictor variables affecting the growth of the decision tree models is shown in Figure 4.10.

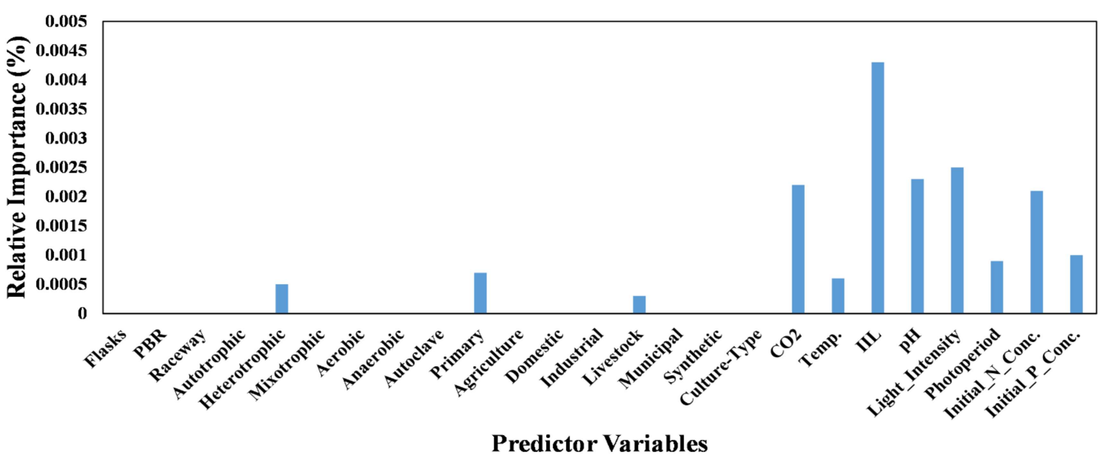

Figure 4.10. Relative importance of the predictor variables on biomass production for class *Trebouxiophyceae*.

In this case, also operating variables were more influential than other variables. IIL is of higher importance among the operating variables than the other operating variables as it vastly affects the microalgae growth in wastewater, particularly during the treatment process. Due to the existence of toxic chemicals and heavy metals in wastewater, the lag phase of the growth curve increases. However, if high IIL is supplied at the start of wastewater treatment, then it readily assists in the acclimatisation of the culture in wastewater. Thus, the lag phase duration decreases [229]. Wang et al. (2010) cultivated *Chlorella sp.* in dairy manure at different initial inoculum concentrations. The authors noticed that with the increase of IIL from 0.14 g/L to 0.18 g/L, there was a simultaneous increase in biomass concentration from 1.4 g/L to 1.8 g/L [416]. However, a very high IIL is not suitable for microalgae growth. It leads to mutual shading, and less light penetrates the culture [520]. Kumar et al. (2014) cultivated *Chlorella sp.* in shrimp-cultured effluent at different IIL ranging from 0.25-12 g/L. It was observed that biomass production also improved with the increase of IIL. Nonetheless, there was no significant increase in biomass concentration above 8-12 g/L. Even in few case studies, biomass concentration started to decrease at a very high concentration of IIL due to autoinhibition [471], [521]. Hence, the selection of optimised IIL for large-scale wastewater treatment is crucial as it reduces the cost of operation [522].

4.3.2. Decision Tree Analysis for class *Chlorophyceae*

4.3.2.1. Nitrogen Removal Efficiency

The optimum tree model representing various combinations of predictor variables leading to the low and high level of NRE for class *Chlorophyceae* is shown in Figure 4.11. Here, primary pretreatment serves as the root of the decision tree.

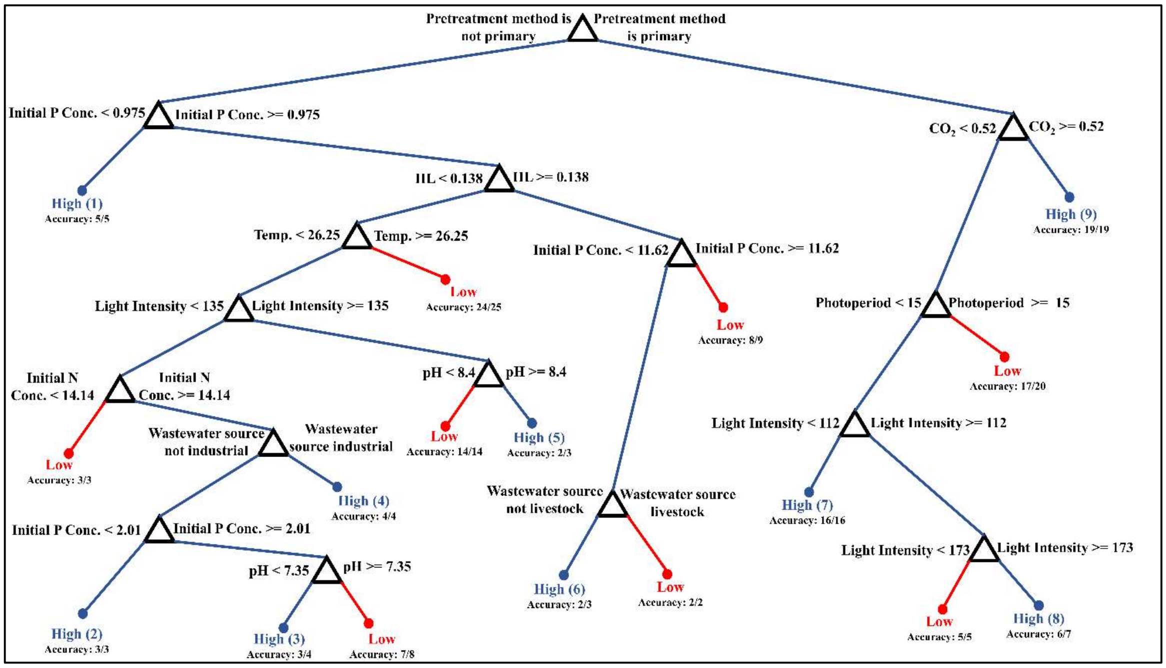

Figure 4.11. Decision tree classification for NRE for class *Chlorophyceae* (refer to Table 4.1. Cultivation parameters used during decision tree analysis. for units).

Here, the primary pretreatment method of the wastewater serves as the root of the tree and nine different paths leading to high NRE for class *Chlorophyceae* were detected. Starting from the left side, when primary pretreatment is not done, the initial P. concentration provides the next level of branching. The decision tree detected six paths leading to high NRE. Amongst these, the sixth path provided the most generalised set of combinations in which the pretreatment method was not primary, initial P. concentration ranged between 0.975 and 11.6 mg/L, IIL ≥ 0.13 g/L with no presence of livestock wastewater. The accuracy of this path was 100% and was strengthened by 13 data points. The right-hand side presents the cases where primary pretreatment was done. Here, CO_2 content provided the next cut, and LI and photoperiod influenced tree growth. There were only three paths on the right-hand side leading to high NRE. Among them, the seventh path provided the most generalised set of rules for increasing NRE whose conditions were primary pretreatment method, CO_2 content < 0.52% (v/v), photoperiod < 15 h and LI < 112 μmol m^{-2} s^{-1}. The accuracy of this path was 100% and was strengthened by 16 data points. This path provided the most reliable set of rules for increasing NRE in the case of class *Chlorophyceae*. The general accuracy of the model was determined by the confusion matrix, as shown in Figure 4.12. The model correctly classified 19 low-data points out of 23 and 15 high-level data points out of 23, with a general accuracy of 73.91%.

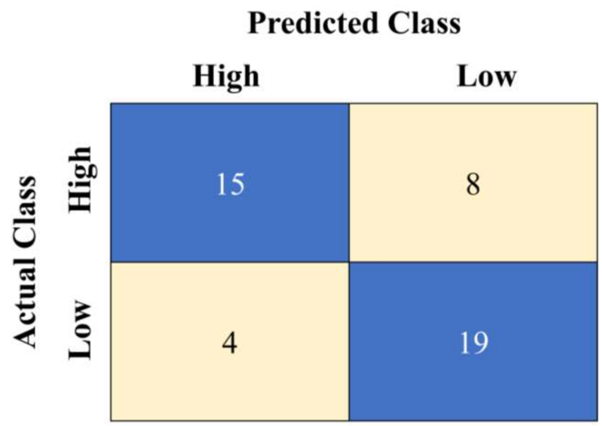

Figure 4.12. Confusion matrix for NRE for class *Chlorophyceae*.

Next, the relative importance of each parameter is shown in Figure 4.13.

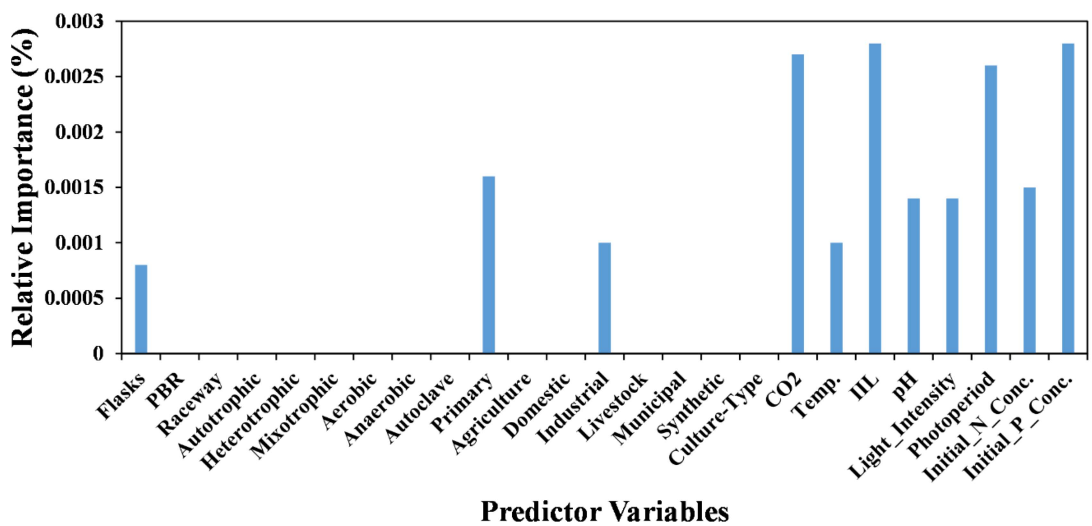

Figure 4.13. Relative importance of the predictor variables on NRE for class *Chlorophyceae*.

Operating variables were more influential than other variables, among which CO_2 content and IIL were more significant. Another critical point may be noted that for class

Chlorophyceae, initial P. concentration was more influential than initial N. concentration as compared to class *Trebouxiophyceae*.

4.3.2.2. Phosphorus Removal Efficiency

The optimum tree model representing various combinations of predictor variables leading to the low and high level of PRE for class *Chlorophyceae* is shown in Figure 4.14.

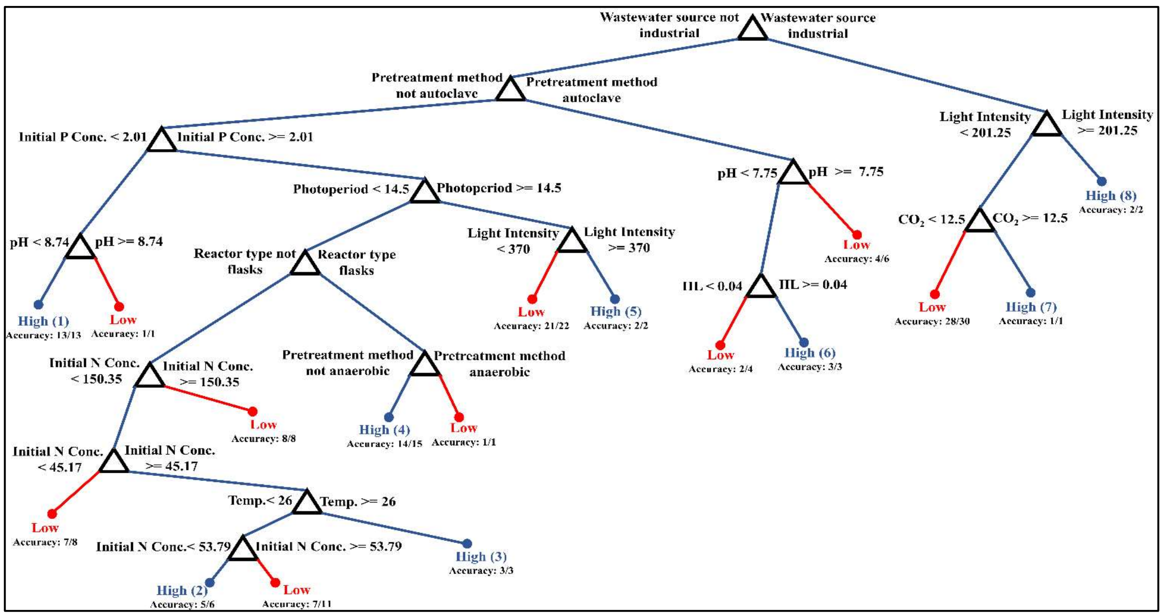

Figure 4.14. Decision tree classification for PRE for class *Chlorophyceae* (refer to Table 4.1 for units).

Here, industrial wastewater source acted as the tree's root, which means left-hand cases will not be applicable for industrial wastewater sources. The decision tree determined eight different paths leading to high PRE in the case of class *Chlorophyceae*. For the left-hand side, when the wastewater source is not industrial, the autoclave pretreatment method provided the second level of branching. Other variables that influenced the tree's growth were initial P. concentration, pH, photoperiod, LI, reactor type, pretreatment method, temp., and initial N. concentration. Total six paths led to high PRE, out of which the sixth path provided the most generalised set of rules. The conditions of the sixth path were wastewater sources other than industrial, autoclave pretreatment method, pH < 7.75 and IIL ≥ 0.04 g/L. The accuracy of this path was 100% and reinforced by 31 data points. The sixth path provided the most reliable set of rules or combination of predictor variables leading to high PRE for the case of class *Chlorophyceae*. A minimal branching level occurred on the right side, where tree started with industrial wastewater source and LI provided the next branching level. Two paths led to high PRE only, out of which the seventh path provided the most generalised set of rules. The conditions of the seventh path were industrial wastewater source, LI < 201.25 μmol m^{-2} s^{-1}, and CO_2 ≥ 12.5 %(v/v). The general accuracy of the model was estimated by a confusion matrix, as shown in Figure 4.15. The model predicted 12 low-level data points out of 19 and 20 high-level data points out 21 correctly with a general accuracy of 80%.

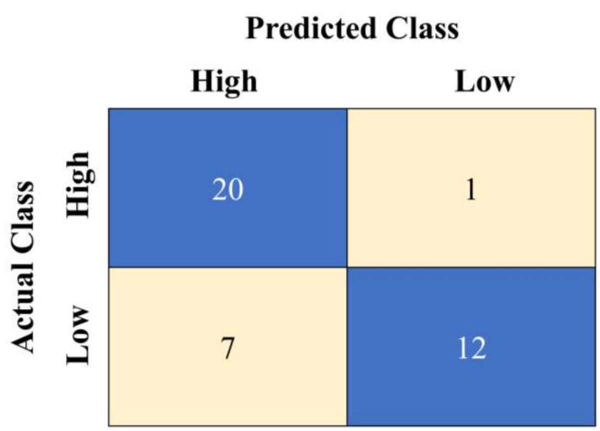

Figure 4.15. Confusion matrix for PRE for class *Chlorophyceae*.

Figure 4.16 represents the relative importance of predictor variables influencing the PRE for class *Chlorophyceae*.

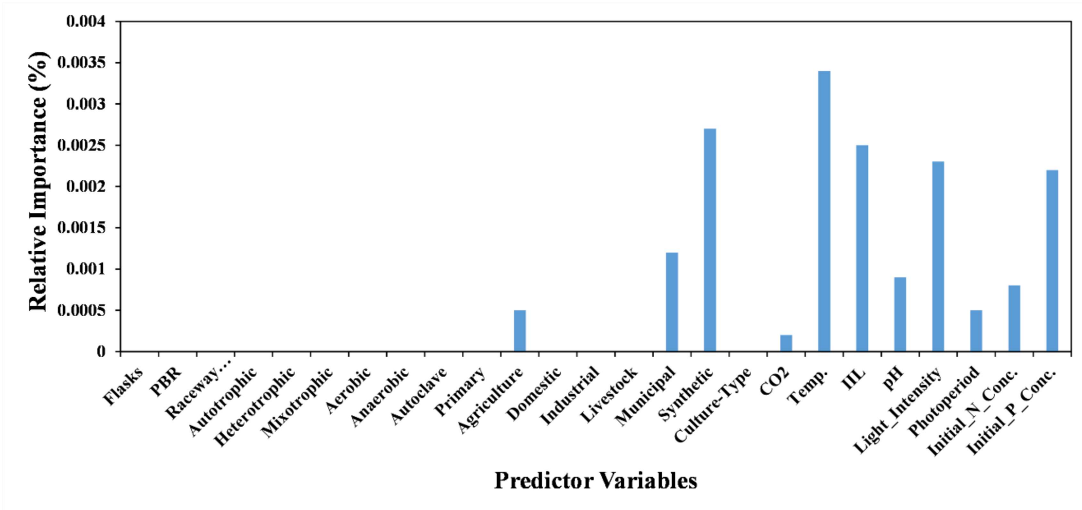

Figure 4.16. Relative importance of the predictor variables on PRE for class *Chlorophyceae*.

As concluded by "Relative Importance" function, among operating variables temperature was more influential than other variables. Effect of temperature on PRE is related to microalgae growth. Ma et al. (2017) cultivated *Scenedesmus sp.* Z-4 belonging to class *Chlorophyceae* in molasses wastewater at different temperature level ranging from 4°C to 25°C. It was observed that as the temperature was increased from 4°C to 25°C, PRE also increased from 44.12 % to 83.3 % [495]. Among the nutrient concentration, initial P. concentration was more influential than initial N. concentration similar to the trend observed in class *Trebouxiophyceae*. This trend was also supported by the study of Ajala and Alexander (2020). They cultivated *Scenedesmus obliqus*, belonging to class *Chlorophyceae*, at different initial concentration of phosphate. PRE decreased from 100 % to 50.17 % as with the increase in phosphate concentration from 17.2 g/L to 145.1 g/L [512].

4.3.2.3. Biomass Production

The optimum tree model representing various combinations of predictor variables leading to the low and high level of biomass production for class *Chlorophyceae* is shown in Figure 4.17.

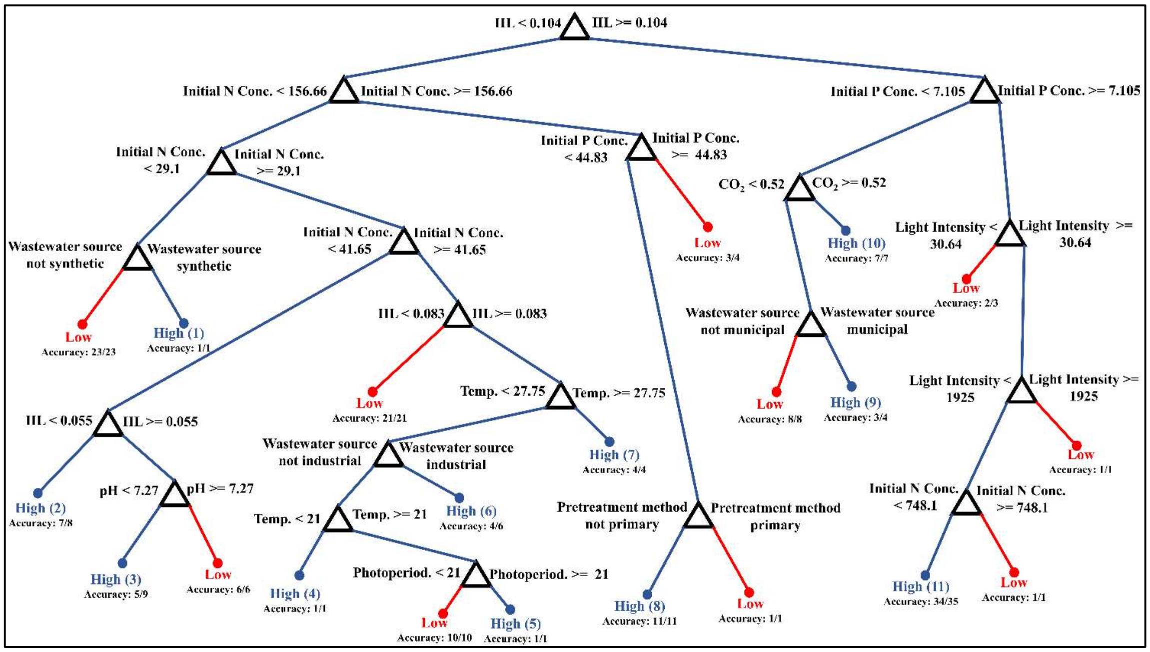

Figure 4.17. Decision tree classification for biomass production for class *Chlorophyceae* (refer to Table 4.1 for units).

Similar to class *Trebouxiophyceae*, here also IIL level divided the tree into two parts serving as the root of the tree and at the left-hand side initial nutrient concentration was more significant. Eleven paths leading to high biomass production were determined for class *Chlorophyceae*. The first eight paths were present on the left-hand side, where the tree starts with IIL < 0.104 g/L. Initial N. concentration provides the next cut, and further branching was influenced by initial P. concentration, pre-treatment method, wastewater source, pH, temp., and photoperiod. Among all the paths, the eighth path provided the most generalised set of rules: IIL> 0.104 g/L, initial N. concentration ≥ 156.66 mg/L, initial P. concentration < 44.83 mg/L, and pretreatment method was other than primary. The accuracy of this path was 100% and reinforced by 11 data points. The remaining three paths (9,10,11) were present on the right-hand side, where the tree starts with IIL ≥ 0.104 g/L. Initial P. concentration provided the next cut. Other factors that influenced the branching were CO_2 content, wastewater source, LI, and initial N. concentration. Path no. 11 provided the most reliable set of rules: IIL≥ 0.104 g/L, initial P. concentration ≥ 7.105 mg/L, LI between 30.64 and 1925 µmol m^{-2} s^{-1}, and initial N. concentration < 748.1 mg/L.

The accuracy of this path was 97.14% and was held by 35 data points. A confusion matrix was used to determine the model's final accuracy, as shown in Figure 4.18. The model correctly classified 14 low-level data points out of 17 and 16 high-level data points out of 24, with a final accuracy of 73.17%.

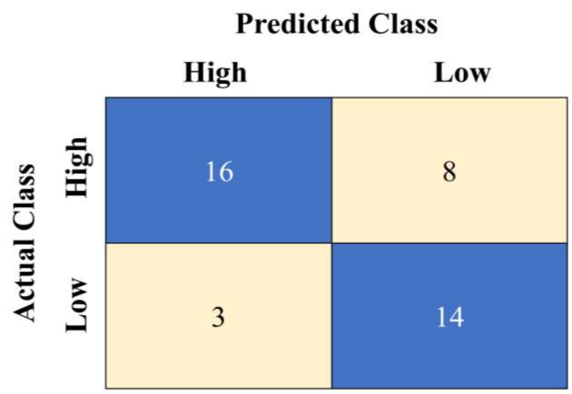

Figure 4.18. Confusion matrix for biomass production for class *Chlorophyceae*.

Next, the relative importance of each parameter influencing the decision tree models is shown in Figure 4.19.

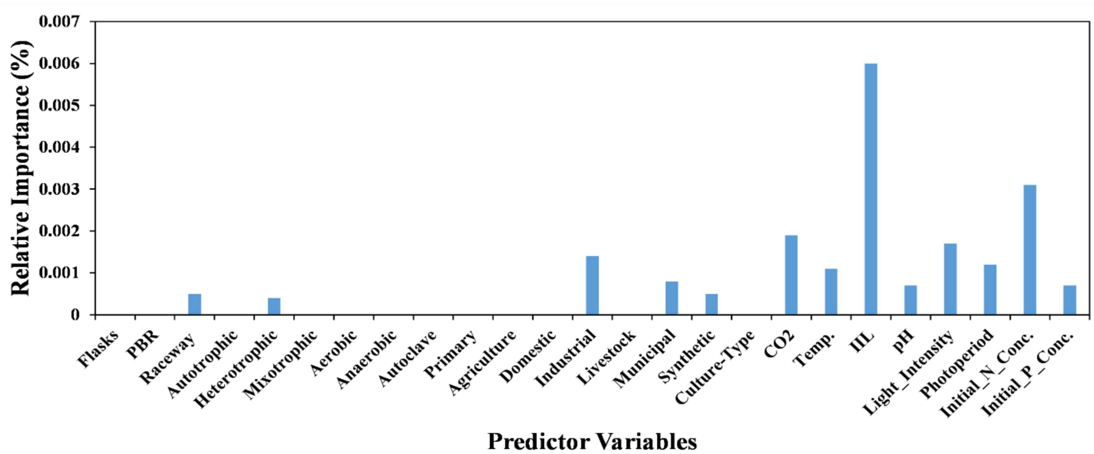

Figure 4.19. Relative importance of the predictor variables on biomass production for class *Chlorophyceae*.

Similar to the case of class *Trebouxiophyceae*, here also, IIL was relatively more vital than other operating variables. Fernando et al. (2021) cultivated *Haematococcus pluvialis* in palm oil effluent at IIL. Likewise, with the rise in IIL from 0.05 mg/L to 0.1 g/L, there was a simultaneous increase in biomass concentration from 0.29 g/L to 0.52 g/L, respectively [488]. In addition to this, initial N. concentration was more significant for both classes than initial P. concentration. N. concentration not only affects the microalgae growth but also regulates the medium's pH, while less than 1% (by weight) of phosphorus is present in microalgae biomass [22]. Ajala et al. (2019) observed that low N. concentration limits biomass production while high N. concentration leads to growth inhibition. Authors cultivated *Scenedesmus obliquus,* and *Oocystis minuta* in secondary treated wastewater at different nitrogen concentrations. In both species, biomass concentration increased as N. concentration was increased up to 377.56 mg/L. Above this value, biomass concentration started declining. A high biomass concentration of 1.3 g/L was observed at 377 mg/L in the case of *Oocystis minuta* [512]. In the present study, decision tree analysis also detected that N. concentrations between 156.66 mg/L and 748.1 mg/L would be ideal for increasing biomass productivity.

4.4. Conclusions

The best combination or paths were determined by accuracy level and no. of data points that support them. The best combination for class *Trebouxiophyceae* was:

- NRE: pretreatment was other than anaerobic, wastewater source is other than domestic, Initial N. concentration < 527 mg/L, Temp. < 24.2 ˚C and photoperiod < 20 h.
- PRE: $CO_2 \geq 0.045\%$ (v/v), pH < 8.35 and initial P. concentration < 12.43 mg/L.
- Biomass: IIL ≥ 0.151 g/L, temp. ≥ 18.25 ˚C, pH < 7.88, and LI < 83.75 $\mu mol\ m^{-2}\ s^{-1}$.

The best combination for class *Chlorophyceae* was:

- NRE: CO_2 content < 0.52% (v/v), photoperiod < 15 h and LI < 112 $\mu mol\ m^{-2}\ s^{-1}$.

- PRE: wastewater source other than industrial, autoclave pretreatment method, pH < 7.75 and IIL ≥ 0.04 g/L.
- Biomass: IIL ≥ 0.104 g/L, initial P. concentration ≥ 7.105 mg/L, LI between 30.64 and 1925 μmol m^{-2} s^{-1}, and initial N. concentration < 748.1 mg/L.

It was also concluded that initial inoculum level and initial nitrogen concentration influenced biomass production for both *Trebouxiophyceae* and *Chlorophyceae* classes.