DEVELOPING PREDICTIVE MODELS FOR BIODIESEL FROM ALGAE USING DATA IN PUBLISHED LITERATURE

by

Ahmet Coşgun

B.S., Chemical Engineering, Boğaziçi University, 2015

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering

Boğaziçi University

2018

DEVELOPING PREDICTIVE MODELS FOR BIODIESEL FROM ALGAE USING DATA IN PUBLISHED LITERATURE

APPROVED BY:

Prof. Ramazan Yıldırı (Thesis Supervisor)	.....
Assoc. Prof. Erdem Gunay (Thesis Co-supervisor)	.....
Assoc. Prof. Burak Alakent	.....
Assoc. Prof. Tuğba Candan-Davran	.....
Assit. Prof. Berat Z. Haznedaroğlu	.....

DATE OF APPROVAL: 03.08.2018

ACKNOWLEDGEMENTS

First and foremost, I would like to express my very great appreciation to my thesis supervisor, Prof. Ramazan Yıldırım, and my co-advisor Assoc. Prof. Erdem Günay for their continuous support of my research, for their patience and motivation.

I am grateful to the jury members Assoc. Prof. Burak Alakent, Assist. Prof. Berat Z. Haznedaroğlu and Assoc. Prof. Tuğba Candan-Davran for their time and consideration.

For a two and a half year, as a member of KB405, I would like to thank my labmates; Elif, Çağla, Barış, Burcu, Beyza, Cansu, for their continuous support. I think of that thesis as a collaborative team work of the entire lab, and would not be that comprehensive if I did not get assistance from them.

The time that I spent in campus for two and a half year as a graduate student was tremendeously high and main reason I enjoyed every minute of it was because of my mates; Ali, Cem, Berkay, Sercan, İlke, Olcay, Ezgi, Uğurcan, Yiğit, Gamze. I want to thank them for being in my life. Time actually moves slow when it comes to having a good time with buddies. I cannot be more grateful to have buddies that I want to spend million hours with; Pelinsu, Gizem Y., Gizem E., Sera, and Sertu. Thank you buddies! As I finally dropping out of school with master's degree, there are still the ones that I want to thank who were there from the start, from first grade; Umut, Sirac, Meriç, Mert, Yaprak, my brothers and sisters.

The last acknowledgement goes to my dear family, my father Sadrettin, my brother Fatih, and my mother Fatmagül, who continuously support me for being a "good" person. Although sometimes the concept of being good was not same for me as for them, they always tried to meet in the middle.

ABSTRACT

DEVELOPING PREDICTIVE MODELS FOR BIODIESEL FROM ALGAE USING DATA IN PUBLISHED LITERATURE

The aim of this thesis was to develop a comprehensive database from published articles about the lipid production from microalgae; then, to use this database for knowledge extraction by employing data mining algorithms to estimate the results of unperformed experiments. A total number of 106 articles were used to construct the database with 5908 instances. Dataset was divided into two groups with respect to reported output variables, which were biomass production (mg/L d), and lipid content (w/w). As the preliminary analysis, the effect of each input variable was investigated by comparing the related articles. Then, for knowledge extraction and prediction-classification purposes, association rule mining, decision tree, and artificial neural network algorithms were applied to both datasets, by using libraries and functions of MATLAB and R. Association rule mining algorithm was implemented to all continuous and categorical variables to examine their effects on output variable, where *Chlorella, Chlorococcum*, and *Nannocholoropsis* species are found to yield high biomass production and high lipid content. Models were compared and evaluated by their accuracy in classification and standard error, root mean square error, and r-squared values in predictive analysis. Parameter tuning was done by randomly dividing the dataset into two sets, as the testing and the training sets, where the training set was used to construct the model, and the testing set was used to calculate the root mean square error and the r-squared values. The optimum models constructed using decision tree algorithm for classification gave 77.8% overall accuracy for biomass production, and 62.2% for lipid content. Artificial neural network algorithm was used for predictive modeling. Absolute error, root mean square error, and r-squared values of the optimum model for biomass production was, 50, 80, and 0.7, and 7, 11, 0.3 for lipid content. Predictive power of the constructed models for lipid content was not as strong as biomass production. The input significance analysis showed that nutritional variables were found to be the most deterministic variables for biomass production, whereas microalgae type was found to be the most deterministic variable for lipid content.

ÖZET

YAYINLANMIŞ MAKALELERDEN ALGLERİN BİODİZEL ÜRETİMİ İLE İLGİLİ ÖNGÖRÜLÜ MODEL GELİŞTİRİLMESİ

Bu tezin amacı mikroalglerden lipit üretimi üzerine yayınlanmış makaleleri inceleyerek kapsamlı bir veri tabanı geliştirmek, bu veri tabanını kullanarak bilgi çıkarımı yapmak ve daha önce yapılmamış deneylerin sonuçlarını tahmin etmek için veri madenciliği algoritmalarını kullanmaktır. Veri tabanı 106 farklı makaleden 5908 veriyle oluşturulmuş olup veritabanı raporlanan sonuç değişkenine göre iki gruba ayrılmıştır. Sonuç değişkenleri biyokütle üretimi (mg / Ld) ve lipit içeriği (w / w) olarak alınmıştır. Giriş değişkenlerinin sonuç değişkenlerine etkisi, aynı giriş değişkeninin etkisiyle ilgilenen makalelerin karşılaştırılması yoluyla ön analiz olarak incelenmiştir. Bilgi çıkarımı ve tahmin-sınıflandırma amaçları için, MATLAB ve R'nin kütüphaneleri ve fonksiyonları kullanılarak her iki veri setine ilişkilendirme kural madenciliği, karar ağacı ve yapay sinir ağı algoritmaları uygulanmıştır. İlişkilendirme kural madenciliği ile *Chlorella, Chlorococcum* ve *Nannochloropsis* mikroalg türlerinin yüksek miktarda biyokütle üretimi ve lipit içeriğine sahip olabileceği bulunmuştur. Sınıflandırma amaçlı modeller doğruluğa, tahmin amaçlı modeller, standart hata, karesel ortalama hata ve determinasyon katsayılarına göre karşılaştırılmış ve değerlendirilmiştir. Veri tabanı rastgele olarak eğitim ve test setine bölünmüş ve eğitim seti model kurmak için kullanılırken test seti karesel ortalama hata ve determinasyon katsayısını bulmak için kullanılmıştır. Sınıflandırma için karar ağacı algoritması kullanılarak oluşturulan optimum modeller, biyokütle üretimi için % 77.8, lipit içeriği için % 62.2 doğruluk ile sonuçlanmıştır. Öngörülü modelleme için yapay sinir ağı algoritması kullanılmıştır. Standart hata, karesel ortalama hata ve determinasyon katsayıları, biyokütle üretimi ve lipit içeriği modelleri için 50, 80 ve 0.7 ve 7, 11, 0.3 şeklinde bulunmuştur. Lipit içeriği için yapılandırılmış modellerin tahmin gücü, biyokütle üretimi kadar güçlü çıkmamıştır. Girdi önem analizi, biyokütle üretimi için besinsel değişkenlerin en belirleyici değişkenler olduğunu, mikroalg tipinin ise lipit içeriği için en belirleyici değişken olduğunu göstermiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

R^2	Coefficient of determination
σ	Mean
μ	Standard deviation

LIST OF ACRONYMS/ABBREVIATIONS

ACS	American Chemical Society
ANN	Artificial Neural Network
ARM	Association Rule Mining
BP	Biomass Production (mg /L d)
CO_2	Carbon Dioxide
DT	Decision Tree
FAMD	Factor Analysis of Mixed Data
LC	Lipid Content (w/w)
Lyo	Lyophilization
MAE	Mean Absolute Error
mg	Milligram
ml	Milliliter
RMSE	Root Mean Square Error
TAG	Triglyceride
vvm	Volume of aeration per unit volume of liquid medium per minute

1. INTRODUCTION

1.1. Microalgae Biodiesel

Among the major energy sources as gaseous fuels and electricity, liquid fuels have certain advantages in storage, transportation and energy density. However, deriving them from fossil resources is becoming controversial with the increased concerns about the environment [1]. A promising alternative for liquid fuels derived from fossil resources is biofuels produced from biomass. A wide range of biofuels are being produced from biomass and can be divided as; solid fuels (biochar), liquid fuels (bioethanol, biodiesel) and gaseous fuels (biogas, biosyngas, biohydrogen) as similar to fossil-based fuels [2]. The biodiesel is produced from food crops, lignocellulosic crops and from microalgae, and named as first, second and third generation biodiesel [2]. First and second generations are open to discussions because of the arable land use for their production and competition with food sources[3]. Production of biodiesel from microalgae stands as a good candidate for a alternative to fossil-based liquid fuels as their growth rate is higher than both the first and second generation resources and the production area does not compete with food production.

Various microalgae strains are studied for their suitability for biodiesel production. The potential of a microalgae strain for biodiesel production is measured by their lipid productivity, which is a combination of biomass productivity and lipid content. Lipid content of different microalgae strains is reported by Rodolfi *et al.*, as being varied from 9.5% to 39.8%. Lipid productivity –as well as biomass production and lipid content- also depends on cultivation conditions and extraction methods aside from microalgae type [4].

Biodiesel is produced from transesterification of lipids in the biomass. However, not all the lipid content in the biomass is convertible to biodiesel. Lipids can be classified as polar and neutral lipids [5]. Polar lipids (phospholipids and glycolipids) are not easy to recover, have stronger bonds, and hard to convert to biodiesel. Neutral lipids can also be divided into two categories as, fatty acid free and fatty acid containing components. The fatty acid free content composes of pigments (carotenes and chlorophylls), hydrocarbons, sterols, wax, sterol esters, and free alcohols, and they are also not suitable for biodiesel

production [6]. Only the fatty acid containing neutral content of lipid, which is mainly triglycerides (TAGs) are the suitable fraction which can be transesterified to form biodiesel [4]. TAGs are compounds that consist of three fatty acid chains, with generally 16 or 18 long carbon chains. TAGs do not have any structural role in cells, instead they are used as storage for carbon and energy in the cytoplasm [7].

Table 1.1. Microalgae Lipid Classes.

	Neutral Lipids	**Polar Lipids**
fatty acid containing components:	Triglycerides (TAGs) Free fatty acids	Phospholipids Glycolipids
fatty acid free components:	Hydrocarbons Sterols Wax Sterol esters Free alcohols Pigments (carotenes, chlorophylls)	

1.2. Microalgae Biodiesel Production Process

The biodiesel production process consists of microalgae cultivation, harvesting (dewatering, thickening), drying, pretreatment for lipid extraction, lipid extraction and transesterification [2]. Transesterification process can be associated with cracking used in petroleum industry. The lipids are transesterified to fatty acid esters to increase the volatility, resulting in production of biodiesel which has similar properties to conventional petroleum diesel [8].

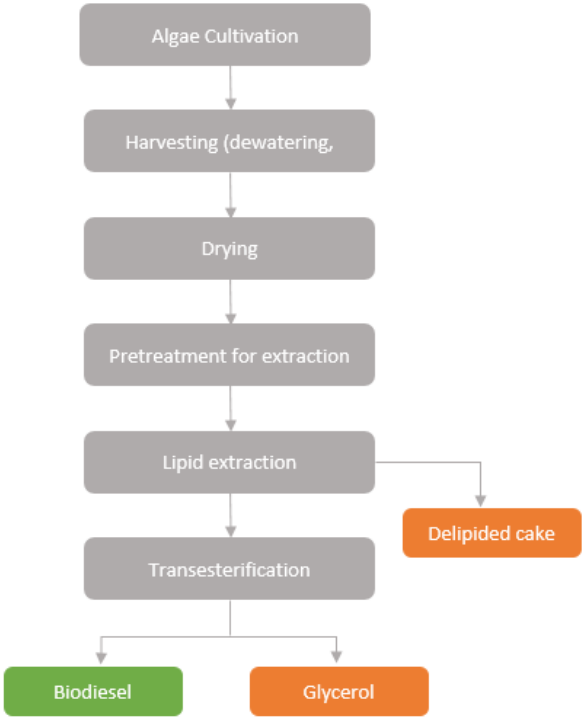

Figure 1.1. Microalgae Biodiesel Production Process.

Microalgae cultivation requires aquatic habitat with appropriate salinity level for selected microalgae. Aquatic habitats can be divided as freshwater, brackish, marine, and hypersaline corresponding to the salinity levels of close to 0%i lower than 3.5%, equal to 3.5%, and higher than 3.5% respectively [9]. Depending on the natural environment of the microalgae strain, the optimum salinity level for microalgal growth and lipid content varies. Freshwater species best grow in salinity level of 0%; whereas, marine species have best biomass productivity at salinity levels around 3.5%. Aside from the salinity level, microalgae requires certain nutrients in the growth medium. Nitrogen and phosphorus are major nutrients for microalgae growth, and other than those, a variety of different compounds are also required in relatively smaller amounts, which are classified as micronutrients and trace elements. Grobbelaar calculated an approximate molecular composition of microalgae as $CO_{0.48}H_{1.83}N_{0.11}P_{0.01}$ [10]. Petkov stated that, although the cost of nutrients required for microalgae cultivation is low compared to other processes, like harvesting and drying, still the cost of nutrients needed to produce 1lt of biodiesel is higher than the price of 1lt petroleum diesel [11]. The nutrients are not the only sources needed for microalgae growth. Also, depending on the cultivation type, a carbon source and an energy source is needed, which in the case of photoautotrophic cultivation is CO_2 as carbon source and light as the

energy source. The optimum amount of salinity, nutrient concentration, light and CO_2 should be satisfied for maximum algal growth. However, in most cases, the conditions for maximum growth do not ensure maximum lipid production Applying stress in cultivation process to increase the lipid content is a common strategy to achieve high lipid productivity. However, in stress conditions although the lipid synthesis is increased in microalgae cell, the growth mechanism slows down. A two-stage process is developed to overcome this problem, in which, the maximum growth is achieved at first, and a stress condition is applied to increase the lipid content of the microalgae before the harvesting [12].

Harvesting and drying of microalgae are processes that are needed to thicken and increase the concentration of microalgae biomass before the lipid extraction. Water content of the microalgae creates a barrier, which makes mass transfer harder during the lipid extraction process and decreases the efficiency of lipid extraction, resulting in low biodiesel yield.

Microalgae has similar density as water and they are usually around 3-30 micrometer in diameter. These two characteristics are the obstacles for both cost effective (20-30% of total biomass producing cost) and efficient recovery for harvesting process [13][14]. Most extraction methods are effective when the water content of dried microalgae is less than 10%. However, harvesting step only reduces this content to about 60%, resulting in a very high energy consumption for drying step (around 89% percent of all the required energy input used in microalgae biodiesel production) [15][16]. Some methods used in harvesting are; screening, flocculation, filtration, gravity sedimentation, flotation and centrifugation [17]. In most of the laboratory scale experiments, centrifugation and microfiltration are the choices of methods. However, they are not cost effective with high energy and capital costs for large-scale processes [3][14]. As the most used method in the literature, centrifugation is tested for its recovery efficiency, and Heasman *et al.* found that the centrifugation efficiency decreases with decreased centrifugation speed [18]. The growth and lipid content do not depend on the harvesting method used.

The cellular water content of microalgae is also needed to be vaporized for effective lipid extraction. Although the wet extraction, which avoids the cost of drying is becoming popular, the research intensity is not dense as drying [19]. Three major methods of drying

are; oven drying, freeze drying and solar drying. Solar drying is slower compared to other methods, but it uses solar energy as the energy source, and stands as an alternative method as the cost effective solutions [3]. Comparison of these methods shows that, the lipid yield of microalgae does not change with the drying method [20][21]. However, lipid quality can be changed with the result of high temperatures caused from oven drying [22].

The cell disruption (pretreatment for lipid extraction) process is mainly used for increasing the efficiency of lipid extraction process. The variety of cell disruption methods can be classified as mechanical and non-mechanical methods [23]. Wet biomass can also be subjected to pretreatment methods; however, some methods are not suitable for wet samples, like grinding or pressing [24].

Lipid extraction process is a mass transfer reaction, where the lipid in microalgae transferred to solvents used to extract. The efficiency of mass transfer primarily depends on the solvent used and the reaction conditions [25]. The methods used for lipid extraction can be categorized as; solvent extraction, accelerated fluid extraction and supercritical fluid extraction. All extraction processes uses polar and/or nonpolar solvents to selectively extract lipids [2]. The primary consideration in extraction is the use of harmful solvents. Chloroform and methanol are the two most common used solvents in extraction processes with harmful effects to environment [26]. Use of organic solvents such as hexane, ethyl acetate, and CO_2 in supercritical fluid extraction is an option to reduce this effect, and studied extensively for their lipid extraction efficiencies in microalgae lipid [27][28][29]. As the lipid content of most microalgae are not too high (15-30%), it must be recovered almost entirely [28]. The methods and solvents used are selected to obtain as much transesterifiable lipid as possible in the microalgae cell. This means the process should not only be lipid specific but also selective to desirable lipid fractions, which is mostly TAGs.[30][31][26].

2. THESIS BACKGROUND AND LITERATURE SURVEY

2.1. Microalgae Species

There are over 100,000 reported microalgae species with different set of characteristics [32]. The ideal strain selection for biodiesel production is a challenging task due to this variety of the characteristics. Biomass productivity and lipid content of the microalgae are two of the most valued ones to compare microalgae for biodiesel production [13]. Unfortunately, high biomass productivity and high lipid content are generally contradictory to each other and if one is found to be high, the other is usually low [33]. Other than those characteristics, high tolerance to environmental stresses, being able to dominate wild strains in open ponds, having limited nutrient requirements, having a high photosynthetic efficiency are other characteristics to choose the ideal strain [14].

The microalgae species that seem to be promising for biodiesel production and being frequently studied are from *Botryococcus, Chlorella, Scenedesmus, Chlamydomonas, Dunaliella,* and *Nannochloropsis* strains because of their suitable characteristics [34].

Lv reported that *Botryococcus braunii* has a high lipid content (50%) and low biomass productivity (28 mg/Ld), and chlorella vulgaris has a low lipid content (20%) and high biomass productivity (doubling time of 19h) [33]. *Dunaliella sp*. are also important because of robustness to high salinity and light, and high growth rate [35]. The composition of the accumulated lipid in the microalgae is another important factor. TAGs are the easily transesterifiable to biodiesel component of the lipid content [36]. The study of Doan reports that the *Nannochloropsis species* have a lipid content of 42-45% where the FAME content is 16-22%[37].

2.2. Cultivation Conditions

High biomass productivity, oil content and eventually high lipid production can be achieved by various techniques including genetic modification of microalgae to use of advanced bioreactors. However, a more direct and effective approach is to optimize the

cultivation conditions. Temperature, irradiance, nutrients, pH, CO_2, salinity, initial inoculum density, reactor type and cultivation type are the variables that affect the growth and lipid content of microalgae, and ultimately affecting the economic feasibility of algal biodiesel [38][39][40]. The conditions for maximum biomass productivity generally yields low lipid content, which may result with low lipid production. To achieve high lipid content, microalgae is generally kept under environmental stress conditions. However, stress conditions lead to low biomass production, which can also cause low lipid production [41]. For most of the cases, increasing the value of an environmental factor increases the biomass production to a certain degree, while a further increase leads to stress condition and results in degradation of growth and even cell death. Maximum tolerances of some species for temperature and CO_2 is given by Ono Cuello [42].

2.2.1. Initial Inoculum Density

In the process of microalgae cultivation, initial inoculum density affects the growth rate of the microalgae. Several articles state that higher initial inoculum density results with high growth rate, increased nutrient removal and high biomass production because of the higher adaptability to the cultivation environment [39][43][44]. However, low inoculum density results in high lipid content, because of the stress condition created for the inoculum (Table 2.1).

2.2.2. Temperature

The optimum temperature for maximum growth is strain dependent, but in most cases 25 ^oC is regarded as the optimum temperature. Most of the microalgae can survive in temperatures, which are 15 ^oC lower than their optimal, a 2-4 ^oC higher temperatures cause a substantial decrease in growth rate or even to the loss of the entire culture [3].

2.2.3. pH

The solubility of CO_2 and other essential nutrients are determined by pH level of the cultivation process [48]. However, because of this dependence with CO_2, the effect of pH levels and CO_2 in growth rate are not fully separated. Some researchers investigated the effect of pH on biomass production by using chemical buffers to uncouple the effects

[49][50][51]. The maximum lipid content is not necessarily achieved at pH level where the maximum biomass production is achieved. Instead, lipid content is maximum in pH values which are slightly lower or higher then the optimum pH value for maximum biomass production. The results obtained from the database are given in Table 2.2

Table 2.1. Effect of Initial Inoculum Density on Biomass Production and Lipid Content.

	Chlorococcum pamirum [45]		Chlorella sp. [39]		Chlamydomonas sp. [46]		Chlorococcum sp. [5]		Scenedesmus abundans [47]	
	BP	LC	BP	LC	BP	LC	BP	LC	BP	LC
	.9	5.5	71.1	18.1	339.9	24.4	58.0	36.3	38.2	15.8
initial biomass conc.	*0.035*	*0.035*	*0.15*	*0.15*	*0.03*	*0.03*				
init. cell. dens. (10^5 cells/mL)							*10*	*10*	*0.02*	*0.02*
	6.3	62.7	75.0	18.7	367.5	30.1	137.3	41.3	38.8	21.6
initial biomass conc.	*0.07*	*0.07*	*0.25*	*0.25*	*0.06*	*0.06*				
init. cell. dens. (10^5 cells/mL)							*100*	*100*	*0.1*	*0.1*
	110.0	56.5	88.3	16.5	384.7	27.2				
initial biomass conc.	*0.3*	*0.3*	*0.35*	*0.35*	*0.09*	*0.09*				
	160.0	46.6			432.9	28.5				
initial biomass conc.	*0.5*	*0.5*			*0.12*	*0.12*				
	217.5	41.4			446.6	28.4				
initial biomass conc.	*0.9*	*0.9*			*0.15*	*0.15*				
	02.5	35.5								
initial biomass conc.	*1.72*	*1.72*								

Table 2.2. Effect of pH on Biomass Production and Lipid Content.

	Scenedesmus sp. [49]	Scenedesmus sp. [50]	Scenedesmus abundans [51]		Scenedesmus sp. [49]	Scenedesmus abundans [51]
pH level	Biomass Production (mg L^{-1} d^{-1})			pH level	Lipid Content (%)	
3	106.7			3	17.2	
4	535.0			4	30.4	
5			70.5	5		17.3
6	558.3		84.0	6	42.7	26.2
7	580.0	79.2	91.5	7	43.6	22.5
8		93.6	124.0	8		21.2
9	565.0	85.2	103.2	9	44.1	21.1
10		75.3		10		
11	555.0			11	41.2	
12	131.7			12	17.7	

2.2.4. Growth Phase

The growth phases of a microalga can be divided as it given in Figure 2.1 [3]:

(i) Lag phase
(ii) Early exponential phase
(iii) Exponential (logarithmic) phase
(iv) Late exponential phase
(v) Linear (early stationary) phase
(vi) Stationary phase
(vii) Decline (late stationary) or death phase

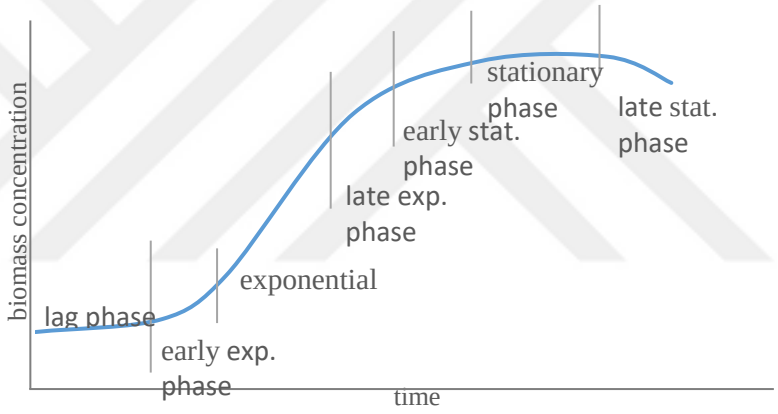

Figure 2.1. Growth Phase of Microalgae.

The length of the lag phase is closely related to adaptation of the microalgae to the cultivation environment, in which a few or no cell division occurs.

Cells in different growth phases have different lipid compositions and lipid contents [52]. Exponential growth phase represents the phase where the maximum growth rate and maximum biomass productivity is achieved [53]. The cell division rates decrease after the exponential phase and approach to zero, which is named as stationary phase [54]. In early stationary phase, microalgae starts to accumulate more lipid, so called the lipid-accumulating phase [55]. Lipid composition generally increases with stationary phase. Polar lipid content is higher in logarithmic phase, but TAG and monounsaturated fatty acid content is higher in stationary phase. Batch cultures are generally harvested at late exponential or stationary phase [54]. Lipid content is increasing towards the late stages, higher than exponential

phases. For production, higher rates are achieved at exponential phases. The effects of growth phase on biomass production and lipid content are presented in Table 2.3 and 2.4 respectively.

Table 2.3. Effect of Growth Phase on Biomass Production.

Cultivation Stage	Isochrysis galbana [56]	Isochrysis sp. [57]	Rhodomonas sp. [57]	Tetraselmis sp. [57]	Scenedesmus abundans [47]	Leptolyngbya sp. [58]	Botryococcus braunii [59]	Botryococcus terribilis [59]	Chlorella vulgaris [59]
early exponential					37.7				
exponential	52.5	49.6	62.3	103.8					
late exponential						129.2	235.0	244.0	310.0
early stationary	81.8	43.3	45.8	65.6					
stationary		17.5	32.8	36.3		77.2			
late stationary					38.2		62.2	60.5	90.2

Table 2.4. Effect of Growth Phase on Lipid Content.

Cultivation Stage	Isochrysis galbana [56]	Neochloris oleoabundans [60]	Isochrysis sp. [57]	Nannochloropsis sp. [57]	Rhodomonas sp. [57]	Tetraselmis sp. [57]	Scenedesmus abundans [47]	Leptolyngbya sp. [58]	Botryococcus braunii [59]	Botryococcus terribilis [59]	Chlorella vulgaris [59]
early exponential							5.8				
exponential	24.4	19.1	23.0	21.6	12.7	11.9					
late exponential								17.9	31.8	30.1	19.5
early stationary	23.2	22.1	29.4	32.1	15.5	14.5					
stationary		35.7	28.0	34.6	13.5	18.2		21.0			
late stationary		37.0					15.8		43.1	41.7	27.0

2.2.5. Cultivation Type

Microalgae can be grown in photoautotrophic, heterotrophic, mixotrophic and photoheterotrophic conditions. The discrimination between cultivation types are done by the energy source and carbon source used as in Table 2.5. [61]:

Table 2.5. Energy and Carbon Source of Different Cultivation Types.

Cultivation Type	**Energy source**	**Carbon source**
Photoautotrophic	Light	Inorganic carbon
Heterotrophic	Organic carbon	Organic carbon
Mixotrophic	Light and organic carbon	Organic and inorganic carbon
Photoheterotrophic	Light	Organic carbon

For optimum growth, sufficient amount of light and inorganic carbon must be available for photoautotrophic condition. Different carbon sources can also affect the growth of microalgae for heterotrophic, mixotrophic and photoheterotrophic conditions making choice of carbon source an important factor. Glucose and acetate are the mainly used carbon sources for heterotrophic and mixotrophic conditions [62]. Heterotrophic condition is reported to be yield higher biomass production compared to photoautotrophic condition. Chlorella vulgaris is reported to grow 4.8 higher [63], whereas Chlorella protothecoides and Chlorella sorokiniana grow 3.4 [64] and 3.3 higher [65] compared to photoautotrophic growth.

Biomass production from mixotrophic growth is not a simple summation of heterotrophic growth and autotrophic growth; instead, it is higher than their combination [62]. Study of Chojnacka and Noworyta showed that mixotrophic condition also shifts the photoinhibitory effect of high light intensities, suggesting better harnessing of high level of light energy. [66].

Although lipid content depends more heavily on other variables, some researchers suggest that autotrophic condition yields higher lipid content. The effects of cultivation type on biomass production and lipid content are presented in Table 2.6 and 2.7 respectively.

Table 2.6. Effect of Cultivation Type on Biomass Production.

Cultivation condition	Chaetoceros sp. [62]	Chlorella sp. [62]	Nannochloropsis sp. [62]	Chlorella vulgaris [67]	Scenedesmus obliquus [68]	Chlorella sorokiniana [69]
autotrophic	21.1	55.9	51.3	256.3		17.3
heterotrophic	29.0	70.3	60.3	338.3	345.3	14.5
mixotrophic	34.6	200.9	160.7	639.3	654.9	
photoheterotrophic						63.3

Table 2.7. Effect of Cultivation Type on Lipid Content.

Cultivation condition	Chaetoceros sp. [62]	Chlorella sp. [62]	Nannochloropsis sp. [62]	Chlorella vulgaris [62]	Scenedesmus obliquus [68]	Chlorella sorokiniana [69]
autotrophic	30.0	30.0	27.9	13.7		26.2
heterotrophic	22.2	21.4	21.6	13.8	14.4	12.9
mixotrophic	25.2	26.4	27.1	8.8	22.8	
photoheterotrophic						20.0

2.3. Light

The properties of light source, as light intensity and wavelength, along with photoperiod are other critical variables that effect microalgae growth and lipid content.

2.3.1. Light Wavelength

Microalgae cells do not uniformly absorb all the wavelengths; instead, certain wavelengths are optimal for the growth of microalgae species. Mostly, blue or blue green (450-475 nm) and red (630-675nm) wavelengths are reported to improve the growth [70][71].

2.3.2. Light Intensity

Light intensity is important for photoautotrophic, photoheterotrophic and mixotrophic conditions. Low light intensity results with insufficient energy input and low biomass production, whereas high light intensity results with photoinhibition/photooxidation [71]. Blanchemain and Grizeau reported that Skeletonema costatum achieves specific growth rate of $0.04h^{-1}$ and lipid content of 44-47% under 50 and 100 micromol photons m^{-2} s^{-1}; however, relatively low (20) and high (400) light intensities lead to specific growth rate of $0.01h^{-1}$ and lipid content of 35-40%[72]. The effect of different light intensities on biomass production and lipid content are presented in Table 2.8 and 2.9 respectively.

Table 2.8. Effect of Light Intensity on Biomass Production.

	Chlorella vulgaris [33]	Dunaliella tertiolecta [73]	Chlorella minutissima [74]	Chlorella sp. [75]	Chlorella sp. [39]	Scenedesmus abundans [51]	Chlorella minutissima [76]	Chlorella sp. [77]	Tetraselmis suecica [78]	Chlorella sorokiniana [69]
BP	50.8	45.0	70.0	165.0	41.7	104.5	90.1	49.7	69.7	5.9
light intensity	*24*	*100*	*100*	*40*	*27*	*40.5*	*60*	*47*	*60*	*15*
BP	188.7	47.5	105.7	239.8*	43.3	109.0	91.8	100.4*	83.7	10.6
light intensity	*60*	*200*	*200*	*200*	*54*	*54*	*80*	*60*	*90*	*35*
BP	190.0*	48.8*	142.9	202.8	83.1	97.5	93.3	97.6	103.1	14.6
light intensity	*120*	*350*	*350*	*400*	*81*	*67.5*	*100*	*80*	*120*	*80*
BP			147.1*		84.2*	119.5*	95.8	96.6	111.4	23.5*
light intensity			*400*		*108*	*81*	*120*	*120*	*150*	*100*
BP					83.4		96.6*		111.7*	20.7
light intensity					*135*		*140*		*180*	*150*
BP							94.2			
light intensity							*160*			

Table 2.9. Effect of Light Intensity on Lipid Content.

	Dunaliella tertiolecta[73]	Chlorella minutissima [74]	Chlorella sp. [75]	Chlorella sp. [39]	Scenedesmus abundans [51]	Chlorella minutissima [76]	Chlorella sp. [77]	Chlorella sorokiniana [69]
LC	21.7	35.7	26.8	26.9	21.2	22.2	41.7*	12.0
light intensity	*100*	*100*	*40*	*27*	*40.5*	*60*	*47*	*15*
LC	22.2	37.1*	33.8	29.8	22.6	23.9	40.5	16.1
light intensity	*200*	*200*	*200*	*54*	*54*	*80*	*60*	*35*
LC	23.4*	32.3	38.3*	31.5	27.1	26.1	26.6	15.7
light intensity	*350*	*350*	*400*	*81*	*67.5*	*100*	*80*	*80*
LC		30.3		33.0	32.8*	28.5	23.6	20.0*
light intensity		*400*		*108*	*81*	*120*	*120*	*100*
LC				33.6*		31.3*		17.0
light intensity				*135*		*140*		*150*
LC						29.3		
light intensity						*160*		

2.3.3. Photoperiod

The light/dark cycle can be manipulated by using artificial light sources to increase microalgae growth. In general, the relationship between photoperiod and biomass productivity is an increasing one. The effect of different light intensities on biomass production and lipid content are presented in Table 2.10. However, light intensity and

photoperiod have a coupled relationship between growth. Higher light intensity together with a longer photoperiod causes a decrease in microalgae growth or even cell death (Table 2.11) [77].

Table 2.10. Effect of Photoperiod on Biomass Production and Lipid Content.

	Chlorella vulgaris [79]	Chlorella protothecoides [80]	Chlorella sp. [77]	Chlorella sorokiniana [69]		Chlorella sp. [77]	Chlorella sorokiniana [69]
BP				5.2	LC		14.0
photoperiod				*8*	*photoperiod*		*8*
BP	15.7	69.5	67.0	15.2	LC	33.4	16.3
photoperiod	*12*	*12*	*12*	*12*	*photoperiod*	*12*	*12*
BP	18.7*				LC		
photoperiod	*14*				*photoperiod*		
BP	16.6	119.5	97.6*	22.1*	LC	30.4	16.0
photoperiod	*16*	*16*	*16*	*16*	*photoperiod*	*16*	*16*
BP	15.1				LC		
photoperiod	*18*				*photoperiod*		
BP	15.5	144.0*	19.4	16.5	LC	39.4*	17.1*
photoperiod	*24*	*24*	*24*	*24*	*photoperiod*	*24*	*24*

Table 2.11. Combined Effect of Light Intensity and Photoperiod on Biomass Production.

light intensity / photoperiod [77]	12	16	24
47	37.3	49.7	112.4
60	48.3	100.4	121.7
80	67.0	97.6	19.4
120	80.3	96.6	

2.4. CO_2

Many microalgae are capable of using inorganic carbon sources from CO_2 and from soluble carbonates [81]. Under natural conditions, the CO_2 content present in the air is between 0.03-0.04% [82]. The concentration of CO_2 used to aerate the microalgae cells has an optimum value for efficient growth while any decrease or increase has a negative effect on growth [43]. It had been reported that increasing the CO_2 content from 0.035% to 0.28% increased the lipid content of Nannochloropsis sp. [83]. Also Hsueh *et al.* reported that increasing CO_2 content from 0.04 to 8% increased both the lipid content and biomass

productivity, however increasing it to 10% resulted the decrease in both [84]. The effect of CO_2 content on biomass production and lipid content are presented in Table 2.12 and 2.13 respectively. Also, the effect of aeration rate on biomass and lipid content are presented in Table 2.14.

Table 2.12. Effect of CO_2 Content on Biomass Production.

	Nannochloropsis oculata [85]	Chlorella vulgaris [33]	Chlorococcum sp. [5]	Chlorella vulgaris [86]	Chlorella vulgaris [59]	Chlorella sp. [87]	Scenedesmus obliquus [88]	Chlorella vulgaris [53]	Chlorella sp. [77]	Chlorella sp. [89]	Tetraselmis suecica [78]
BP										21.2	
CO2 %										*0*	
BP	43.0	192.1*	53.0	34.0	43.9	194.1	94.5	107.4	131.0*	17.7	93.2
CO2 %	*0.03*	*0.5*	*0.04*	*0.03*	*2.5*	*0.03*	*6*	*0.03*	*10*	*0.03*	*0.04*
BP	184.3*	187.0	104.0	212.0	66.7	224.7	144.3	116.0	92.5	27.0*	105.8
CO2 %	*2*	*1*	*1*	*2*	*5*	*1*	*12*	*5*	*20*	*3*	*5*
BP	16.0	143.3	103.0	216.0	71.0*	284.7	205.9*	184.9*	111.6		114.8*
CO2 %	*5*	*6*	*3*	*4*	*10*	*2*	*15*	*15*	*30*		*10*
BP	4.5	126.4	132.0*	247.0*	42.5	328.2	163.3				76.0
CO2 %	*10*	*12*	*6*	*8*	*20*	*3*	*20*				*15*
BP	4.5		23.0	232.0		408.2*					
CO2 %	*15*		*10*	*16*		*5*					
BP						224.7					
CO2 %						*10*					

Table 2.13. Effect of CO_2 Content on Lipid Content.

	Chlorococcum sp. [5]	Chlorella vulgaris [86]	Chlorella vulgaris [59]	Chlorella sp. [87]	Scenedesmus obliquus [88]	Chlorella vulgaris [53]	Chlorella sp. [77]	Chlorella sp. [89]
lipid content (%)								25.1*
CO2 %								*0*
lipid content (%)	10.3	11.1	35.7	19.1	17.1	6.5	36.8*	14.5
CO2 %	*0.04*	*0.03*	*2.5*	*0.03*	*6*	*0.03*	*10*	*0.03*
lipid content (%)	16.2*	11.2	36.4	24.6*	19.0	12.4	27.9	24.9
CO2 %	*1*	*2*	*5*	*1*	*12*	*5*	*20*	*3*
lipid content (%)	14.5	11.5	37.3*	24.2	20.8*	26.0*	35.8	
CO2 %	*3*	*4*	*10*	*2*	*15*	*15*	*30*	
lipid content (%)	14.6	12.0*	34.8	7.2	15.4			
CO2 %	*6*	*8*	*20*	*3*	*20*			
lipid content (%)	6.1	10.7		5.6				
CO2 %	*10*	*16*		*5*				
lipid content (%)				4.8				
CO2 %				*10*				

Table 2.14. Effect of Aeration Rate on Biomass Production and Lipid Content.

	Chlorella sp. [39]	Scenedesmus obliquus [88]	Chlorella sp. [77]		Chlorella sp. [39]	Scenedesmus obliquus [88]	Chlorella sp. [77]
BP	58.5	139.3	100.0	LC	15.1	17.9	33.4*
aeration rate (vvm)	*0.067*	*0.05*	*0.1*	*aeration rate (vvm)*	*0.067*	*0.05*	*0.1*
BP	73.5	220.2	107.1	LC	32.0*	21.4	32.9
aeration rate (vvm)	*0.133*	*0.1*	*0.5*	*aeration rate (vvm)*	*0.133*	*0.1*	*0.5*
BP	95.7*	233.9*	143.0*	LC	28.1	22.8*	32.3
aeration rate (vvm)	*0.2*	*0.15*	*0.9*	*aeration rate (vvm)*	*0.2*	*0.15*	*0.9*
BP	49.1	225.9		LC	24.8	21.4	
aeration rate (vvm)	*0.267*	*0.2*		*aeration rate (vvm)*	*0.267*	*0.2*	
BP	29.7	194.4		LC	24.1	18.7	
aeration rate (vvm)	*0.333*	*0.25*		*aeration rate (vvm)*	*0.333*	*0.25*	

2.5. Nutrients

The required nutrients in the growth medium for microalgae cultivation can be classified as macronutrients, micronutrients, trace elements and vitamins (depending on the auxotrophy for vitamin B) [90].

The key essential macronutrients are nitrogen and phosphorus. The concentration of these elements influences the growth rate and lipid synthesis of the microalgae. Micronutrients (such as sulfur, iron, magnesium, and calcium) and trace elements (such as manganese, zinc, molybdenum, cobalt, copper and boron) are also required for efficient growth of microalgae in relatively small amounts compared to macronutrients. Also, a great amount of microalgae are auxotrophs for vitamin B, and require vitamin B12 (cobalamin), vitamin B1 (thiamin), and vitamin B7 (biotin) [91].

2.5.1. Nitrogen

The nitrogen concentration as well as the source of nitrogen is important for microalgae growth. The nitrogen source is important as it controls the pH of the medium. Nitrogen can be supplied in three forms; as nitrate (NO_3), as ammonia (NH_4), or as urea [92]. The effect of different nitrogen sources on biomass production and lipid content are presented in Table 2.15 and 2.16 respectively.

Table 2.15. Effect of Nitrogen Source on Biomass Production.

Nitrogen source	Neochloris oleoabundans [93]	Scenedesmus sp.[49]	Monoraphidium sp. [94]	Nannochloropsis gaditana [95]	Chlorella sp. [96]
$NaNO_3$	389.5	578.3		89.6	33.8
KNO_3			93.0		
$(NH_2)_2CO$	333.0	496.7		52.1	33.6
NH_4Cl			49.5		12.4
NH_4HCO_3	169.2				
NH_4NO_3		55.0	35.0		

Table 2.16. Effect of Nitrogen Source on Lipid Content.

Nitrogen source	Neochloris oleoabundans [93]	Scenedesmus sp. [49]	Nannochloropsis gaditana [95]	Isochrysis galbana [97]
$NaNO_3$	38.0	43.3	34.7	19.9
KNO_3				18.1
$(NH_2)_2CO$	17.5	7.6	23.0	24.8
NH_4Cl				30.9
NH_4HCO_3	19.0			36.1*
NH_4NO_3		22.0		31.8

The optimum concentration of nitrogen for highest biomass production varies with other parameters, like microalgae type, but the general trend is that in low concentrations the microalgae growth is limited due to nitrogen deprivation, and in high concentrations, the inhibitive effect of nitrogen limits the cell growth. The effect of NO_3 concentration on biomass production and lipid content are presented in Table 2.17 and 2.18 respectively.

2.5.2. Phosphorus

The molar composition of microalgae contains less than 1% P by weight. Although the content of P in the molar composition is low, it is significantly required for the microalgae growth [92]. The normal trend is higher biomass production with the increase in phosphate concentration in the growth medium. However, some studies showed a decrease in biomass production with further increase in phosphate [51][105]. The effect of PO_4 concentration on biomass production and lipid content are presented in Table 2.19 and 2.20 respectively.

Table 2.17. Effect of NO_3 Concentration on Biomass Production.

	Neochloris oleoabundans [93]	Nannochloropsis oculata [98]	Chlorococcum sp. [5]	Scenedesmus obliquus [99]	Chlorococcum pamirum [45]	Scenedesmus sp. [100]	Chlorella sp. [87]	Chlorella vulgaris [101]	Chlorella minutissima [102]	Chlorella vulgaris [41]	Chlamydomonas sp. [46]	Chlorella sp. [103]	Scenedesmus sp. [104]
BP				47.5	28.6	8.3	9.4		20.7			19.0	11.8
NO_3 conc.				*0*	*0*	*0*	*0*		*0*			*0*	*0*
BP	287.8	103.5	26.0	153.0	132.0	21.4	19.1	429.6	80.2	30.2	292.6	26.0	106.3
NO_3 conc.	*0.255*	*0.075*	*0.017*	*0.300*	*0.100*	*0.015*	*0.187*	*0.100*	*0.425*	*0.001*	*0.188*	*0.375*	*0.750*
BP	375.0	104.6	51.0	173.8	242.0	23.2	38.4	571.8	88.3	84.0	364.2	42.0	134.9
NO_3 conc.	*0.425*	*0.150*	*0.100*	*0.600*	*0.250*	*0.031*	*0.280*	*1.500*	*0.850*	*0.021*	*0.375*	*0.750*	*1.500*
BP	602.4	127.0	48.0	172.5	353.0	25.2	83.8	498.0	95.8	130.1	327.0	46.0	131.6
NO_3 conc.	*0.850*	*0.300*	*0.750*	*0.900*	*1.000*	*0.062*	*0.375*	*3.750*	*1.275*	*0.043*	*0.750*	*1.500*	*2.250*
BP	552.4	132.6	63.0	162.5	284.0	26.7	75.7		98.3	131.0	318.8	33.0	128.1
NO_3 conc.	*1.275*	*0.375*	*1.500*	*1.500*	*1.500*	*0.124*	*0.467*		*1.700*	*0.085*	*1.125*	*3.000*	*3.000*
BP	514.0	142.2	47.0			23.8			96.7	121.2	281.6		
NO_3 conc.	*1.700*	*0.750*	*3.000*			*0.247*			*2.125*	*0.165*	*1.500*		
BP		138.3											
NO_3 conc.		*1.500*											

Table 2.18. Effect of NO_3 Concentration on Lipid Content.

	Neochloris oleoabundans [93]	Nannochloropsis oculata [98]	Chlorococcum sp. [5]	Scenedesmus sp. [49]	Chlorella vulgaris [67]	Nannochloropsis gaditana [95]	Scenedesmus sp. [100]	Chlorella minutissima [102]	Chlorella vulgaris [41]	Chlamydomonas sp. [46]	Scenedesmus sp. [104]
LC					23.9		27.9	17.1			33.7
NO_3 conc.					*0*		*0*	*0*			*0*
LC	38.0	15.9	43.0	52.7	19.6	40.2	20.0	33.1		19.5	29.1
NO_3 conc.	*0.255*	*0.075*	*0.017*	*0.200*	*0.150*	*0.075*	*0.015*	*0.425*		*0.188*	*0.750*
LC	33.0	13.0	32.0	48.7	12.1	35.6	20.3	32.6		30.1	25.7
NO_3 conc.	*0.425*	*0.150*	*0.100*	*0.400*	*1.500*	*0.925*	*0.031*	*0.850*		*0.375*	*1.500*
LC	16.0	7.9	18.0	43.3		32.5	19.7	32.0	30.8	28.9	21.8
NO_3 conc.	*0.850*	*0.300*	*0.750*	*0.600*		*1.775*	*0.062*	*1.275*	*0.043*	*0.750*	*2.250*
LC	8.0	15.3	13.0	30.3		30.6	19.8	30.3	31.4	28.1	18.1
NO_3 conc.	*1.275*	*0.375*	*1.500*	*0.800*		*2.625*	*0.124*	*1.700*	*0.085*	*1.125*	*3.000*
LC	8.0	14.4	9.0	13.8			18.9	29.8	15.3	29.2	
NO_3 conc.	*1.700*	*0.750*	*3.000*	*1.000*			*0.247*	*2.125*	*0.165*	*1.500*	
LC		5.9									
NO_3 conc.		*1.500*									

Table 2.19. Effect of PO_4 Concentration on Biomass Production.

	Scenedesmus obliquus [106]	Dunaliella tertiolecta [107]	Chlorococcum pamirum [45]	Chlorella sp. [87]	Scenedesmus abundans [51]	Chlorella minutissima [102]	Chlorella sp. [103]	Chlorella regularis var. [108]
BP	8.5		35.1				10.0	
PO_4 conc.	*0*		*0*				*0*	
BP	44.4	18.7	160.0	8.4	39.5	25.0	24.0	540.9
PO_4 conc.	*3.16×10^{-4}*	*1.92×10^{-5}*	*4.38×10^{-5}*	*3.97×10^{4}*	*1.15×10^{4}*	*3.97×10^{4}*	*8.76×10^{5}*	*1.75×10^{4}*
BP	38.1	28.3	347.0	15.7	73.5	68.1	28.0	574.2
PO_4 conc.	*6.32×10^{-4}*	*9.61×10^{-5}*	*1.75×10^{-4}*	*5.50×10^{4}*	*2.30×10^{4}*	*8.45×10^{4}*	*1.75×10^{4}*	*7.97×10^{4}*
BP	38.1	29.0	433.0	27.5	134.0	93.4	33.0	720.5
PO_4 conc.	*1.05×10^{-3}*	*4.81×10^{-4}*	*3.51×10^{-4}*	*7.08×10^{4}*	*3.44×10^{4}*	*1.30×10^{3}*	*2.63×10^{4}*	*1.43×10^{3}*
BP		26.1	432.0	89.0	110.5	95.9	34.0	
PO_4 conc.		*1.76×10^{-3}*	*8.76×10^{-4}*	*1.02×10^{3}*	*4.59×10^{4}*	*1.75×10^{3}*	*3.51×10^{4}*	
BP		16.0	217.0	69.9		99.9		
PO_4 conc.		*1.20×10^{-2}*	*1.75×10^{-3}*	*1.32×10^{3}*		*2.19×10^{3}*		
BP				60.5		96.7		
PO_4 conc.				*1.63×10^{3}*		*2.65×10^{3}*		
BP						94.9		
PO_4 conc.						*3.10×10^{3}*		

Table 2.20. Effect of PO_4 Concentration on Lipid Content.

	Scenedesmus obliquus [106]	Dunaliella tertiolecta [107]	Scenedesmus abundans [51]	Chlorella minutissima [102]	Chlorella regularis var. [108]
LC	20.9*				
PO_4 conc.	*0*				
LC	6.8	43.5*	24.7*	22.8	24.8
PO_4 conc.	*3.16×10^{-4}*	*1.92×10^{-5}*	*1.15×10^{-4}*	*3.97×10^{-4}*	*1.75×10^{-4}*
LC	13.1	35.0	23.3	32.8*	35.7
PO_4 conc.	*6.32×10^{-4}*	*9.61×10^{-5}*	*2.30×10^{-4}*	*8.45×10^{-4}*	*7.97×10^{-4}*
LC	10.8	31.1	23.0	30.7	42.5*
PO_4 conc.	*1.05×10^{-3}*	*4.81×10^{-4}*	*3.44×10^{-4}*	*1.30×10^{-3}*	*1.43×10^{-3}*
LC		35.5	22.2	29.0	
PO_4 conc.		*1.76×10^{-3}*	*4.59×10^{-4}*	*1.75×10^{-3}*	
LC		32.0		27.6	
PO_4 conc.		*1.20×10^{-2}*		*2.19×10^{-3}*	
LC				26.8	
PO_4 conc.				*2.65×10^{-3}*	
LC				25.5	
PO_4 conc.				*3.10×10^{-3}*	

2.6. Choice of reactor

Microalgae can be cultivated in flasks, photobioreactors, or in open ponds. Controlling the culture conditions in open ponds is hard and not suitable for finding optimal conditions for microalgae growth, but valuable for scale-up experiments because of their low cost [109]. Pulz compared the biomass concentrations during production in open ponds and photobioreactors and stated that biomass production can be as high as 0.1-0.2 g/l and 2-8 gl/L, respectively in those two kind of reactors [110]. Photobioreactors and flasks are commonly used reactors in experiments. The effect of oxygen transfer and uniformity of light are the variables determining the effectiveness of the flask or photobioreactors. Light uniformity can achieved by smaller reactor diameter, as in photobioreactor where the diameter is generally smaller than 0.1m [111]. The suitable agitation is required for flasks whereas in photobioreactors due to the small diameter, agitation is not needed [112].

2.7. Stress Conditions

Some microalgae species significantly increase their lipid content under stress conditions. The total lipid content may vary from 1% to 85% in microalgae, while values higher than 40% is achieved typically in stress conditions [113]. Stress conditions can be classified as; imposed by chemical stimuli (nutrient starvation, salinity, pH), physical stimuli (temperature, light intensity), and growth phase and aging [7]. Stress is imposed mainly to increase the lipid content, however while increasing the lipid content, stress conditions also decrease the growth rate. The optimal stress condition is where the maximum lipid production is achieved. A two-stage cultivation process is also another approach, where the maximum biomass production is achieved at first, and the stress condition is created to favor the lipid synthesis [12].

Limiting major nutrients such as nitrogen and phosphorus is commonly used method to achieve stress conditions, as it is easy to manipulate. The lipid content and also lipid composition in the favor of TAGs are reported to be increased by limiting nutrients [114][52]. The growth rate does not instantly drop down after the nutrient is limited; instead, the intracellular nitrogen reservoir creates a buffer period which postpones the decrease in the cell growth [41].

Microalgae can be divided into three groups with respect to their natural habitat, as freshwater, and marine microalgae, and the halotolerants, which can grow both in the absence and in presence of salts [115]. Stress conditions can be achieved by changing the natural habitat's salt content to enhance the lipid accumulation in the microalgae cells. Salt stress are easy to achieve, as the NaCl is cheap and, the concentration can be easily controlled [116]. Takagi *et al.* achieved 67.8% lipid content in 1M NaCl culture, and 60.6% in 0.5 M [117]. Kaewkannetra et al reported lipid content of 36% in 0.3M NaCl, and 9.5% in absence of NaCl [118]. The effect of salinity concentration on biomass production and lipid content are presented in Table 2.21 and 2.22 respectively.

Table 2.21. Effect of Salinity on Biomass Production.

	marine species						halotolerant				freshwater species							
	mixed sp. [119]	Dunaliella tertiolecta [107]	Nannochloropsis sp. [120]	Amphora subtropica [121]	Dunaliella sp. [121]	Dunaliella salina [122]	Leptolyngbya sp. [58]	Nannochloropsis oceanica [123]	Chlorella vulgaris [111]	Dunaliella tertiolecta [124]	Chlorococcum sp. [5]	mixed sp. [125]	Chlorococcum pamirum [45]	mixed sp. [126]	Scenedesmus sp. [127]	mixed sp. [111]	Scenedesmus sp. [115]	Acutodesmus dimorphus [116]
BP	6.8	29.7	316.0	21.5	24.2	38.4	82.1	714.1	120.0	96.4	60.0	54.9	335.0	45.0	22.8	134.3	235.2	16.5
salinity	*10*	*29*	*14*	*1*	*1*	*29*	*0*	*0*	*0*	*29*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
BP	9.5	34.8	296.0	31.4	40.4	38.4	60.6	654.1	111.0	102.5	65.0	56.7	235.0	45.5	19.7	13.3	206.5	13.7
salinity	*15*	*88*	*27*	*29*	*29*	*58*	*35*	*27*	*19*	*58*	*5*	*1*	*1*	*10*	*3*	*19*	*9*	*1*
BP	9.4	26.9	135.0	44.7	50.0	55.7		531.3	87.0	88.9	64.0	53.2	185.0	41.5	15.3	13.7	188.9	14.5
salinity	*20*	*146*	*54*	*59*	*59*	*88*		*40*	*38*	*88*	*10*	*3*	*10*	*20*	*6*	*38*	*14*	*3*
BP	9.6	17.8	48.5	39.4	75.3	59.0				87.0	14.0	47.7	135.0	35.2	9.1		172.2	12.5
salinity	*25*	*205*	*81*	*117*	*117*	*117*				*117*	*20*	*5*	*20*	*30*	*9*		*23*	*6*
BP	10.6	1.8		13.3	93.3	51.9				75.3		44.1			5.6		159.3	9.7
salinity	*30*	*263*		*176*	*176*	*146*				*146*		*7*			*12*		*29*	*12*
BP	11.3			10.1	73.3					52.3		32.5			3.9			
salinity	*35*			*293*	*293*					*175*		*9*			*23*			
BP										38.2								
salinity										*205*								

Table 2.22. Effect of Salinity on Lipid Content.

	marine species					halotolerant			freshwater species							
	mixed sp. [119]	Dunaliella tertiolecta [107]	Nannochloropsis sp. [120]	mixed sp. [121]	Dunaliella salina [122]	Leptolyngbya sp. [58]	Chlorella vulgaris [111]	Dunaliella tertiolecta [124]	Chlorococcum sp. [5]	mixed sp. [125]	Chlorococcum pamirum [45]	mixed sp. [126]	Scenedesmus sp. [127]	mixed sp. [111]	Scenedesmus sp. [115]	Acutodesmus dimorphus [116]
LC	29.0	30.2	23.3	10.1	6.35	20	15.3	38.9	10.3	11.6	35.3	19.2	19	20.5	22.4	14.4
salinity	*10.0*	*29.2*	*13.5*	*1.5*	*29*	*0*	*0.0*	*29*	*0*	*0.0*	*0*	*0*	*0*	*0*	*0*	*0*
LC	29.2	27.5	22.7	15.3	8.04	16	26.4	34.9	15.4	18.9	39.5	28.1	22	39.7	38.9	17.7
salinity	*15.0*	*87.7*	*27.0*	*29.3*	*58*	*35*	*19.0*	*58*	*5*	*1.0*	*1*	*10*	*3*	*19*	*9*	*1*
LC	28.9	27.1	41.8	22.0	14.7		33.4	34.3	17.7	29.6	50.9	33.6	23.4	37.5	41.1	20.4
salinity	*20.0*	*146.1*	*54.0*	*58.5*	*88*		*38*	*88*	*10*	*3.0*	*10*	*20*	*6*	*38*	*14*	*3*
LC	28.6	30.6	36.3	39.1	20.7			23.1	29.8	38.8	56.2	37.1	24.5		45.7	18.1
salinity	*25.0*	*204.5*	*81*	*117.0*	*117*			*117*	*20*	*5.0*	*20*	*30*	*9*		*23*	*6*
LC	30.5	74.8		19.8	22			16.1		28.6			24.6		52.1	27.8
salinity	*30.0*	*263*		*175.5*	*146*			*146*		*7.0*			*12*		*29*	*12*
LC	30.4			19.4				11.2		21.2			33.1			
salinity	*35*			*293*				*175*		*9*			*23*			
LC								12.2								
salinity								*205*								

2.8. Cell Disruption Methods

Preprocessing microalgae by cell disruption before lipid extraction is reported to increase the lipid content recovered from microalgae. Various articles investigated the effect of different cell disruption methods with comparing the lipid recovered from microalgae with different methods and with non-disruption. Table 2.24 shows that, almost all the cell disruption methods recover more lipid compared to non-disrupted cases. Cell disruption with microwave is reported to be the most efficient cell disruption method in most articles. More advanced methods as liquid nitrogen and steam explosion is also found to be effective.

2.9. Lipid Extraction Solvents

Microalgae lipid is mostly recovered with solvents. The nature of the lipid is important to selectively extract the desired lipid content from microalgae paste. Various polar and non-polar, organic and inorganic solvents are studied in the literature. Although there are differences in the lipid recovered, the lipid composition is also important in biodiesel studies. For this work, solvents that extracts the most lipid are investigated. In most studies, $CHCl_3$ with an alcohol performs highly efficiently (Table 2.23).

Table 2.23. Effect of Different Solvents and Mixtures on Lipid Content.

	Scenedesmus sp. [128]	Nannochloropsis gaditana [129]	Nannochloropsis oculata [130]
$CHCl_3$-CH_3OH	6.0	10.6	8.5
n-Hexane	0.8	0.7	5.7
CH_3OH		33.0	
Ethanol			20.1

2.10. Data Mining Methods

2.10.1. Association Rule Mining

Association rule mining is interested in finding rules where an antecedent results with a consequent. It is usually shown as:

Table 2.24. Effect of Different Cell Disruption Methods on Lipid Content.

Cell Disruption Method	Botryococcus sp. [131]	Chlorella vulgaris [131]	Scenedesmus sp. [131]	Chlorella sp. [132]	Nostoc sp. [132]	Tolypothrix sp. [132]	Chlorella vulgaris [133]	Scenedesmus sp. [20]	Chlorococcum sp. [134]	Chlorella sorokiniana [134]	Nannochloropsis gaditana [135]	Chlorella sorokiniana [135]
Microwave	28.5* (+ lyo)	10.1* (+ lyo)	10.3* (+ lyo)	18.0 (+ lyo)	16.0 (+ lyo)	16.0* (+ lyo)	18.0	28.6*	24.0*	34.3	11.1	14.5 (+ lyo)
Bead milling	28.1 (+ lyo)	7.7 (+ lyo)	8.4 (+ lyo)	16.0 (+ lyo)	12.1 (+ lyo)	13.0 (+ lyo)	10.0					
Osmotic Shock	10.8 (+ lyo)	7.9 (+ lyo)	7.0 (+ lyo)	15.0 (+ lyo)	13.0 (+ lyo)	14.0 (+ lyo)			18.2	36.0*		
Autoclaving	11.9 (+ lyo)	9.7 (+ lyo)	5.5 (+ lyo)	20.5* (+ lyo)	18.0* (+ lyo)	14.0 (+ lyo)			14.3	33.4	10.8	14.4 (+ lyo)
Sonication	8.9 (+ lyo)	6.1 (+ lyo)	7.4 (+ lyo)	11.0 (+ lyo)	10.0 (+ lyo)	8.1 (+ lyo)	15.0	18.8			10.6	14.2 (+ lyo)
none	7.9 (+ lyo)	5.0 (+ lyo)	2.2 (+ lyo)	8.0 (+ lyo)	7.6 (+ lyo)	3.6 (+ lyo)	3.0		12.5	28.8	9.8	11.3 (+ lyo)
Liquid nitrogen							28.4*					
Steam explosion											18.2*	18.4* (+ lyo)

$$X \rightarrow Y \tag{2.1}$$

where, X is called antecedent and Y is called consequent. It can easily be generalized to more than two items with introducing more than one antecedent.

Association rule mining has three common arguments that are widely used to assess the effectiveness of the rule; support, confidence and lift. Calculations of these arguments are done as Equation 2.1-2.3follows:

$$\text{Support}(X \rightarrow Y) = P(X,Y) = \frac{\text{number of instances with both X and Y}}{\text{total number of instances}} \tag{2.2}$$

$$\text{Confidence}(X \rightarrow Y) = P(Y|X) = \frac{\text{number of instances with both X and Y}}{\text{number of instances with X}} \tag{2.3}$$

$$\text{Lift}(X \rightarrow Y) = \frac{\text{confidence(X}\rightarrow\text{Y)}}{\text{proportion of instances with Y to total number of instances}} \tag{2.4}$$

Confidence is the conditional probability of finding an antecedent (X) and a consequent (Y) together. The strength of the rule is quantified by confidence and to treat a rule as strong it should be close to 1, and significantly higher than sole probability of a consequent. The second argument is formulated and named as lift by the proportion of confidence to the probability of a consequent. Lift values higher than 1 suggests that antecedent (X) resulting with consequent (Y) is more likely; whereas, values lower than 1 suggests the opposite. Support of a rule, on the other hand, is the measurement for the statistical significance. Low support shows that there are not enough instances to accept the rule as significant [136].

2.10.2. Decision Tree

Decision tree method is a nonparametric supervised learning method, which can be used both for classification and regression. Decision tree is a hierarchical model, where the output is predicted with a sequence of serial splits. The first node with all the instances is called the root of the decision tree, the final nodes are called terminal leaves and the nodes in between are internal decision nodes. Each split has a criterion defined on a specific input

value. The prediction is done by taking the suitable branch and continued until hitting to a leaf node. Figure 2.2 shows a sample decision tree.

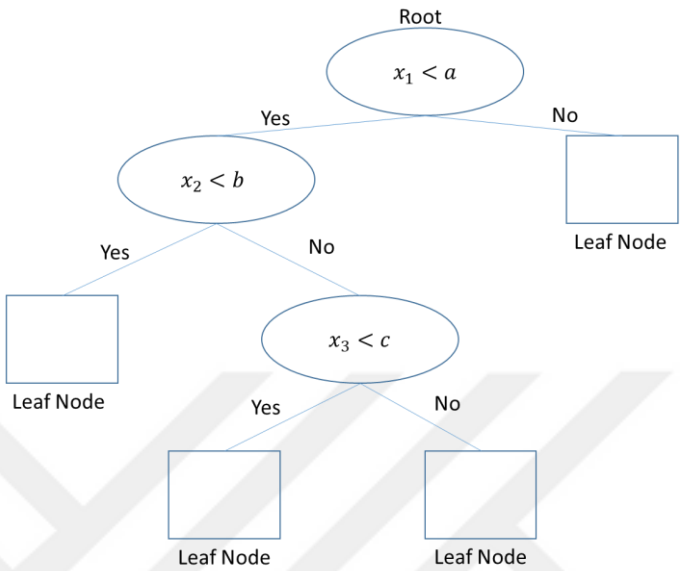

Figure 2.2. A simple Decision Tree Model.

The performance of a decision tree heavily depends on the goodness of the split. The goodness of the split is measured by impurity measures. A split is successful if the nodes are pure after the split. Mainly three functions are used to measure impurity, namely; Entropy, Gini index, and Misclassification error [136].

$$Entropy = \theta(p, 1-p) = -p \log_2 p - (1-p) \log_2 (1-p) \tag{2.5}$$

$$Gini\ Index = \theta(p, 1-p) = 2p(1-p) \tag{2.6}$$

$$Misclassification\ Error = \theta(p, 1-p) = 1 - \max(p, 1-p) \tag{2.7}$$

where, p is the purity in two class-problems.

2.10.3. Artificial Neural Network

The basic processing element of artificial neural network is perceptron, which has inputs, connection weights associated with each input and the output. In the simplest case,

the output is the sum of the inputs multiplied by their weights. A simple perceptron model is given in Figure 2.3, in which:

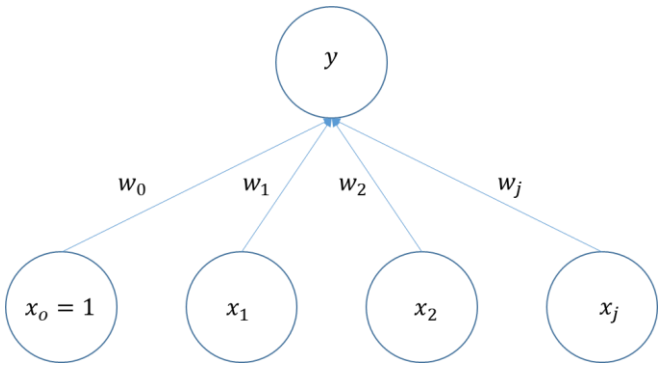

Figure 2.3. A Simple One-Layer Perceptron Model.

x_j's are the inputs, w_j's are connection weights of inputs to the output, and y is the output. w_o and x_0 covers the intercept value, which is added to the model to make it more general. x_o is equal to 1, and w_0 is just the intercept value. Described simplistic case, where the output is the weighted sum of inputs can be written as:

$$y = \sum x_j w_j + w_0 \tag{2.8}$$

Given the input variables, the artificial neural network algorithm needs to learn the weights to predict the output variable. A training set with both input variables and output variable should be given to the model to learn the weights.

Artificial neural network can be used to both classification and prediction problems. For classification problems, the output is divided into several classes. For the cases where the output is more than 1, an equal number of parallel perceptrons are added to the model.

Single layered perceptrons can only be used to predict linear functions. Multilayer perceptrons (MLP), in which inputs are not directly connected to output, instead, one or more additional layers are added to the model, where the output of one layer is used as the input of the following layer, can be used in highly nonlinear models. An example of two-layer perceptron is given in Figure 2.4, where x_j's are inputs, z_h's are the hidden units, w_j's are the weight of the first layer, v_h's are the weights of the second layer, and y is the output.

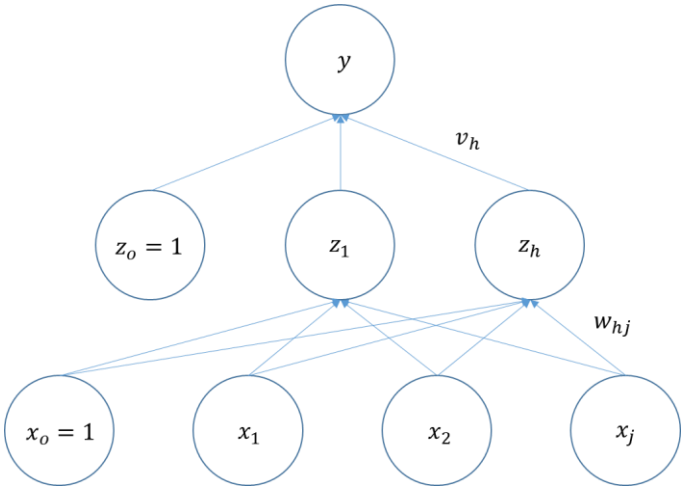

Figure 2.4. A Two-Layered Perceptron Model.

Although there is not a limitation for the number of hidden layers, one hidden layer is generally used in practice. However, in some practices multiple hidden layers can be preferred to have a narrow and long networks instead of short and fat ones [136].

3. COMPUTATIONAL DETAILS

3.1. Experimental Data Collection

3.1.1. Dataset

Dataset was constructed from 113 articles published between 1994 to 2017. However, after preprocessing articles used in database is decreased to 106 articles, with total of 5356 instances. Figure 3.1 shows the increasing interest in the microalgae biodiesel through years. The dataset included 41 input variables (categorical and numeric) and three output variables (biomass production, lipid content and lipid production). Although data for lipid production was also collected, it was not modeled or studied because of the strong correlation with the other two outputs. The effect of input variables in biomass production and lipid content were analyzed separately. However, instance number of these outputs are different because in some research papers, only one of the output is reported. The number of data points for biomass production and lipid content are 4989 and 2572, respectively. Figure 3.2 illustrates two database in terms of number of data points.

The number of articles published about microalgae biodiesel over years is presented in Figure 3.1.

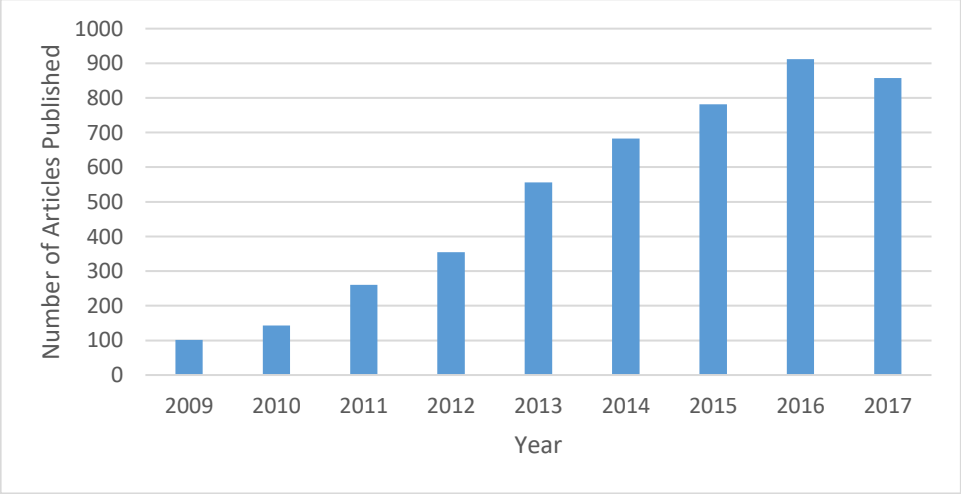

Figure 3.1. Number of Articles Published on Microalgae Biodiesel Through Years.

Lipid content of the microalgae biomass is determined after the lipid extraction step. However, for simplicity, only microalgae cultivation step is considered in analysis and modeling.

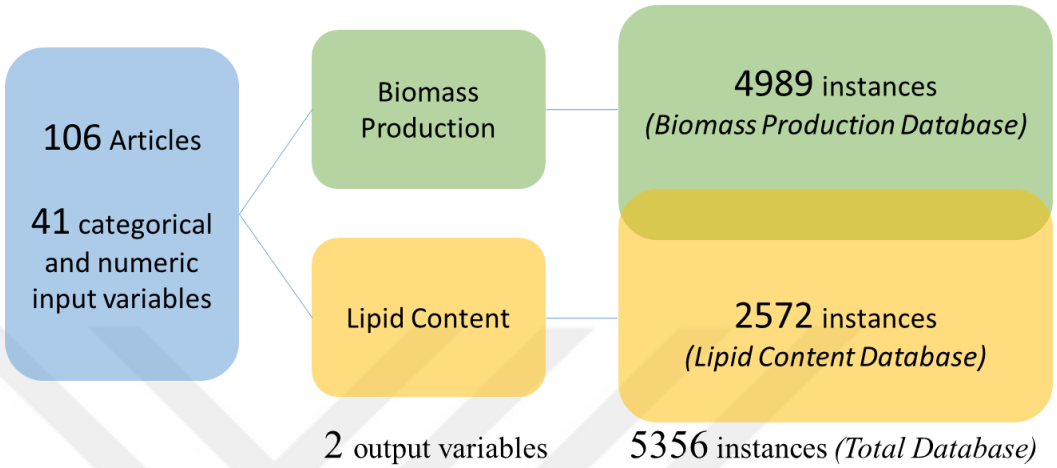

Figure 3.2. Number of Instances in the Database with respect to Output Variable.

3.1.2. Variables

Variables affecting the biomass production and lipid content of microalgae in cultivation stage are composed of the ones that are analyzed previously in the literature, and shown to affect the outputs. Some other variables are also included into database although the affect is not studied in the articles used in the database. The variables can be categorized as microalgae type, water type, reactor type and volume, cultivation type, temperature and pH, growth phase and time, light, nutrients and CO_2.

Different microalgae species are studied throughout these articles; these species are given in Table 3.1, and classification is done starting from the very top of the classification tree:

Figure 3.3 summarizes the overall dataset in terms of microalgae. The band widths are related with the number of data in the overall dataset of that microalgae. Taxonomies of all microalgae are gathered from AlgaeBase website. Domain, phylum, class, family, genus, and species names are noted and number of data of all taxonomical classes is given in following chart. If the number of data of the related branch is lower than 10, it is not plotted

in the graph. 90% of data belong to Eukaryota. 80% of Eukaryota's belongs to Chlorophyta. 57% of Chlorophyta's belongs to Trebouxiophyceae, and so on. Most common microalgae specie, *Chlorella Vulgaris*, has 1007 data points, which account for 21% of all data.

Table 3.1. Number of Different Microalgae Types in the Dataset.

classification	Total	Eukaryota	Prokaryota
Domain	2	1	1
Phylum	10	9	1
Class	23	22	1
Family	61	54	7
Genus	87	78	9
Species	190	181	9

Microalgae can also be divided into two groups as marine and freshwater microalgae. The number of data is also given in Table 3.2.

Table 3.2. Water Types and Their Data Number in the Dataset.

Input Variable	Categories	Number of Data
Water type	freshwater	3778
	marine	1578

The reactor used for cultivation is divided into two categories; flask and photobioreactor. Photobioreactors are optimized reactors, where light and CO_2 are uniformly distributed for optimum growth. Open pond is another type of a reactor used for large scale experiments, however the control level of these reactors are low, and are not reliable in comparative analysis, so they are not included in the dataset. The range of cultivation reactor volume and the number of data points for each reactor type is given in Table 3.3. In Figure 3.4, distribution of reactor volume is shown in more detail.

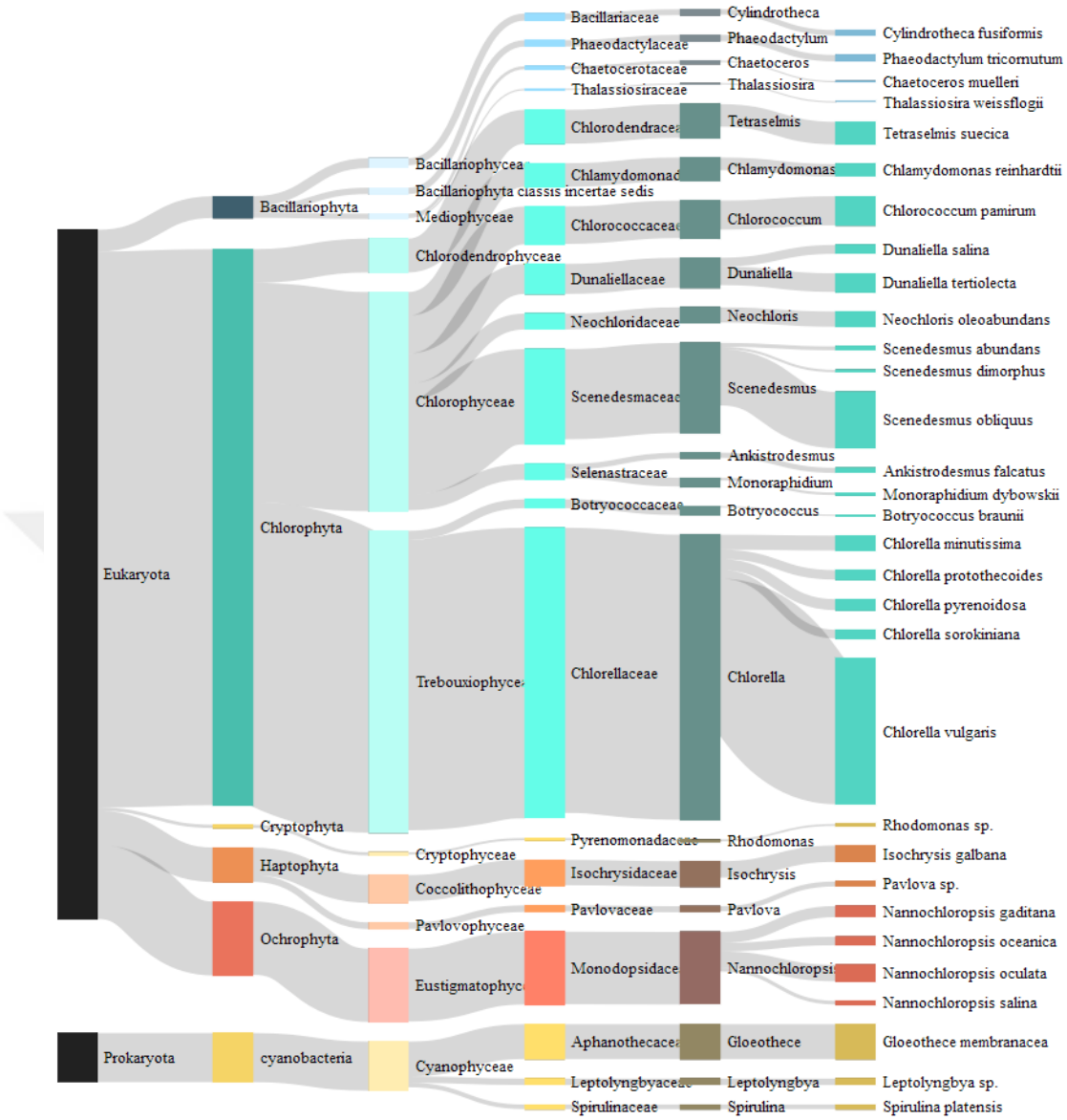

Figure 3.3. Classification of Microalgae in the Dataset represented by Number of Instances.

Table 3.3. Reactor Type and Volume.

Input Variable	Categories	Range	Number of Data
Type of Reactor	flask photobioreactor		2695 2661
Culture medium volume (mL)		10 - 10000	

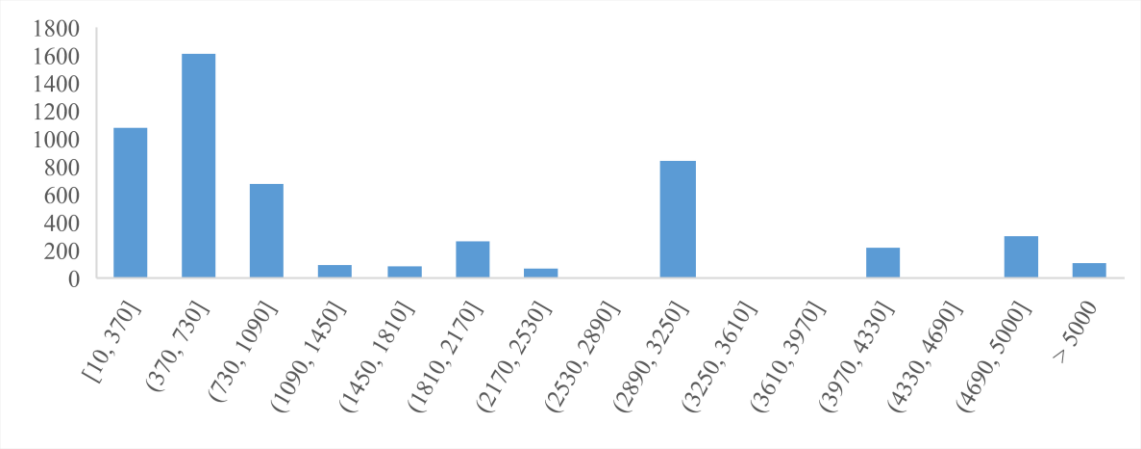

Figure 3.4. Distribution of Culture Volume in Database.

Cultivation type of microalgae is categorized into four categories; autotrophic, heterotrophic, mixotrophic, and photoheterotrophic with respect to energy and carbon source available for microalgae growth. Table 3.4 shows the number of data for each category in the dataset.

Table 3.4. Microalgae Cultivation Types in Database.

Input Variable	Categories	Number of Data
Cultivation type	autotrophic	4632
	heterotrophic	151
	mixotrophic	443
	photoheterotrophic	130

Temperature and pH of the cultivation medium are important variables affecting the biomass production and lipid content of microalgae. Although temperature of the medium can be held constant, constant pH value is achieved with the help of buffer solutions because of the effect of changes in nutrient concentrations and CO_2 level on pH level. In some of the data, buffer solutions are used to achieve constant pH value. Table 3.5 shows the range of temperature and pH values, and number of data of pH control. Following graphs shows the distribution of temperature and pH in more detail.

Table 3.5. Temperature and pH range in Database.

Input Variable	Categories	Range	Number of Data
Temperature (°C)		10 - 50	
pH		3 - 12	
pH control	constant-pH varying pH initial		302 5054

The difference in the stage of microalgae when the cultivation is stopped affects the output variables. Growth stage is categorized as; lag phase, early exponential phase, exponential phase, late exponential phase, early stationary phase, stationary phase, and late stationary phase. Time required to reach the stage where the cultivation ceased is also recorded into dataset. Table 3.6 shows the number of data for each cultivation stages and range of cultivation time.

Table 3.6. Cultivation Stages in Database.

Input Variable	Categories	Range	Number of Data
Cultivation time (days)		0.2 - 68	
Cultivation Stage	early exponential early stationary exponential lag late exponential late stationary stationary		491 558 1198 262 501 424 1922

Nine different light sources used in the articles composes the dataset; white LEDs, blue LEDs, green LEDs, yellow LEDs, orange LEDs, red LEDs, fluorescent, white lamps, and sunlight. The number of data for each light source is given in Table 3.7, with the range of light intensity and photoperiod. In graph and, the distribution of light intensity and photoperiod is shown in more detail.

Table 3.7 Light wavelengths in Database with ranges of light intensity and photoperiod.

Input Variable	Categories	Range	Number of Data
μmol photons m^{-2} s^{-1}		0 - 700	
Light wavelength	380-760nm (white LEDs)		224
	400-800nm (fluorescent)		4454
	460-475nm (blue LEDs)		77
	515-540nm (green LEDs)		77
	587-595nm (yellow LEDs)		77
	590-610nm (orange)		77
	620-645nm (red LEDs)		79
	none		131
	Sunlight		39
	white lamps		121
Photoperiod(h)		0 - 24	

For simplicity and for better comparison, nutrients are converted to molar concentration, if reported otherwise, as individual components instead of molecules. Only nitrogen and carbon sources added, aside from their total concentration, to dataset for analyzing the effect of different nitrogen and carbon sources. Table 3.8 shows the most common molecules used in the cultivation mediums and their corresponding nutritional components.

Table 3.8. Major Nutritional Components of Corresponding Molecules.

Molecule	Components
KH_2PO_4	PO_4 - K
K_2HPO_4	PO_4 - K
Na_2HPO_4	PO_4 - Na
NaH_2PO_4	PO_4 - Na
$(NH_4)_2HPO_4$	PO_4 - N
NH_4Cl	N - Cl
$(NH_2)_2CO$	N
$NH4HCO_3$	N - C

Table 3.9. Major Nutritional Components of Corresponding Molecules. (cont.)

KNO_3	N	-	K		
$Ca(NO_3)_2$	N	-	Ca		
NH_4NO_3	N				
$NaNO_3$	N	-	Na		
$NaHCO_3$	C	-	Na		
Na_2CO_3	C	-	Na		
acetate	C				
acetic acid	C				
Sodium acetate	C	-	Na		
Glucose	C				
Citric acid	C				
Ferric Citrate	Fe	-	C		
Ferric Ammonium Citrate	Fe	-	N	-	C
$MgSO_4$	Mg	-	SO_4		
$MgCl_2$	Mg	-	Cl		
NaCl	Na	-	Cl		
$CaCl_2$	Ca	-	Cl		
$FeCl_3$	Fe	-	Cl		
$FeSO_4$	Fe	-	SO_4		
$Fe(NH_4)_2(SO_4)_2$	Fe	-	N	-	SO_4
Na_2SiO_3	Si	-	Na		
$MnCl_2$	Mn	-	Cl		
$MnSO_4$	Mn	-	SO_4		
$ZnCl_2$	Zn	-	Cl		
$ZnSO_4$	Zn	-	SO_4		
Na_2MoO_4	Mo	-	Na		
$(NH_4)_6Mo_7O_{24}$	Mo	-	N		
MoO_3	Mo				
$CoCl_2$	Co	-	Cl		
$CoSO_4$	Co	-	SO_4		
$Co(NO_3)_2$	Co	-	N		
$CuSO_4$	Cu	-	SO_4		
$CuCl_2$	Cu	-	Cl		
H_3BO_3	B				

The concentration of most important elements required for microalgae growth is given in Table 3.9, with their range. The range of the concentrations are higher in macronutrients, following by micronutrients. Trace elements are required in relatively low concentrations (up to 10^{-4}mol/L) but equally essential for high growth and lipid content.

For autotrophic and mixotrophic modes, microalgae require inorganic carbon as carbon source, mostly carbon dioxide is used by aerating the cultivation medium. Feed gas flow and CO_2 content of the feed gas are important variables affecting the microalgae growth and lipid content. VVM*CO_2 is a combined variable of two variables. The range of these variables are summarized in Table 3.10.

Cell Disruption and extraction solvents are other important input parametrs for lipid content database. Different methods for cell disruption are considered in database (Figure 3.5). Also, lipid extraction solvents are considered as mixtures. Seven different mixtures are considered in database (Table 3.11).

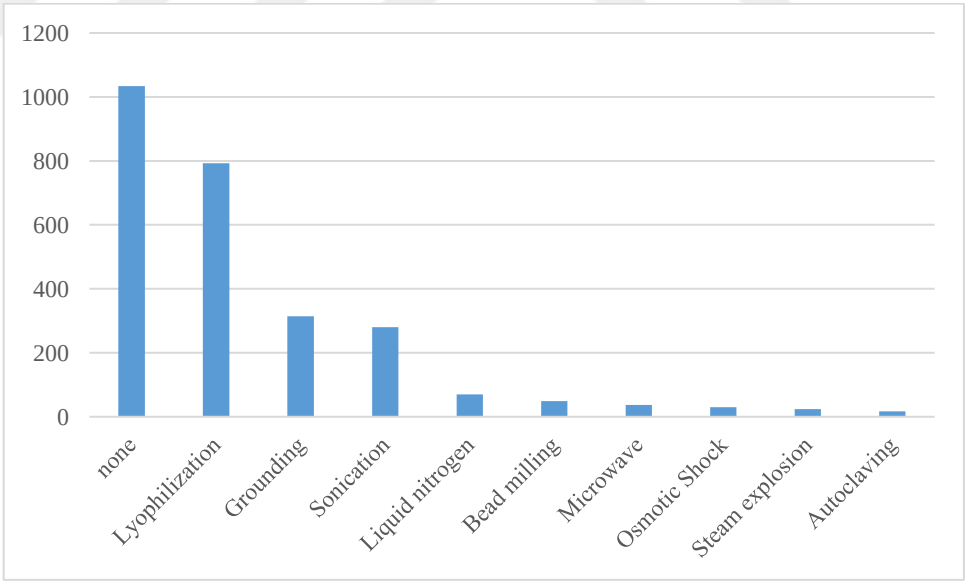

Figure 3.5 Cell Disruption Methods in Database.

The variables investigated in the articles used in the construction of the database are summarized in Figure 3.6 , where 40 different articles out of 106 used in the database are investigated the effect of microalgae species on biomass production and/or lipid content, and

38 different articles are investigated the effect of NO_3 concentration on biomass production and/or lipid content and so on.

Table 3.10. Range of Nutrients used in Database.

	Input Variable	Range
Nutrients (mol/L)	PO_4	0 - 0.012
	N	0 - 0.237
	NH_4	0 - 0.063
	NO_3	0 - 0.237
	C	0 - 3.331
	Cautotrphic	0 - 0.200
	Cheterotrophic	0 - 3.331
	$NAHCO_3$	0 - 0.200
	$NA2CO_3$	0 - 4E-04
	Acetic acid	0 - 0.140
	Sodium acetate	0 - 0.561
	Glucose	0 - 3.330
	Tris base	0 - 0.040
	EDTA	0 - 0.002
	SO_4	0 - 0.045
	Mg	0 - 0.054
	K	0 - 0.054
	Na	0 - 5.007
	Ca	0 - 0.012
	Fe	0 - 2E-04
	Si	0 - 1E-04
	Mn	0 - 4E-04
	Zn	0 - 3E-04
	Mo	0 - 5E-05
	Co	0 - 2E-05
	Cu	0 - 6E-05
	B	0 - 0.002
	Vitamin B12	0 - 3
	Vitamin B1	0 - 3
	Vitamin B7	0 - 0.08

Table 3.11. Range of Feed Gas Flow, CO_2 Content, and Feed CO_2 Flow in Database.

Input Variable	Range
feed gas flow (vvm)	0 - 2.915
CO_2 content (%)	0 - 30
Feed CO_2 flow (vvm)	0 - 0.27

Table 3.12 Lipid Extraction Solvents in Database.

Parameter	Input Variables	Number of Data
Lipid Extraction Solvents	$CHCl_3$-CH_3OH	1726
	n-Hexane	168
	CH_3OH-Dichloromethane	39
	Ethyl Ether	21
	n-Hexane-Isopropanol	16
	CH_3OH	13
	Ethanol	3

Table 3.13. Range of Biomass Production.

Output Variable	Range
Biomass Production (mg/L d)	-14 - 1193

Table 3.14. Range of Lipid Content.

Output Variable	Range
Lipid Content (%)	0 - 74.8

3.1.3. Preprocessing

The articles with unique methods or materials were discarded. Some articles reported the microalgae growth in other forms that cannot be converted into biomass production in terms of $mgL^{-1}d^{-1}$ without any further information for calibration, like cell density, and

optical density measurements. The lipid content measurements in some articles also had the same problem, like fluorescence intensity. Although fluorescence intensity has a meaning within the experiment to compare the lipid contents, it is not reliable to compare the fluorescence intensity of different experiments. The articles with outputs, which were not going to be reliable to compare between articles were discarded.

A clustering analysis was used to detect the outliers in the dataset. Some values were so off that they created a cluster on their own, as shown in Table 3.14. The instances in clusters 2 and 4 were also deleted from the dataset.

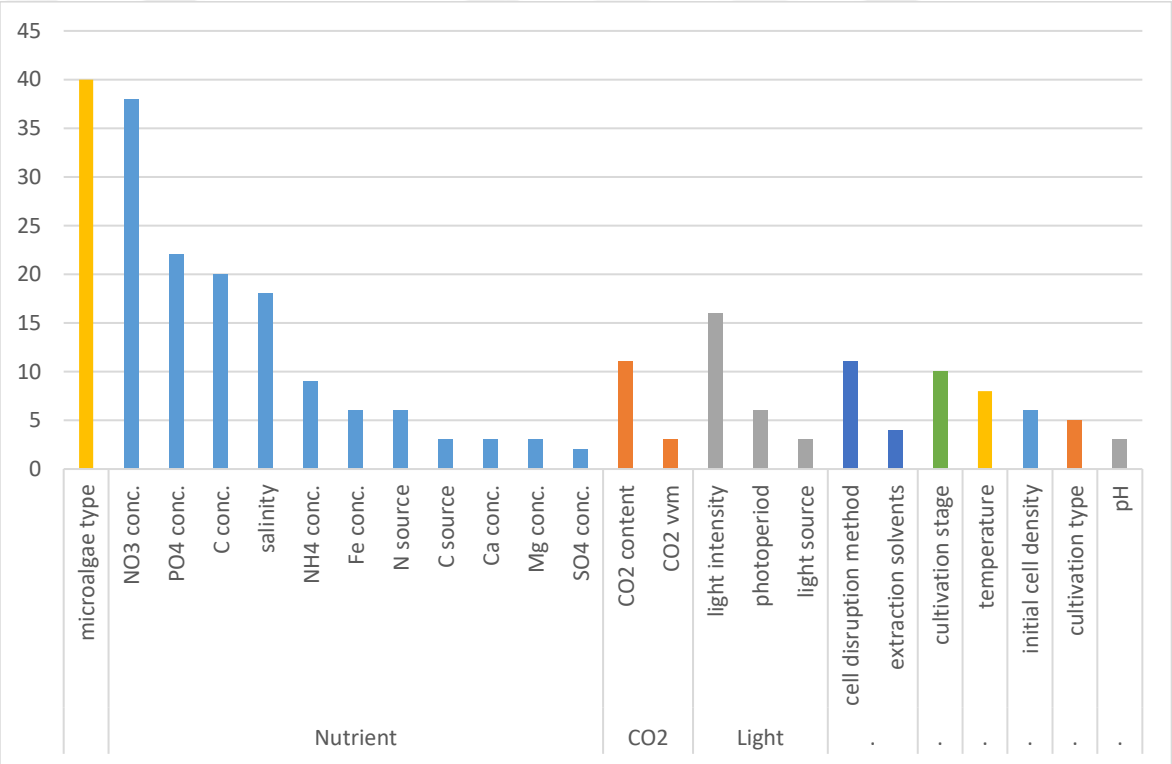

Figure 3.6. Variables Investigated in the Articles used in the Dataset.

Table 3.15. Number of Data in Each Cluster in Cluster Analysis.

Cluster	# of data in clusters
1	1541
2	9
3	4011
4	20

In case of a missing data, the average value of the missing variable was substituted. For some variables, the mean values are calculated from subcategories of other variables that are accepted to be related with the variable of interest. Also for conversion of numeric values into same unit, some conversion assumptions were made, like converting lux to μmol photons m^{-2} s^{-1}. Tables 3.15-3.23 summarize the assumptions made to fill the missing values.

Table 3.16. Assumed Temperature Values of Different Microalgae Types.

variable	subcategory	value
Temperature (^oC)	Chlorococcum sp.	25
	Chlorella sp.	25
	Phaeodactylum tricornutum	24
	Nannochloropsis gaditana	24
	Chaetoceros calcitrans	23
	Scenedesmus sp.	25
	Scenedesmus acutus	25
	Desmodesmus abundans	26
	Desmodesmus sp.	26
	Desmodesmus intermedius	26
	Scenedesmus obtusus	25
	Scenedesmus pectinatus var	25
	Nannochloropsis oculata	23
	Chlorella sorokiniana	26
	Neochloris oleoabundans	27

Table 3.17. Conversion Factors of Some Variables.

variable	conversion	Multiplier
Fe-EDTA solution	liter to gram	multiplied by 5.11
Light intensity	lux to μmol photons m^{-2} s^{-1}	multiplied by 0.0135
Light intensity	W/m^2 To Photons	multiplied by 4.59

Table 3.18. NaCl Concentration of Different Water Types.

variable	subcategory	value
NaCl	Distilled water	0.025
	Filtered natural seawater	27

Table 3.19. Assumed Reactor Sizes of Different Reactor Types.

variable	subcategory	value
culture medium(ml)	photobioreactor	3000
	flask	1000

Table 3.20. Assumed Light Intensity and Photoperiod Values for Autotrophic Growth.

variable	subcategory	value
µmol photons m^{-2} s^{-1}	autotrophic	160
Photoperiod(h)	autotrophic	24

Table 3.21. Assumed Aeration Rates for Different Reactor Types.

variable	subcategory	value
feed gas flow (vvm)	General	1
	flask	0.5
	photobioreactor	1.2

Table 3.22. Assumed Days of Cultivation for Different Growth Phases.

variable	subcategory	value
Cultivation time (days)	early exponential	3
	exponential	5
	late exponential	7.6
	early stationary	8.6
	stationary	12.4
	late stationary	16.5

Table 3.23. Assumed pH Values of Different Microalgae Types.

variable	subcategory	value
pH	GENERAL	7.16
	Isochrysis galbana	8.3
	Bracteacoccus grandis	6.6
	Neochloris oleoabundans	6.8
	Phaeodactylum tricornutum	7.5
	Nannochloropsis sp.	7.3
	Chlorella vulgaris	6.9
	Nannochloropsis oculata	7.6
	Isochrysis sp.	8.3
	Tetraselmis sp.	7.8
	Rhodomonas sp.	8.3
	Chlorococcum sp.	7.4
	Ankistrodesmus falcatus	6.8
	Scenedesmus sp.	7
	Cylindrotheca fusiformis	8
	Pavlova sp.	7.5
	Tetraselmis suecica	8.2
	Scenedesmus abundans	7
	Isochrysis sphacrica	8.3
	Nannochloropsis gaditana	7.1
	Chaetoceros calcitrans	7.5
	Chaetoceros muelleri	7.5
	Dunaliella salina	7.7
	Chlorella sp.	7
	Nannochloropsis oceanica	7.4
	Scenedesmus acutus	7
	Chlorococcum sp.	7.4
	Botryococcus sp.	7
	Chlorella sorokiniana	6.7
	Scenedesmus obliquus	7
	Scenedesmus incrassatulus	7
	Desmodesmus spinosus	7
	Dunaliella tertiolecta	7.7
	Chlorella minutissima	6.8

Table 3.24. Assumed CO2 Content for Photobioreactors.

variable	subcategory	value
CO_2 content (%)	photobioreactor	3

The correlation coefficients between the variables were investigated. Highly correlated variables were eliminated from the dataset. High degree of correlation was assumed to be higher than 0.7. Figure 3.7 shows the correlation between input variables.

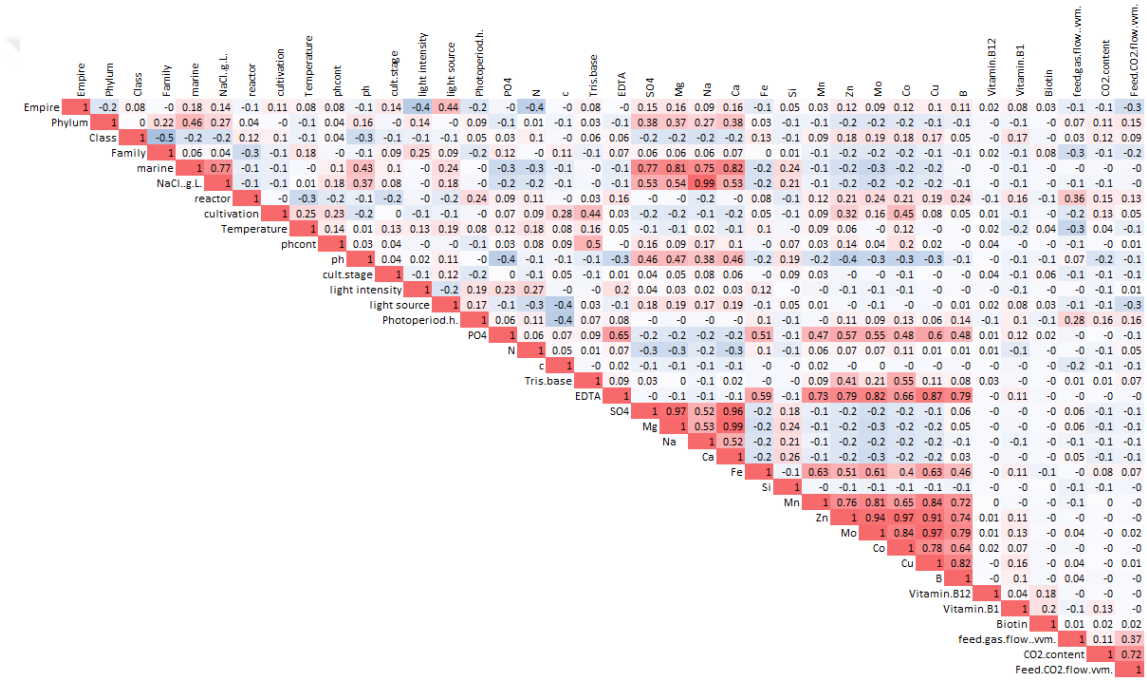

Figure 3.7. Correlation Table of the Input Variables.

Correlation between Mn- Zn- Mo- Co- Cu- B and EDTA was found to be higher than 0.7 in most cases. The Mn data was selected as representative of those seven components and named as “trace elements”.

Na- Ca- SO_4 and Mg had also high degree of correlation. These nutrients are abundantly available in sea salt, creating a correlation with water type used in the culture medium as well. These nutrients were all deleted from dataset, and Mg was kept as the representative for these salts.

The relation between these variables were also tested by Factor analysis of mixed data (FAMD) analysis. The graph shows the contribution of quantitative variables to the first two “new” dimensions created by FAMD algorithm. The direction of the arrows indicate the relation between the variables, where the same direction means correlation. The same variables are observed as related.

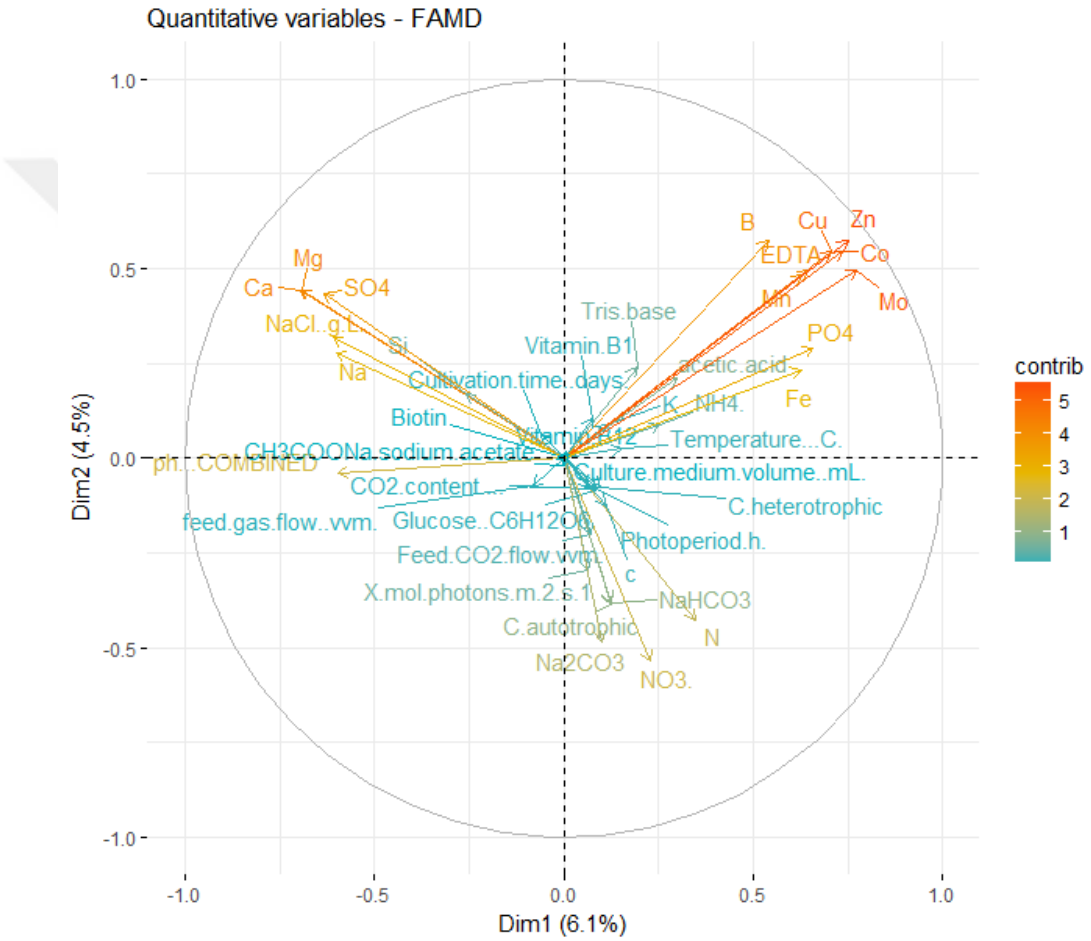

Figure 3.8. FAMD Analysis of the Input Variables.

For decision tree and artificial neural network algorithms, k-fold cross validation is implemented to achieve more stable results. Algorithms like decision tree and neural network are supervised learning algorithms, in which dataset is divided into testing and training sets. The training set is used to build the model and testing set is used to test the performance of the model. However, although the split between train and test instances are

random, there may still class imbalance problem. The instances of one group of a variable may be present significantly higher in either group which lowers the stability of the model. K-fold cross validation overcomes this problem by dividing dataset into k folds, uses k-1 folds for training and one fold for testing, and repeats that process by changing the test set with another fold.

3.1.3. Association Rule Mining

Association rule mining algorithm was employed to dataset with use of arules package of R. The function apriori in the arules package was used to find the rules. The tuned parameters in the algorithms was; minimum number of variables considered to construct the rule (minlen), maximum number of variables considered to construct the rule (maxlen), support value of the rule (supp), and confidence value of the rule (conf). Individual and combined effects of the input variables on the output variable were investigated through tuning minlen and maxlen parameters. Then reliability of a rule was decided by their support and confidence levels. A rule was accepted as reliable if the support value is higher than 0.01 and confidence level is higher than 0.5.

3.1.4. Decision Tree

Decision tree algorithm was constructed by using the fitctree function of MATLAB®. The tuned parameters were maximum number of splits in the tree (MaxNumSplits), and minimum number of data allowed in the leaf nodes (MinLeafSize). Although low MinLeafSize resulted in better models, values lower than 1% of number of data used to construct the model was not favored, because of the possible misleading caused by data points coming from a single article. MaxNumSplits were also limited for having simple, understandable and easily interpretable trees. All possible candidate trees yielding high accuracy were compared, with respect to overall accuracy, accuracy for predicting high, medium, low classes, and error in predicting data points as high which were actually low, and vice versa. For minimizing deviations in the error and accuracy values, the averages of a hundred trees with same parameters were used in the comparative analysis. The errors and accuracies were computed from confusion matrix as follows:

Table 3.25. A Sample Confusion Matrix.

		Predicted		
		Low	Medium	High
Actual	Low	A	B	C
	Medium	D	E	F
	High	G	H	I

$$= \frac{Overall\ accuracy}{\frac{number\ of\ data\ correctly\ predicted\ into\ its\ true\ class}{total\ number\ of\ data}}$$

$$= \frac{A+E+I}{A+B+C+D+E+F+G+H+I} \quad (3.1)$$

$$= \frac{Accuracy\ of\ predicting\ high\ class}{\frac{number\ of\ data\ correctly\ predicted\ into\ high\ class}{actual\ number\ of\ data\ in\ high\ class}} = \frac{I}{G+H+I} \quad (3.2)$$

$$= \frac{Accuracy\ of\ predicting\ medium\ class}{\frac{number\ of\ data\ correctly\ predicted\ into\ medium\ class}{actual\ number\ of\ data\ in\ medium\ class}}$$

$$= \frac{E}{D+E+F} \quad (3.3)$$

$$= \frac{Accuracy\ of\ predicting\ low}{\frac{number\ of\ data\ correctly\ predicted\ into\ low\ class}{actual\ number\ of\ data\ in\ low\ class}}$$

$$= \frac{A}{A+B+C} \quad (3.4)$$

$$Error\ rate = (1 - Overall\ accuracy) * 100 \quad (3.5)$$

5-fold cross validation was applied, in which data was divided into five equal sized subclasses randomly, and five trees constructed with changing the test set in each tree with these subclasses by turns. The error and accuracy values were calculated as being the average of these trees.

In decision tree algorithm, standardization of variables is not a necessary step before modeling, as all variables are evaluated within themselves for finding the optimum split. The differences of means and variances between all variables do not affect the model structure and performance.

The criterion used for splitting was specified as Gini diversity index. The importance of the input variables in the model was determined by predictorImportance function of MATLAB®.

3.1.5. Artificial Neural Network

Before applying artificial neural network algorithm to dataset, the input variables under consideration were standardized to eliminate the bias stemming from the differences in the variances of the input variables. The "normalizeData" function of R was used for this purpose, where the mean and variance of all input variables were changed to 0 and 1, respectively. The standardization was done as follows:

$$x_{new} = \frac{x - \mu}{\sigma} \tag{3.6}$$

k-fold cross validation (CV) method was applied for constructing and training of the neural network model. In CV method, one fold of the k-folds was reserved as the test set, and the remaining k-1 folds were used for training and constructing the model. This procedure was applied k times, at which test set was selected from different fold at each time. k was chosen as 10 in the artificial neural network models in this study.

The "mlp" function of RSNNS package of R was used to built artificial neural network model. Error back propagation is the default function used for training set. Other input arguments for "mlp" function that were tuned for optimum model are; size, maxit, hiddenActfunc, and learnFunc.

Size represents the number of units in the hidden layer(s). Maxit represents the maximum iteration to learn. HiddenActFunc is the activation function of all hidden units. And learnFunct is the learning function to be used.

The choice of the activation function is very crucial for the model efficiency. Each activation function performs differently for different kind of dataset and finding the best one for the dataset is an important step for modeling. For this purpose, 47 activation functions available in the RSNNS package was applied to the models. Together with activation

function, learning function has the same characteristics for model efficiency. Eight learning function available in the RSNNS package was also applied to the models. Table 3.25 summarizes the activation and learning functions tried for finding the optimum models.

Table 3.26. Hidden Activation and Learning Functions Employed in ANN algorithm.

hiddenActFunct's			
Act_Logistic	Act_Elliott	Act_BSB	Act_TanH
Act_TanH_Xdiv2	Act_Perceptron	Act_Signum	Act_Signum0
Act_Softmax	Act_StepFunc	Act_HystStep	Act_BAM
Logistic_notInhibit	Act_MinOutPlusWeight	Act_Identity	Act_IdentityPlusBias
Act_LogisticTbl	Act_RBF_Gaussian	Act_RBF_MultiQuadratic	Act_RBF_ThinPlateSpline
Act_less_than_0	Act_at_most_0	Act_at_least_2	Act_at_least_1
Act_exactly_1	Act_Product	Act_ART1_NC	Act_ART2_Identity
Act_ART2_NormP	Act_ART2_NormV	Act_ART2_NormW	Act_ART2_NormIP
Act_ART2_Rec	Act_ART2_Rst	Act_ARTMAP_NCa	Act_ARTMAP_NCb
Act_ARTMAP_DRho	Act_LogSym	Act_CC_Thresh	Act_Sinus
Act_Exponential	Act_TD_Logistic	Act_TD_Elliott	Act_Euclid
Act_Component	Act_RM	Act_TACOMA	
learnFunct's			
Std_Backpropagation	BackpropBatch	BackpropMomentum	
Quickprop	Rprop	RpropMAP	
BackpropWeightDecay	SCG		

The models with different size, maxit, hiddenActFunc, and learnFunc were compared with each other with respect to mean absolute error (MAE), root mean square error (RMSE), and r-squared (R^2) values. The one with minimum MAE, and RMSE, with maximum R^2 value was chosen to be optimum model, which are calculated as follows:

$$\text{Mean Absolute Error (MAE)} = \frac{1}{n}\sum_{i=1}^{n}|p_i - r_i| \tag{3.7}$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - r_i)^2} \tag{3.8}$$

$$\text{Rsquared } (R^2) = 1 - \frac{\sum_{i=1}^{n}(p_i - r_i)^2}{\sum_{i=1}^{n}(r_i - \bar{r})^2} \tag{3.9}$$

Where p_i and r_i are predicted and real value of the output variable of the i^{th} data point, respectively. n is the number of data, and $\bar{r}$ is the mean vale of real values.

After artificial neural network model was tuned, analysis of relative importance of variables was done by the use of "garson" function of R. Importance is evaluated as either increase in accuracy or decrease in node impurity by the use of the selected variable.

Standardized residual analysis is an important indicator for model adequacy. The relation between error and an input variable should be unrelated. If it is not the case, it means that the selected input variable has an effect in the error, which suggests inadequate modeling. Standardized residuals were calculated difference between real and predicted values divided by standard deviation of real values.

4. RESULTS AND DISCUSSION

The dataset gathered from the published data in the literature was used mainly for three purposes. First, data was used to make simple analysis to find direct relations of input variables to output variables. This was done by using principal component analysis (PCA) and presented in the first section of this part. Secondly, decision tree (DT) algorithm was implemented to dataset for both biomass production and lipid content. The aim was to present heuristics for high output, which is presented in the second section. Finally, artificial neural network (ANN) algorithm is implemented to find relatively significant inputs for the process, and also asses the prediction power of the ANN model, which is presented in the third section.

4.1. Association Rule Mining

Dataset gathered from literature composed of continuous and discrete variables. For association rule mining, continuous variables are also needed to be discretized. Output variables, biomass production and lipid content were also discretized into three classes, each having equal number of instances. The discretization is summarized in Table 4.1 and Table 4.2.

Table 4.1. Discretized Classes of Input Variables.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Temperature	[10,22)	[22,26)	[26,31)	[31,50]			
Light intensity	[0]	[11,42.4)	[42.4,80)	[80,125)	[125,216)	[216,700]	
Photoperiod	[0]	[8,14)	[14,16)	[16,21)	[21,24]		
CO2 content (%)	[0]	[0.03,0.28)	[0.28,1)	[1,3)	[3,8)	[8,20)	[20,30]
PO_4 conc. (mol/L)	[0,8.76×10^{-5})	[8.76×10^{-5} ,0.000344)	[0.000344,0.000845)	[0.000845,0.001615)	[0.001615,0.009185)	[0.009185,0.012018]	
N conc. (mol/L)	[0,0.000442)	[0.000442,0.001471)	[0.001471,0.002929)	[0.002929,0.005602)	[0.005602,0.012186)	[0.012186,0.026477)	[0.026477,0.237204]
Aeration rate (vvm)	[0]	[0.000025,0.15)	[0.15,0.4)	[0.4,0.8333)	[0.8333,1.5)	[1.5,2.915]	
CO2 (aeration rate*CO2 content)	[0]	[7.5×10^{-9} ,0.000002)	[0.000002,0.000025)	[0.00025,0.004)	[0.004,0.025)	[0.025,0.06)	[0.06,0.27]

Table 4.2. Discretized Classes of Output Variables.

	Low	Medium	High
Biomass production (mg/ L day)	[-14,60)	[60,160)	[160,990]
Lipid content (%)	[0,13)	[13,25)	[25,74.8]

The rules gathered from association rule mining algorithm are interpreted with their support, confidence and lift values. To interpret a rule as significant, high confidence and support values, together with lift value higher or lower than 1 is expected.

4.1.1. Association Rule Mining for Biomass Production

Some of the general rules are given in the Table 4.3 and Table 4.4 to reach both high and low biomass production rate.

Table 4.3. Selected ARM Rules Yielding High Biomass Production.

Input Variable *Feature A*	**Biomass productivity (mg/L d)** *Class B*	**support** *Fraction of all data to data that is in class B and has feature A*	**confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**lift** *Fraction of confidence to data that is in class B*	**count** *Number of data that has feature A and in class B*
CO_2%= Class 4	HIGH	0.07	0.78	2.28	331
VVM* CO_2%= Class 4	HIGH	0.08	0.69	2.02	396
VVM=Class 3	HIGH	0.10	0.55	1.63	482
CO_2%=Class 6	HIGH	0.05	0.54	1.59	234
CO_2%=Class 3	HIGH	0.04	0.53	1.54	218
Cultivation type=mixotrophic	HIGH	0.05	0.51	1.50	227
VVM* CO_2%=Class 5	HIGH	0.04	0.50	1.47	223
Cultivation Stage=early exponential	HIGH	0.05	0.49	1.43	233
Type of Reactor=photobioreactor	HIGH	0.25	0.49	1.43	1265
Cultivation Stage=exponential	HIGH	0.10	0.47	1.37	494
Photoperiod =Class 5	HIGH	0.29	0.45	1.32	1428
CO_2%=Class 5	HIGH	0.07	0.44	1.29	335
Type of Reactor=photobioreactor Photoperiod=Class 5	HIGH	0.24	0.58	1.70	1174
Type of Reactor=photobioreactor Photoperiod=Class 5 CO_2%=Class 4	HIGH	0.05	0.98	2.88	259

Table 4.4. Selected ARM Rules Yielding Low Biomass Production.

Input Variable *Feature A*	**Biomass productivity (mg/L d)** *Class B*	**support** *Fraction of all data to data that is in class B and has feature A*	**confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**lift** *Fraction of confidence to data that is in class B*	**count** *Number of data that has feature A and in class B*
Photoperiod = Class 2	LOW	0.11	0.81	2.46	530
N conc.=Class 1	LOW	0.05	0.72	2.18	256
NO_3 conc.=Class 2	LOW	0.08	0.62	1.88	403
N conc.=Class 2	LOW	0.09	0.62	1.86	426
Cultivation Stage=late stationary	LOW	0.05	0.61	1.85	258
PO_4 conc.=Class 1	LOW	0.06	0.53	1.62	275
NO_3 conc.= Class 1	LOW	0.07	0.52	1.57	335
VVM= Class 1	LOW	0.14	0.52	1.57	713
VVM *CO_2% = Class 1	LOW	0.14	0.51	1.54	719
CO_2%= Class 2	LOW	0.26	0.50	1.53	1277
VVM *CO_2%= Class 2	LOW	0.08	0.49	1.50	383
Type of Reactor=flask	LOW	0.23	0.48	1.44	1134
Cultivation Stage=lag	LOW	0.02	0.42	1.28	108
NH_4 conc.= Class 1	LOW	0.20	0.38	1.14	1007
NH_4 conc.= Class 1, CO_2%= Class 2	LOW	0.17	0.70	2.13	841
Type of Reactor=flask, CO_2%= Class 2	LOW	0.19	0.58	1.76	944
NH_4 conc.= Class 1, Glucose conc.=Class 1, CO_2%=Class 2	LOW	0.17	0.71	2.14	827

The support, confidence, and lift values are computed as the following equations:

$$\text{Support}(X \rightarrow Y) = P(X,Y) = \frac{\text{number of instances with both antecedent and consequent}}{\text{total number of instances}}$$

$$\text{Confidence}(X \rightarrow Y) = P(Y|X) = \frac{\text{number of instances with both antecedent and consequent}}{\text{number of instances with antecedent}}$$

$$\text{Lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{proportion of number of instances with consequent to total number of instances}}$$

In the case when reactor type used is photobioreactor; support, confidence and lift values for high biomass production is calculated as follows:

$$\text{Support}(X \rightarrow Y) = P(X,Y) = \frac{1265}{4989} = 0.2536$$

$$\text{Confidence}(X \rightarrow Y) = P(Y|X) = \frac{1265}{2606} = 0.4854$$

$$\text{Lift}(X \rightarrow Y) = \frac{0.4854}{1699/4989} = 1.4254$$

The effect of the variables is analyzed individually with all its classes, and the trends are investigated if there is any in the following sections. Then the most common microalgae strains in the dataset are analyzed separately to compare the variable classes of nitrogen and phosphorus concentration that lead to high or low biomass production.

4.1.1.1. Cultivation Stage. Figure 4.1 shows that highest biomass production rates are achieved at early exponential, exponential, late exponential and early stationary phases. Exponential phase is described as the stage where maximum growth rate is achieved, which is in agreement with the findings. In lag stage almost no cell growth occurs, whereas in late stationary phase even cell death can be observed, and both have high lift values for low biomass production. This result is in consistency with literature, where the biomass production is reported to be in its maximum level in exponential phases (Figure 2.3) [57][58][59].

4.1.1.2. Cultivation Type. In the database, 86% of the experiments were done in autotrophic cultivation conditions. The lift values of autotrophic condition was found to be close to 1 because of this reason. However mixotrophic and heterotrophic cultivation conditions yield high biomass production. Figure 4.2 shows higher lift values for best biomass production rates, and shows lower lift values for worst production rates. Belotti *et al.* also reported biomass production of 338mg L^{-1} d^{-1}, 639 mg L^{-1} d^{-1}, 256 mg L^{-1} d^{-1} for heterotrophic, mixotrophic, and autotrophic cultivation respectively [67], which are in agreement with findings in Figure 4.2 .

4.1.1.3. Light Wavelength. Microalgae absorb light with different efficiencies in different wavelengths. The instances where red and blue light is used as light source, high biomass production values are achieved, with lift values around 2 (Figure 4.3). Wang *et al.* and Teo *et al.* also reported that wavelengths corresponding to blue and red lights results with high biomass production [70] [71].

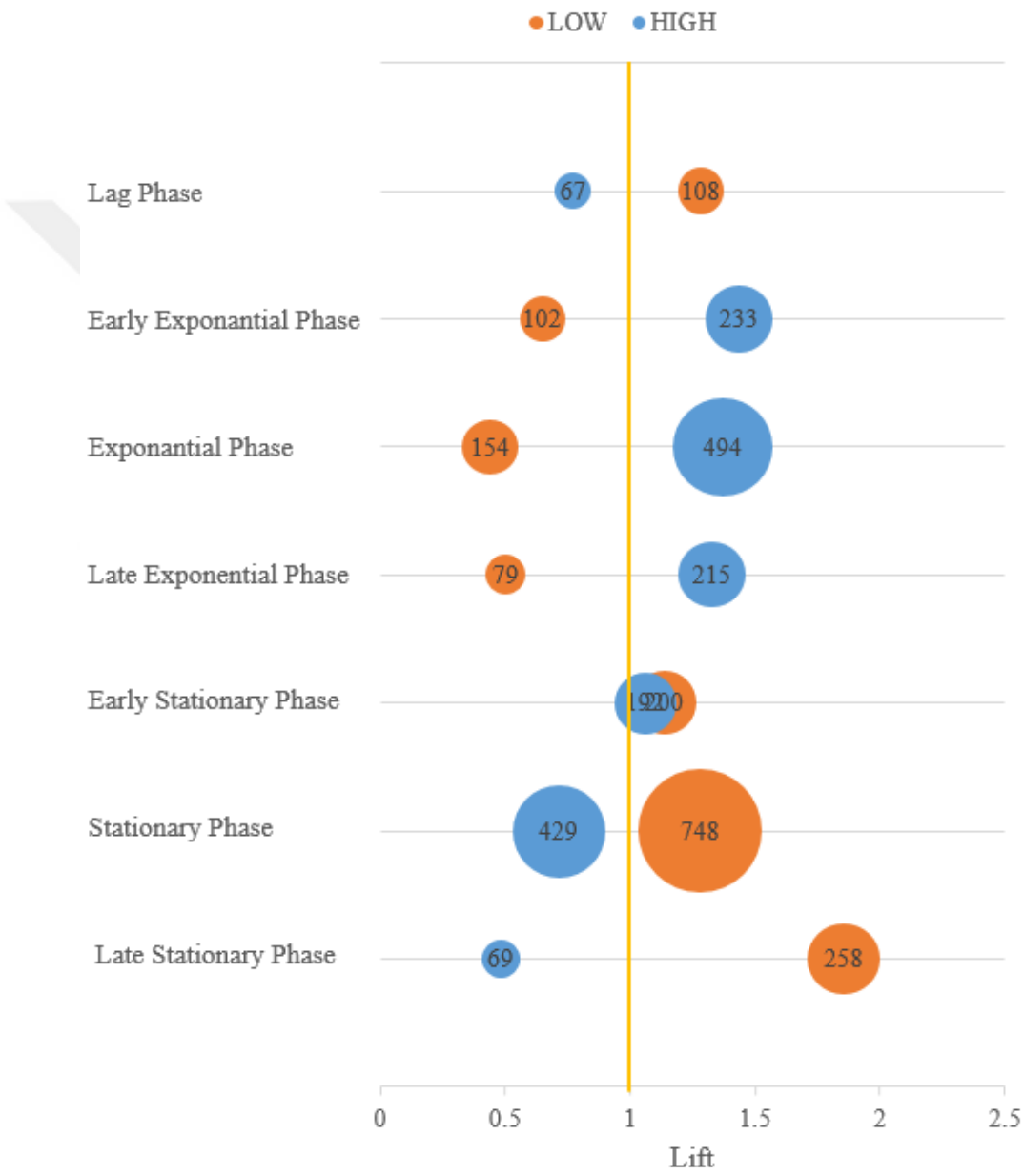

Figure 4.1. Lift and Count Values of Different Stages for Biomass Production.

4.1.1.4. Light Intensity. As seen in Figure 4.4, the lowest level of light intensity is 0, which is heterotrophic cultivation condition. The high biomass production of lowest light intensity

class is related to cultivation condition rather than light intensity effect. In following two classes, the possibility of having low biomass production is favorable compared to having high biomass production. With increased light intensity lift is also increased for high biomass production, and decreased for low biomass production. The articles [33] [39] and [69] report that biomass production was steadily increased with increasing light intensity.

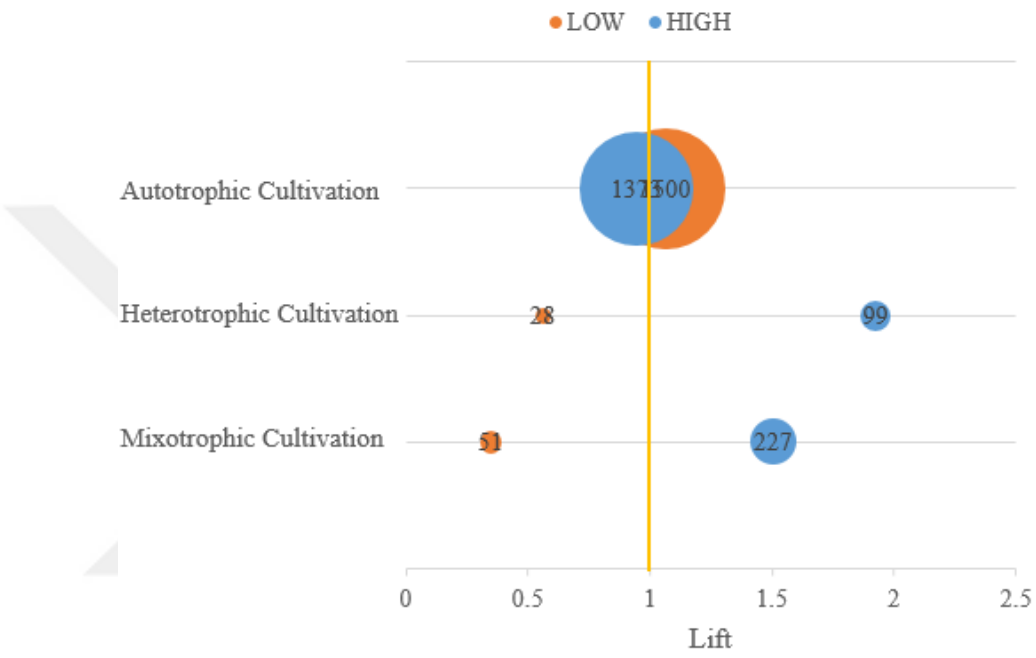

Figure 4.2. Lift and Count Values of Different Cultivation Types for Biomass Production.

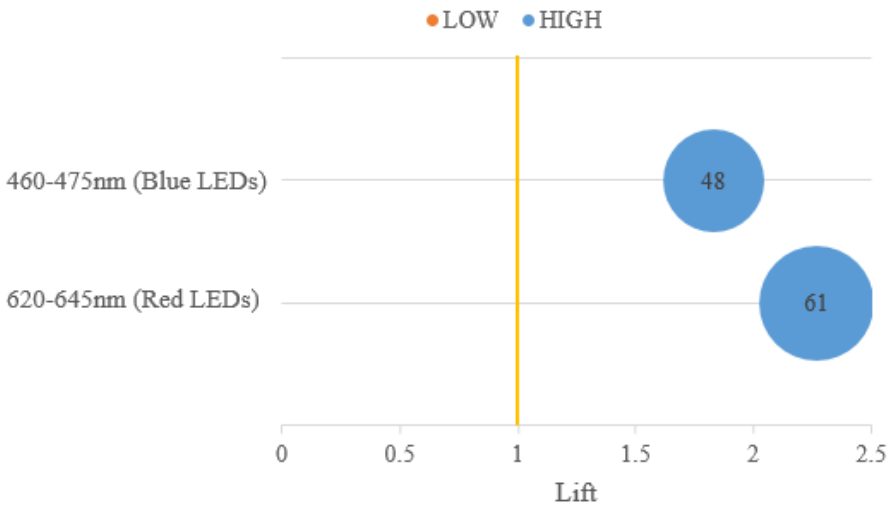

Figure 4.3. Lift and Count Values of Different Light Wavelengths for Biomass Production.

4.1.1.5. CO_2 Content. Figure 4.5 shows that in lowest two concentration levels, biomass production is in the favor of low production rates. When the concentration is increased, the result is in favor of high production rates. The highest lift value is achieved at fourth concentration level, which is in agreement with the trend that biomass production increases with increased CO_2 concentration until a certain value, in which cell growth is inhibited.

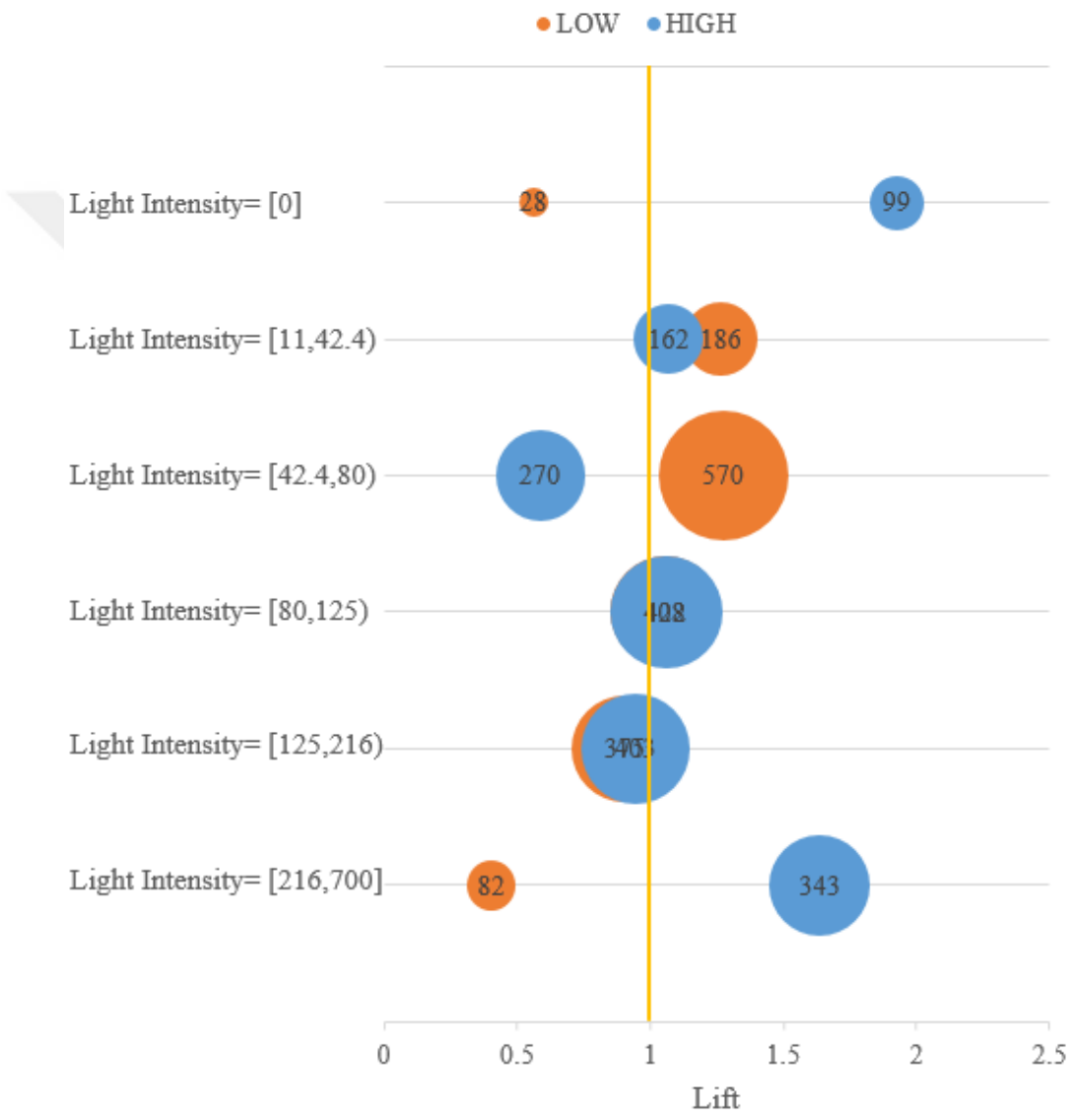

Figure 4.4. Lift and Count Values of Different Light Intensities for Biomass Production.

4.1.1.7. N concentration. Figure 4.6 shows that increasing the concentration increases the lift of high biomass production and decreases the lift of low biomass production. After fourth level, the trend slows down and stabilizes. Qi *et al.*, Converti *et al.*, Pancha *et al.*, and

Chakraborty *et al.* were studied the effect of nitrogen concentration on biomass production and also found that increasing concentration of nitrogen in the culture medium result with increased biomass production rate [41] [98] [100] [102].

4.1.1.8. PO4 concentration. Aside from the third level PO_4 concentration, Figure 4.7 shows that biomass production is increases with increasing PO_4 concentration until a certain level than starts to decrease.

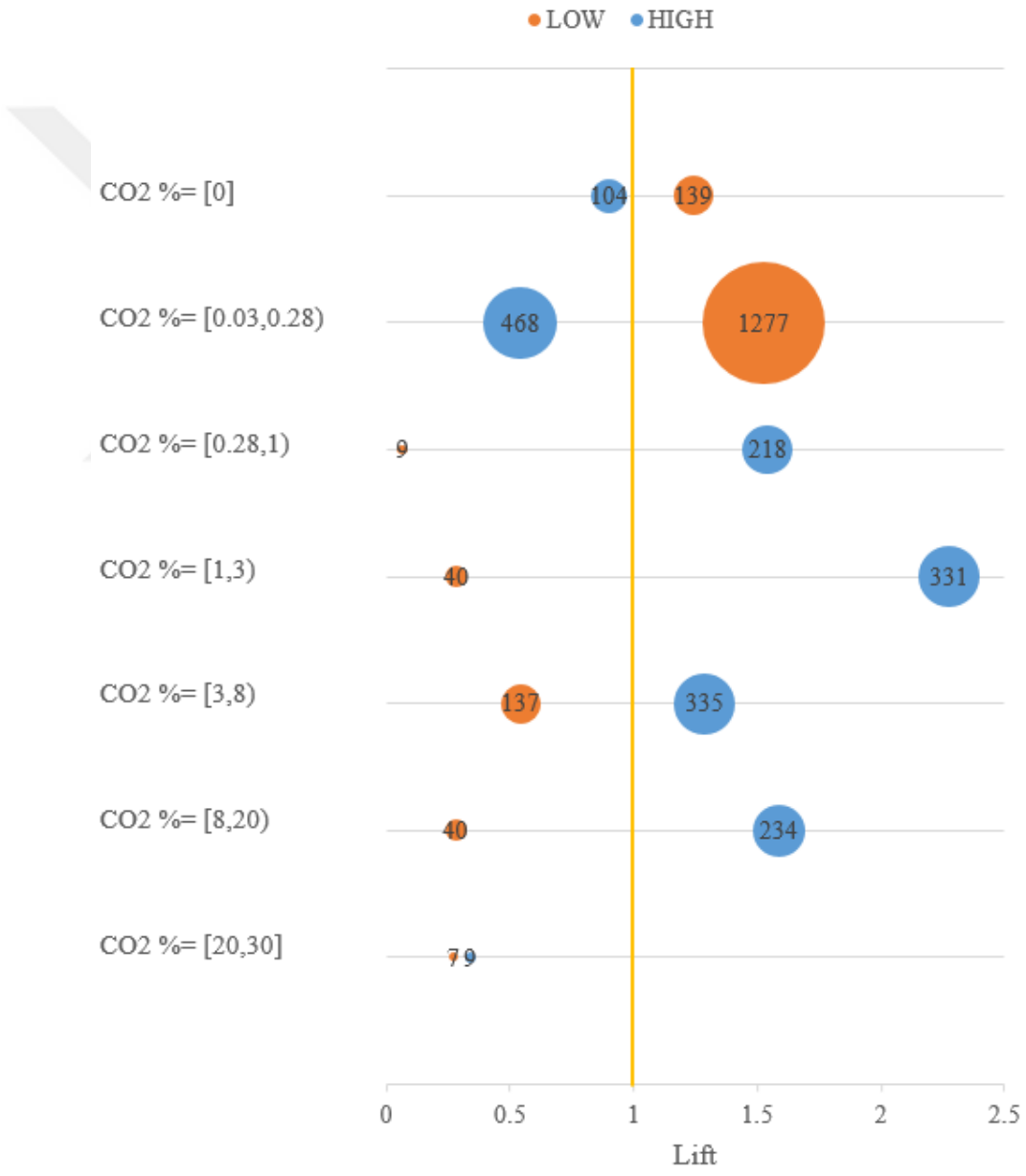

Figure 4.5. Lift and Count Values of Different CO2 Contents for Biomass Production.

Table 4.5. ARM Results of Selected Microalgae Types for Biomass Production.

Microalgae Type *Feature A*	**Biomass productivity (mg L^{-1} day^{-1})** *Class B*	**support** *Fraction of all data to data that is in class B and has feature A*	**confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**lift** *Fraction of confidence to data that is in class B*	**count** *Number of data that has feature A and in class B*
Chlamydomonas	LOW	0.017	0.512	1.548	87
	MEDIUM	0.011	0.318	0.966	54
	HIGH	0.006	0.171	0.501	29
Chlorella	LOW	0.101	0.257	0.778	506
	MEDIUM	0.151	0.382	1.160	751
	HIGH	0.143	0.361	1.061	711
Chlorococcum	LOW	0.012	0.276	0.834	62
	MEDIUM	0.008	0.187	0.568	42
	HIGH	0.024	0.538	1.579	121
Dunaliella	LOW	0.034	0.835	2.526	172
	MEDIUM	0.003	0.063	0.192	13
	HIGH	0.004	0.102	0.299	21
Isochrysis	LOW	0.018	0.650	1.967	91
	MEDIUM	0.002	0.086	0.261	12
	HIGH	0.007	0.264	0.776	37
Nannochloropsis	LOW	0.028	0.305	0.922	142
	MEDIUM	0.023	0.249	0.757	116
	HIGH	0.042	0.446	1.311	208
Scenedesmus	LOW	0.044	0.341	1.032	221
	MEDIUM	0.045	0.346	1.051	224
	HIGH	0.041	0.313	0.920	203
Tetraselmis	LOW	0.023	0.487	1.474	115
	MEDIUM	0.018	0.377	1.147	89
	HIGH	0.006	0.136	0.398	32

4.1.1.9. Temperature. Figure 4.8 shows that the microalgae biomass production is optimum at second and third temperature levels. In first and fourth levels, extremes, the lift of low production is around 1.5. All of the instances tested with highest temperature class resulted with low biomass production, whereas there are still high class instances available in the lowest temperature class. This observation is also reported by Mata *et al.*, which states that microalgae can tolerate temperatures lower than its optimum values, however even a slight increase in temperature can result in cell death [3].

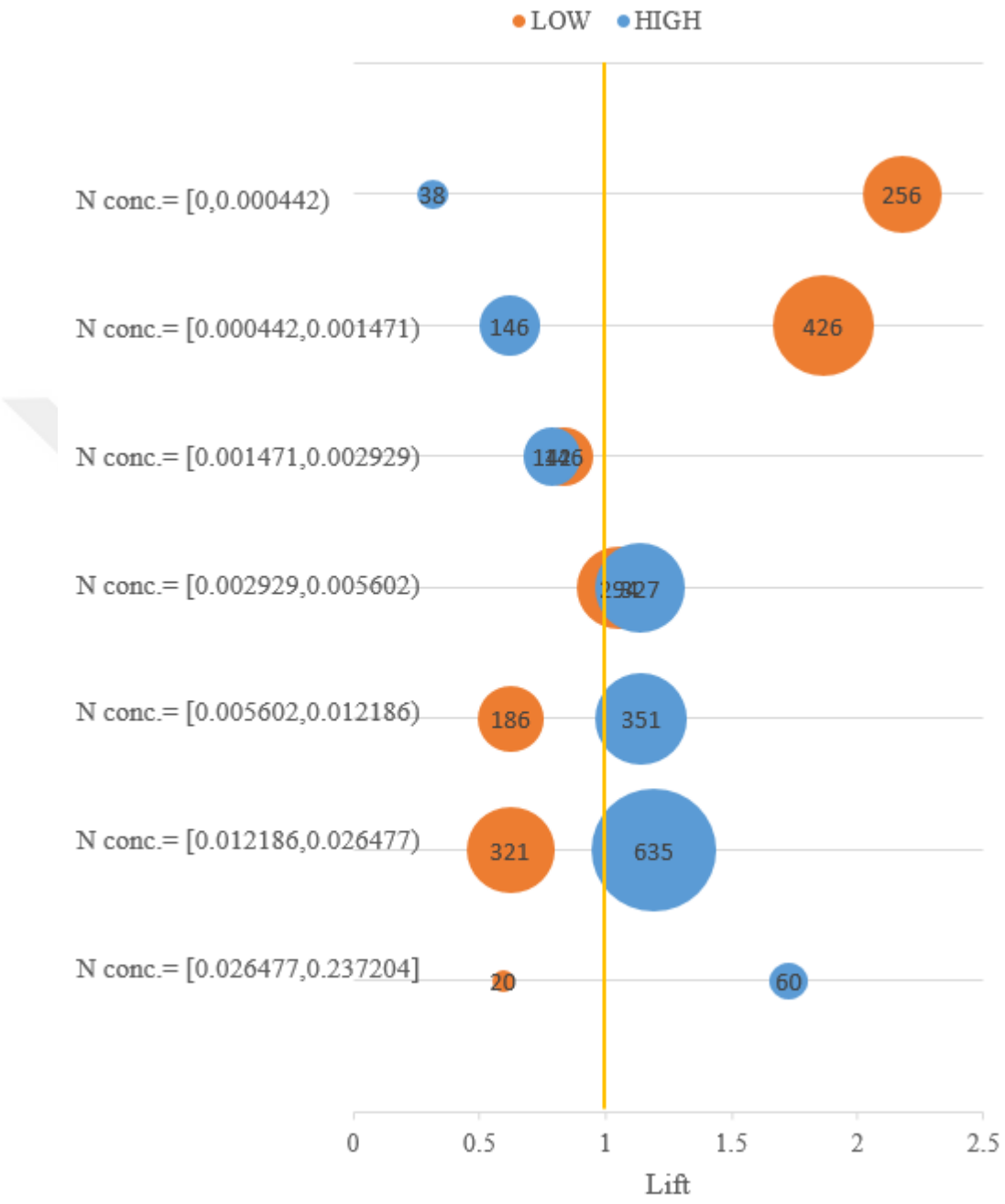

Figure 4.6. Lift and Count Values of Different Nitrogen Concentrations for Biomass Production.

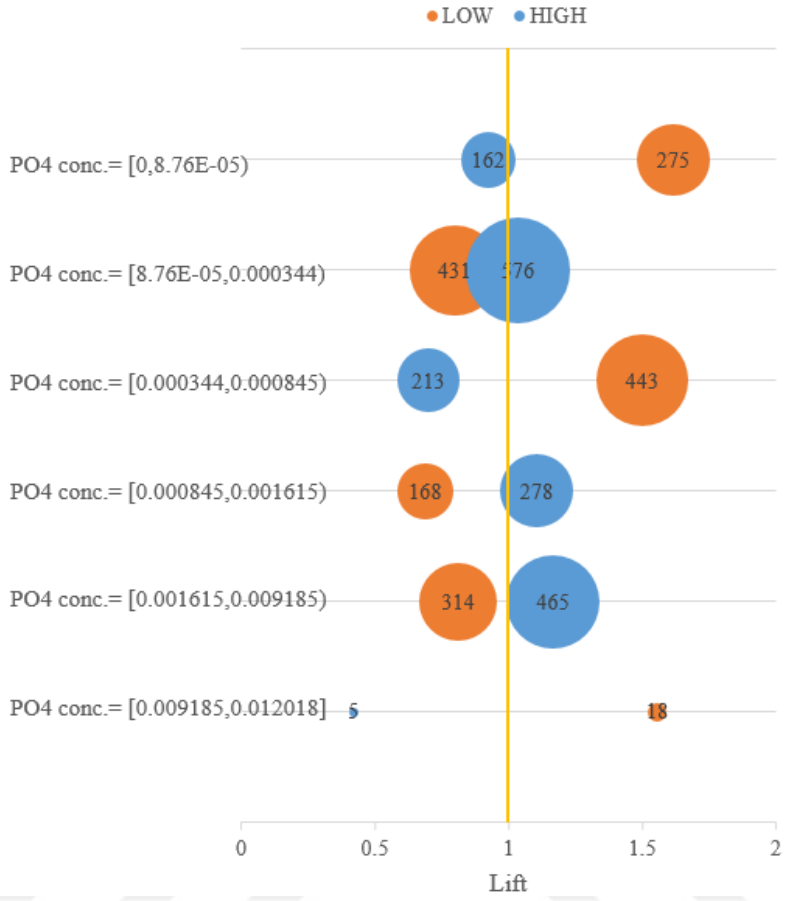

Figure 4.7. Lift and Count Values of Different Phosphorus Concentrations for Biomass Production.

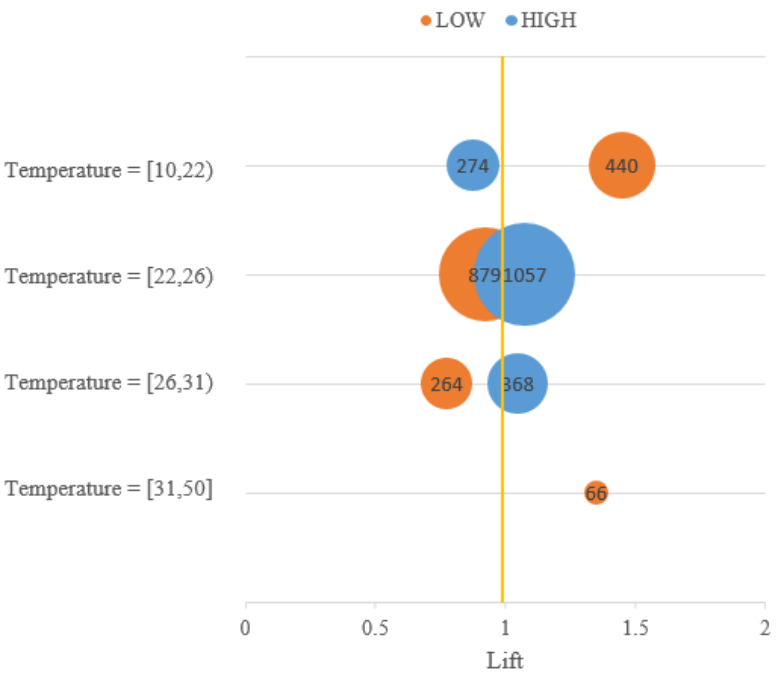

Figure 4.8. Lift and Count Values of Different Temperatures for Biomass Production.

4.1.2. Association Rule Mining for Lipid Content

Some of the general rules are given in Table 4.6 and Table 4.7 to reach both high and low lipid content.

Table 4.6. Selected ARM Rules Yielding High Lipid Content.

Input Variable *Feature A*	**Lipid content (%)** *Class B*	**support** *Fraction of all data to data that is in class B and has feature A*	**confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**lift** *Fraction of confidence to data that is in class B*	**count** *Number of data that has feature A and in class B*
N conc.=Class 1	HIGH	0.053	0.757	2.399	140
CO_2%=Class 7	HIGH	0.020	0.720	2.282	54
EDTA conc.=Class 1	HIGH	0.044	0.712	2.256	116
Salinity level =freshwater high	HIGH	0.011	0.612	1.941	30
VVM =Class 6	HIGH	0.032	0.585	1.855	86
Fe conc.=Class 5	HIGH	0.048	0.574	1.820	128
VVM*CO_2 =Class 6	HIGH	0.028	0.544	1.725	74
C conc.=Class 3	HIGH	0.087	0.481	1.525	230
LightIntensity =Class 4	HIGH	0.061	0.478	1.515	162
Salinity level=marine high	HIGH	0.017	0.473	1.500	44
PO_4 conc.=Class 2	HIGH	0.114	0.434	1.375	301

Table 4.7. Selected ARM Rules Yielding Low Lipid Content.

Input Variable *Feature A*	**Lipid content (%)** *Class B*	**support** *Fraction of all data to data that is in class B and has feature A*	**confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**lift** *Fraction of confidence to data that is in class B*	**count** *Number of data that has feature A and in class B*
CO_2%=Class 3	LOW	0.065	0.637	1.847	172
LightIntensity =Class 6	LOW	0.072	0.525	1.522	190
Cultivation Stage=exponantial	LOW	0.078	0.504	1.460	207
VVM =Class 3	LOW	0.076	0.476	1.381	202

The effect of the variables is analyzed individually with all its classes, and the trends are investigated if there is any in the following sections.

4.1.2.1. Cultivation Stage. From literature, the lipid content is reported to be increased with culture age. From association rule mining of the lipid content data, lift values shows that lipid content is lower in lag, and exponential phases, and higher in stationary phases (Figure 4.9). In Table 2.4 the same trend can be seen in results gathered from [47] [57]-[60].

Figure 4.9. Lift and Count Values of Different Cultivation Stages for Lipid Content.

4.1.2.2. Fe concentration. Figure 4.10 shows a correlation with increasing Fe concentration in the culture medium and lipid content.

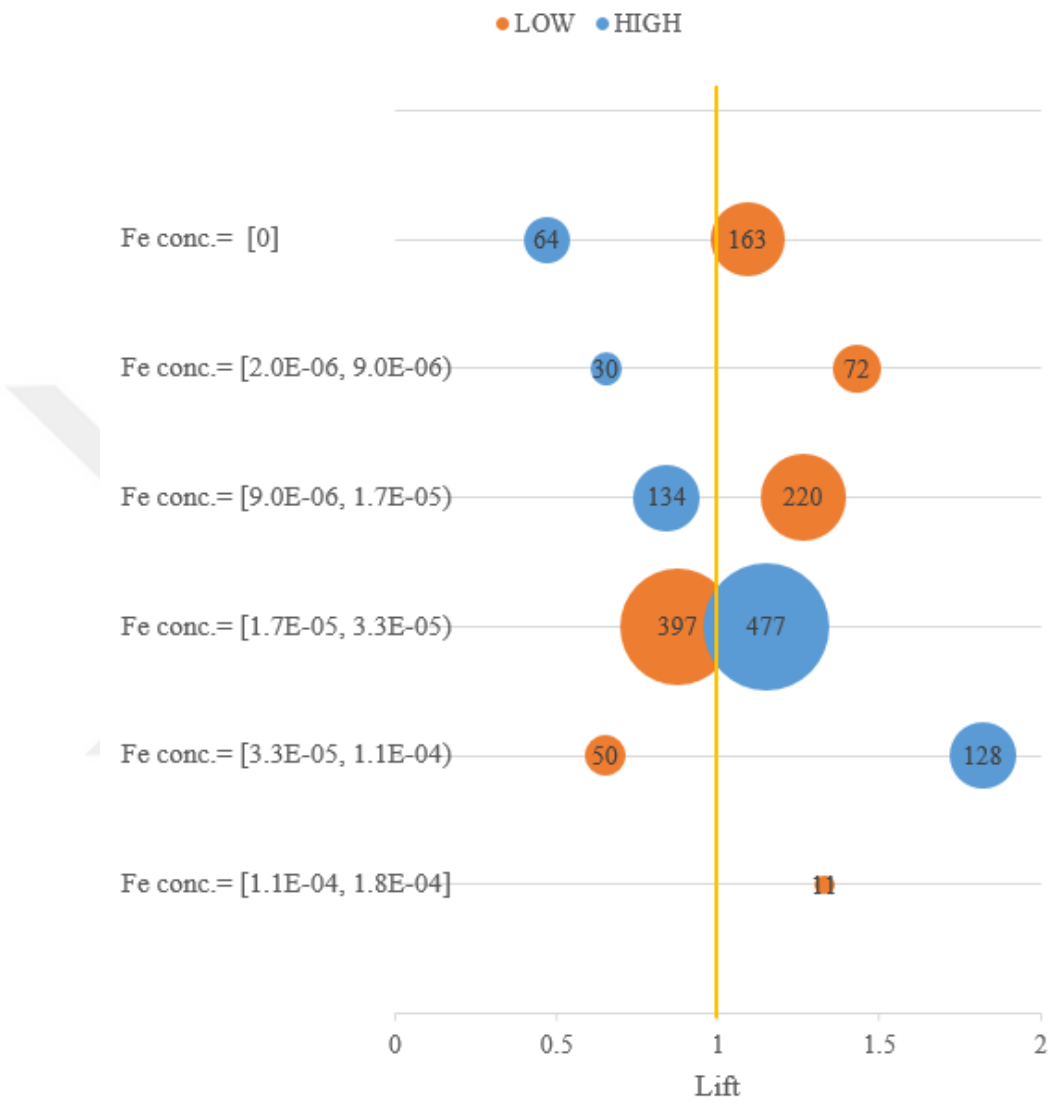

Figure 4.10. Lift and Count Values of Different Iron Concentrations for Lipid Content.

4.1.2.3. N concentration. N limitation is a common technique used to create stress condition for microalgae. As seen in Figure 4.11, N limitation results with increased lipid content, although low nitrogen concentration decreases the growth rate. First level of nitrogen concentration's lift value is found to be around 2.5 from association rule mining algorithm. This result is in agreement with the experiments done by Li *et al.*, Converti et al, and Pancha et al, which are investigated the effect of nitrogen concentration on lipid content, and also found that the nitrogen limitation triggers the lipid accumulation mechanism [93] [98] [100].

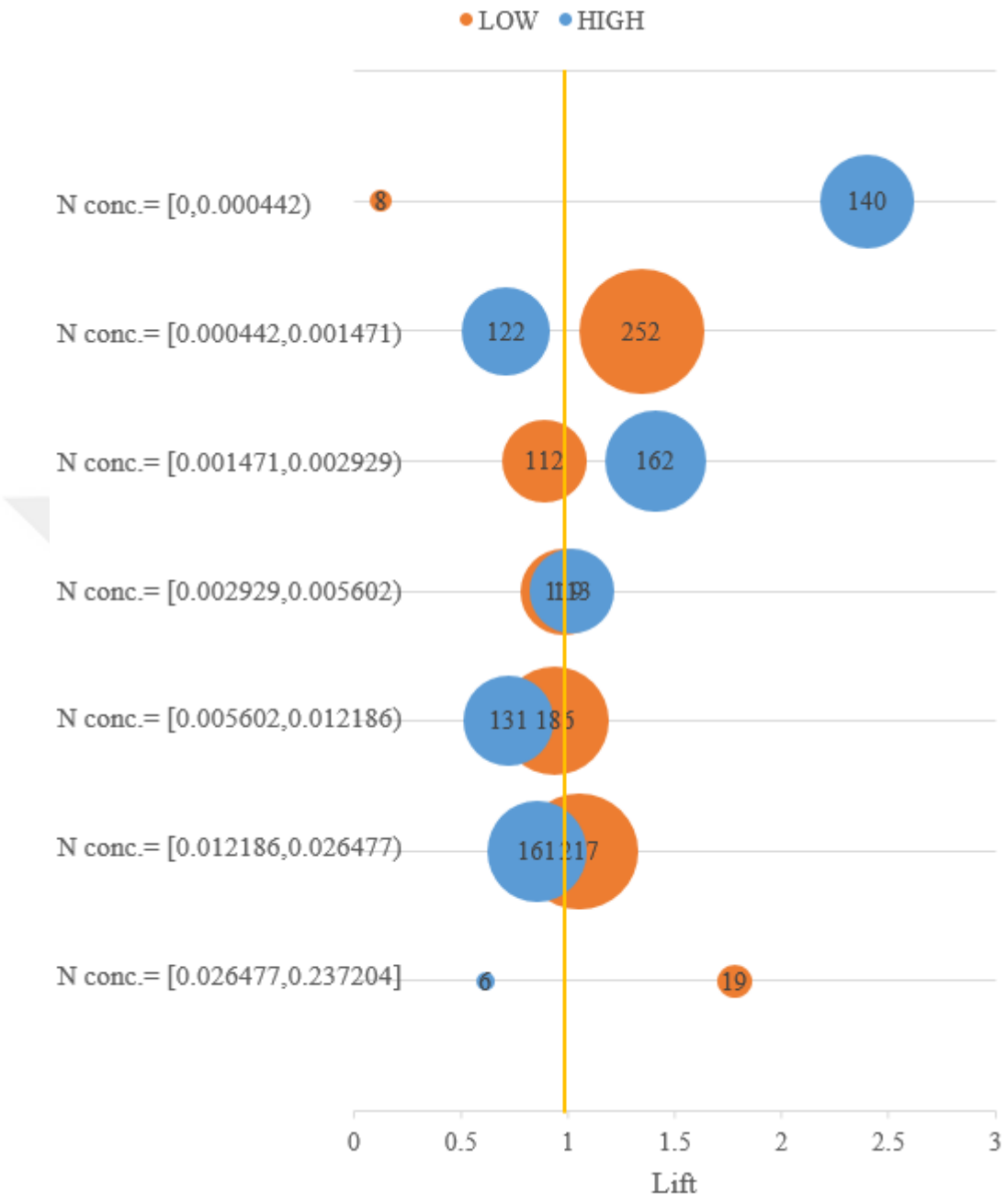

Figure 4.11. Lift and Count Values of Different Nitrogen Concentrations for Lipid Content.

4.1.2.4. PO_4 concentration. Phosphorus is another major nutrient required for microalgae growth, with nitrogen. High values of phosphorus yields high biomass concentration. Nevertheless, phosphorus limitation increases the lipid content of the microalgae. Figure 4.12 shows that highest lipid content is achieved in the lowest second PO_4 concentration level gathered from the dataset.

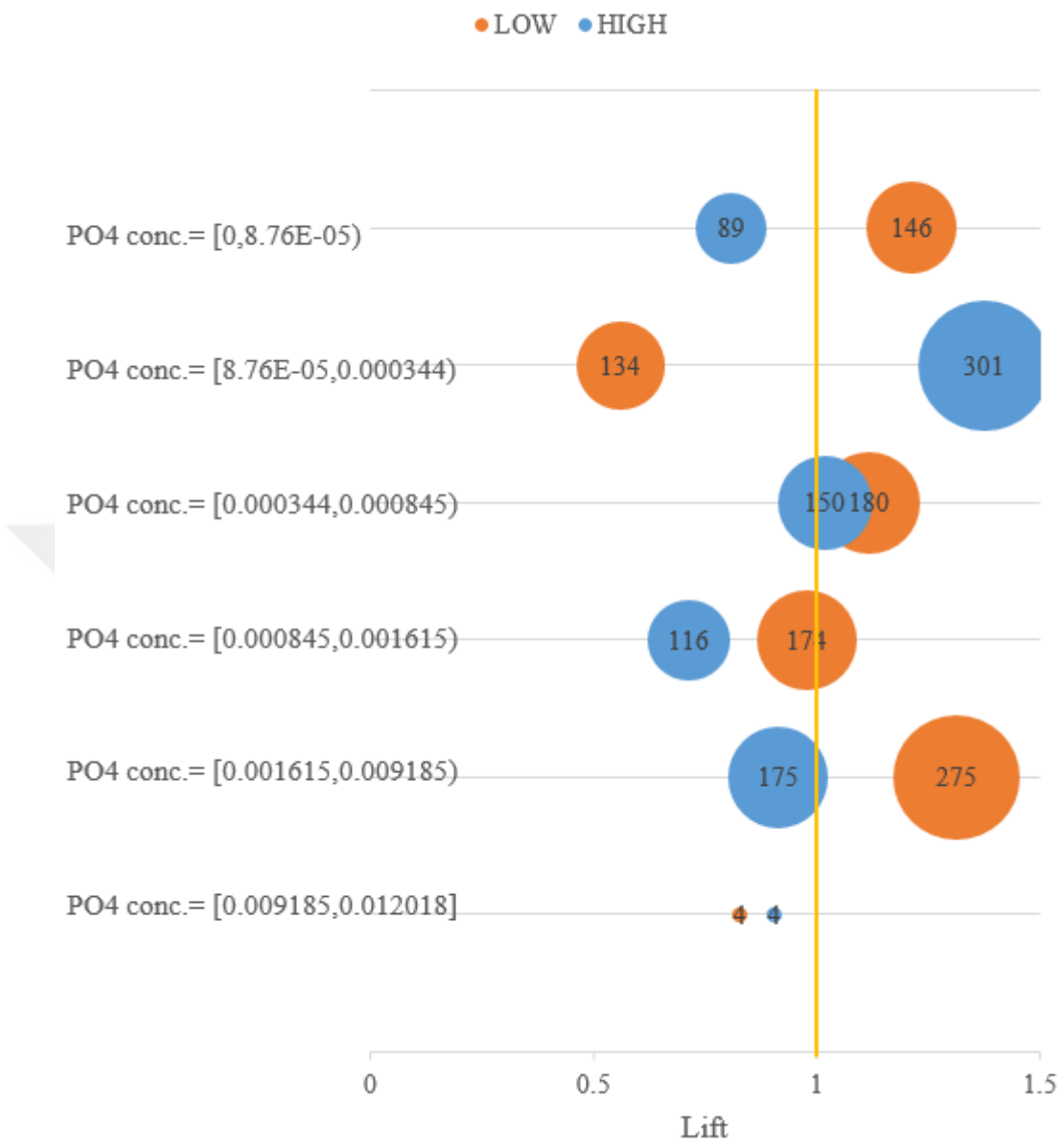

Figure 4.12. Lift and Count Values of Different Phosphorus Concentrations for Lipid Content.

4.1.2.5. Salinity Stress. Salinity stress is another method used to create stress and force microalgae to accumulate lipid. Although number of data for salinity stress was limited, the lift values of high salt levels for both marine and freshwater species is found as 1.5 and 2 (Figure 4.13). Higher biomass production in high salinity for both freshwater and marine microalgae is also observed in the works done by Harwati et al, Arora *et al.*, Dahmen *et al.*, and Ahmed *et al.*[5] [115] [121] [122].

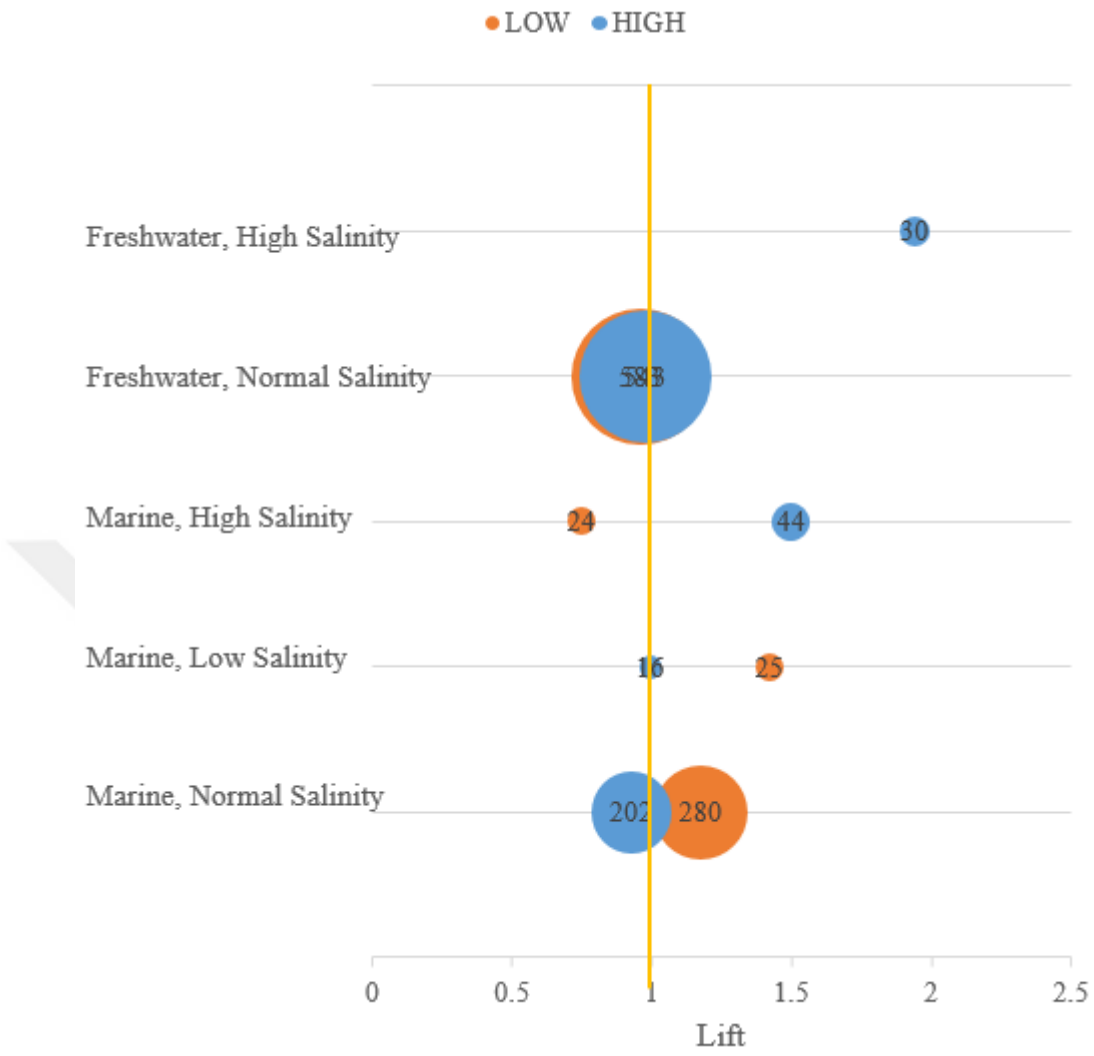

Figure 4.13. Lift and Count Values of Different Salinity Levels for Lipid Content.

4.1.2.6. Species. From data gathered from the literature most common eight microalgae species are tested with association rule mining to investigate their lipid content. Some microalgae species are found to have high lipid content, like Chlorococcum and Nannochloropsis species, which have lift value around 2 for high lipid content, as seen in Table 4.8. On the other side, Tetraselmis, Isochrysis, and Chlamydomonas species have high lift values for low lipid content.

4.1.2.7. VVM*CO_2 content. The combined variable CO_2 rate shows that lift values are increased with increased level for high lipid content and decreases for low lipid content(Figure 4.14).

Table 4.8. Lift and Count Values of Selected Microalgae Species for Lipid Content.

Microalgae Species *Feature A*	**Lipid Content (%)** *Class B*	**Support** *Fraction of all data to data that is in class B and has feature A*	**Confidence** *Fraction of data that has feature A to data that is in class B and has feature A*	**Lift** *Fraction of confidence to data that is in class B*	**Count** *Number of data that has feature A and in class B*
Chlamydomonas	LOW	0.025	0.434	1.259	66
	MEDIUM	0.020	0.349	1.027	53
	HIGH	0.012	0.217	0.688	33
Chlorella	LOW	0.073	0.263	0.763	192
	MEDIUM	0.108	0.392	1.154	286
	HIGH	0.095	0.345	1.094	252
Chlorococcum	LOW	0.008	0.171	0.496	20
	MEDIUM	0.009	0.205	0.604	24
	HIGH	0.028	0.624	1.978	73
Dunaliella	LOW	0.017	0.387	1.121	46
	MEDIUM	0.013	0.294	0.866	35
	HIGH	0.014	0.319	1.012	38
Isochrysis	LOW	0.016	0.430	1.247	43
	MEDIUM	0.015	0.410	1.207	41
	HIGH	0.006	0.160	0.507	16
Nannochloropsis	LOW	0.017	0.178	0.515	46
	MEDIUM	0.021	0.212	0.625	55
	HIGH	0.060	0.610	1.934	158
Scenedesmus	LOW	0.077	0.394	1.142	204
	MEDIUM	0.077	0.394	1.160	204
	HIGH	0.042	0.212	0.673	110
Tetraselmis	LOW	0.017	0.880	2.551	44
	MEDIUM	0.002	0.080	0.236	4
	HIGH	0.000	0.000	0.000	0

4.1.2.8. Cell Disruption Method. As given in Figure 4.15, any of the cell disruption method employed before the lipid extraction method, generally favors (lift value higher than 1) high lipid content. The instance numbers are very low to make a comparative analysis within the methods used for cell disruption, instead comparison with non-disrupted case can be made. The high lift values for high lipid content class suggests that more extraction is possible with cell disruption.

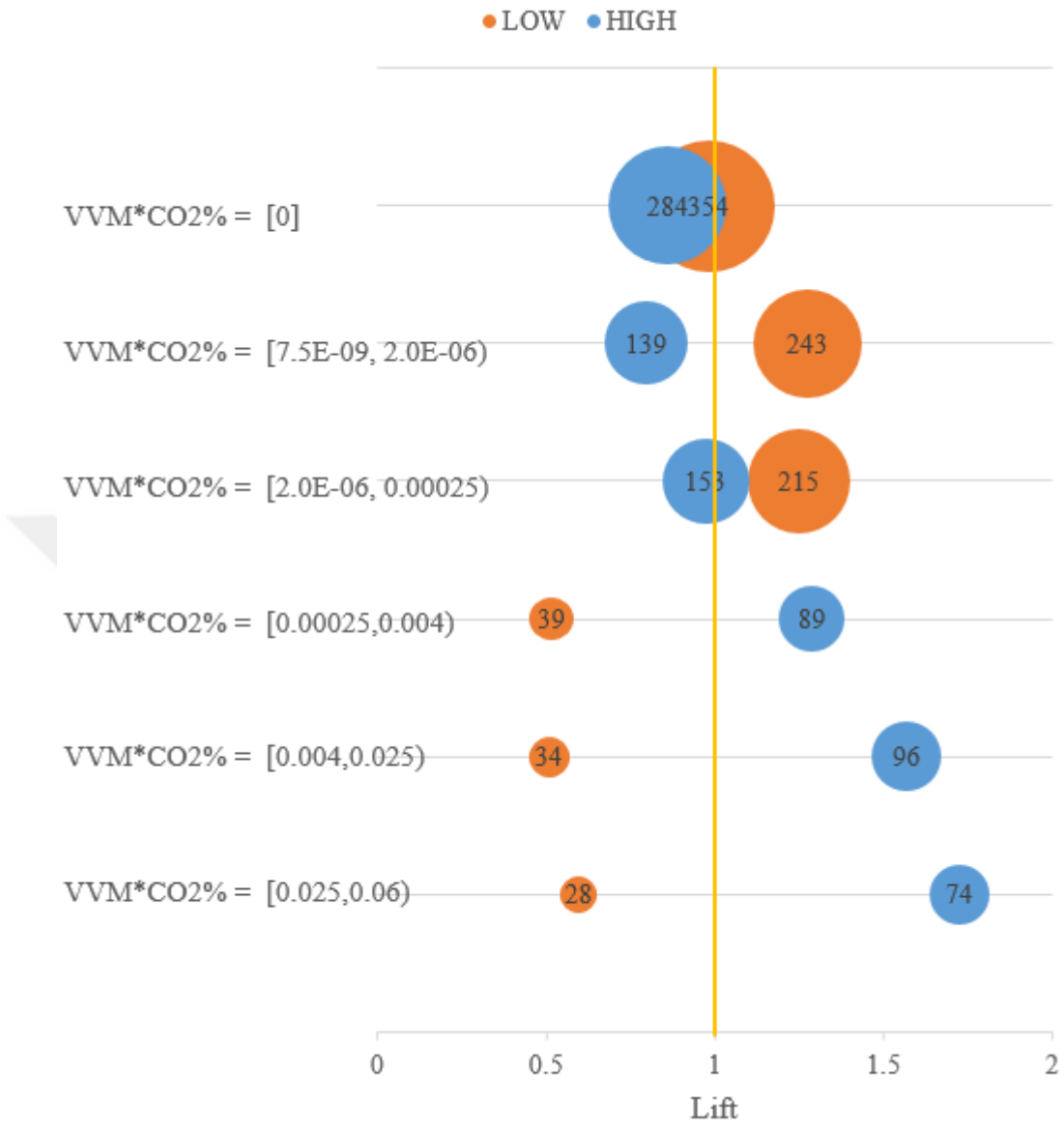

Figure 4.14. Lift and Count Values of Different CO2 Flows for Lipid Content.

4.1.2.9. Solvents Used for Extraction. From lift values given in Figure 4.16, n-hexane has tendency to yield low lipid content. Experiments done by Ranjan *et al.*, Menendez *et al.*, and Servaes *et al.* also shows that n-hexane recovers low lipid that other solvents that are commonly used like methanol-chloroform mixture [128] [129] [130]. Majority of instances were using $CHCl_3$-CH_3OH mixture as solvent. Other mixtures are rarely available in the dataset for making statistically significant rules.

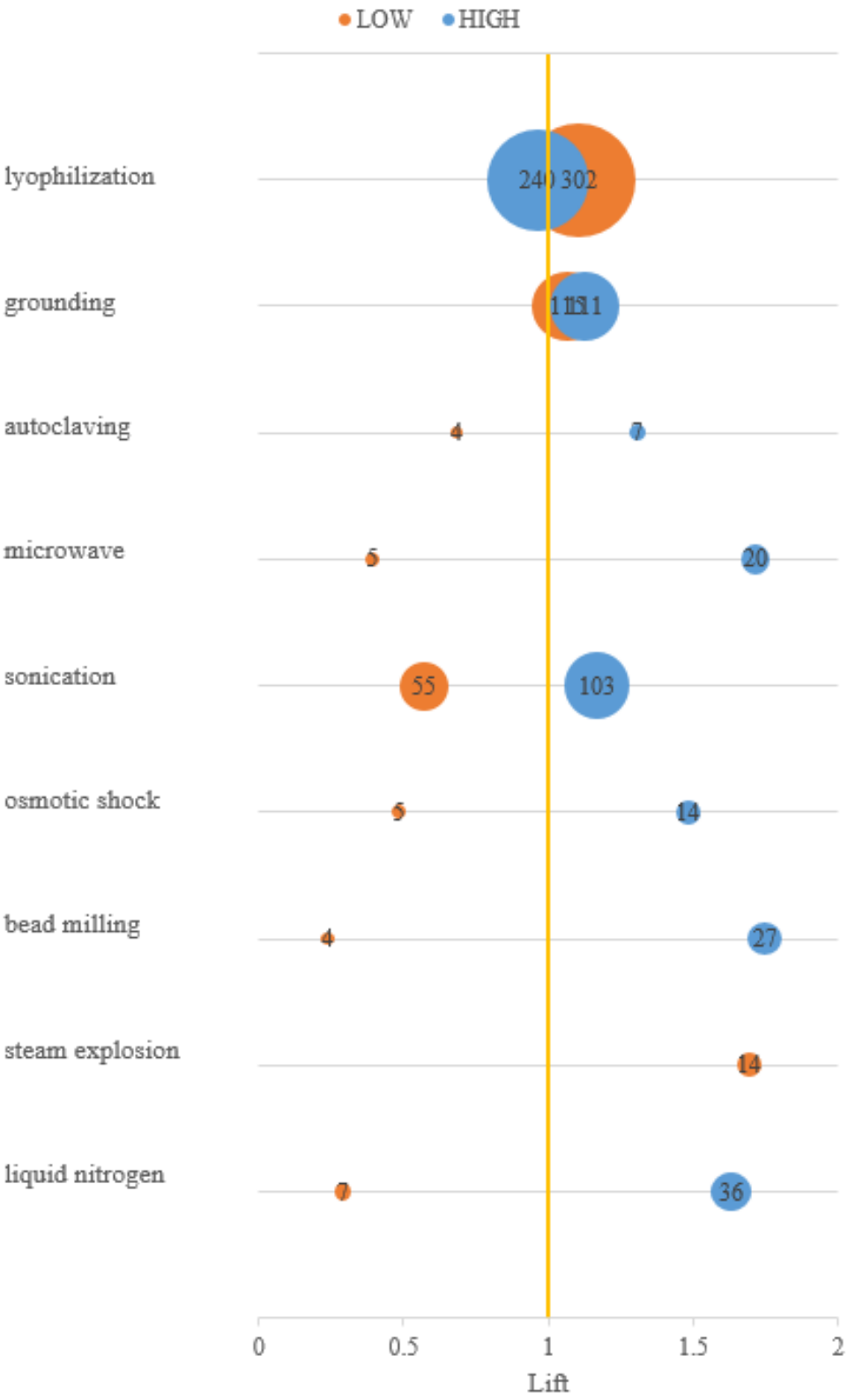

Figure 4.15. Lift and Count Values of Different Cell Disruption Methods for Lipid Content.

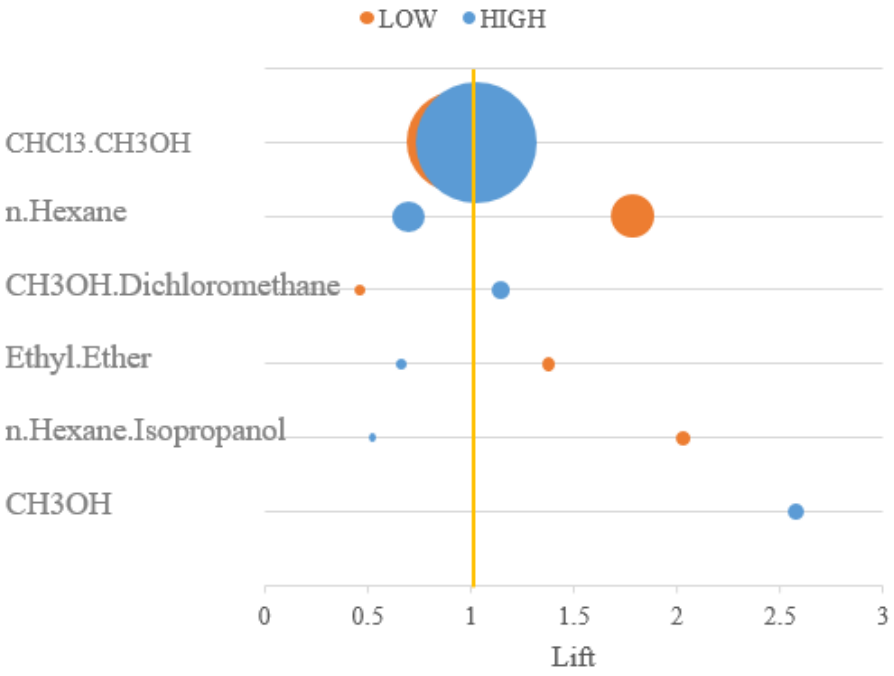

Figure 4.16. Lift and Count Values of Different Solvents for Lipid Content.

4.2. Decision Tree

Two output variables, biomass production and lipid content, are modeled with decision tree algorithm for classification purposes. In both cases, the output is categorized into three classes as low, medium and high. Then the decision tree model is constructed to classify a given instance with its known input variables. The model is constructed with "fitctree" function of MATLAB®. Two model parameters, maxnumsplit and minleafsize, are tuned with using a train set to build the model and testing the performance of this model by using a test set. Performance is quantified by overall accuracy, accuracy of low class, accuracy of medium class, and accuracy of high class.

4.2.1. Decision Tree for Biomass Production

Decision tree analysis for biomass production is performed for the subset of autotrophic growth, as the 86% of data belong to that subset, in which the resulting tree would give more simplistic heuristics for autotrophic growth then heuristics for all growth conditions.

Number of nodes and errors for both training and testing sets are noted. Number of nodes in the decision tree model is regarded as the complexity of the model. The output variable, biomass production ($mgL^{-1}day^{-1}$) is divided into three classes, as low, medium and high biomass production. The division is done by both to yield equal-sized classes to avoid class imbalance problem and also to have a meaningful division within the nature of the dataset. Classes, their ranges, and number of data points are summarized in Table 4.9.

Table 4.9. Class Ranges of Biomass Production for Decision Tree Analysis.

Class	Range	Number of Data
Low	[-14,56)	1566
Medium	[56,155)	1669
High	[155,990)	1754

From numerous trials for estimating the biomass production, number of nodes and error values are plotted for comparison of the models. Figure 4.17 shows that when the number of nodes (complexity of the model) increases, error decreases. However, after a certain complexity, error stabilizes and even starts to increase in test case, which suggest overfitting.

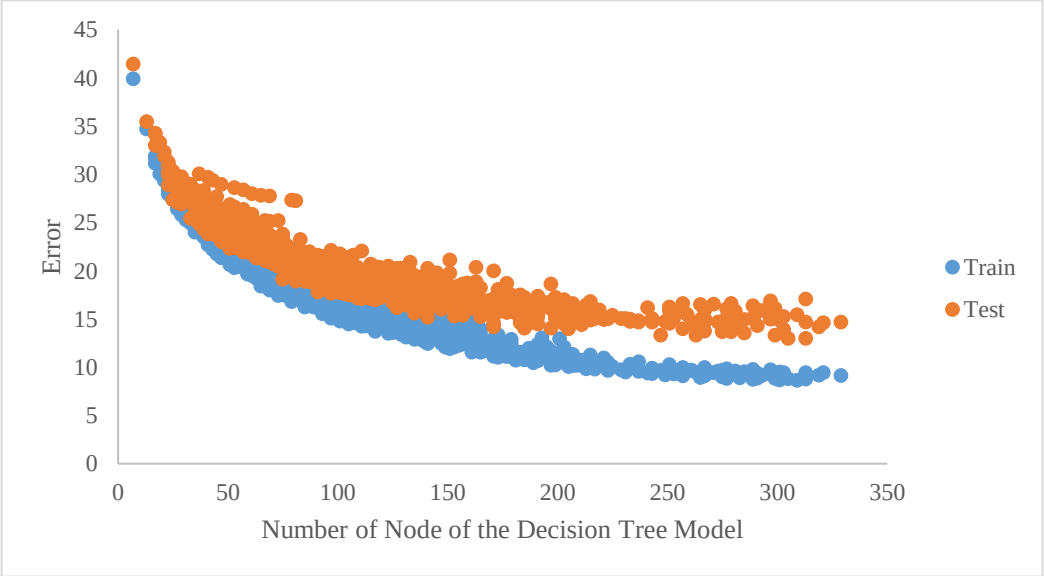

Figure 4.17. Error of Test and Train Sets for Biomass Production Prediction with Different Number of Decision Tree Nodes.

Although error values are minimized at around 150 nodes, for the sake of simplicity of the model, node number is selected to be around 50 nodes, in which difference between the error values for train and test set starts to increase. Tree with 250 nodes may have significant smaller errors, however, this implies overlearning, which is not an improvement of the tree performance. Also, a large tree can hinder the practical results that become too complex for practical use. The best possible models are given inFigure 4.18, in which the model parameters; maxnumsplit and minleafsize are tuned to give high accuracies in predicting high biomass production, low biomass production and overall dataset.

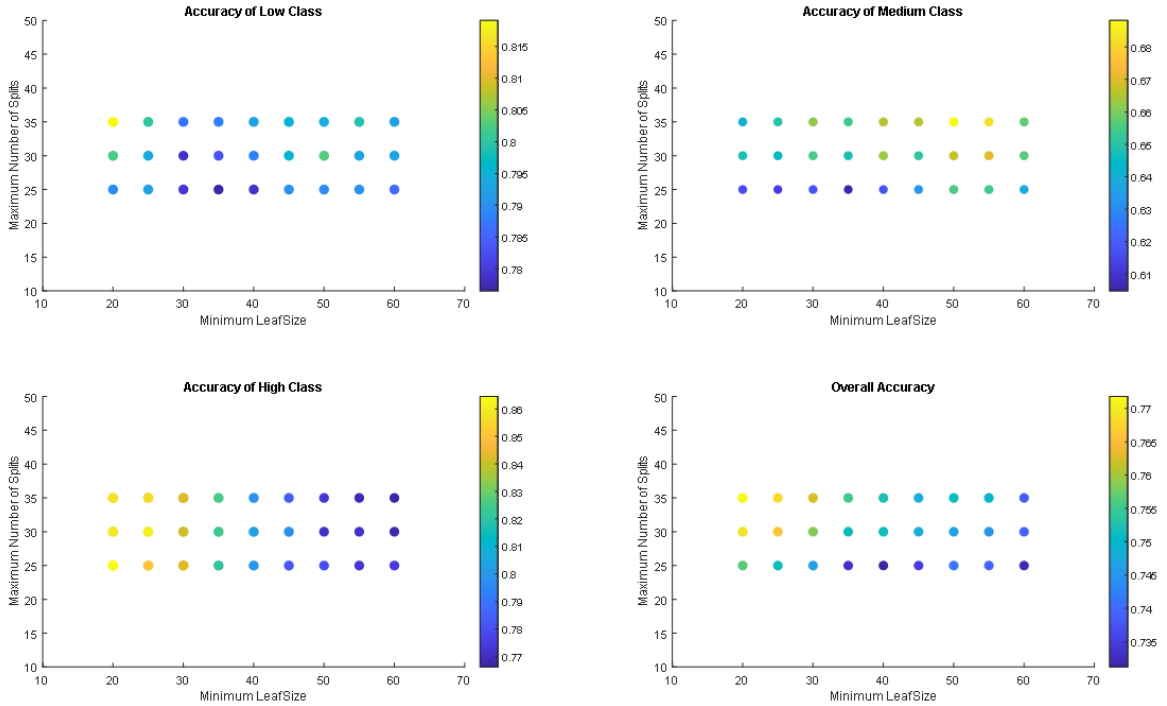

Figure 4.18. Test Set Accuracy of Overall Set and Individual Classes with Different Parameter Values for Biomass Production.

Accuracy of medium class was treated as not vitally important, as finding a rule for high biomass production and avoiding a way that results with low biomass production is more important. As can be seen from Figure 4.18, accuracy of low class was not changing too much between tested parameters. Maximizing accuracy of high class while keeping the overall accuracy in a reasonable value and keeping the model as simple as possible, maxnumsplit is selected as 30 and minleafsize is selected as 30 as well.

Confusion matrix of all test results from cross validation is given in Table 4.10. The rows of the confusion matrix symbolize the actual data number for each class, whereas columns represents number of data predicted to be in each class.

Table 4.10. Confusion Matrix of Test Set for Biomass Production.

		Predicted		
		Low	Medium	High
Actual	Low	719	177	43
	Medium	98	685	149
	High	23	137	789

The optimized decision tree model is employed to the whole dataset and resulting tree is given in Figure 4.19. The first split is 17h photoperiod is the split value calculated by the decision tree algorithm. Although it is not an exact limit, it is valid for the instances in the dataset. The splits of the numerical variables should be thought as empirical approximations. Overall error is 22.2%.

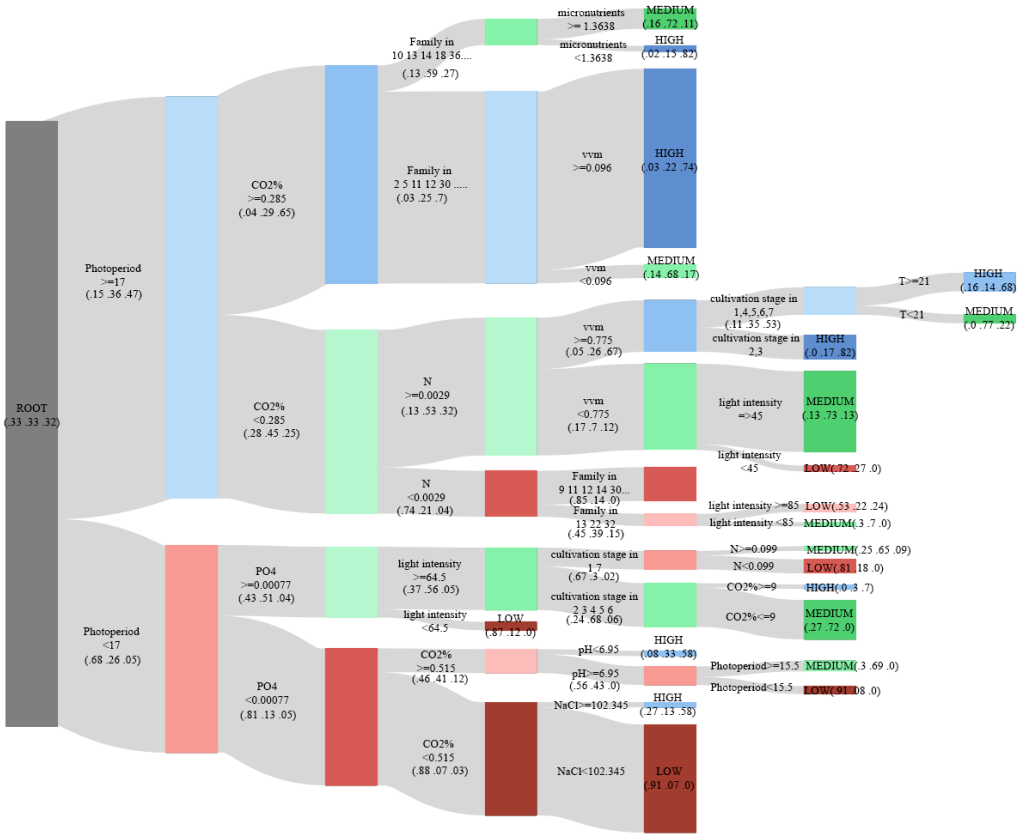

Figure 4.19. Decision Tree of Whole Dataset for Biomass Production.

In relative importance of input variables (Figure 4.20), obtained from decision tree algorithm, most important variables can be named as CO2%, vvm, photoperiod, and nitrogen conc., where the first two are main carbon sources for microalgae growth, third is main energy source, and fourth is the most common element found in microalgae cells.

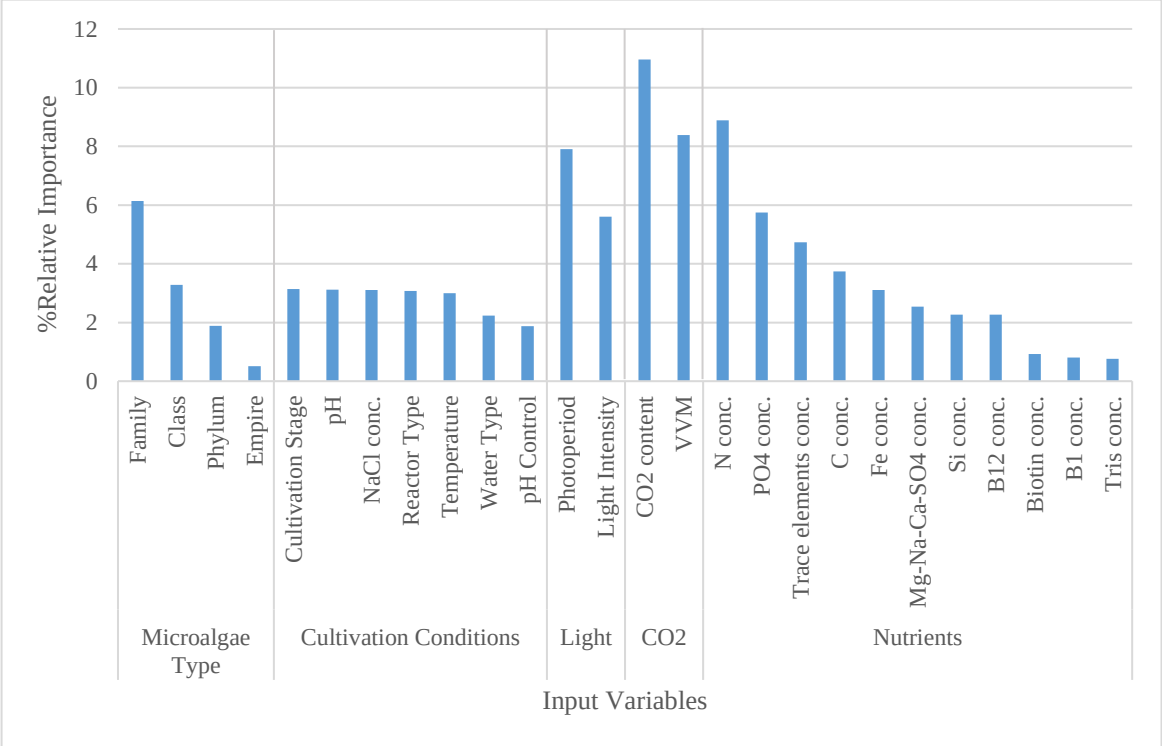

Figure 4.20. Predictor Importance Estimates of Whole Dataset for Biomass Production.

Table 4.11. Confusion Matrix of Whole Dataset for Biomass Production.

		Predicted		
		Low	Medium	High
Actual	Low	1126	225	73
	Medium	143	912	356
	High	20	124	1247

Table 4.11 shows the confusion matrix of test and training set. It can be seen that above 75% of the low and high classes are correctly classified. Also, the majority of the misplaced instances are located in the neighboring classes. The low accuracy of the medium class is expected, and can be reasoned as due to the leakage of the instances to both sides.

The low accuracy of the medium class is expected, which can be rationalized by saying that the medium classes are prone to impurity because of the leakage of instance from both neighboring classes.

4.2.2. Decision Tree for Lipid Content

Decision tree algorithm was also used for classification of lipid content (w/w). Categorization of the output, lipid content, was done by dividing the dataset into three equal-instanced categories. Categories are summarized in Table 4.12:

Table 4.12. Class Ranges of Lipid Content for Decision Tree Analysis.

Class	Range	Number of Data
Low	[0,13)	857
Medium	[13,24.5)	858
High	[24.5,74.8]	857

The error rates of train and test sets are plotted with respect to complexity of the model in Figure 4.21, where number of nodes corresponds to complexity. The optimum number of nodes of the model is decided to be around 30. Further increasing the complexity of the model slightly decreases the error values.

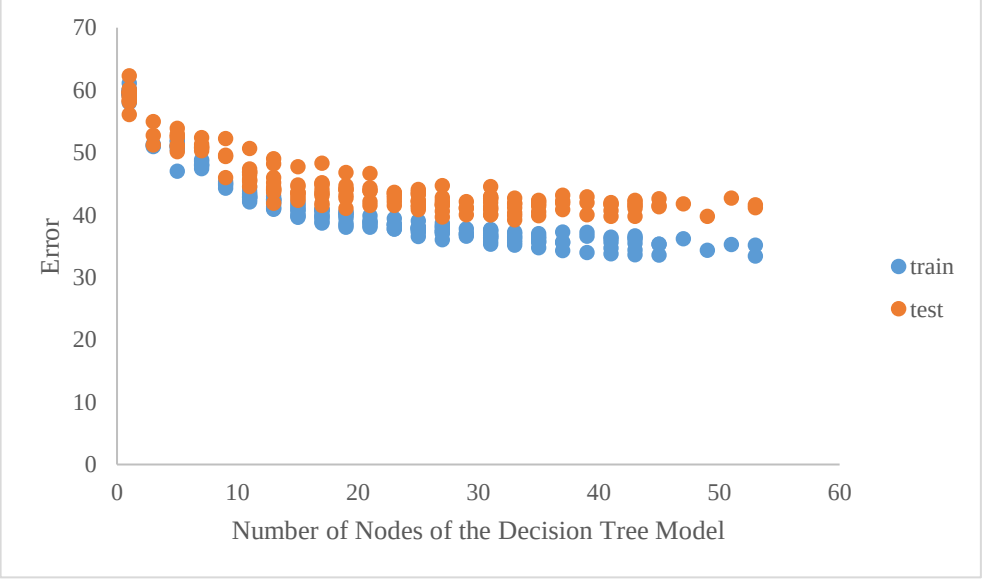

Figure 4.21. Error of Test and Train Sets for Lipid Content Prediction with Different Number of Decision Tree Nodes.

Maxnumsplit and minleafsize parameters are tuned to find the optimum tree having node number around 30. Figure 4.22 shows the results of accuracies of each class together with the accuracy of the overall model for the test set. The optimum values for maxnumsplit and minleafsize are decided to be 30 and 65, which maximizes overall accuracy, without losing accuracies in low and high class predictions.

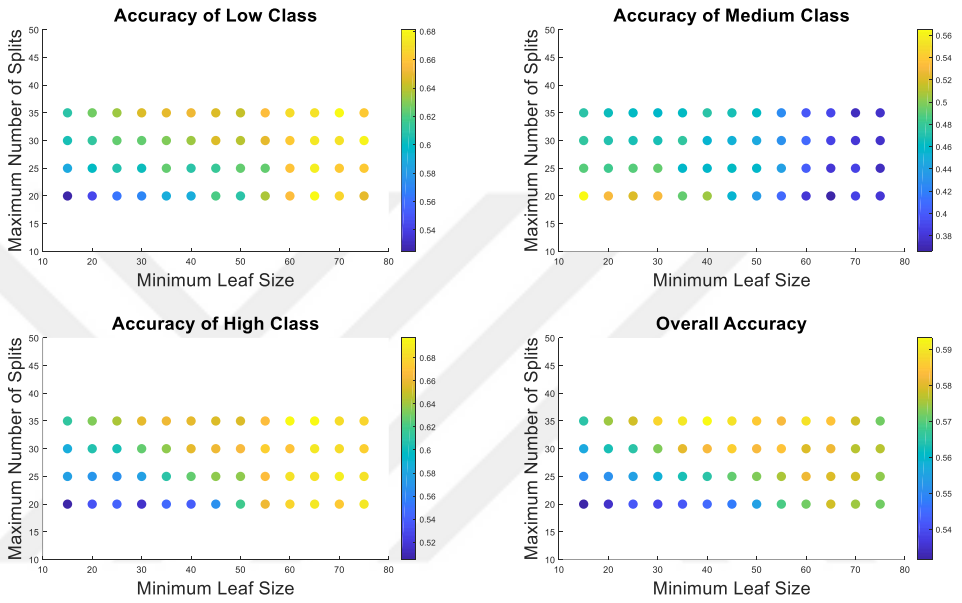

Figure 4.22. Test Set Accuracy of Overall Set and Individual Classes with Different Parameter Values for Lipid Content.

Confusion matrix of all test results is given in Table 4.13.

Table 4.13. Confusion Matrix of Test Set for Lipid Content.

		Predicted		
		Low	Medium	High
Actual	Low	194	26	61
	Medium	102	110	95
	High	33	45	191

The optimized decision tree model is employed to the whole dataset. As seen in Figure 4.23, first and second splits are made with microalgae family, as lipid content of

microalgae strongly depends on their metabolism. Error of the optimized decision tree model using whole databasewas 37.8%.

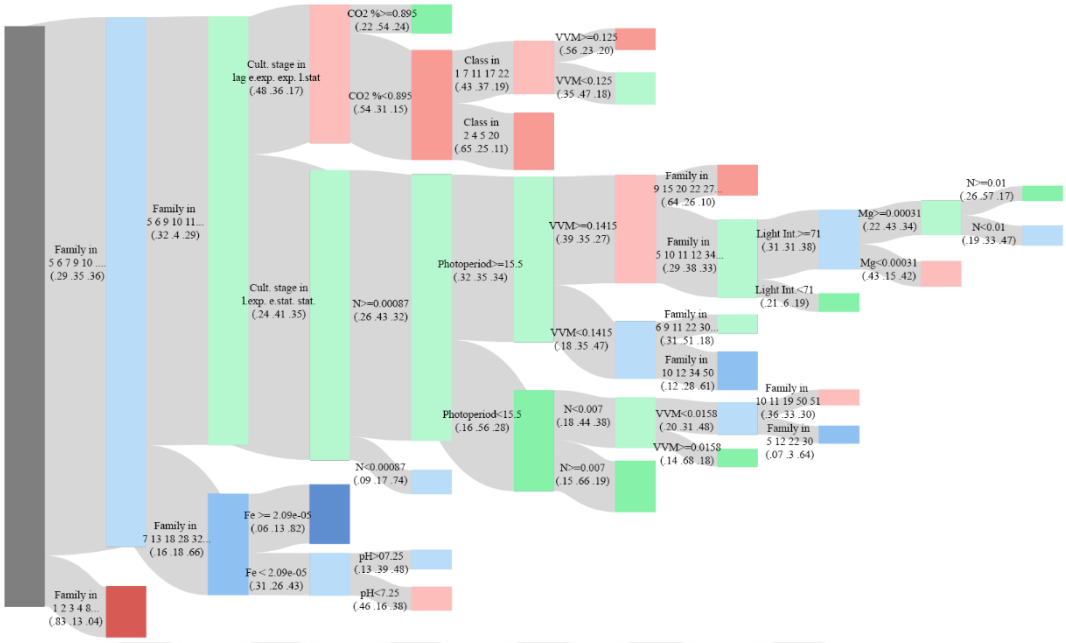

Figure 4.23. Decision Tree of Whole Dataset for Lipid Content.

In relative importance of input variables (Figure 4.24), obtained from decision tree algorithm, most important variables can be named as microalgae family, iron concentration, and nitrogen concentration. Microalgae family is the most specific classification parameter given into decision tree algorithm. This importance of microalgae family on lipid content suggest that some microalgae are more likely to accumulate more/less lipid than others. Unlike biomass production, lipid content is highly related with microalgae type. The importance of iron concentration was also observed in association rule mining, where high concentrations of iron results with high lipid content. Nitrogen concentration was also found to be highly important, which can be explained by stress conditions created by nitrogen limitation that results with lipid accumulation in microalgae cells.

The confusion matrix of test and training set shows that higher than 60% of the low and high classes are correctly classified. Also, the majority of the misplaced instances are located in the neighboring classes, as same as biomass production.

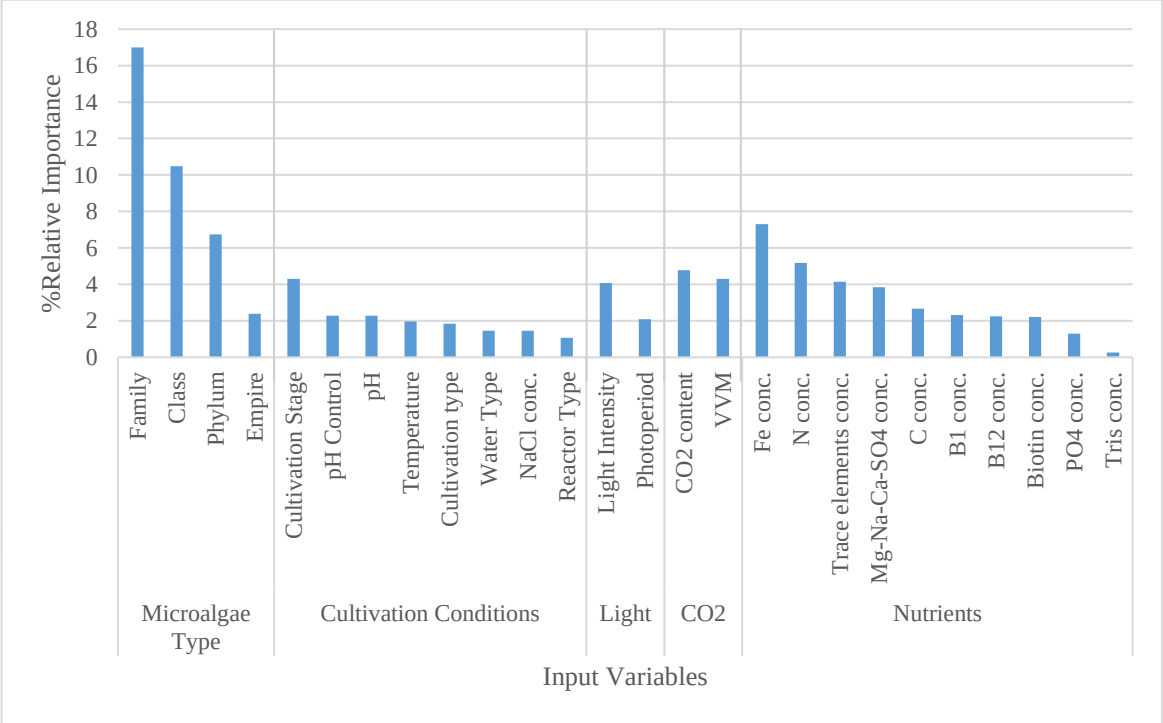

Figure 4.24. Predictor Importance Estimates of Whole Dataset for Lipid Content.

Table 4.14. Confusion Matrix of Whole Dataset for Lipid Content.

		Predicted		
		Low	Medium	High
Actual	Low	607	181	77
	Medium	204	466	183
	High	174	154	525

4.3. Artificial Neural Network

4.3.1. Artificial Neural Network for Biomass Production

For artificial neural network algorithm, RSNNS package of R was used. The mlp function of this package requires one hidden activation function and one learning function for modeling. Forty-three different activation functions and eight different learning functions were tested for their performance quantified by their error, RMSE, and R^2 values. The possible candidates for reasonable models were studied further by increasing the maximum iteration number. Number of nodes is the tuned parameter in the mlp function. Number of

neurons were changed between 2 and 50. The data was filtered to only include autotrophic cultivation type. Data was divided into 434 groups, in which each group represents data from one experiment, to prevent the accuracies which are higher than the true potential of the model.

All combinations of hidden activation and learning functions are tested for their performance, and best four model combinations are selected for further analysis. The selected functions were as shown in Table 4.15.

Table 4.15. Hidden Activation and Learning Function Combinations Used for ANN optimization for Biomass Production.

models	hidden activation functions	learning function
1	Act_Logistic	SCG
2	Act_LogisticTbl	
3	Act_Sinus	
4	Act_TD_Logistic	

Four most promising models were compared to each other with respect to error, RMSE, and R^2 values. Figure 4.25-4.27 show the comparison with respect to different node numbers.

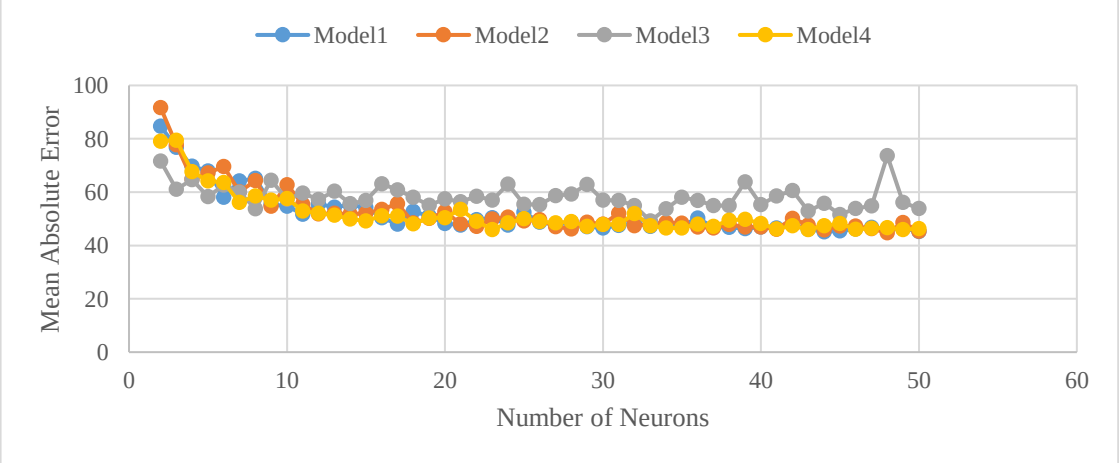

Figure 4.25. Error Value of Different Function Combinations for Biomass Production.

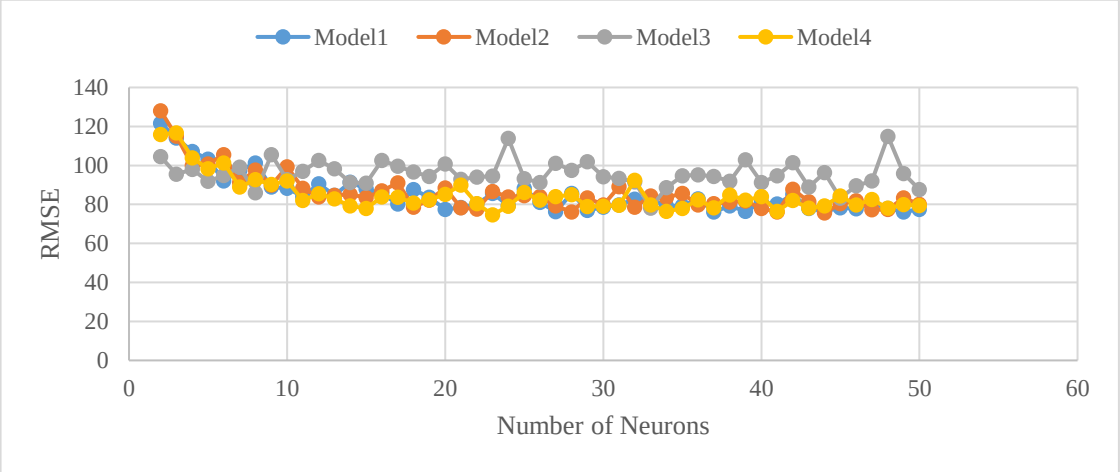

Figure 4.26. RMSE Value of Different Function Combinations for Biomass Production.

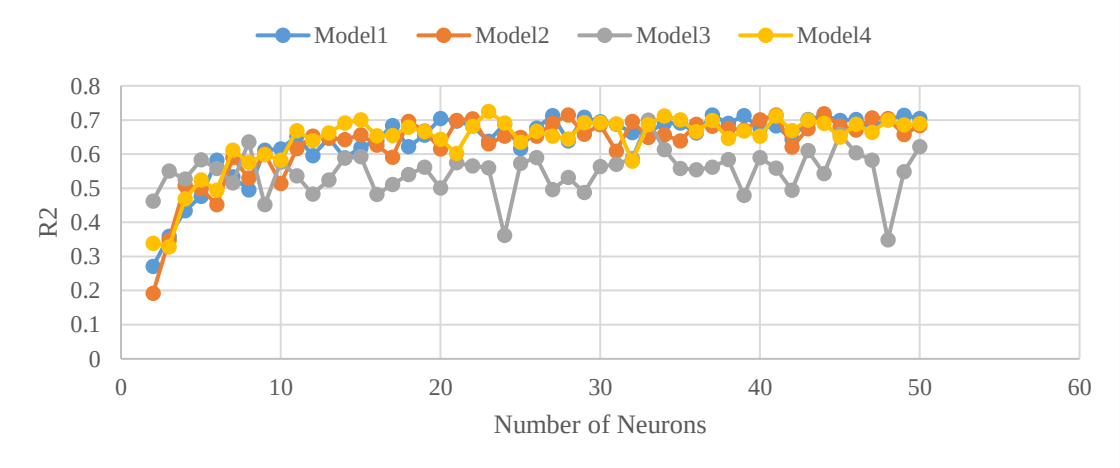

Figure 4.27. R2 Value of Different Function Combinations for Biomass Production.

Model1 (where hidden activation function is Act_Logistic and learning function is SCG) was selected as the model which describes data best. The values of error, RMSE, and R^2 stabilizes after 30 neuron, which is where the model is accepted and further analyzed. Maximum number of iterations were tuned to be 100, as the higher values result with overlearning.

The observed values of the output -biomass production- was plotted with the predicted values of the model inFigure 4.28. Different colors show different experiments in the test set.

Relative importances of variables were determined by garson function in R. Figure 4.29 shows that volume, reactor type, photoperiod, and cultivation time is most important

four variables. Although in which stage of microalgae growth the microalgae cultivated is more explanatory, data for cultivation day was also become important, in the sense that dataset composes of instances that microalgae growth was monitored for days, in which although the growth stage does not change, the biomass production changes in small values. For capturing that trend, cultivation day becomes important variable that minimizes error.

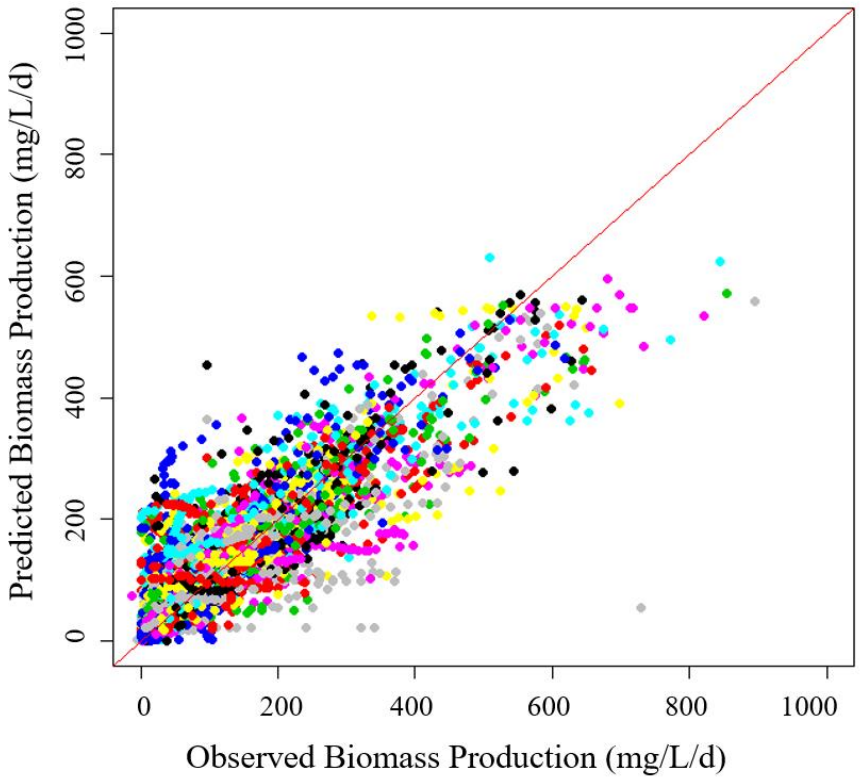

Figure 4.28. Test Results of MLP for Biomass Production with leave-one-out Method.

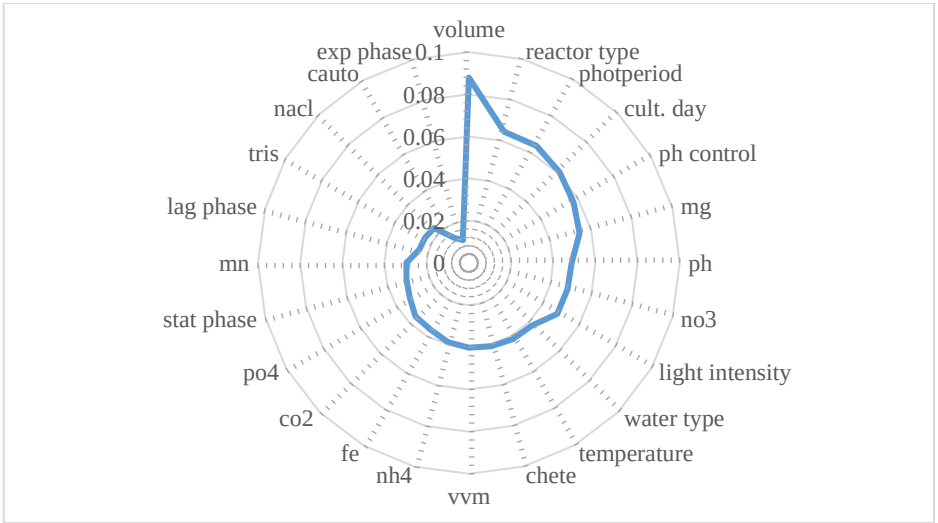

Figure 4.29. Relative Importance of Variables for Biomass Production.

The predictive power of the model was varient from experiment to experiment. Some experiments were poorly predicted. The poorly predicted experiments were individually examined to investigate the inadequacy of the model to predict those experiments. It was mainly resulting from three major problems. The most common problem was including experiments which investigate the effect of an input variable which is not included in the model, because of high number of categories of that input (like microalgae type) and because most of the other experiments did not mentioned that variable (like initial cell density). Other problem was stemming from underestimating the effect of an input variable, mostly autotrophic carbon source. Last problem was stemming from differences in the optimum culture condition for different microalgae species. If one article is investigating extreme conditions for one microalgae species whereas this condition was not an extreme in the general dataset, prediction gets poorer. This last problem can also be associated with the general problem that prediction is not as strong as where the input space is highly populated. Table 4.16. shows individual RMSE values for articles used in the database.

4.3.2. Artificial Neural Network for Lipid Production

All combinations of hidden activation and learning functions are tested for their performance, and best eighteen model combinations are selected for further analysis. The selected functions were as shown in Table 4.17.

Table 4.16. Hidden Activation and Learning Function Combinations Used for ANN optimization for Lipid Content.

models	hidden activation functions	learning function
1-2-3 4-5-6 7-8-9 10-11-12 13-14-15 16-17-18	Act_Logistic Act_Elliott Act_TanH Act_LogisticTbl Act_Sinus Act_Exponential	SCG BackPropBatch Rprop

Table 4.17. RMSE Values for Different Articles Used in the Dataset for Biomass Production.

Article	# of Data	# of Experiment	RMSE	Article	# of Data	# of Experiment	RMSE
Renaud *et al.*	18	3	4.3	Wu	15	1	69.9
Zhu	24	2	176.4	Qv	35	7	33.2
Illman	10	1	87.0	Teo	55	5	38.7
Takagi	6	1	150.2	Wu	140	10	41.9
Renaud	20	4	25.2	Feng	180	16	104.1
Hu	20	2	41.8	Singh	45	4	34.7
Li *et al.*	53	8	64.6	Abdelaziz	16	2	63.5
Converti	13	1	68.7	Ren	24	4	95.6
Chiu *et al.*	34	5	130.4	Muthuraj	21	4	66.5
Mandal	83	8	21.4	Ma	99	9	63.5
Rodolfi	30	1	89.7	Solovchenko	18	2	106.8
Lee *et al.*	18	1	88.5	Pancha	6	1	14.6
Arora	6	1	107.5	Roldan	16	4	195.0
Chokshi	5	1	103.4	Montoya	157	5	98.0
Lv	98	12	63.2	Xia	8	1	130.0
Huerlimann	35	11	35.1	Karpagam	26	6	13.0
Tang *et al.*	6	1	30.3	Vidyashankar	32	1	15.4
Chen	21	3	80.5	Tripathi	23	4	52.1
Prabakaran	18	1	8.7	Krzeminska	66	6	30.7
Shanab	5	5	17.4	Ra	200	20	12.9
Tang *et al.*	10	3	92.3	He	36	6	88.2
Zheng	15	2	89.4	Demirel	4	1	13.2
Xin	4	1	19.6	Ghosh	60	6	73.7
Wang	13	1	55.4	Han	169	13	33.3
Pal	36	4	232.7	Pancha	6	1	20.1
Cheirslip	6	1	22.0	Huang	9	1	154.4
Harwati	16	3	47.0	Slocomobe	117	1	14.9
Ho	14	3	149.8	Bagchi	56	8	30.5
Liu	18	3	102.5	Mohsenpour	462	36	45.1
Mallick	10	1	20.4	Mandotra	13	3	39.5
Griffiths	388	22	58.1	Qi	69	5	52.4
Wu	19	3	50.9	Mondal	42	6	19.7
Chellamboli	68	5	158.6	Dahmen	12	2	56.7
Frumento	5	1	27.6	Tan	29	4	78.0
Belotti	245	26	128.6	Arora	60	6	26.6
Gorain	9	1	12.9	Li	131	8	38.8
Song	20	1	128.7	Mondal	96	22	17.1
Ratha	31	4	10.3	Ahmed	5	1	38.8
Zhou	16	1	68.5	Luangpipat	24	1	65.9
Welter	10	1	39.4	Liang	37	4	23.8
Gao	32	4	45.8	Sivamakrishnan	45	9	39.0
Chu	12	1	45.4	Mondal	28	5	36.5
Nascimento	45	7	212.5				

Best model was achieved with the parameters; maximum number of iterations of 50, hidden activation function was selected as Act_Logistic, learning function was selected as SCG, number of neurons was 10 with one layer. The resulting models error was 7.73, RMSE is 10.05, and R^2 of 0.32.

The observed values of the output –lipid content- is plotted with the predicted values of the model in Figure 4.30. Different colors show different folds in the test set.

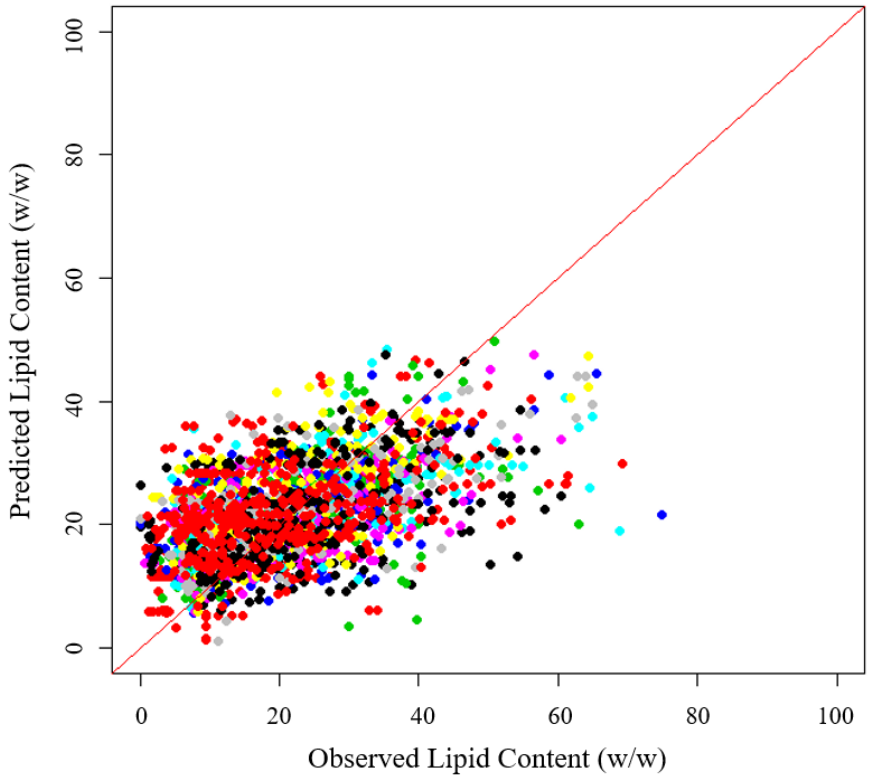

Figure 4.30. Test Results of MLP for Lipid Content.

Figure 4.31 shows that there is a trend with the output variable and residuals, which tells incomplete modeling. Maybe from latent variables that we did not get into account.

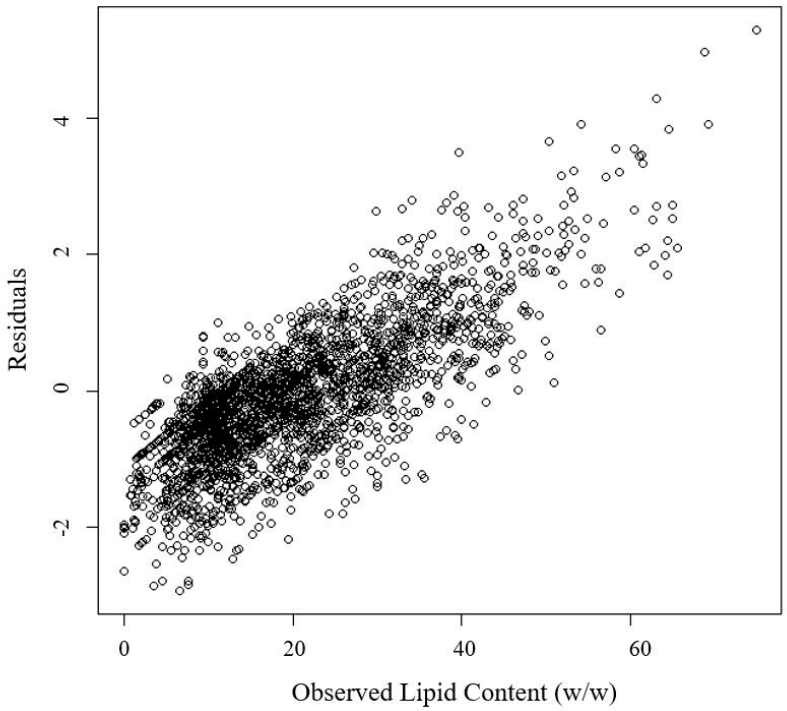

Figure 4.31. Residual versus Output (Lipid Content) of ANN model.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

In this thesis, a comprehensive database was constructed from articles published in the literature from1994 to 2017. 113 articles were gathered to form the dataset with 5908 instances. After cleaning and preprocessing the dataset, the number of articles and the number of instances used in the analyses were reduced to 106 and 5356, respectively. Three output variables, biomass production (mg/L d), lipid content (w/w), and lipid production were available in the dataset. Some articles reported these output variables in uncommon units. Unit conversion of outputs were made if possible, however some could not be converted into desired units without any further information for calibration. When the unit conversion was not possible, the instances were deleted in preprocessing step. Same procedure was also performed for input variables. Filling missing values, removing outliers, and reducing number of input variables are other steps involved in preprocessing.

Two of the three output variables were independent from each other. However, lipid production was simply calculated as multiplication of other two output variables, biomass production and lipid content. To examine the effect of input variables to two independent output variables, only biomass production and lipid content were studied. Database was divided into two subsets, where one had instances with reported biomass productions, and other with reported lipid contents. Association rule mining, decision tree, and artificial neural network analyses were applied to both database. K-fold cross validation was used in all models constructed.

In association rule mining analysis for biomass production and lipid content, algorithm needs minimum support and confidence values for compiling rules, and 0.01 and 0.3 for these values were selected respectively. Although more than one antecedents can be employed into the algorithm, the most useful, practical, and general rules are gathered with one antecedent cases. Some of the general rules gathered from association rule mining for biomass production and lipid content are given in Table 5.1. and Table 5.2.

Table 5.1. General Rules extracted from ARM for Biomass Production.

Cultivation Stage:	Unfavorable production in lag phase Probability of high production increases until Exponential phase Maximum (1.5) at Exponential Phase Probability of high production starts to decrease after Exponential Phase Unfavorable production after Early Stationary Phase
Cultivation Type:	Favorable production in heterotrophic and mixotrophic cultivation types
Light Wavelength:	Favorable production in blue and red light wavelengths
Light Intensity:	Favorable production in high light intensities Unfavorable production in low light intensities
CO2 Content:	Unfavorable production in low CO2 content Maximum at moderate concentration levels Unfavorable production in highest CO2 content
Microalgae Species:	Favorable production in Chlorella, Chlorococcum, and Nannochloropsis species Unfavorable production in Chlamydomonas, Dunaliella, Isochrysis, and Tetraselmis species
N concentration:	Gradual increase in the probability of high production from lowest to highest N concentration levels
PO4 concentration:	Gradual increase in the probability of high production until the highest PO4 concentration level
Temperature:	Unfavorable production in low and high temperature levels

Table 5.2. General Rules extracted from ARM for Lipid Content.

Cultivation Stage:	Unfavorable lipid content in lag, early exponential, and exponential phases Favorable lipid content in late exponential, early stationary, stationary, and late stationary phases.
Fe concentration:	Gradual increase in the probability of high lipid content with increasing Fe concentration Unfavorable lipid content in lowest three Fe concentration levels Favorable lipid content in moderate Fe concentration levels
N concentration:	Gradually decrease in the probability of high lipid content from lowest to highest N concentration levels Unfavorable lipid content in highest three N conc. levels
Salinity Stress:	Favorable lipid content in high salinity
Microalgae Species:	Favorable lipid content in Chlorella, Chlorococcum,and Nannochloropsis species Unfavorable lipid content in Chlamydomonas, Isochrysis, Scenedesmus and Tetraselmis species.
VVM*CO2:	Gradual increase in the probability of high lipid content with increasing CO2 flow rate Favorable lipid content in highest three levels
Cell Disruption Method:	Almost all cell disruption methods favors high lipid recovery
Solvent used in Extraction:	n-Hexane is not a good candidate for high lipid recovery

Decision tree analysis was done for classification purposes. Both of the output variables were categorized into three categories, and named as low, medium and high. Decision tree analyses were done in MATLAB using “fitctree” function, where two parameters were tuned for finding the optimum model; “maxnumsplit” and “minleafsize”. For biomass production these values were 30, 30 and for lipid content they were 30, 65. Performance of the decision tree for classification was quantified by model accuracy. Accuracies of both models are given in Table 5.3. The decision tree model for biomass production was highly powerful in classifying high results, whereas model for lipid content was more powerful in classifying low results. Overall accuracy of lipid content was not high as biomass production. It was found that cultivation conditions are the most significant variables in biomass production model, whereas microalgae type is the most significant variable in lipid content model.

Table 5.3. Accuracies of the Decision Tree Models.

	Overall Accuracy	Accuracy of High Class	Accuracy of Medium Class	Accuracy of Low Class
Biomass Production	77.7%	89.6%	64.6%	79.1%
Lipid Content	62.2%	61.5%	54.6%	70.2%

Artificial neural network analyses were performed for predictive purposes. Analyses were done in R, using “mlp” function, where four parameters were tuned for finding the optimum model; maximum number of iterations, number of neurons, learning function, and activation function. For biomass production database these parameters were tuned as 500, 20, SCG, Act_Logistic, and for lipid content database 50, 10, SCG, Act_Logistic. The analyses are done by implementing 10-fold cross validation, with ensuring instances from same experiments would be in the same folds. Performance of artificial neural network was quantified by mean absolute error, root mean square error, and r-squared values. The results were summarized in Table 5.4. The lipid content model was not reliable as residual analysis showed that there was a correlation between residuals and output.

Table 5.4. Results of the artificial neural network models with distribution of the outputs.

	Average of the output	Standard deviation of the output	Mean Absolute Error	Root Mean Square Error	R^2
Biomass Production	141.31	140.54	48.24	70.61	0.71
Lipid Content	20.38	12.21	7.73	10.05	0.32

5.2. Recommendations

Recommendations for more detailed and accurate knowledge extractions and for better models can be offered as follows:

- More data will improve the models. In some regions of the input space, instances got very sparse, which lowers the prediction ability of the models.
- In lipid content models, it is obvious that some variables affecting the output is not considered. More detailed database would result with better models. In addition, quantification of lipid content, and the concept of lipid differs from article to article. Construction of the database with clearly described experiments to include these variables will improve the model.
- Input variables that are already known to affect the output variable, like initial cell density, can be added to the model if sufficient number of experiments presents the information about those variables.. Species of microalgae can also be considered in modeling, either by modeling separately modeling each specie or by using high performance computers to handle large number of categories present as microalgae species.
- Other data mining algorithms can be applied to reach better prediction results, or more detailed information about the process.

REFERENCES

1. P. R. Mooij, G. R. Stouten, J. Tamis, M. C. M. van Loosdrecht, and R. Kleerebezem, "Survival of the fattest," *Energy Environ. Sci.*, vol. 6, no. 12, p. 3404, 2013.

2. M. Mubarak, A. Shaija, and T. V. Suchithra, "A review on the extraction of lipid from microalgae for biodiesel production," *Algal Research*, vol. 7. pp. 117–123, 2015.

3. T. M. Mata, A. A. Martins, and N. S. Caetano, "Microalgae for biodiesel production and other applications: A review," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 1. pp. 217–232, 2010.

4. L. Rodolfi *et al.*, "Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor," *Biotechnol. Bioeng.*, vol. 102, no. 1, pp. 100–112, 2009.

5. T. U. Harwati, T. Willke, and K. D. Vorlop, "Characterization of the lipid accumulation in a tropical freshwater microalgae Chlorococcum sp.," *Bioresour. Technol.*, vol. 121, pp. 54–60, 2012.

6. Y. C. Sharma, B. Singh, and J. Korstad, "A critical review on recent methods used for economically viable and eco-friendly development of microalgae as a potential feedstock for synthesis of biodiesel," *Green Chem.*, vol. 13, no. 11, p. 2993, 2011.

7. Q. Hu *et al.*, "Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances," *Plant Journal*, vol. 54, no. 4. pp. 621–639, 2008.

8. S. Development, *Biofuels: Biotechnology, Chemistry, and Sustainable Development*, vol. 2010, no. 5. 2010.

9. J. ZHU, J. RONG, and B. ZONG, "Factors in mass cultivation of microalgae for biodiesel," *Chinese J. Catal.*, vol. 34, no. 1, pp. 80–100, 2013.

10. J. U. Grobbelaar, “Algal Nutrition – Mineral Nutrition,” *Handb. Microalgal Cult.*, pp. 95–115, 2007.

11. G. Petkov, A. Ivanova, I. Iliev, and I. Vaseva, “A critical look at the microalgae biodiesel,” *European Journal of Lipid Science and Technology*, vol. 114, no. 2. pp. 103–111, 2012.

12. L. Gouveia and A. C. Oliveira, “Microalgae as a raw material for biofuels production,” *J. Ind. Microbiol. Biotechnol.*, vol. 36, no. 2, pp. 269–274, 2009.

13. M. J. Griffiths, R. P. van Hille, and S. T. L. Harrison, “Lipid productivity, settling potential and fatty acid profile of 11 microalgal species grown under nitrogen replete and limited conditions,” *J. Appl. Phycol.*, vol. 24, no. 5, pp. 989–1001, 2012.

14. N. Sanyano, P. Chetpattananondh, and S. Chongkhong, “Coagulation-flocculation of marine Chlorella sp. for biodiesel production,” *Bioresour. Technol.*, vol. 147, pp. 471–476, 2013.

15. H. Taher, S. Al-Zuhair, A. H. Al-Marzouqi, Y. Haik, and M. Farid, “Effective extraction of microalgae lipids from wet biomass for biodiesel production,” *Biomass and Bioenergy*, vol. 66, pp. 159–167, 2014.

16. J. Ferrell and V. Sarisky-Reed, “National Algal Biofuels Technology Roadmap,” *U.S. Dep. Energy*, no. May, p. 140, 2010.

17. K. Y. Show and D. J. Lee, “Algal Biomass Harvesting,” in *Biofuels from Algae*, 2013, pp. 85–110.

18. M. Heasman, J. Diemar, W. O’Connor, T. Sushames, and L. Foulkes, “Development of extended shelf-life microalgae concentrate diets harvested by centrifugation for bivalve molluscs - a summary,” *Aquac. Res.*, vol. 31, no. 8–9, pp. 637–659, 2000.

19. L. Brennan and P. Owende, “Biofuels from microalgae-A review of technologies for production, processing, and extractions of biofuels and co-products,” *Renewable and Sustainable Energy Reviews*, vol. 14, no. 2. pp. 557–577, 2010.

20. A. Guldhe, B. Singh, I. Rawat, K. Ramluckan, and F. Bux, “Efficacy of drying and cell disruption techniques on lipid recovery from microalgae for biodiesel production,” *Fuel*, vol. 128, pp. 46–52, 2014.

21. R. K. Balasubramanian, T. T. Yen Doan, and J. P. Obbard, “Factors affecting cellular lipid extraction from marine microalgae,” *Chem. Eng. J.*, vol. 215–216, pp. 929–936, 2013.

22. A. Widjaja, C. C. Chien, and Y. H. Ju, “Study of increasing lipid production from fresh water microalgae Chlorella vulgaris,” *J. Taiwan Inst. Chem. Eng.*, vol. 40, no. 1, pp. 13–20, 2009.

23. M. Mubarak, A. Shaija, and T. V. Suchithra, “Ultrasonication: An effective pre-treatment method for extracting lipid from Salvinia molesta for biodiesel production,” *Resour. Technol.*, vol. 2, no. 3, pp. 126–132, 2016.

24. J. Kim *et al.*, “Methods of downstream processing for the production of biodiesel from microalgae,” *Biotechnology Advances*, vol. 31, no. 6. pp. 862–876, 2013.

25. G. S. Araujo *et al.*, “Extraction of lipids from microalgae by ultrasound application: Prospection of the optimal extraction method,” *Ultrason. Sonochem.*, vol. 20, no. 1, pp. 95–98, 2013.

26. W. Lu, Z. Wang, and Z. Yuan, “Characteristics of lipid extraction from Chlorella sp. cultivated in outdoor raceway ponds with mixture of ethyl acetate and ethanol for biodiesel production,” *Bioresour. Technol.*, vol. 191, pp. 433–437, 2015.

27. R. Halim, B. Gladman, M. K. Danquah, and P. a. Webley, “Oil extraction from microalgae for biodiesel production,” *Bioresour. Technol.*, vol. 102, no. 1, pp. 178–185, 2011.

28. K. Ramluckan, K. G. Moodley, and F. Bux, “An evaluation of the efficacy of using selected solvents for the extraction of lipids from algal biomass by the soxhlet extraction method,” *Fuel*, vol. 116, pp. 103–108, 2014.

29. Y. Li *et al.*, “A comparative study: the impact of different lipid extraction methods on current microalgal lipid research.,” *Microb. Cell Fact.*, vol. 13, no. 1, p. 14, 2014.

30. M. D. Macías-Sánchez, C. Mantell, M. Rodríguez, E. Martínez de la Ossa, L. M. Lubián, and O. Montero, “Supercritical fluid extraction of carotenoids and chlorophyll a from Synechococcus sp.,” *J. Supercrit. Fluids*, vol. 39, no. 3, pp. 323–329, 2007.

31. J. Wu, M. A. Alam, Y. Pan, D. Huang, Z. Wang, and T. Wang, “Enhanced extraction of lipids from microalgae with eco-friendly mixture of methanol and ethyl acetate for biodiesel production,” *J. Taiwan Inst. Chem. Eng.*, vol. 71, pp. 323–329, 2017.

32. J. Sheehan and J. Sheehan, “A Look Back at the U.S. Department of Energy’s Aquatic Species Program—Biodiesel from Algae,” *Program*, no. July, 1998.

33. J. M. Lv, L. H. Cheng, X. H. Xu, L. Zhang, and H. L. Chen, “Enhanced lipid production of Chlorella vulgaris by adjustment of cultivation conditions,” *Bioresour. Technol.*, vol. 101, no. 17, pp. 6797–6804, 2010.

34. S.-H. Ho, X. Ye, T. Hasunuma, J.-S. Chang, and A. Kondo, “Perspectives on engineering strategies for improving biofuel production from microalgae — A critical review,” *Biotechnol. Adv.*, vol. 32, no. 8, pp. 1448–1459, 2014.

35. E. M. Fakhry and D. M. El Maghraby, “Fatty Acids Composition and Biodiesel Characterization of Dunaliella salina,” *J. Water Resour. Prot.*, vol. 2013, no. September, pp. 894–899, 2013.

36. Y. M. Gong and M. L. Jiang, “Biotechnology Letters” *Biodiesel production with microalgae as feedstock: from strains to biodiesel*, vol. 33, no. 7. pp. 1269–1284, 2011.

37. T. T. Y. Doan, B. Sivaloganathan, and J. P. Obbard, “Screening of marine microalgae for biodiesel feedstock,” *Biomass and Bioenergy*, vol. 35, no. 7, pp. 2534–2544, 2011.

38. L. Wang *et al.*, “Anaerobic digested dairy manure as a nutrient supplement for cultivation of oil-rich green microalgae Chlorella sp.,” *Bioresour. Technol.*, vol. 101, no. 8, pp. 2623–2628, 2010.

39. F. Han, H. Pei, W. Hu, M. Song, G. Ma, and R. Pei, “Optimization and lipid production enhancement of microalgae culture by efficiently changing the conditions along with the growth-state,” *Energy Convers. Manag.*, vol. 90, pp. 315–322, 2015.

40. M. J. Griffiths and S. T. L. Harrison, “Lipid productivity as a key characteristic for choosing algal species for biodiesel production,” *J. Appl. Phycol.*, vol. 21, no. 5, pp. 493–507, 2009.

41. F. Qi, H. Pei, G. Ma, S. Zhang, and R. Mu, “Improving productivity and quality of biodiesel from Chlorella vulgaris SDEC-3M through customized process designs,” *Energy Convers. Manag.*, vol. 129, pp. 100–107, 2016.

42. E. Ono and J. L. Cuello, “Design parameters of solar concentrating systems for CO2-mitigating algal photobioreactors,” *Energy*, vol. 29, no. 9–10, pp. 1651–1657, 2004.

43. S. Y. Chiu, C. Y. Kao, C. H. Chen, T. C. Kuan, S. C. Ong, and C. S. Lin, “Reduction of CO2 by a high-density culture of Chlorella sp. in a semicontinuous photobioreactor,” *Bioresour. Technol.*, vol. 99, no. 9, pp. 3389–3396, 2008.

44. P. S. Lau, N. F. Y. Tam, and Y. S. Wong, “Effect of algal density on nutrient removal from primary settled wastewater,” *Environ. Pollut.*, vol. 89, no. 1, pp. 59–66, 1995.

45. P. Feng, Z. Deng, Z. Hu, Z. Wang, and L. Fan, “Characterization of Chlorococcum pamirum as a potential biodiesel feedstock,” *Bioresour. Technol.*, vol. 162, pp. 115–122, 2014.

46. C. H. Tan *et al.*, "Strategies for enhancing lipid production from indigenous microalgae isolates," *J. Taiwan Inst. Chem. Eng.*, vol. 63, pp. 189–194, 2016.

47. C. Chellamboli and M. Perumalsamy, "Application of response surface methodology for optimization of growth and lipids in Scenedesmus abundans using batch culture system," *RSC Adv.*, vol. 4, no. 42, pp. 22129–22140, 2014.

48. A. Juneja, R. M. Ceballos, and G. S. Murthy, "Effects of environmental factors and nutrient availability on the biochemical composition of algae for biofuels production: A review," *Energies*, vol. 6, no. 9. pp. 4607–4638, 2013.

49. H. Y. Ren, B. F. Liu, C. Ma, L. Zhao, and N. Q. Ren, "A new lipid-rich microalga Scenedesmus sp. strain R-16 isolated using Nile red staining: Effects of carbon and nitrogen sources and initial pH on the biomass and lipid production," *Biotechnol. Biofuels*, vol. 6, no. 1, pp. 1–10, 2013.

50. R. Tripathi, J. Singh, and I. S. Thakur, "Characterization of microalga Scenedesmus sp. ISTGA1 for potential CO2sequestration and biodiesel production," *Renew. Energy*, vol. 74, pp. 774–781, 2015.

51. S. K. Mandotra, P. Kumar, M. R. Suseela, S. Nayaka, and P. W. Ramteke, "Evaluation of fatty acid profile and biodiesel properties of microalga Scenedesmus abundans under the influence of phosphorus, pH and light intensities," *Bioresour. Technol.*, vol. 201, pp. 222–229, 2016.

52. F. J. L. Gordillo, M. Goutx, F. L. Figueroa, and F. X. Niell, "Effects of light intensity, CO2 and nitrogen supply on lipid class composition of Dunaliella viridis," *J. Appl. Phycol.*, vol. 10, no. 2, pp. 135–144, 1998.

53. S. F. Mohsenpour and N. Willoughby, "Effect of CO2aeration on cultivation of microalgae in luminescent photobioreactors," *Biomass and Bioenergy*, vol. 85, pp. 168–177, 2016.

54. G. A. Dunstan, J. K. Volkman, S. M. Barrett, and C. D. Garland, “Changes in the lipid composition and maximisation of the polyunsaturated fatty acid content of three microalgae grown in mass culture,” *J. Appl. Phycol.*, vol. 5, no. 1, pp. 71–83, 1993.

55. N. Arora, A. Patel, P. a. Pruthi, and V. Pruthi, “Synergistic dynamics of nitrogen and phosphorous influences lipid productivity in Chlorella minutissima for biodiesel production,” *Bioresour. Technol.*, vol. 213, pp. 79–87, 2015.

56. C. J. Zhu, Y. K. Lee, and T. M. Chao, “Effects of temperature and growth phase on lipid and biochemical composition of Isochrysis galbana TK1,” no. 1989, pp. 451–457, 1997.

57. R. Huerlimann, R. de Nys, and K. Heimann, “Growth, lipid content, productivity, and fatty acid composition of tropical microalgae for scale-up production,” *Biotechnol. Bioeng.*, vol. 107, no. 2, pp. 245–257, 2010.

58. J. Singh, R. Tripathi, and I. S. Thakur, “Characterization of endolithic cyanobacterial strain, Leptolyngbya sp. ISTCY101, for prospective recycling of CO2and biodiesel production,” *Bioresour. Technol.*, vol. 166, pp. 345–352, 2014.

59. I. A. Nascimento, I. T. D. Cabanelas, J. N. Dos Santos, M. a. Nascimento, L. Sousa, and G. Sansone, “Biodiesel yields and fuel quality as criteria for algal-feedstock selection: Effects of CO2-supplementation and nutrient levels in cultures,” *Algal Res.*, vol. 8, pp. 53–60, 2015.

60. C. M. Gatenby, D. M. Orcutt, D. a. Kreeger, B. C. Parker, V. a. Jones, and R. J. Neves, “Biochemical composition of three algal species proposed as food for captive freshwater mussels,” *J. Appl. Phycol.*, vol. 15, no. 1, pp. 1–11, 2003.

61. C. Y. Chen, K. L. Yeh, R. Aisyah, D. J. Lee, and J. S. Chang, “Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: A critical review,” *Bioresour. Technol.*, vol. 102, no. 1, pp. 71–81, 2011.

62. B. Cheirsilp and S. Torpee, “Enhanced growth and lipid production of microalgae under mixotrophic culture condition: Effect of light intensity, glucose concentration and fed-batch cultivation,” *Bioresour. Technol.*, vol. 110, pp. 510–516, 2012.

63. Y. Liang, N. Sarkany, and Y. Cui, “Biomass and lipid productivities of Chlorella vulgaris under autotrophic, heterotrophic and mixotrophic growth conditions,” *Biotechnol. Lett.*, vol. 31, no. 7, pp. 1043–1049, 2009.

64. X. M. Shi, F. Chen, J. P. Yuan, and H. Chen, “Heterotrophic production of lutein by selected Chlorella strains,” in *Journal of Applied Phycology*, 1997, vol. 9, no. 5, pp. 445–450.

65. Y. Zheng, Z. Chi, B. Lucker, and S. Chen, “Two-stage heterotrophic and phototrophic culture strategy for algal biomass and lipid production,” *Bioresour. Technol.*, vol. 103, no. 1, pp. 484–488, 2012.

66. K. Chojnacka and A. Noworyta, “Evaluation of Spirulina sp. growth in photoautotrophic, heterotrophic and mixotrophic cultures,” *Enzyme Microb. Technol.*, vol. 34, no. 5, pp. 461–465, 2004.

67. G. Belotti, M. Bravi, B. De Caprariis, P. De Filippis, and M. Scarsella, “Effect of Nitrogen and Phosphorus Starvations on Chlorella vulgaris Lipids Productivity and Quality under Different Trophic Regimens for Biodiesel Production,” *Am. J. Plant Sci.*, vol. 4, no. December, pp. 44–51, 2013.

68. S. K. Bagchi and N. Mallick, “Carbon dioxide biofixation and lipid accumulation potential of an indigenous microalga Scenedesmus obliquus (Turpin) Kützing GA 45 for biodiesel production,” *RSC Adv.*, vol. 6, no. 36, pp. 29889–29898, 2016.

69. M. Mondal *et al.*, “Influence of carbon sources and light intensity on biomass and lipid production of Chlorella sorokiniana BTA 9031 isolated from coalfield under various nutritional modes,” *Energy Convers. Manag.*, vol. 145, pp. 247–254, 2017.

70. C. Y. Wang, C. C. Fu, and Y. C. Liu, “Effects of using light-emitting diodes on the cultivation of Spirulina platensis,” *Biochem. Eng. J.*, vol. 37, no. 1, pp. 21–25, 2007.

71. C. L. Teo, M. Atta, A. Bukhari, M. Taisir, A. M. Yusuf, and A. Idris, “Enhancing growth and lipid production of marine microalgae for biodiesel production via the use of different LED wavelengths,” *Bioresour. Technol.*, vol. 162, pp. 38–44, 2014.

72. A. Blanchemain and D. Grizeau, “Eicosapentaenoic acid content of Skeletonema costatum as a function of growth and irradiance; relation with chlorophyll a content and photosynthetic capacity,” *J. Exp. Mar. Bio. Ecol.*, vol. 196, no. 1–2, pp. 177–188, 1996.

73. H. Tang, N. Abunasser, M. E. D. Garcia, M. Chen, K. Y. Simon Ng, and S. O. Salley, “Potential of microalgae oil from Dunaliella tertiolecta as a feedstock for biodiesel,” *Appl. Energy*, vol. 88, no. 10, pp. 3324–3330, 2011.

74. H. Tang, M. Chen, M. E. D. Garcia, N. Abunasser, K. Y. S. Ng, and S. O. Salley, “Culture of microalgae Chlorella minutissima for biodiesel feedstock production,” *Biotechnol. Bioeng.*, vol. 108, no. 10, pp. 2280–2287, 2011.

75. Q. He, H. Yang, L. Wu, and C. Hu, “Effect of light intensity on physiological changes, carbon allocation and neutral lipid accumulation in oleaginous microalgae,” *Bioresour. Technol.*, vol. 191, pp. 219–228, 2015.

76. S. Chakraborty, D. Mohanty, S. Ghosh, and D. Das, “Improvement of lipid content of Chlorella minutissima MCC 5 for biodiesel production,” *J. Biosci. Bioeng.*, vol. 122, no. 3, pp. 294–300, 2016.

77. T. Thawechai, B. Cheirsilp, Y. Louhasakul, P. Boonsawang, and P. Prasertsan, “Mitigation of carbon dioxide by oleaginous microalgae for lipids and pigments production: Effect of light illumination and carbon dioxide feeding strategies,” *Bioresour. Technol.*, vol. 219, pp. 139–149, 2016.

78. F. Li, S. C. Srivatsa, W. Batchelor, and S. Bhattacharya, “A study on growth and pyrolysis characteristics of microalgae using Thermogravimetric Analysis-Infrared Spectroscopy and synchrotron Fourier Transform Infrared Spectroscopy,” *Bioresour. Technol.*, vol. 229, pp. 1–10, 2017.

79. C. Y. Chen, K. L. Yeh, H. M. Su, Y. C. Lo, W. M. Chen, and J. S. Chang, “Strategies to enhance cell growth and achieve high-level oil production of a Chlorella vulgaris isolate,” *Biotechnol. Prog.*, vol. 26, no. 3, pp. 679–686, 2010.

80. I. Krzemińska, A. Nawrocka, A. Piasecka, P. Jagielski, and J. Tys, “Cultivation of Chlorella protothecoides in photobioreactors: The combined impact of photoperiod and CO2 concentration,” *Eng. Life Sci.*, vol. 15, no. 5, pp. 533–541, 2015.

81. B. Wang, Y. Li, N. Wu, and C. Q. Lan, “CO2 bio-mitigation using microalgae,” *Appl. Microbiol. Biotechnol.*, vol. 79, no. 5, pp. 707–718, 2008.

82. L. M. Brown, “Uptake of carbon dioxide from flue gas by microalgae,” *Energy Convers. Manag.*, vol. 37, no. 6–8, pp. 1363–1367, 1996.

83. H. Hu and K. Gao, “Optimization of growth and fatty acid composition of a unicellular marine picoplankton, *Nannochloropsis* sp., with enriched carbon sources.,” *Biotechnol. Lett.*, vol. 25, no. 5, pp. 421–5, 2003.

84. H. T. Hsueh, W. J. Li, H. H. Chen, and H. Chu, “Carbon bio-fixation by photosynthesis of Thermosynechococcus sp. CL-1 and Nannochloropsis oculta,” *J. Photochem. Photobiol. B Biol.*, vol. 95, no. 1, pp. 33–39, 2009.

85. S. Y. Chiu, C. Y. Kao, M. T. Tsai, S. C. Ong, C. H. Chen, and C. S. Lin, “Lipid accumulation and CO2 utilization of Nannochloropsis oculata in response to CO2 aeration,” *Bioresour. Technol.*, vol. 100, no. 2, pp. 833–838, 2009.

86. E. Y. Ortiz Montoya, A. a. Casazza, B. Aliakbarian, P. Perego, A. Converti, and J. C. M. De Carvalho, “Production of Chlorella vulgaris as a source of essential fatty acids

in a tubular photobioreactor continuously fed with air enriched with CO2 at different concentrations," *Biotechnol. Prog.*, vol. 30, no. 4, pp. 916–922, 2014.

87. S. Ghosh, S. Roy, and D. Das, "Improvement of Biomass Production by Chlorella sp. MJ 11/11 for Use as a Feedstock for Biodiesel," *Appl. Biochem. Biotechnol.*, vol. 175, no. 7, pp. 3322–3335, 2015.

88. S. K. Bagchi and N. Mallick, "Carbon dioxide biofixation and lipid accumulation potential of an indigenous microalga Scenedesmus obliquus (Turpin) Kützing GA 45 for biodiesel production," *RSC Adv.*, vol. 6, no. 36, pp. 29889–29898, 2016.

89. M. Mondal *et al.*, "Mixotrophic cultivation of Chlorella sp. BTA 9031 and Chlamydomonas sp. BTA 9032 isolated from coal field using various carbon sources for biodiesel production," *Energy Convers. Manag.*, vol. 124, pp. 297–304, 2016.

90. C. Yusuf, "Biodiesel from microalgae," *Biotechnol. Adv.*, vol. 25, no. 3, pp. 294–306, 2007.

91. M. Ghafari, B. Rashidi, and B. Z. Haznedaroglu, "Effects of macro and micronutrients on neutral lipid accumulation in oleaginous microalgae," *Biofuels*, vol. 9, no. 2, pp. 147–156, 2018.

92. A. Richmond, "Handbook of microalgal culture: biotechnology and applied phycology/edited by Amos Richmond.," *Orton.Catie.Ac.Cr*, p. 472, 2004.

93. Y. Li, M. Horsman, B. Wang, N. and Wu, and C. Q. Lan, "Effects of nitrogen sources on cell growth and lipid accumulation of green alga Neochloris oleoabundans," *Appl. Microbiol. Biotechnol.*, vol. 81, no. 4, pp. 629–636, 2008.

94. L. F. Wu, P. C. Chen, and C. M. Lee, "The effects of nitrogen sources and temperature on cell growth and lipid accumulation of microalgae," *Int. Biodeterior. Biodegradation*, vol. 85, pp. 506–510, 2013.

95. M. Ren and K. Ogden, “Cultivation of Nannochloropsis gaditana on mixtures of nitrogen sources,” *Environ. Prog. Sustain. Energy*, vol. 33, no. 2, pp. 551–555, 2014.

96. M. Muthuraj, V. Kumar, B. Palabhanvi, and D. Das, “Evaluation of indigenous microalgal isolate Chlorella sp. FC2 IITG as a cell factory for biodiesel production and scale up in outdoor conditions,” *J. Ind. Microbiol. Biotechnol.*, vol. 41, no. 3, pp. 499–511, 2014.

97. A. Roopnarain, S. Sym, and V. M. Gray, “Effect of nitrogenous resource on growth, biochemical composition and ultrastructure of Isochrysis galbana (Isochrysidales, Haptophyta),” *Phycol. Res.*, vol. 63, no. 1, pp. 43–50, 2015.

98. A. Converti, A. a. Casazza, E. Y. Ortiz, P. Perego, and M. Del Borghi, “Effect of temperature and nitrogen concentration on the growth and lipid content of Nannochloropsis oculata and Chlorella vulgaris for biodiesel production,” *Chem. Eng. Process. Process Intensif.*, vol. 48, no. 6, pp. 1146–1151, 2009.

99. H. Wu and X. Miao, “Biodiesel quality and biochemical changes of microalgae Chlorella pyrenoidosa and Scenedesmus obliquus in response to nitrate levels,” *Bioresour. Technol.*, vol. 170, pp. 421–427, 2014.

100. I. Pancha *et al.*, “Nitrogen stress triggered biochemical and morphological changes in the microalgae Scenedesmus sp. CCNM 1077,” *Bioresour. Technol.*, vol. 156, pp. 146–154, 2014.

101. C. Li, Y. Yu, D. Zhang, J. Liu, N. Ren, and Y. Feng, “Combined effects of carbon, phosphorus and nitrogen on lipid accumulation of *Chlorella vulgaris* in mixotrophic culture,” *J. Chem. Technol. Biotechnol.*, vol. 91, no. 3, pp. 680–684, 2016.

102. S. Chakraborty, D. Mohanty, S. Ghosh, and D. Das, “Improvement of lipid content of Chlorella minutissima MCC 5 for biodiesel production,” *J. Biosci. Bioeng.*, vol. 122, no. 3, pp. 294–300, 2016.

103. M. Mondal, A. Ghosh, G. Oinam, O. N. Tiwari, K. Gayen, and G. N. Halder, "Biochemical responses to bicarbonate supplementation on biomass and lipid productivity of Chlorella Sp. BTA9031 isolated from Coalmine area," *Environ. Prog. Sustain. Energy*, vol. 36, no. 5, pp. 1498–1506, 2017.

104. R. Sivaramakrishnan and A. Incharoensakdi, "Enhancement of total lipid yield by nitrogen, carbon, and iron supplementation in isolated microalgae," *J. Phycol.*, vol. 53, no. 4, pp. 855–868, 2017.

105. K. Liang, Q. Zhang, M. Gu, and W. Cong, "Effect of phosphorus on lipid accumulation in freshwater microalga *Chlorella* sp.," *J. Appl. Phycol.*, vol. 25, no. 1, pp. 311–318, 2013.

106. S. Mandal and N. Mallick, "Microalga Scenedesmus obliquus as a potential source for biodiesel production," *Appl. Microbiol. Biotechnol.*, vol. 84, no. 2, pp. 281–291, 2009.

107. X. Qv, Y. Guo, and J. Jiang, "Assessment of the e ff ects of nutrients on biomass and lipid accumulation in Dunaliella tertiolecta using a response surface methodology," *RSC Adv.*, vol. 4, pp. 42202–42210, 2014.

108. L. Fu *et al.*, "Excessive phosphorus enhances Chlorella regularis lipid production under nitrogen starvation stress during glucose heterotrophic cultivation," *Chem. Eng. J.*, vol. 330, no. August, pp. 566–572, 2017.

109. S. Park and Y. Li, "Integration of biological kinetics and computational fluid dynamics to model the growth of Nannochloropsis salina in an open channel raceway," *Biotechnol. Bioeng.*, vol. 112, no. 5, pp. 923–933, 2015.

110. O. Pulz, "Photobioreactors: Production systems for phototrophic microorganisms," *Applied Microbiology and Biotechnology*, vol. 57, no. 3. pp. 287–293, 2001.

111. T. Luangpipat and Y. Chisti, “Biomass and oil production by Chlorella vulgaris and four other microalgae — Effects of salinity and other factors,” *J. Biotechnol.*, vol. 257, pp. 47–57, 2017.

112. S. Bouyam, W. Choorit, S. Sirisansaneeyakul, and Y. Chisti, “Heterotrophic production of Chlorella sp. TISTR 8990—biomass growth and composition under various production conditions,” *Biotechnol. Prog.*, vol. 33, no. 6, pp. 1589–1600, 2017.

113. M. A. Borowitzka and L. J. Borowitzka, *Micro-Algal Biotechnology*. 1988.

114. D. Pal, I. Khozin-Goldberg, Z. Cohen, and S. Boussiba, “The effect of light, salinity, and nitrogen availability on lipid production by Nannochloropsis sp.,” *Appl. Microbiol. Biotechnol.*, vol. 90, no. 4, pp. 1429–1441, 2011.

115. N. Arora *et al.*, “Insights into the Enhanced Lipid Production Characteristics of a Fresh Water Microalga under High Salinity Conditions,” *Ind. Eng. Chem. Res.*, vol. 56, no. 25, pp. 7413–7421, 2017.

116. K. Chokshi, I. Pancha, A. Ghosh, and S. Mishra, “Salinity induced oxidative stress alters the physiological responses and improves the biofuel potential of green microalgae Acutodesmus dimorphus,” *Bioresour. Technol.*, vol. 244, pp. 1376–1383, 2017.

117. M. Takagi, Karseno, and T. Yoshida, “Effect of salt concentration on intracellular accumulation of lipids and triacylglyceride in marine microalgae Dunaliella cells,” *J. Biosci. Bioeng.*, vol. 101, no. 3, pp. 223–226, 2006.

118. P. Kaewkannetra, P. Enmak, and T. Chiu, “The effect of CO2 and salinity on the cultivation of Scenedesmus obliquus for biodiesel production,” *Biotechnol. Bioprocess Eng.*, vol. 17, no. 3, pp. 591–597, 2012.

119. S. M. Renaud and D. L. Parry, “Microalgae for use in tropical aquaculture II: Effect of salinity on growth, gross chemical composition and fatty acid composition of three species of marine microalgae,” *J. Appl. Phycol.*, vol. 6, no. 3, pp. 347–356, 1994.

120. a. J. Martínez-Roldán, H. V. Perales-Vela, R. O. Cañizares-Villanueva, and G. Torzillo, “Physiological response of Nannochloropsis sp. to saline stress in laboratory batch cultures,” *J. Appl. Phycol.*, vol. 26, no. 1, pp. 115–121, 2014.

121. I. BenMoussa-Dahmen, H. Chtourou, F. Rezgui, S. Sayadi, and A. Dhouib, “Salinity stress increases lipid, secondary metabolites and enzyme activity in Amphora subtropica and Dunaliella sp. for biodiesel production,” *Bioresour. Technol.*, vol. 218, pp. 816–825, 2016.

122. R. A. Ahmed *et al.*, “Bioenergy application of Dunaliella salina SA 134 grown at various salinity levels for lipid production,” *Sci. Rep.*, vol. 7, no. 1, 2017.

123. A. Solovchenko, A. Lukyanov, O. Solovchenko, S. Didi-Cohen, S. Boussiba, and I. Khozin-Goldberg, “Interactive effects of salinity, high light, and nitrogen starvation on fatty acid and carotenoid profiles in Nannochloropsis oceanica CCALA 804,” *Eur. J. Lipid Sci. Technol.*, vol. 116, no. 5, pp. 635–644, 2014.

124. M. H. Liang, X. Y. Qv, H. Chen, Q. Wang, and J. G. Jiang, “Effects of Salt Concentrations and Nitrogen and Phosphorus Starvations on Neutral Lipid Contents in the Green Microalga Dunaliella tertiolecta,” *J. Agric. Food Chem.*, vol. 65, no. 15, pp. 3190–3197, 2017.

125. P. C. Gorain, S. K. Bagchi, and N. Mallick, “Effects of calcium, magnesium and sodium chloride in enhancing lipid accumulation in two green microalgae,” *Environ. Technol.*, vol. 34, no. 13–14, pp. 1887–1894, 2013.

126. R. Karpagam, K. J. Raj, B. Ashokkumar, and P. Varalakshmi, “Characterization and fatty acid profiling in two fresh water microalgae for biodiesel production: Lipid enhancement methods and media optimization using response surface methodology,” *Bioresour. Technol.*, vol. 188, pp. 177–184, 2015.

127. I. Pancha *et al.*, “Salinity induced oxidative stress enhanced biofuel production potential of microalgae Scenedesmus sp. CCNM 1077,” *Bioresour. Technol.*, vol. 189, pp. 341–348, 2015.

128. A. Ranjan, C. Patil, and V. S. Moholkar, “Mechanistic assessment of microalgal lipid extraction,” *Ind. Eng. Chem. Res.*, vol. 49, no. 6, pp. 2979–2985, 2010.

129. J. M. Bermúdez Menéndez *et al.*, “Optimization of microalgae oil extraction under ultrasound and microwave irradiation,” *J. Chem. Technol. Biotechnol.*, vol. 89, no. 11, pp. 1779–1784, 2014.

130. K. Servaes, M. Maesen, B. Prandi, S. Sforza, and K. Elst, “Polar Lipid Profile of Nannochloropsis oculata Determined Using a Variety of Lipid Extraction Procedures,” *J. Agric. Food Chem.*, vol. 63, no. 15, pp. 3931–3941, 2015.

131. J. Y. Lee, C. Yoo, S. Y. Jun, C. Y. Ahn, and H. M. Oh, “Comparison of several methods for effective lipid extraction from microalgae,” *Bioresour. Technol.*, vol. 101, no. 1 SUPPL., pp. S75–S77, 2010.

132. P. Prabakaran and a. D. Ravindran, “A comparative study on effective cell disruption methods for lipid extraction from microalgae,” *Lett. Appl. Microbiol.*, vol. 53, no. 2, pp. 150–154, 2011.

133. H. Zheng, J. Yin, Z. Gao, H. Huang, X. Ji, and C. Dou, “Disruption of chlorella vulgaris cells for the release of biodiesel-producing lipids: A comparison of grinding, ultrasonication, bead milling, enzymatic lysis, and microwaves,” *Appl. Biochem. Biotechnol.*, vol. 164, no. 7, pp. 1215–1224, 2011.

134. S. Rakesh *et al.*, “Cell disruption methods for improving lipid extraction efficiency in unicellular microalgae,” *Eng. Life Sci.*, vol. 15, no. 4, pp. 443–447, 2015.

135. E. Lorente, X. Farriol, and J. Salvadó, “Steam explosion as a fractionation step in biofuel production from microalgae,” *Fuel Process. Technol.*, vol. 131, pp. 93–98, 2015.

136. E. Alpaydin, *Introduction to Machine Learning*. 2004.

137. R. Karpagam, R. Preeti, B. Ashokkumar, and P. Varalakshmi, “Enhancement of lipid production and fatty acid profiling in Chlamydomonas reinhardtii, CC1010 for biodiesel production,” *Ecotoxicol. Environ. Saf.*, vol. 121, pp. 253–257, 2015.

138. C. J. Zhu, Y. K. Lee, and T. M. Chao, “Effects of temperature and growth phase on lipid and biochemical composition of Isochrysis galbana TK1,” in *Journal of Applied Phycology*, 1997, vol. 9, no. 5, pp. 451–457.

139. Z. Demirel, E. Imamoglu, and M. Conk Dalay, “Fatty acid profile and lipid content of Cylindrotheca closterium cultivated in air-lift photobioreactor,” *J. Chem. Technol. Biotechnol.*, vol. 90, no. 12, pp. 2290–2296, 2015.

140. A. M. Illman, A. H. Scragg, and S. W. Shales, “Increase in Chlorella strains calorific values when grown in low nitrogen medium,” *Enzyme Microb. Technol.*, vol. 27, no. 8, pp. 631–635, 2000.

141. M. Song, H. Pei, W. Hu, and G. Ma, “Evaluation of the potential of 10 microalgal strains for biodiesel production,” *Bioresour. Technol.*, vol. 141, pp. 245–251, 2013.

142. M. Takagi, K. Watanabe, K. Yamaberi, and T. Yoshida, “Limited feeding of potassium nitrate for intracellular lipid and triglyceride accumulation of Nannochloris sp. UTEX LB1999,” *Appl. Microbiol. Biotechnol.*, vol. 54, no. 1, pp. 112–117, 2000.

143. S. K. Ratha, S. Babu, N. Renuka, R. Prasanna, R. B. N. Prasad, and A. K. Saxena, “Exploring nutritional modes of cultivation for enhancing lipid accumulation in microalgae,” *J. Basic Microbiol.*, vol. 53, no. 5, pp. 440–450, 2013.

144. S. M. Renaud, L.-V. Thinh, G. Lambrinidis, and D. L. Parry, “Effect of temperature on growth, chemical composition and fatty acid composition of tropical Australian microalgae grown in batch cultures,” *Aquaculture*, vol. 211, no. 1–4, pp. 195–214, 2002.

145. X. P. Zhou, L. Xia, H. M. Ge, D. L. Zhang, and C. X. Hu, “Feasibility of biodiesel production by microalgae Chlorella sp. (FACHB-1748) under outdoor conditions,” *Bioresour. Technol.*, vol. 138, pp. 131–135, 2013.

146. C. Welter, J. Schwenk, B. Kanani, J. Van Blargan, and J. M. Belovich, “Minimal medium for optimal growth and lipid production of the microalgae Scenedesmus dimorphus,” *Environ. Prog. Sustain. Energy*, vol. 32, no. 4, pp. 937–945, 2013.

147. H. Hu and K. Gaol, “Optimization of growth and fatty acid composition of a unicellular marine picoplankton, Nannochloropsis sp., with enriched carbon sources,” *Biotechnol. Lett.*, vol. 25, no. 5, pp. 421–425, 2003.

148. Y. Gao, M. Yang, and C. Wang, “Nutrient deprivation enhances lipid content in marine microalgae,” *Bioresour. Technol.*, vol. 147, pp. 484–491, 2013.

149. Y. Huang, J. Cheng, H. Lu, R. Huang, J. Zhou, and K. Cen, “Simultaneous enhancement of microalgae biomass growth and lipid accumulation under continuous aeration with 15% CO 2,” *RSC Adv.*, vol. 5, no. 63, pp. 50851–50858, 2015.

150. F. F. Chu, P. N. Chu, P. J. Cai, W. W. Li, P. K. S. Lam, and R. J. Zeng, “Phosphorus plays an important role in enhancing biodiesel productivity of Chlorella vulgaris under nitrogen deficiency,” *Bioresour. Technol.*, vol. 134, pp. 341–346, 2013.

151. S. P. Slocombe *et al.*, “Unlocking nature’s treasure-chest: Screening for oleaginous algae,” *Sci. Rep.*, vol. 5, pp. 1–17, 2015.

152. I. A. Nascimento *et al.*, “Screening Microalgae Strains for Biodiesel Production: Lipid Productivity and Estimation of Fuel Quality Based on Fatty Acids Profiles as Selective Criteria,” *Bioenergy Res.*, vol. 6, no. 1, pp. 1–13, 2013.

153. C. Li, Y. Yu, D. Zhang, J. Liu, N. Ren, and Y. Feng, “Combined effects of carbon, phosphorus and nitrogen on lipid accumulation of Chlorella vulgaris in mixotrophic culture,” *J. Chem. Technol. Biotechnol.*, no. August 2014, 2015.

154. C. L. Teo, H. Jamaluddin, N. A. M. Zain, and A. Idris, “Biodiesel production via lipase catalysed transesterification of microalgae lipids from Tetraselmis sp.,” *Renew. Energy*, vol. 68, pp. 1–5, 2014.

155. T. T. Nguyen, X. T. Bui, M. D. Pham, W. Guo, and H. H. Ngo, “Effect of Tris-(hydroxymethyl)-amino methane on microalgae biomass growth in a photobioreactor,” *Bioresour. Technol.*, vol. 208, pp. 1–6, 2016.

156. Y. Tang, Y. Zhang, J. N. Rosenberg, N. Sharif, M. J. Betenbaugh, and F. Wang, “Efficient lipid extraction and quantification of fatty acids from algal biomass using accelerated solvent extraction (ASE),” *RSC Adv.*, vol. 6, no. 35, pp. 29127–29134, 2016.

157. A. E. M. Abdelaziz, D. Ghosh, and P. C. Hallenbeck, “Characterization of growth and lipid production by Chlorella sp. PCH90, a microalga native to Quebec,” *Bioresour. Technol.*, vol. 156, pp. 20–28, 2014.

158. Y. Du, B. Schuur, S. R. a. Kersten, and D. W. F. Brilman, “Microalgae wet extraction using N-ethyl butylamine for fatty acid production,” *Green Energy Environ.*, vol. 1, no. 1, pp. 79–83, 2016.

159. Y. Ma, Z. Wang, C. Yu, Y. Yin, and G. Zhou, “Evaluation of the potential of 9 Nannochloropsis strains for biodiesel production,” *Bioresour. Technol.*, vol. 167, pp. 503–509, 2014.

160. R. a. I. Abou-Shanab, I. a. Matter, S. N. Kim, Y. K. Oh, J. Choi, and B. H. Jeon, “Characterization and identification of lipid-producing microalgae species isolated from a freshwater lake,” *Biomass and Bioenergy*, vol. 35, no. 7, pp. 3079–3085, 2011.

161. L. Xin, H. Hong-ying, and Z. Yu-ping, “Growth and lipid accumulation properties of a freshwater microalga Scenedesmus sp. under different cultivation temperature,” *Bioresour. Technol.*, vol. 102, no. 3, pp. 3098–3102, 2011.

162. B. Wang and C. Q. Lan, “Optimising the lipid production of the green alga Neochloris oleoabundans using box-behnken experimental design,” *Can. J. Chem. Eng.*, vol. 89, no. 4, pp. 932–939, 2011.

163. L. Xia, S. Song, Q. He, H. Yang, and C. Hu, “Selection of microalgae for biodiesel production in a scalable outdoor photobioreactor in north China,” *Bioresour. Technol.*, vol. 174, pp. 274–280, 2014.

164. S. H. Ho, C. Y. Chen, and J. S. Chang, “Effect of light intensity and nitrogen starvation on CO_2 fixation and lipid/carbohydrate production of an indigenous microalga *Scenedesmus oliquus* CNW-N,” *Bioresour. Technol.*, vol. 113, pp. 244–252, 2012.

165. S. Vidyashankar *et al.*, “Characterization of fatty acids and hydrocarbons of chlorophycean microalgae towards their use as biofuel source,” *Biomass and Bioenergy*, vol. 77, pp. 75–91, 2015.

166. J. Liu, C. Yuan, G. Hu, and F. Li, “Effects of light intensity on the growth and lipid accumulation of microalga Scenedesmus sp. 11-1 under nitrogen limitation,” *Appl. Biochem. Biotechnol.*, vol. 166, no. 8, pp. 2127–2137, 2012.

167. R. Valdez-Ojeda *et al.*, “Characterization of five fresh water microalgae with potential for biodiesel production,” *Algal Res.*, vol. 7, pp. 33–44, 2015.

168. N. Mallick, S. Mandal, A. K. Singh, M. Bishai, and A. Dash, “Green microalga Chlorella vulgaris as a potential feedstock for biodiesel,” *J. Chem. Technol. Biotechnol.*, vol. 87, no. 1, pp. 137–145, 2012.

169. L. F. Wu, P. C. Chen, A. P. Huang, and C. M. Lee, “The feasibility of biodiesel production by microalgae using industrial wastewater,” *Bioresour. Technol.*, vol. 113, pp. 14–18, 2012.

170. C. H. Ra, C. H. Kang, N. K. Kim, C. G. Lee, and S. K. Kim, “Cultivation of four microalgae for biomass and oil production using a two-stage culture strategy with salt stress,” *Renew. Energy*, vol. 80, pp. 117–122, 2015.

171. M. Sakarika and M. Kornaros, “Kinetics of growth and lipids accumulation in Chlorella vulgaris during batch heterotrophic cultivation: Effect of different nutrient limitation strategies,” *Bioresour. Technol.*, vol. 243, pp. 356–365, 2017.

172. M. Bekirogullari, I. S. Fragkopoulos, J. K. Pittman, and C. Theodoropoulos, “Production of lipid-based fuels and chemicals from microalgae: An integrated experimental and model-based optimization study,” *Algal Res.*, vol. 23, pp. 78–87, 2017.

173. C. Chellamboli and M. Perumalsamy, “Application of response surface methodology for optimization of growth and lipids in Scenedesmus abundans using batch culture system,” *RSC Adv.*, vol. 4, no. 42, pp. 22129–22140, 2014.

174. D. Frumento, A. A. Casazza, S. Al Arni, and A. Converti, “Cultivation of Chlorella vulgaris in tubular photobioreactors: A lipid source for biodiesel production,” *Biochem. Eng. J.*, vol. 81, pp. 120–125, 2013.

APPENDIX A: ARTICLES INVOLVED IN DATABASE

Table A.1. Articles Involved in Database

Number of Articles	Reference	Number of Articles	Reference	Number of Articles	Reference
1	[119]	37	[67]	72	[137]
2	[138]	38	[125]	73	[139]
3	[140]	39	[141]	74	[87]
4	[142]	40	[143]	75	[39]
5	[144]	41	[145]	76	[130]
6	[60]	42	[146]	77	[127]
7	[147]	43	[148]	78	[149]
8	[93]	44	[129]	79	[135]
9	[98]	45	[150]	80	[151]
10	[85]	46	[152]	81	[88]
11	[106]	47	[94]	82	[153]
12	[4]	48	[107]	83	[53]
13	[22]	49	[154]	84	[155]
14	[131]	50	[99]	85	[156]
15	[33]	51	[45]	86	[51]
16	[57]	52	[58]	87	[102]
17	[128]	53	[157]	88	[41]
18	[73]	54	[95]	89	[158]
19	[79]	55	[96]	90	[77]
20	[132]	56	[159]	91	[89]
21	[160]	57	[123]	92	[121]
22	[74]	58	[100]	93	[46]
23	[133]	59	[120]	94	[55]
24	[161]	60	[86]	95	[78]
25	[162]	61	[163]	96	[103]
26	[114]	62	[59]	97	[122]
27	[62]	63	[134]	98	[111]
28	[5]	64	[126]	99	[124]
29	[164]	65	[165]	100	[104]
30	[166]	66	[167]	101	[108]
31	[168]	67	[50]	102	[69]
32	[13]	68	[80]	103	[115]
33	[169]	69	[170]	104	[171]
34	[49]	70	[75]	105	[172]
35	[173]	71	[97]	106	[116]
36	[174]				