MACHINE LEARNING ANALYSIS OF DATA COLLECTED FROM PUBLISHED LITERATURE ON PHOTOCATALYTIC REFORMING OF GLYCEROL

by

Rüveyda Karakoyun

B.S., Chemical Engineering, Boğaziçi University, 2019

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

2023

ACKNOWLEDGEMENTS

This thesis was concluded with the support of many people. I am very grateful for their help.

At the beginning, I owe a debt of gratitude to my thesis supervisor, Prof. Ramazan Yıldırım for his wisdom, support, patience, and guidance during my thesis period. With his help and encouragement, I completed my graduate studies.

I would like to express my thanks to my thesis committee, Prof. Ahmet Erhan Aksoylu and Assoc. Prof. Mehmet Erdem Günay for their valuable time to evaluate my thesis.

Moreover, I would like to thank Burcu Oral for her guidance. She always helped me with her experience and knowledge when she was in the excessive busyness of her PhD studies.

I would like to express my deepest appreciation to my family. My husband, Ahmet Karakoyun, always supported and motivated me with all his heart. I would like to thank my beloved parents, my father Özcan Demir and my mother Fatma Demir, who raised me until these days and supported me with their endless love. Also, I am very thankful to my husband's parents Abdulkadir Karakoyun and Güllü Karakoyun. They always supported me during my studies with their kind hearts.

I am grateful to my siblings Emre Demir, Şule Demir, Yağmur Karakoyun, Eylül Nisa Karakoyun, and my sweet little niece Eslem Demir. They embellished my days and nights and kept me fully motivated.

ABSTRACT

MACHINE LEARNING ANALYSIS OF DATA COLLECTED FROM PUBLISHED LITERATURE ON PHOTOCATALYTIC REFORMING OF GLYCEROL

In this thesis, the aim is to extract knowledge from the data that was collected from published literature about photocatalytic reforming of glycerol. 791 data points were collected from 93 articles. This data was cleaned, organized, and prepared for the machine learning methods. Random forest and ANN (Artificial Neural Network) were used as machine learning techniques. By using them, the models for band gap and hydrogen production rates were constructed. Cross validation was applied to all models to prevent overfitting. For hydrogen production rate model, the missing values for band gap were filled with the predicted values of ANN of band gap. In random forest, feature importance was determined and the variables with the highest effect on the result were found. For band gap, the most important variables were weight percent of cocatalyst, percent of semiconductor and calcination temperature and duration. For hydrogen production rate, the most significant variables were photocatalyst load, band gap, glycerol concentration, weight percent of cocatalyst and pH. In random forest, the best model was determined by changing test/train split and k values in k-fold cross validation for various tree number and number of samples in a leaf node. For band gap model, 0.25 test/train split and 4-fold with 41 trees and 1 sample was the best model with RMSE (Root Mean Square Error) of 0.234 and R-squared of 0.73. For hydrogen production rate model, 0.25 test/train split and 5-fold with 81 trees and 2 samples was the best model with RMSE of 1.09×10^4 and R-squared value of 0.71. For ANN, test/train split ratio, k value for k-fold cross validation, the number of neurons and activation function were changed to find the best model. For band gap, 52 neurons and ReLU function gave the best model with RMSE of 0.282 and R-squared value of 0.70 with 0.3 test/train split and 4-fold cross validation. For hydrogen production rate model, 0.25 test/train split ratio, 7-fold cross validation, 63 neurons and ReLU function gave the best model with RMSE of 1.47×10^4 and R-squared value of 0.60.

ÖZET

GLİSEROLÜN FOTOKATALİTİK REFORMLAMASI HAKKINDA LİTERATÜRDEN TOPLANAN VERİLER İLE MAKİNE ÖĞRENMESİ ANALİZİ

Bu tezin amacı, gliserolün fotokatalitik reformlanması hakkında yayınlanmış literatürden toplanan verideki bilgiyi açığa çıkarmaktır. 93 makaleden 791 veri noktası toplanmıştır. Bu veriler makine öğrenmesi metotları için temizlenmiş, düzenlenmiş ve hazırlanmıştır. Rastgele orman ve YSA (Yapay Sinir Ağı) makine öğrenmesi teknikleri olarak kullanılmıştır. Bu teknikleri kullanarak, bant aralığı ve hidrojen üretim hızları için modeller oluşturulmuştur. Aşırı öğrenmeyi önlemek için tüm modellere çapraz doğrulama uygulanmıştır. Hidrojen üretim hızı modeli için, bant aralığında eksik olan değerler, bant aralığının YSA modelinde tahmin edilen değerler ile doldurulmuştur. Rastgele ormanda, özelliklerin önemi belirlenmiş ve sonuca en yüksek etkisi olan değişkenler bulunmuştur. Bant aralığı için en önemli olan değişkenler kokatalizörün ağırlık yüzdesi, yarıiletkenin yüzdesi ve kalsinasyon sıcaklığıdır. Hidrojen üretim hızı için en önemli değişkenler fotokatalizör yüklemesi, bant aralığı, gliserol konsantrasyonu, kokatalizörün ağırlık yüzdesi ve pH'dır. Rastgele ormanda, ağaç sayısı ve düğümün içerdiği örnek sayısı için k-katmanlı çapraz doğrulamada test/eğitim seti ayırma oranı ve k değerleri değiştirilerek en iyi model belirlenmiştir. Bant aralığı için 0.25 test/eğitim seti ayırma oranı, 4-katmanlı çapraz doğrulama, 41 ağaç ve 1 örnek, 0.234 RMSE ve 0.73 R-kare ile en iyi sonucu vermiştir. Hidrojen üretim hızı modeli için ise 0.25 test/eğitim seti ayırma oranı, 5-katmanlı çapraz doğrulama, 81 ağaç ve 2 örnek, 1.09×10^4 RMSE ve 0.71 R-kare değeri ile en iyi model olmuştur. YSA için, nöron sayısı, aktivasyon fonksiyonu, test/eğitim seti ayırma yüzdesi ve k değerleri değiştirilerek en iyi model belirlendi. Bant aralığı için 0.3 test/eğitim seti ayırma oranı, 4-katmanlı çapraz doğrulama, 52 nöron ve ReLU aktivasyon fonksiyonu, 0.282 RMSE ve 0.70 R-kare değeri ile en iyi modeli vermiştir. Hidrojen üretim hızı için ise 0.25 test/eğitim seti ayırma oranı, 7-katmanlı çapraz doğrulama, 3 nöron ve ReLU aktivasyon fonksiyonu RMSE 1.47×10^4 ve R-kare değeri 0.60 ile en iyi modeli vermiştir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

ÖZET v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF SYMBOLS xiii

LIST OF ACRONYMS/ABBREVIATONS xiv

1. INTRODUCTION 1

2. LITERATURE REVIEW 4

2.1. Photocatalytic Reforming of Glycerol 4

2.1.1. Glycerol and Its Importance 4

2.1.2. Photocatalytic Reforming of Glycerol 5

2.1.3. TiO_2-based Photocatalysts 7

2.1.4. Other Photocatalysts and Details 7

2.2. Machine Learning 8

2.2.1. Machine Learning Methods 8

2.2.2. Random Forests 10

2.2.3. Artificial Neural Network 10

3. COMPUTATIONAL DETAILS 13

3.1. Data Collection 13

3.2. Random Forest 15

3.3. Artificial Neural Network 17

4. RESULTS AND DISCUSSION 18

4.1. Pre-analysis of Dataset 18

4.2. Analysis of Band Gap Models 26
4.2.1. Random Forest Model 26
4.2.2. Artificial Neural Network Model 30
4.3. Analysis of Hydrogen Production Rate Models 32
4.3.1. Random Forest Model 32
4.3.2. Artificial Neural Network Model 37
5. CONCLUSION AND RECOMMENDATIONS 40
5.1. Conclusion 40
5.2. Recommendations 42
REFERENCES 43
APPENDIX A: ARTICLES IN THE DATASET 61
APPENDIX B: COPYRIGHT PERMISSION FOR FIGURES 65

LIST OF FIGURES

Figure 2.1. Published literature of conversion of glycerol into hydrogen and value-added products [5]. 5

Figure 2.2. Chemical reactions for hydrogen production from glycerol. 5

Figure 2.3. Biomass reforming, biomass oxidation and water splitting representation [13]. 6

Figure 2.4. The model of ANN [24]. 11

Figure 4.1. The distribution of calcination temperature of semiconductors. 21

Figure 4.2. The distribution of calcination temperature of photocatalysts. 22

Figure 4.3. Weight percent of cocatalyst vs H_2 production rate. 23

Figure 4.4. Photocatalyst load vs H_2 production rate. 24

Figure 4.5. Glycerol concentration vs H_2 production rate. 24

Figure 4.6. The significance of variables for band gap model. 28

Figure 4.7. The comparison of validation set results for random forest model of band gap with original variables for 41 trees and 1 sample in the leaf node. ... 28

Figure 4.8. The comparison of test set results for random forest model of band gap with original variables for 41 trees and 1 sample in the leaf node. 29

Figure 4.9. The comparison of validation set results for random forest model of band gap with the most effective variables for 41 trees and 1 sample in the leaf node. …….. 29

Figure 4.10. The comparison of test set results for random forest model of band gap with the most effective variables for 41 trees and 1 sample in the leaf node. …… 30

Figure 4.11. The comparison of validation set results for ANN model of band gap with the most effective variables for 52 neurons and ReLU as activation function. ……... 31

Figure 4.12. The comparison of test set results for ANN model of band gap with the most effective variables for 52 neurons and ReLU as activation function. 31

Figure 4.13. The significance of variables for hydrogen production rate model. …….. 34

Figure 4.14. The comparison of validation set results for random forest model of hydrogen production rate with original variables for 81 trees and 2 samples in the leaf node. ……... 34

Figure 4.15. The comparison of test set results for random forest model of hydrogen production rate with original variables for 81 trees and 2 samples in the leaf node. …….. 35

Figure 4.16. The comparison of validation set results for random forest model of hydrogen production rate with the most effective variables for 81 trees and 2 samples in the leaf node. ……. 35

Figure 4.17. The comparison of test set results for random forest model of hydrogen production rate with the most effective variables for 81 trees and 2 samples in the leaf node. ……... 36

Figure 4.18. The comparison of validation set results for ANN model of hydrogen production rate with the most effective variables for 63 neurons and ReLU function. …… 37

Figure 4.19. The comparison of test set results for ANN model of hydrogen production rate with the most effective variables for 63 neurons and ReLU function. …… 38

Figure B.1. Copyright permission for Figure 2.1. ……... 65

Figure B.2. Copyright permission for Figure 2.3. ……... 66

Figure B.3. Copyright permission for Figure 2.3. ……... 67

Figure B.4. Copyright permission for Figure 2.4. ……... 68

LIST OF TABLES

Table 3.1. Input variables of database. …….. 14

Table 4.1. The number of common semiconductors. …… 19

Table 4.2. The number of common photocatalysts. ……... 19

Table 4.3. Preparation method for semiconductors. …….. 20

Table 4.4. Preparation method of photocatalysts. ……..... 20

Table 4.5. Calcination temperature and time ranges. …… 21

Table 4.6. Crystal structure of TiO_2-based photocatalysts. ……... 22

Table 4.7. Light and lamp types of the experiments. ……. 23

Table 4.8. Reaction temperature of the experiments. …… 25

Table 4.9. pH values of the experiments. …….. 26

Table 4.10. The R^2 and RMSE results of random forest models constructed with all variables and significant variables. …….. 30

Table 4.11. The R^2 and RMSE result of ANN model with 52 neurons & ReLU as activation function. …….. 32

Table 4.12. The RMSE results of random forest model with 81 trees and 2 samples in the leaf node. …… 36

Table 4.13. The RMSE result of ANN model with 63 neurons and ReLU activation function. ……... 38

Table 4.14. The R^2 and RMSE results of all models. ……... 39

Table A.1. Article references. …… 61

LIST OF SYMBOLS

b_j	Bias
e	Error
E	Loss Function
g	Activation Function
n	Number of Experiments
p	Predicted Value
t	Target Value
v_j	Net Input
w_p	Weight of Artificial Neural Network
x_p	Inputs of Artificial Neural Network
y	Output of Artificial Neural Network

LIST OF ACRONYMS/ABBREVIATONS

ANN	Artificial Neural Network
CART	Classification and Regression Trees
ReLU	Rectified Linear Unit
RMSE	Root Mean Square Error

1. INTRODUCTION

The energy demand increases tremendously with the growing population, increasing use of energy and concerns for global warming and environmental protection; as a result of this, the need for alternative fuels becomes quite significant and obvious to see. Biofuels are clean, environmentally friendly, and efficient renewable energy sources [1]; for instance, biodiesel is one of the most common biofuels. It can be produced from biomass with a significant amount of glycerol as byproduct.

The biodiesel industry is growing every single day [2]. Glycerol can be used in different industries such as pharmaceuticals, cosmetics, food industry and many others. However, the growth in biodiesel industry may cause the over-production of [3]. It is expected in future that the excess glycerol will be a significant ecological problem if it is not utilized in an effective and environmentally friendly manner. Thus, various technologies for the utilization of glycerol have been investigated in recent years [1].

Glycerol can be converted into valuable products such as hydrogen, dihydroxyacetone, lactic acid, ethylene glycol, glyceraldehyde, glyceric acid, hydroxypyruvic acid, tartronic acid, oxalic acid, and formic acid [4,5]. To obtain these value-added products, various methods such as carboxylation, oxidation, hydrolysis, esterification, transesterification, pyrolysis, and etherification can be used [3]. The number of publications about the conversion of glycerol into hydrogen and value-added products increase continuously [5].

Hydrogen is a clean, storable, efficient, and environmentally friendly fuel. Mostly, it is obtained from fossil fuels such as natural gas and coal [6]. With a promising technology, hydrogen can be obtained from glycerol; for instance, the photocatalytic reforming of glycerol can convert solar energy into chemical energy through producing hydrogen [7].

In photocatalytic reforming of glycerol, semiconductors with a cocatalyst are used; in some studies, semiconductor composites which include more than one semiconductor are also utilized. The variables such as semiconductor and cocatalyst types, crystalline structure, preparation method of semiconductor, preparation method of photocatalyst, calcination

temperature, calcination time, and weight percent of cocatalyst affects the performance of the process [5].

In general, TiO_2 based semiconductors are used in photocatalytic reforming of glycerol since TiO_2 is cost effective, nontoxic, photoactive, stable, and accessible. The crystal structure of TiO_2 can affect the hydrogen production rate. It can be anatase, rutile, brookite, or the combination of them. Mostly, anatase-rutile structure is used in the studies since it has high photocatalytic efficiency [5].

Also, there are other semiconductors such as ZnO, CdS, g-C_3N_4, $La_{0.2}Na_{0.98}TaO_3$ that are used in photocatalytic reforming. Mostly, semiconductors are prepared by sol-gel and hydrothermal methods. To increase the photocatalytic activity, cocatalysts are deposited on semiconductors mostly by photodeposition method. In general, noble metals are used as cocatalyst [5].

Since the research on photocatalytic reforming on glycerol increases, it is possible to collect large amount of data from literature, and analyze using machine learning techniques, which use statistics, mathematics, and computer programming to reveal the patterns inside data and construct models [8]. There are different machine learning functions such as classification, regression, clustering, association, etc. [9], and these functions can be performed using various techniques such as decision trees, random forest, and artificial neural networks. For instance, decision trees are one of the most popular classification and regression techniques; the rules used to classify the data can be extracted and used for future works to decide the experimental conditions [10]. Additionally, random forest, which is the combination of decision trees, can be used for more accurate predictions [11]. Another method for machine learning is artificial neural networks (ANN). It imitates the function of the human brain, and it can learn, analyze, and adapt to a changing environment [12].

In this thesis, the data for photocatalytic reforming of glycerol was collected, and a dataset was created from the publications between 2005 and 2022 in Web of Science. First, the dataset was cleaned, organized, and prepared for machine learning algorithms. Then, the models for band gap and hydrogen production rate were constructed by implementing random forest, and artificial neural network algorithms with this dataset. The model

performances for random forest and artificial neural network were measured by root mean square error (RMSE). With these algorithms, the effects of variables on band gap values and hydrogen production rates were investigated.

This thesis is comprised of 5 chapters. In Literature Survey (Chapter 2), the photocatalytic reforming of glycerol and machine learning methods were reviewed. In Computational Details (Chapter 3), the creation of the database and the construction of models were explained. In Results and Discussion (Chapter 4), the results were discussed briefly. In the end, the conclusions of this study were presented in Conclusion and Recommendations (Chapter 5).

2. LITERATURE REVIEW

2.1. Photocatalytic Reforming of Glycerol

2.1.1. Glycerol and Its Importance

With growing population and increasing energy consumption and growing concern for environment has been increasing tremendously leading to a search for new energy resources as alternative to fossil fuels. Biofuels are among the most efficient, clean, and environmentally friendly, renewable energy sources [1], and biodiesel is one of the most commonly studied biofuels. The biodiesel industry expands every day; in 2018, 36.8 million liters of biodiesel was produced, and it is estimated to be 44 million liters production in 2027 [2].

During the conversion of biomass to biodiesel, glycerol is produced as the major byproduct; hence, the growth of biodiesel industry will cause immense glycerol production [3]. Excess glycerol is expected to create serious problems for the disposal or utilization. Since glycerol disposal is costly, wasteful, and not ecofriendly, its utilization has been investigated extensively in recent years [1]. Limited amount of glycerol can be used in various industries such as cosmetics, pharmaceutical, food industry and many others [3]. The remaining can be converted into value-added products such as hydrogen, dihydroxyacetone, lactic acid, ethylene glycol, glyceraldehyde, glyceric acid, hydroxypyruvic acid, tartronic acid, oxalic acid, and formic acid [4,5]. There are some methods to convert glycerol to value-added products by catalytic conversion processes such as carboxylation, oxidation, hydrolysis, esterification, transesterification, pyrolysis, and etherification [3].

There is an increasing interest in conversion of glycerol into hydrogen and other value-added products. Since the growth in biodiesel industry causes excess glycerol, the utilization of glycerol becomes significant. In Figure 2.1, the number of published articles in literature is given in terms of years. Research on hydrogen conversion is more popular than other value-added products whose production process is more complicated [5].

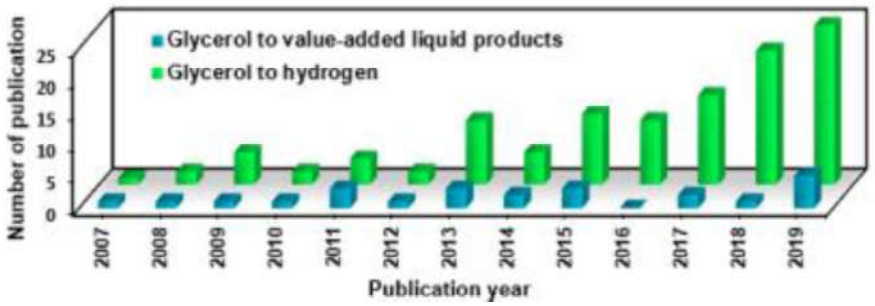

Figure 2.1. Published literature of conversion of glycerol into hydrogen and value-added products [5].

Hydrogen is a clean, storable, efficient, and environmentally friendly fuel. It exists in hydrocarbons, water and biomass. Hydrogen is mostly produced from fossil fuels such as natural gas and coal [6]. Hydrogen can be also obtained by photocatalytic reforming of glycerol which is a promising technology that converts solar energy into chemical energy. At atmospheric conditions, 1 mol of glycerol is converted to 7 moles of H_2 [7]. Chemical reactions for hydrogen production from glycerol are given in Figure 2.2.

$$C_3H_8O_3 \rightarrow 3CO + 4H_2 \qquad Decomposition\ Reaction$$
$$H_2O + CO \rightarrow H_2 + CO_2 \qquad Water\ Gas\ Shift\ Reaction$$
$$C_3H_8O_3 + 3H_2O \rightarrow 7H_2 + 3CO_2 \qquad Steam\ Reforming\ Reaction$$

Figure 2.2. Chemical reactions for hydrogen production from glycerol.

2.1.2. Photocatalytic Reforming of Glycerol

Photocatalytic reforming of glycerol is a feasible technology, which can convert solar energy into chemical energy. It is a process between photooxidation and photocatalytic water splitting. In water splitting, H_2 and O_2 are produced under appropriate conditions by using photocatalyst. It is an endergonic reaction. In photooxidation, if the same photocatalyst is used in a solution which involves organic substance such as glycerol, oxidation would appear directly under aerobic conditions. Photocatalytic reforming is between these two reactions, which takes place in the presence of an organic substance and absence of oxygen. In Figure

2.3, photocatalytic reforming between photooxidation and photocatalytic water splitting is given [13].

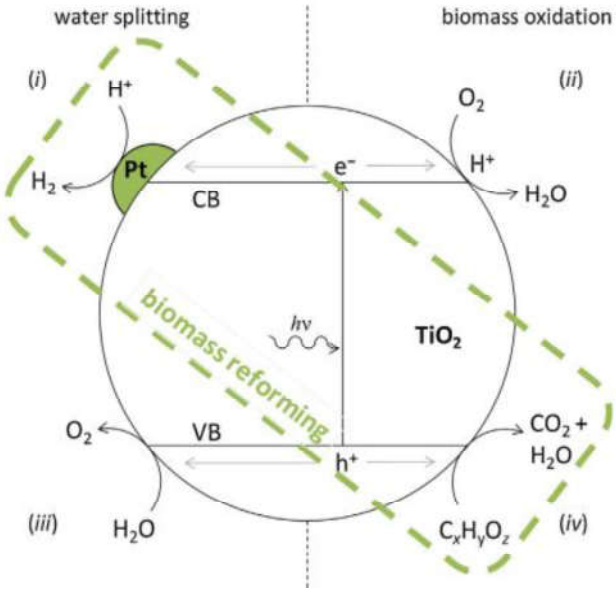

Figure 2.3. Biomass reforming, biomass oxidation and water splitting representation [13].

In photocatalytic reforming of glycerol, solid semiconductors, as pristine form or dopped with some elements, are used together with a cocatalyst. Semiconductor composites, which are composed of more than one semiconductor, are also used in some studies. There are various parameters that affect the comparison such as semiconductor and cocatalyst types, crystalline structure, preparation method of semiconductor, preparation method of photocatalyst, calcination temperature, calcination time, and weight percent of cocatalyst that deposited on semiconductor [5].

In literature, the interest in photocatalytic reforming of glycerol increases every day. There are various studies about this topic. One of them is the study of Estahbanati et al. in 2020. In the study of them, an extensive review of different experiments for hydrogen production with photocatalytic reforming of glycerol was provided. Most of the experiments were conducted by using TiO_2 as semiconductor and noble metals cocatalyst. Also, they reviewed the production of other valuable products by using photocatalytic reforming of glycerol [5].

Kumar et al. uses TiO_2 as semiconductor and noble metal as cocatalyst as well. In their study, the effects of photocatalyst load, weight percent of cocatalyst and concentration of glycerol on hydrogen production rate were investigated [14]. In Tahir and Fajrina's study, TiO_2 was used with another semiconductor as composite. In this study, the effect of pH was also investigated [15]. Like these studies, generally, TiO_2 is used as semiconductor in the literature with cocatalyst or another semiconductor.

2.1.3. TiO_2-based Photocatalysts

In general, TiO_2 based photocatalysts are used in the literature. TiO_2 is advantageous in many aspects. It is cost efficient, nontoxic, photoactive, thermally stable and accessible. TiO_2 can be found in different crystal structures such as anatase, rutile and brookite [5]. Anatase and rutile structures have smaller band gaps and particle radius rather than brookite. The band gap and surface area of anatase is larger than rutile structure [16]. Anatase structure has higher photocatalytic activity. By heating metastable anatase phase to high temperature, thermodynamically stable rutile phase is obtained. Mixed phase of anatase and rutile improved photocatalytic efficiency. P25, which is comprised of 80% anatase and 20% rutile phases, is the most common photocatalyst used in photocatalytic reforming. It has high photocatalytic efficiency under solar light due to the existence of anatase and rutile phases [5].

2.1.4. Other Photocatalysts and Details

There are other semiconductors such as ZnO, CdS, g-C_3N_4, $La_{0.2}Na_{0.98}TaO_3$ that are used in photocatalytic reforming of glycerol in recent years. The most common photocatalyst preparation methods are sol-gel and hydrothermal methods in the literature. There are also other methods such as hydrolysis, modified sol-gel, ionothermal, etc. Calcination temperature and duration play a significant role in photocatalyst preparation. In general, 350 – 500 °C for calcination temperature and 2 – 5 h for calcination duration are used. Preparation method and calcination temperature affect the structure of semiconductor which can be nanoparticle, nanotube, nanorod, nanosphere or hallow sphere [5].

By deposition of cocatalyst, the activity of photocatalyst increases. Reactions occur at the border between cocatalyst and semiconductor. Mostly, noble metals are used as cocatalyst in literature. Pt is the most common cocatalyst in photocatalytic reforming of glycerol. Additionally, Cu, Au, Ag, Ni and Pd are also used frequently, and they have positive effects in photocatalytic activity. Cu is the second most commonly used cocatalyst probably because the noble metals are costlier. The common deposition methods of cocatalysts on semiconductors are photodeposition, impregnation, and chemical synthesis [5].

In addition, composite semiconductors are used for photocatalytic reforming. Composite semiconductors are comprised of more than one semiconductor. The efficiency of photocatalyst improves with the composition of semiconductors [17]. The stability of composite semiconductors is better than single semiconductors. In photocatalytic reforming of glycerol, nanocomposites seem to be promising alternatives even though there are limited number of studies in literature [5].

2.2. Machine Learning

2.2.1. Machine Learning Methods

In recent years, the interest in machine learning has increased continuously. Large amounts of data accumulated in literature for every field can be organized, analyzed, and modeled by using machine learning techniques. Machine learning uses statistics, mathematics, and computer programming to see the patterns in data and construct models to make predictions or to extract rules [8]. It aims to create computer programs which can learn from data. Larger datasets are more suitable to be analyzed with machine learning algorithms efficiently [9].

Machine learning techniques can be grouped into two categories as supervised and unsupervised learning techniques. The main difference between these two categories is supervised learning uses labeled data which has a relation between input and output variables. Classification and regression are examples of supervised learning. On the other

hand, clustering and association are examples for unsupervised learning which use unlabeled datasets [9].

Classification is one of the supervised techniques of machine learning. The aim of classification is splitting the data into classes by using the output variable. The output variable should be categorical or suitable to be categorical. There are some classification techniques such as k-nearest neighbor algorithm, decision trees, support vector machines, artificial neural network, etc. [9].

Regression is another type of supervised learning. Regression determines the relation between independent and dependent variables. Basically, regression indicates the change of dependent variable based on the independent variables. Linear regression and logistic regression are the types of regression [18]. Linear regression is used for estimation of continuous dependent variables by using independent variables. The aim is to find the line that minimizes the error for each point. Logistic regression is similar to linear regression. The main difference between these two methods is that dependent variable in logistic regression is not continuous. It can be discrete or categorical [18]. There are also other methods like artificial neural networks, random forest, support vector machines.

Clustering is one of the unsupervised learning algorithms of machine learning. The purpose of clustering is grouping the data based on the similarities in variables. In a cluster, the members are similar to each other, but they are different from the other clusters [9].

Machine learning techniques can be applied to different studies. For instance, algorithms can be used for photocatalytic and photoelectrochemical reactions in various publications [9]. In the study of Oral et al. in 2020, machine learning tools were applied to photoelectrochemical water splitting [19]. Additionally, tools can be used for the systems such as solar cells and biofuels. In the study of Yılmaz and Yıldırım, machine learning methods in perovskite solar cells were reviewed [20]. In the study of Odabaşı and Yıldırım, they also studied the analysis of perovskite solar cells by using machine learning techniques [21].

2.2.2. Random Forests

Random forest is an algorithm that combines the series of decision trees. Random forest provides high accuracy for classification and regression. It is one of the most common machine learning methods [22]. Random forest algorithm works better than linear regression with categorical and nonlinear data. Additionally, random forest estimates the error rate more accurately than decision trees because the error rate converges with the increasing number of trees [11].

2.2.3. Artificial Neural Network

Artificial neural networks (ANN) are basically imitating the function of the human brain. As the human brain, it can learn, analyze, and adapt to a changing environment. The human brain is comprised of neurons and neural networks. Neurons receive signal and transfer this signal to other neurons. Also, in ANN, there are nodes that are connected to each other as human neural system. The neurons in ANN imitate the behavior of neurons in the human brain [23].

There are different types of ANN such as multi-layer feed-forward perceptron network, probabilistic neural network, recurrent networks, etc. The most common type is multi-layer feed-forward perceptron network that is comprised of an input layer, hidden layers and an output layer. Inputs are received by input layer, passes through hidden layers to output layer. Output is received as a response of ANN from output layer [23]. The layers are connected with weights. By using a training dataset, weights are optimized and used for estimation of the output of a new input variable [12].

To construct a network model, the number of hidden layers, the number of neurons, activation function and learning rate should be optimized by repeating the training procedure. Specifying the number of neurons in hidden layers is important to determine the design of the neural network. If there are few neurons in hidden layers, it will cause under-fitting and if there are too many neurons, it will cause over-fitting. Under-fitting causes inaccuracy to identify signals in a complex dataset. Over-fitting causes insufficient training for all neurons in hidden layers since there is limited information in the training set [12].

ANN has the ability to reveal non-linear relations. To construct ANN model, there are some main components such as inputs, weights, and outputs. Inputs which are represented as $x_1, \dots, x_p$ are received by neurons from the external environment or other neurons. The received signal is modified by weight, which is represented as $\boldsymbol{w} = (w_1, \dots, w_p)$, is a vector of synaptic weights. The interception or threshold of a neuron is called as bias and represented as b_j. The equation of the net input is given as [24]

$$v_j = b_j + \sum_{j=1}^{p} w_{ij} x_j. \tag{2.1}$$

The net input determines the activation of the neuron. The activation of the neuron depends on the activation function. The equation for the activation function is given as [24]

$$y_j = g(v_j). \tag{2.2}$$

In the above equation, g represents the activation function. If the activation function is a unit step, the output is 1 for the net input which is greater than 0. Else, the output will be 0. There are other activation functions such as liner, sigmoid, hyperbolic tangent, etc. [24].

In Figure 2.4, the general artificial neural network model is given [24].

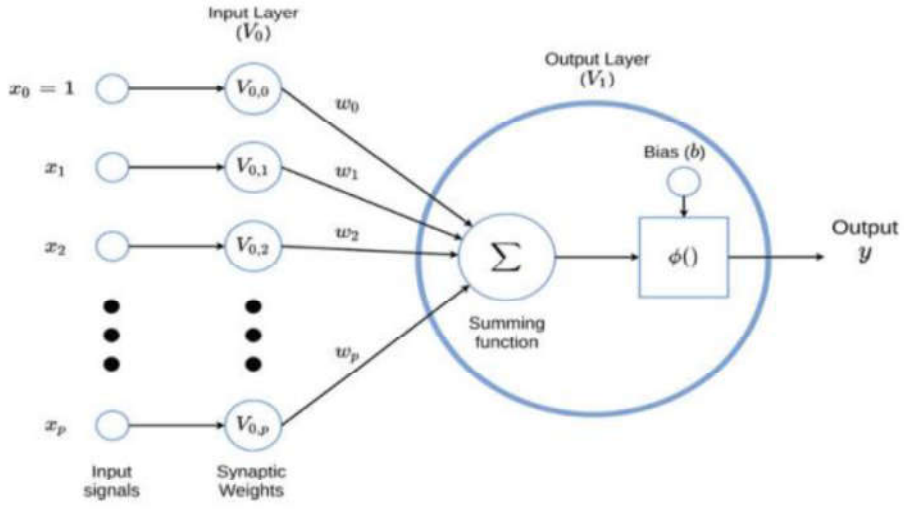

Figure 2.4. The model of ANN [24].

The backpropagation algorithm is used to improve the performance of the neural network by reducing the error from the output. This algorithm is used for training by an iterative method. It is suitable for large datasets. The backpropagation method includes feed-forward of input training pattern, calculation of error and adjustment of weights [25].

The backpropagation algorithm adjusts the weights until error becomes insignificant. Error can be calculated as

$$E = \frac{1}{2}\sum_{i}(t_i - y_i)^2 = \frac{1}{2}\sum_{i} e_i^2. \tag{2.3}$$

In this equation, E represents the loss function, t_i represents the target value and y_i represents the output value and e is the error. The loss function is used to optimize weights [24].

3. COMPUTATIONAL DETAILS

3.1. Data Collection

In this thesis, the aim was to extract knowledge from published literature for photocatalytic reforming of glycerol using machine learning. For these 150 articles, which were published between 2007 and 2022 about photocatalytic reforming of glycerol, and selected with relevance search in Web of Science database were reviewed. While constructing the database, 93 of 150 articles were used since the remaining articles were not suitable for data extraction; 791 data points were taken from these 93 articles. In dataset, the type of semiconductor and cocatalyst, weight percent of cocatalyst, structure and form of the semiconductor, preparation methods, calcination temperature and duration, light type, experimental conditions such as solution volume, reaction temperature, band gap were entered as descriptors (input variables) while the hydrogen production rate were taken as output variable: the band gap was also used as output variable for the models developed to predict the bandgap from other structural properties.. Python was used for analyzing and modeling the dataset. The categorical variables such as preparation methods for semiconductor and photocatalyst, form of photocatalyst were converted to numeric variables by using one hot encoding in Python library. For the photocatalysts that are not calcinated, calcination temperature was taken as 25 °C and calcination duration was taken as 3 hours as the average value of the variable. The details of the dataset are given in Table 3.1.

Initially, the dataset was analyzed in Excel as pre-analysis before machine learning application. By using bubble graphs and scatter plots, the effects of variables on band gap and H_2 production rate were investigated. As a result of this analysis, some of the variables such as power of light, type of light, and solution volume were found to be ineffective on band gap and H_2 production rate. Then the data is divided into two groups. One of these two groups is organized according to the band gap and the variables that are effective on this property. Another group is constructed for H_2 production rate with the relevant variables.

In Table 3.1, the variables that are used to construct the database and their ranges are given.

Table 3.1. Input variables of database.

Input Variables	**Ranges for numeric or sub-categories for categoric variables**
Semiconductor	AC, Ag_2O, Bi_2O_3, $Cd_{1-x}Zn_xS$, CdS, CNT, CS, $Cu_{0.02}Ti_{0.98}O_2$, Cu_2O, $CuFe_2O_4$, CuO, CuO_{1-x}, CuS, FCNT, Fe_2O_3, g-C_3N_4, GO, Graphene, hex-CdS, $In_{0.5}WO_3$, In_2O_3, $La_{0.2}Na_{0.98}TaO_3$, Montmorillonite, N719, Ni-Ni(OH)$_2$, Ni(OH)$_2$, NiO, NiO_x, $PbTiO_3$, PdS, PHPT, PRC1, PRC2, PRC3, rGO, $SrAl_2O_4$, TiO_2, Zn_2TiO_4, ZnO, ZnO-ZnS, ZnS, γ-Al_2O_3
Co-catalyst	Ag, Ag-QD, Au, Co, CP-1, Cr, Cu, Fe, La, N, Nd_2O_3, Ni, Pd, Pt, $SnCl_2$
Semiconductor Preparation Method	Chemical Synthesis, Hydrolysis, Hydrothermal, Impregnation, Precipitation, Sol-Gel
Calcination Temperature (Semiconductor)	300-700 Celsius
Calcination Time (Semiconductor)	0-8 hours
Photocatalyst Preparation Method	Chemical Synthesis, Hydrothermal, Impregnation, Ionothermal, Molten Salt Synthesis, Photodeposition, Precipitation, Sol-Gel, Solid State Dispersion
Calcination Temperature (Photocatalyst)	200-1000 °Celsius

Table 3.1. Input variables of database (cont.).

Input Variables	Ranges for numeric or sub-categories for categoric variables
Calcination Time (Photocatalyst)	0-5 hours
Light Type	Fluorescent, Halogen, Hg, Hg-Xe, LED, Solar, UV lamp, Xe
Power	3-500 W
Filter	300-768 nm
pH	0 - 14
Crystal. Structure of Semiconductor	Anatase, Rutile, Brookite
Band Gap (eV)	1.51 - 4.69

3.2. Random Forest

To construct a predictive model, random forest was used. In Python, random forest models were created by *scikit-learn* and *pandas* libraries [26, 27]. Initially, the dataset was read from an Excel file. The columns were filled with *Nothing* for the categorical variables if there were no usage of them. Then, one hot encoding was applied to the categorical variables such as semiconductor and cocatalyst, form of semiconductor, light, and preparation methods of semiconductor and photocatalyst. By using one hot encoding, these categorical variables were converted to numerical variables.

There were some missing values in the variables of the dataset such as photocatalyst load, glycerol concentration, and weight percent of cocatalyst, etc. Since these missing values were significant for the sake of model performance, they were filled with the average mean of the training sets of these variables. The details of training and test sets were given in the following paragraphs.

In band gap, there were some missing variables. The band gap model was constructed by using the existing band gap values as output variables. After the model was constructed, these missing variables were predicted. The predicted band gap values were used in hydrogen production rate model as input variables.

Standardization was applied to the variables since their ranges were quite different from each other. For instance, the range of weight percent of cocatalyst was between 0 to 50 but the range of calcination temperature was between 25 to 1000. During the analyses, this difference could affect the importance of the variables. Thus, the variables were normalized. *StandardScaler* was used for standardizing the variables [26].

After all these preparation steps, the dataset was ready for modeling. First of all, the dataset was divided by *train_test_split* in Python [26]. With *test_size*, the size of testing data was found, and this testing data was put aside for testing after model was constructed [26]. K-fold cross validation was applied to the remaining part. The training dataset was divided into k folds. One of them was called as validation set and the remaining k-1 sets were training sets. The model was constructed by using training set and validation set was tested by using this model. This process was repeated for k times and error was found for each time. At the end, average error was calculated. The train and test split ratio and k of the k-fold cross validation were determined by analyzing different split ratios and k values for different number of trees and various number of samples in the leaf node. The test/train split ratio was changed between 20% to 30%. The k value was changed to 3, 4, 5, 7 and 10. The number of trees was changed between 1 to 300 by 5. The sample number of a leaf was changed between 1 to 5. The model with the lowest root mean square error (RMSE) for validation was chosen and tested with the testing data. The equation of RMSE is given as [28]

$$e_i = p_i - t_i, \tag{3.1}$$

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}|e_i|^2\right]^{1/2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}|e_i|^2}. \tag{3.2}$$

In Equation 3.1 and 3.2, e_i is error for each step which is calculated by predicted value p_i and target value t_i. RMSE is calculated by addition of each error where n is the number of experiments [28,29].

The most significant variables were determined. *Feature_importances_* function can determine the important variables for a model [26]. For band gap model, semiconductor and cocatalyst types, structure of semiconductor, calcination temperature and duration, and weight percent of cocatalyst are the examples of the most important features. For hydrogen production rate model, photocatalyst load, band gap, glycerol concentration were the most important variables. For band gap models, 2nd cocatalyst and percent ratio of cocatalysts were eliminated. Additionally, preparation methods for semiconductor and cocatalyst were eliminated as well since there were a lot of missing information. For hydrogen production rate, form of the semiconductor was also eliminated. After determining the important features, the model was reconstructed with these variables. The differences between both models were given in the Results & Discussion section.

3.3. Artificial Neural Network

Artificial Neural Network was used to construct a predictive model for band gap and hydrogen production rate. In Python, artificial neural networks were constructed by *numpy* and *tensorflow* libraries [30,31].

The variables were prepared for modeling as in the previous section. One hot encoding, normalization and feature importance were implemented. Validation was applied for various test/train split ratios and k values. Additionally, the neuron numbers were analyzed for the range between 2 to 100 by 5. Linear, sigmoid, tanh and ReLU (Rectified Linear Unit) functions were taken as activation functions. In a Python loop, all of these variables were tried. For the values which can create model with the lowest RMSE were chosen.

4. RESULTS AND DISCUSSION

4.1. Pre-analysis of Dataset

Before the machine learning modeling, the general structure of the dataset will be described with the activities performed to prepare the data for machine learning analysis. Initially, the inappropriate data, which have no information about the hydrogen production rate (target variable) was eliminated. After that, the missing information in the input variables was completed if it is possible such as filling the missing values with the mean of training set of the corresponding variable or some other value that may reflect the physical situation better. For example, the missing calcination temperature usually refers to the case that the semiconductor/photocatalyst was not calcined. Hence, it was meaningful to take room temperature (25 °C) as calcination temperature. In such cases, the calcination time was taken as 3 hours, which is average in dataset. There were other missing values for significant variables such as preparation methods of semiconductors and photocatalysts, weight percent of cocatalyst, photocatalyst load, and band gap. For preparation method, if there was lack of information and the method could not be categorized under a specific method, it was assumed as chemical synthesis. The missing bandgap values, however, were handled with a completely different approach because some of the information related to semiconductor can be used to predict bandgap as well. Hence, a model was constructed to predict the unknown band gaps from the known values in the dataset. Then the predicted values were used to complete the dataset so that the predicted values can be used, with the values already in the dataset for the prediction of hydrogen production rate.

The most commonly used semiconductors in the dataset are presented in Table 4.1 with the total number in the dataset. The semiconductors whose total number in the dataset is small, were put into the group of *others*. As it can be seen from Table 4.1, the most commonly used semiconductor is TiO_2 with 69%. The remaining 23% of the dataset is mainly composed of ZnO, g-C_3N_4, $La_{0.2}Na_{0.98}TaO_3$, CdS and $SrAl_2O_4$ while the others constitute the remaining 8% of data.

Table 4.1. The number of common semiconductors.

Semiconductor Type	Total Number in the Dataset
TiO_2	551
ZnO	69
g-C_3N_4	47
$La_{0.2}Na_{0.98}TaO_3$	27
CdS	25
$SrAl_2O_4$	14
Others	62

After specifying the most commonly used semiconductors, the most frequently utilized cocatalyst-semiconductor pairs were determined. As the TiO_2 is the most common semiconductor, the photocatalysts involving TiO_2 and Pt is also the most common pair as expected; Pt, which has a high photocatalytic activity as cocatalyst, is used in almost 40% of total data. Noble metals such as Pt, Ag, and Au are also frequently used as cocatalyst since they have high activity even at low temperature and they are more stable [27]. However, since the cost of noble metals is high, the second most used cocatalyst is Cu, which can enhance the photocatalytic activity and is appropriate for electron transfer like noble metals [33].

Table 4.2. The number of common photocatalysts.

Photocatalyst	Total Number in the Dataset
Pt- TiO_2	198
Cu- TiO_2	65
Ni- TiO_2	33
Ag- TiO_2	32
Au- TiO_2	28
Others	154

The preparation methods of semiconductors and photocatalyst are also influential for the performance. In Table 4.3, the frequency of the preparation methods is given while the methods used to load the cocatalyst on semiconductor surface are given in Table 4.4. If there is lack of information for preparation method, it is written under chemical synthesis. Since

there was a lot of missing information filled with a general method, preparation methods were not taken into account during modeling.

Table 4.3. Preparation method for semiconductors.

Preparation Method	**Total Number in the Dataset**
Chemical Synthesis	484
Hydrolysis	7
Hydrothermal	153
Impregnation	32
Modified Hummer's Method	1
Precipitation	18
Sol-gel	104

For semiconductors, the most commonly used preparation method is hydrothermal treatment while, the most used method for photocatalyst preparation is photodeposition. Impregnation is also used quite frequently to load the cocatalyst on semiconductor surface.

Table 4.4. Preparation method of photocatalysts.

Preparation Method	**Total Number in the Dataset**
Chemical Synthesis	76
Hydrothermal	70
Impregnation	194
Ionothermal	3
Molten Salt Synthesis	2
Photodeposition	242
Precipitation	93
Sol-gel	81
Solid State Dispersion	39

The calcination temperature and time used during semiconductors or photocatalysts preparation also have significant effects on catalytic performance. For the data that does not include information about the calcination, the calcination temperature is taken as 25 °C and

calcination time is taken as 3 hours. The range of calcination temperature and duration for both semiconductors and photocatalysts are given in Table 4.5.

Table 4.5. Calcination temperature and time ranges.

Type	Temperature Range (°C)	Time Range (h)
Semiconductor	25 – 700	0 – 8
Photocatalyst	25 – 1000	0 – 5

The distribution of calcination temperature used for the preparation of semiconductors and photocatalysts, which refer to semiconductor+cocatalyst after cocatalyst loading, are presented Figure 4.1 and 4.2, respectively.

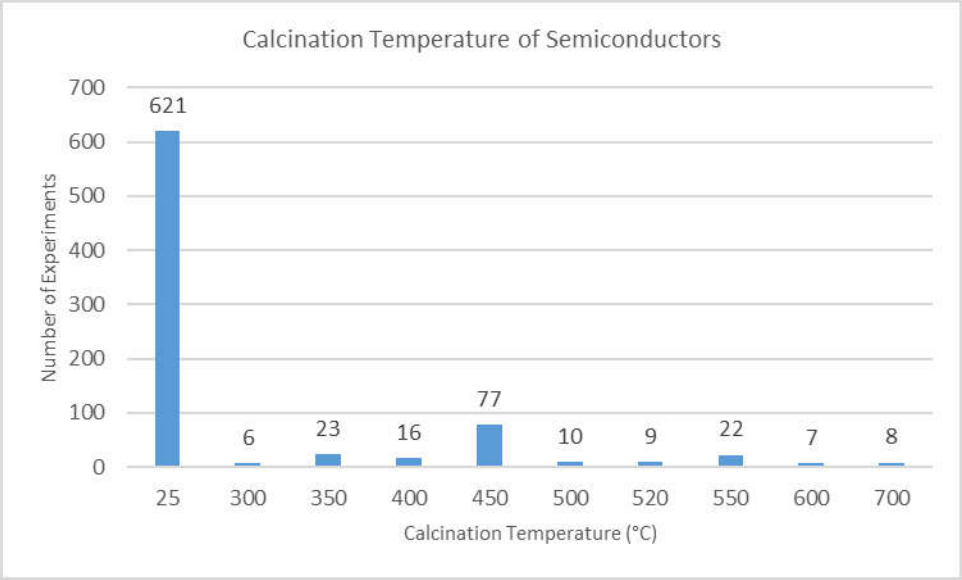

Figure 4.1. The distribution of calcination temperature of semiconductors.

For the photocatalysts that contain TiO_2 as semiconductor, which is the larger portion of the dataset, the crystal structure has an important effect on the results of experiments. Table 4.6 presents the frequency of data containing various crystal structures of TiO_2. In general, anatase-rutile structure is preferred since it improves photocatalytic efficiency.

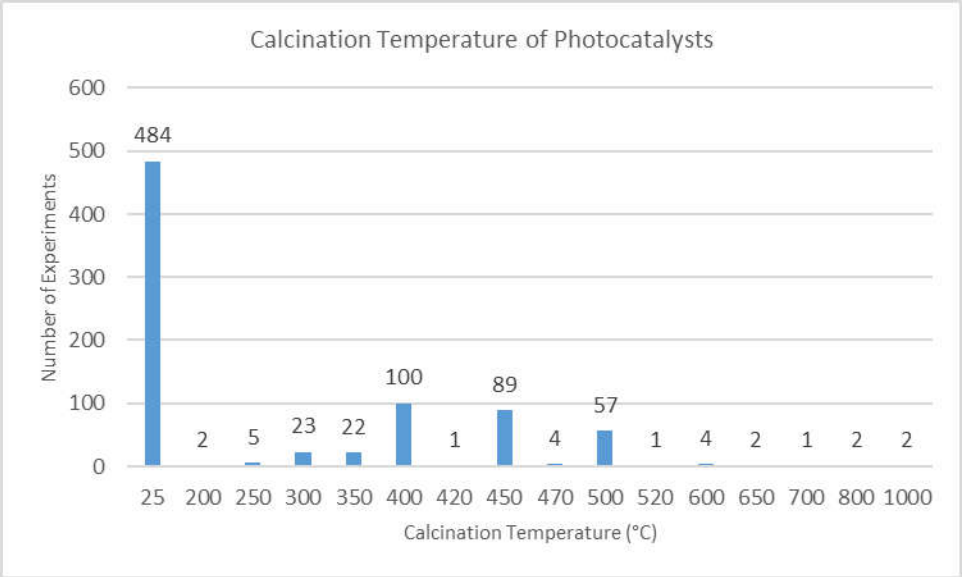

Figure 4.2. The distribution of calcination temperature of photocatalysts.

Table 4.6. Crystal structure of TiO_2-based photocatalysts.

Crystal Structure	Total Number in the Dataset
Anatase	194
Rutile	5
Brookite	2
Anatase-Rutile	355
Anatase-Brookite	25
Anatase-Rutile-Brookite	5

Another important variable is the light type. H_2 production may vary under different UV and visible light irradiation. On the other hand, the lamp type does not seem to have significant effects as it was discussed later in more details. The distribution of data according to the light and lamp types used in experiments is given in Table 4.7.

Another important variable is the weight percent of cocatalyst in a photocatalyst. The range of the weight percent changes between 0 and 50. The trend in the weight percentage use is given in Figure 4.3 against hydrogen production rate. Surprisingly, hydrogen production decreases if the cocatalyst loading is higher than a certain limit.

Table 4.7. Light and lamp types of the experiments.

Light Type	Lamp Type	Total Number in the Dataset
UV	Fluorescent	9
	Hg	230
	LED	17
	UV Lamp	62
	Xe	28
Visible	Fluorescent	1
	Hg-Xe	26
	LED	1
	Solar	200
	Xe	225

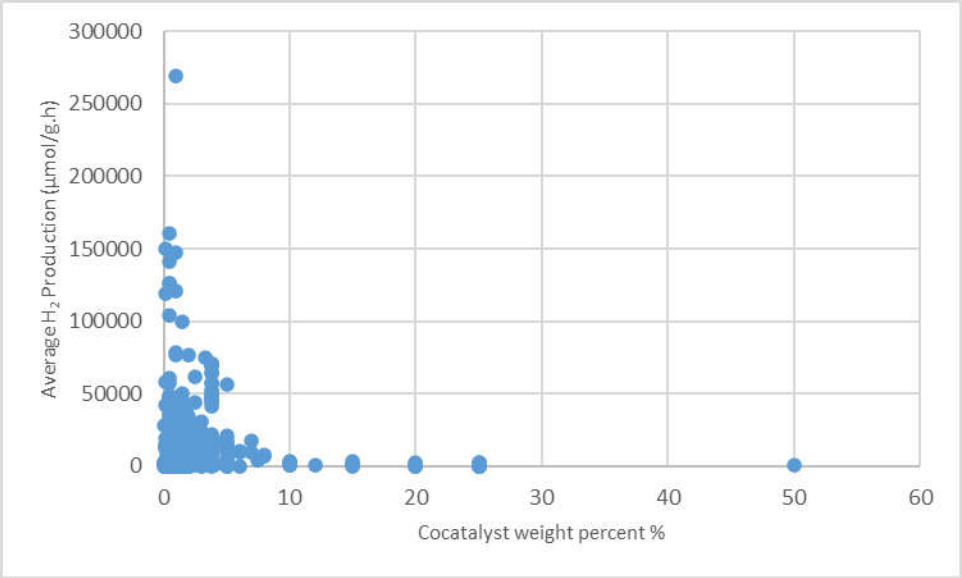

Figure 4.3. Weight percent of cocatalyst vs H_2 production rate.

Similar to the weight percentage of cocatalyst, the amount of photocatalyst used per volume also has similar effects on H_2 production, which is higher at the low photocatalyst loading probably due to better efficiency of photocatalyst in absorbing irradiation.

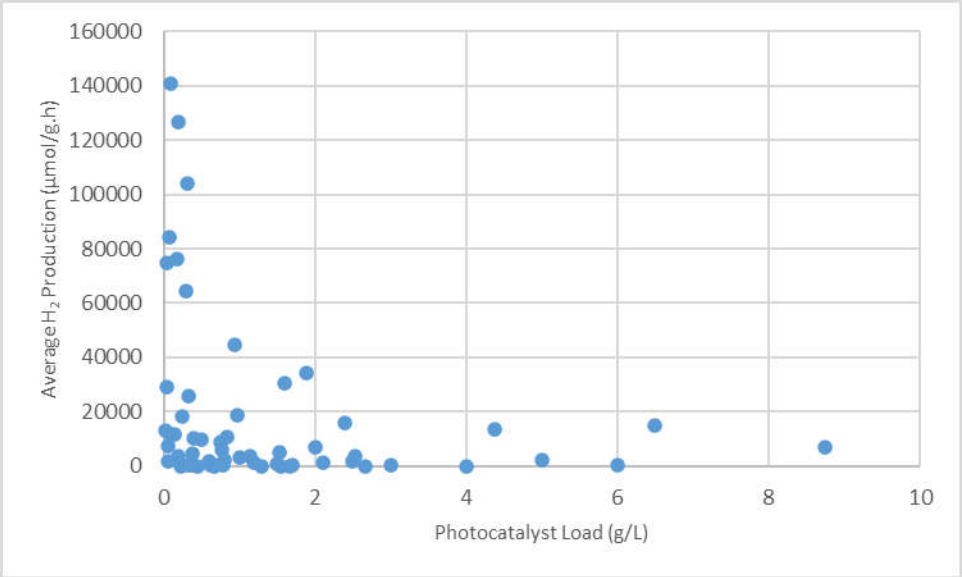

Figure 4.4. Photocatalyst load vs H_2 production rate.

The effects of glycerol concentration on hydrogen production also have the same trend as presented in Figure 4.5; the hydrogen production is generally higher in lower glycerol concentration (below 20%).

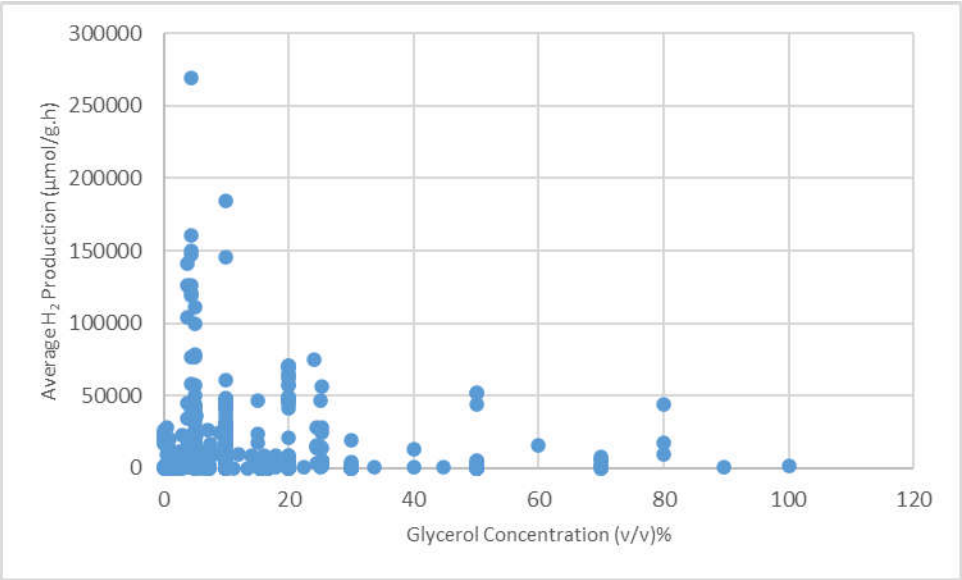

Figure 4.5. Glycerol concentration vs H_2 production rate.

Table 4.8. Reaction temperature of the experiments.

Reaction Temperature (°C)	Total Number in the Dataset
5	3
10	4
15	3
24	4
25	499
30	57
40	21
50	7
60	14
80	14

The temperature, pH and solution volume are the other variables that are also adjusted to improve the hydrogen production rate; the distribution of temperature and pH are presented in Table 4.8 and 4.9, respectively. In general, the experiments take place under ambient conditions while the most commonly used pH range was 6-7. The range of solution volume is also changed between 5 ml to 1200 ml; however, there is no significant effect of solution volume on H_2 production rate.

To sum up, some variables that are given in the dataset have significant effects on H_2 production rate while some semiconductor/ photocatalyst related variables also affect the band gap. While constructing the models, the significance of the variables was considered according to the above analyses. For band gap models, the types of semiconductors and cocatalysts, the structure of the photocatalysts, weight percentage of cocatalysts, preparation methods, calcination time and duration are used due to their effects on band gap value. For H_2 production rate models, on the other hand, the operational variables (except lamp type, solution volume, power of the light source and light intensity, which have no significant effects on H_2 production) were also used.

Table 4.9. pH values of the experiments.

pH	Total Number in the Dataset
0 – 2	14
2 – 3	2
3 – 4	20
4 – 5	15
5 – 6	23
6 – 7	128
7 – 8	7
8 – 9	2
9 – 10	17
10 – 11	11
11 – 12	8

4.2. Analysis of Band Gap Models

To construct a band gap model, two different machine learning techniques were used. Random forest and artificial neural network methods give significant results. In the following sections, the details of each method and results were given. Also, the results were discussed.

4.2.1. Random Forest Model

After all preparation and optimization steps, the results for random forest were given in the following part. For random forest model of band gap, 0.25 test/train split ratio, 4-fold cross validation, 41 trees and 1 sample in a leaf node were used as a result of Python. According to these parameters, R-square and root mean square error were calculated. Initially, these variables are used to construct model by using the original dataset that the preparation was explained in Section 3.2. The model was constructed by the training set and this constructed model was tested by the validation set. In Figures 4.7 and 4.9, predicted and experimental values were compared for the results of validation set. After that, the model was tested with the test set that was separated from the dataset at the first step. In Figures 4.8 and 4.10, the results of test set were given. As it can be understood from these Figures,

the constructed model gave the similar results for both validation and test sets, which explained the model performance was good.

Feature importance was applied during these steps, and the most effective features were determined as explained in Section 3.2. As a result of feature importance, second cocatalyst and the percentage of cocatalyst in cocatalyst composite were eliminated. The most important variables were the weight percent of cocatalyst, percentage of semiconductor, and calcination temperature. In Figure 4.6, the most significant variables were given.

In Figures 4.7, 4.8, 4.9 and 4.10, the effect of feature importance was determined. In Figures 4.7 and 4.8, the results were given without using feature importance (i.e. using all input variables). In Figures 4.9 and 4.10, the results were given for the dataset which feature importance was applied, and unimportant features was eliminated. By using the most important variables in the model, the performance of model was increased since the least effective variables were eliminated. In Table 4.10, the fitness results for both models were given, and it can be seen that the model which was constructed by using only the significant variables gave better results with higher R^2 and lower RMSE values. Feature importance affects the model performance in a positive way. With the elimination of insignificant variables, the performance of the model was improved. Thus, R-squared value was increased, and root mean square error was decreased.

It can be said that determining the important variable in the dataset is effective for the model performance. With the models that are constructed with significant variables, better results can be obtained in terms of R-squared and Root Mean Square Error values. Also, the same result can be seen in Figure 4.7, 4.8 and Figure 4.9, 4.10 The results are much closer to the trendline in Figures 4.9 and 4.10, the model with significant variables.

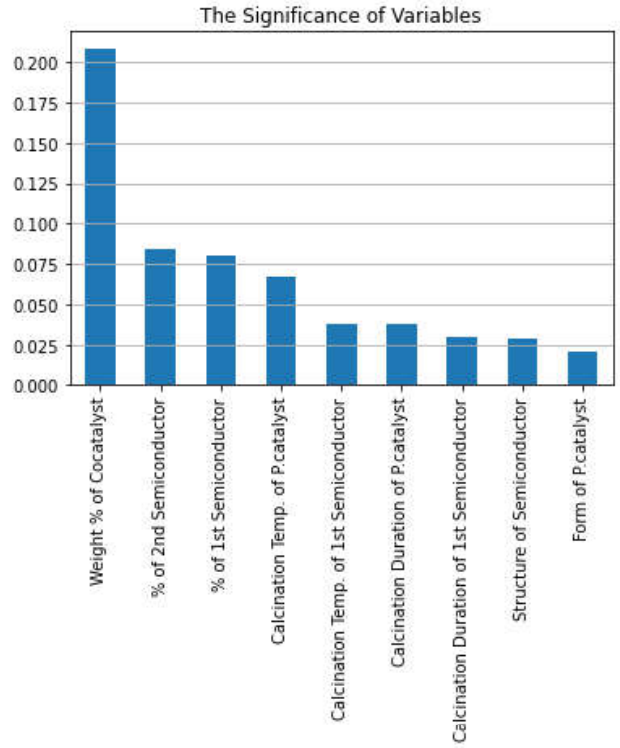

Figure 4.6. The significance of variables for band gap model.

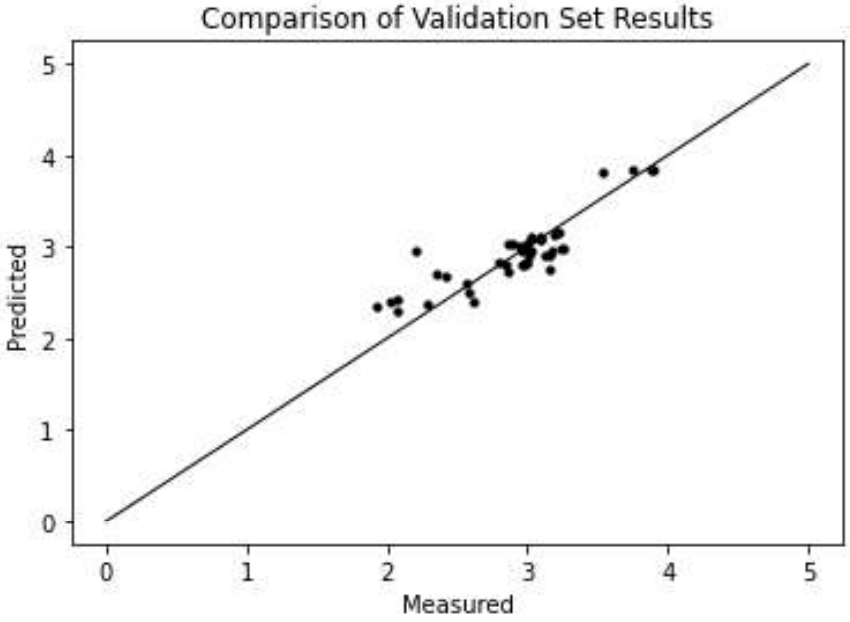

Figure 4.7. The comparison of validation set results for random forest model of band gap with original variables for 41 trees and 1 sample in the leaf node.

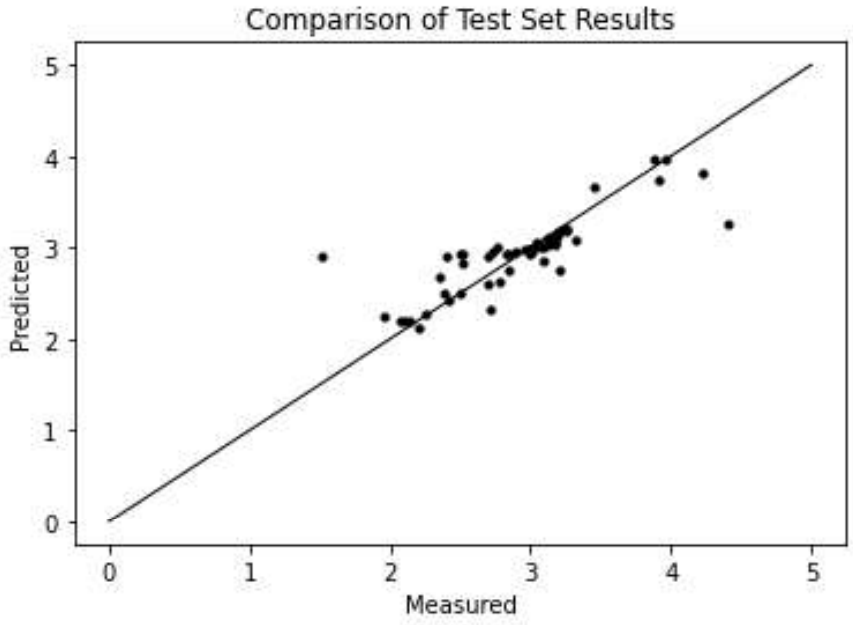

Figure 4.8. The comparison of test set results for random forest model of band gap with original variables for 41 trees and 1 sample in the leaf node.

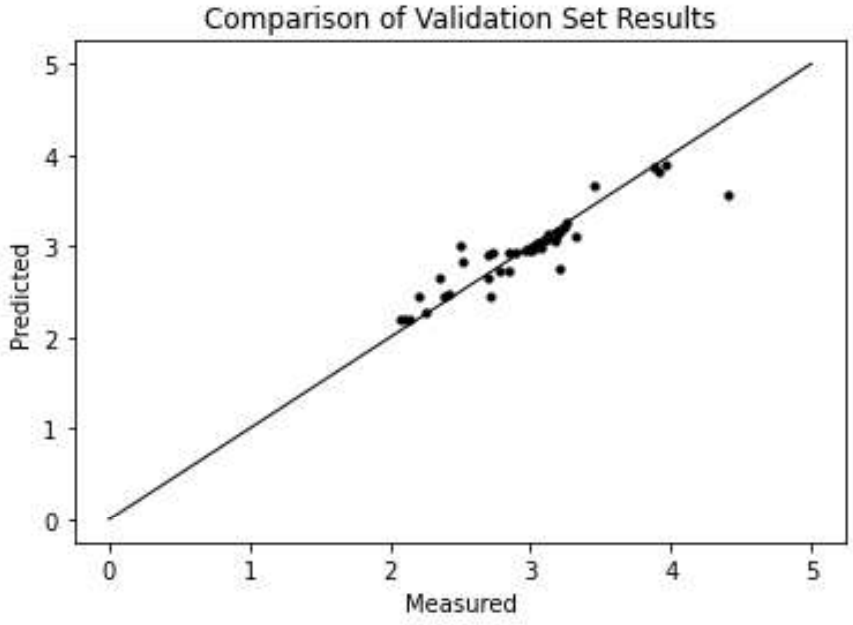

Figure 4.9. The comparison of validation set results for random forest model of band gap with the most effective variables for 41 trees and 1 sample in the leaf node.

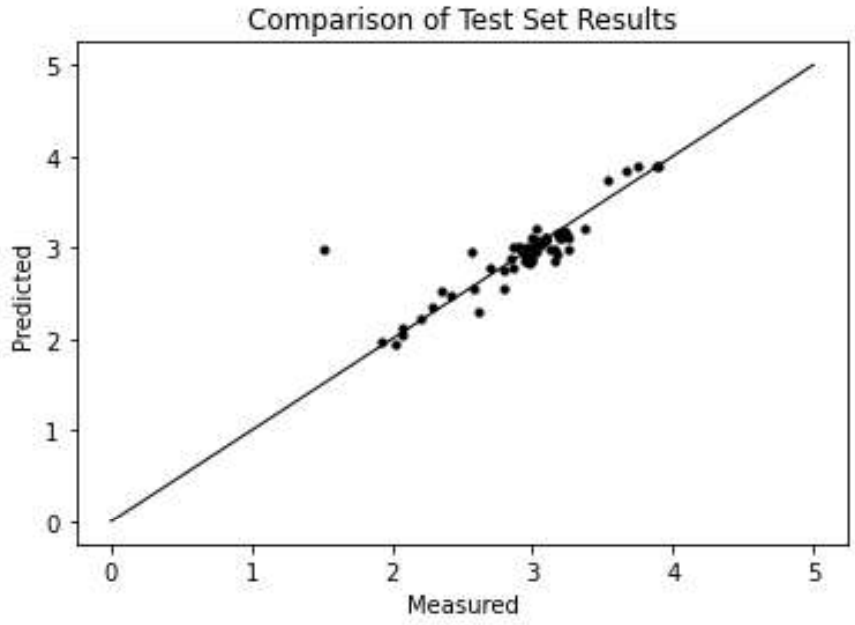

Figure 4.10. The comparison of test set results for random forest model of band gap with the most effective variables for 41 trees and 1 sample in the leaf node.

Table 4.10. The R^2 and RMSE results of random forest models constructed with all variables and significant variables.

Results	**R^2**	**RMSE**
Results of validation set with all variables (Figure 4.7)	0.77	0.216
Results of test set with all variables (Figure 4.8)	0.68	0.305
Results of validation set with significant variables (Figure 4.9)	0.83	0.206
Results of test set with significant variables (Figure 4.10)	0.73	0.234

4.2.2. Artificial Neural Network Model

In Artificial Neural Network model, the results for each combination were collected and the best model was determined. It was found as 0.3 test/train split ratio, 4-fold cross validation, 52 neuron numbers and ReLU activation function gave the best model with the highest R-squared value and minimum Root Mean Square Error.

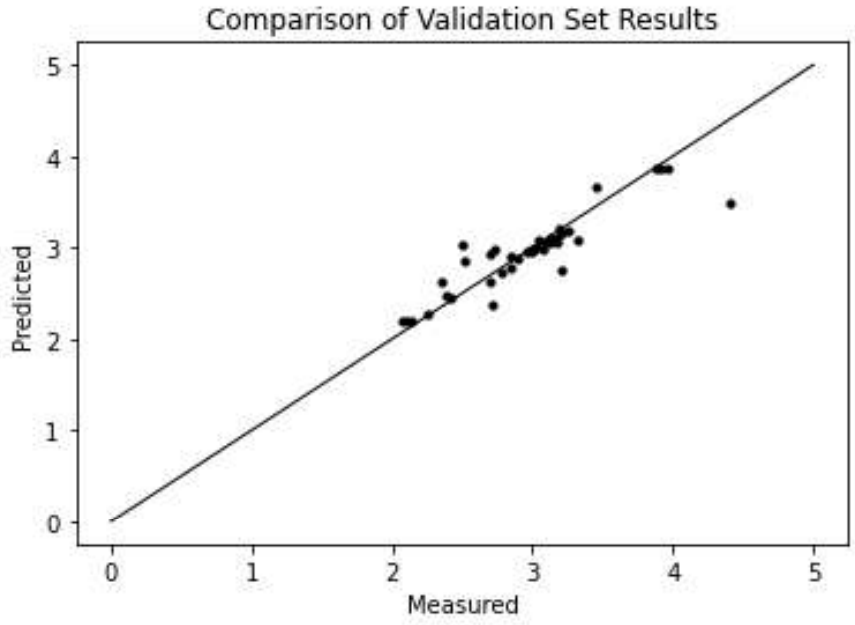

Figure 4.11. The comparison of validation set results for ANN model of band gap with the most effective variables for 52 neurons and ReLU as activation function.

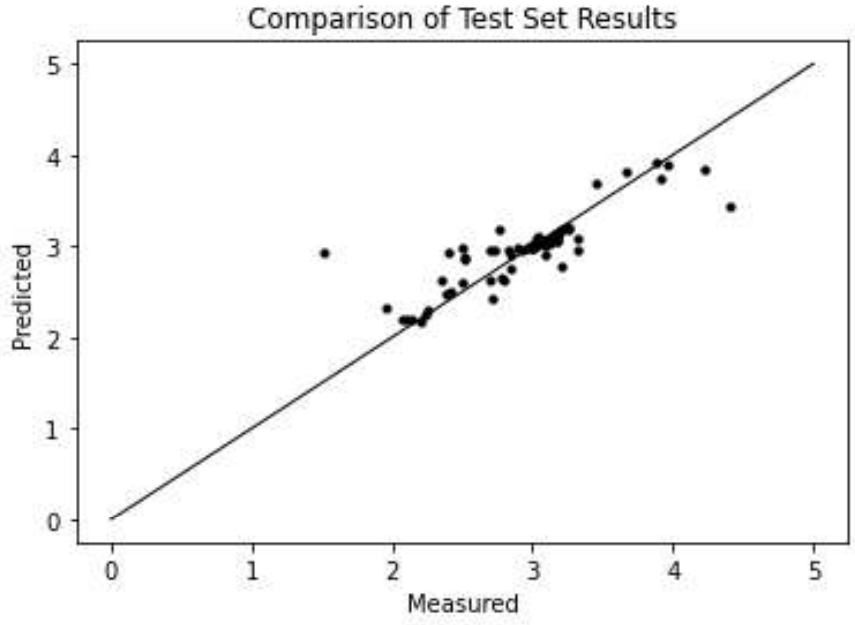

Figure 4.12. The comparison of test set results for ANN model of band gap with the most effective variables for 52 neurons and ReLU as activation function.

Like in the random forest model, the validation set, and test set gave similar results. The validation set result was slightly better than test set result since the model was constructed by using the training set of cross validation.

Table 4.11. The R^2 and RMSE result of ANN model with 52 neurons & ReLU as activation function.

Results	**R^2**	**RMSE**
Validation set results for significant variables (Figure 4.11)	0.80	0.223
Test set results for significant variables (Figure 4.12)	0.70	0.282

The neuron number was chosen as 52 because the lower neuron numbers cause underfitting and higher neuron numbers cause overfitting. Underfitting causes inaccuracy to identify signals in a complex dataset and overfitting causes insufficient training for all neurons in hidden layer because there is limited information in the training set.

By using the above model with 52 neurons and Rectified Linear Unit activation function, the missing values in the output were predicted. It is a very significant step for the remaining parts of this thesis. For hydrogen production rate model, these predicted band gap values were used to construct the model. The predicted values of band gap were kept in an Excel file to fill the missing values in the dataset for further analysis.

4.3. Analysis of Hydrogen Production Rate Models

Two different machine learning techniques were used for hydrogen production rate as well. The results for random forest and artificial neural network methods were given below.

4.3.1. Random Forest Model

The best model with highest R-squared and lowest RMSE was obtained for 81 trees and 2 samples in a leaf node. 0.25 test/train split ratio, 5-fold cross validation gave the best model.

As it can be seen, tree numbers of band gap model and hydrogen production rate model are quite different. Since there are more variables in hydrogen production rate model, the number of trees is higher than the band gap model.

Like in the band gap model, the effect of feature importance was investigated in hydrogen production rate model as well. In Figure 4.13, the most significant variables were given. As it can be seen from Figure 4.13, photocatalyst load, band gap, glycerol concentration, weight percent of cocatalyst, pH, calcination temperature, weight percent of semiconductor were the most significant variables. The percent of cocatalyst, second cocatalyst and the form of the photocatalyst were not effective as these variables. Thus, these variables were eliminated. The difference between the models, which were constructed by using all variables and by using the only significant variables were given in Figures 4.14-4.15 and 4.16-4.17 respectively.

In hydrogen production rate models, the validation and test set results were compared as well. In Figures 4.14 and 4.16, the results for validation set which were constructed by using all variables and by using the significant variables were given, respectively. In Figures 4.15 and 4.17, test set results were given. Validation set results were slightly better than the test set results since the size of validation set was slightly smaller than the size of test set.

In hydrogen production rate model, since the range for hydrogen production rate was larger than band gap, a specific area was given in Figures 4.14, 4.15, 4.16, 4.17, 4.18 and 4.19. The data points were collapsed in the range of 0 to 20,000 μmol/g.h. Thus, in each figure, the range between 0 to 20,000 μmol/g.h was given in a smaller figure to see the distribution better.

As it can be seen from both the graphs and the table, feature importance affects the results significantly. The model constructed with the most effective variables gives better results in terms of R-square and Root Mean Square Error values.

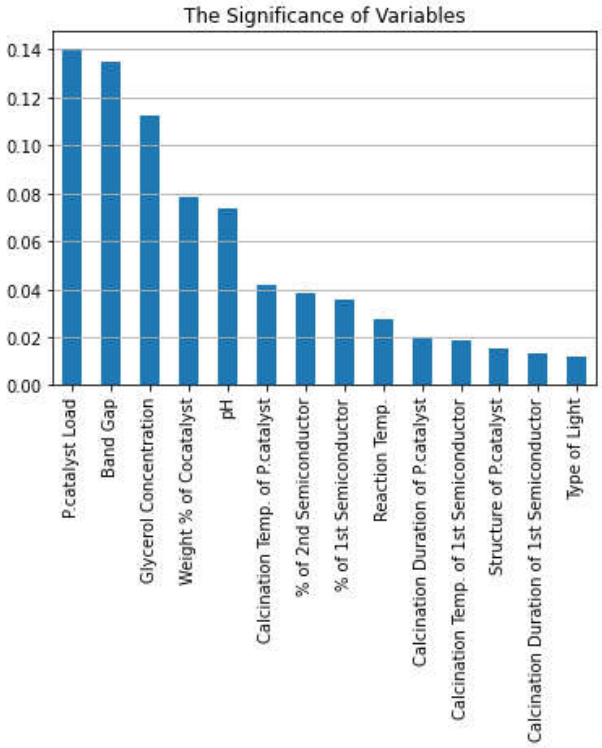

Figure 4.13. The significance of variables for hydrogen production rate model.

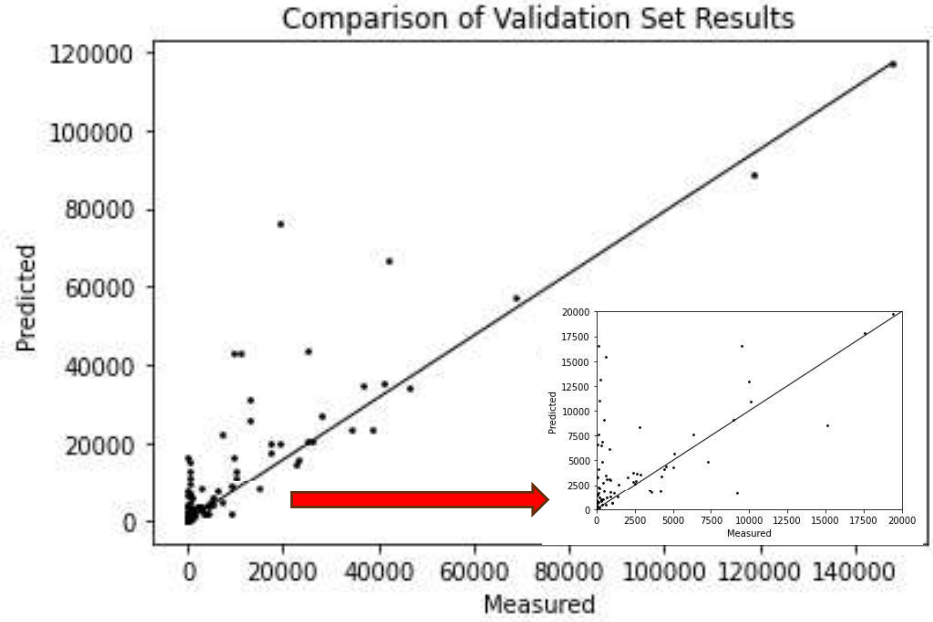

Figure 4.14. The comparison of validation set results for random forest model of hydrogen production rate with original variables for 81 trees and 2 samples in the leaf node.

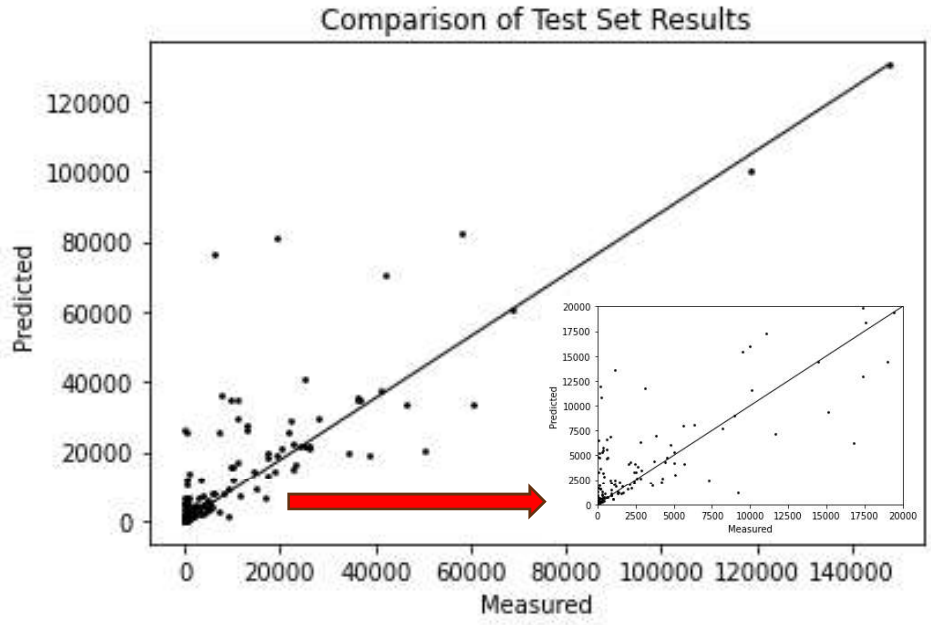

Figure 4.15. The comparison of test set results for random forest model of hydrogen production rate with original variables for 81 trees and 2 samples in the leaf node.

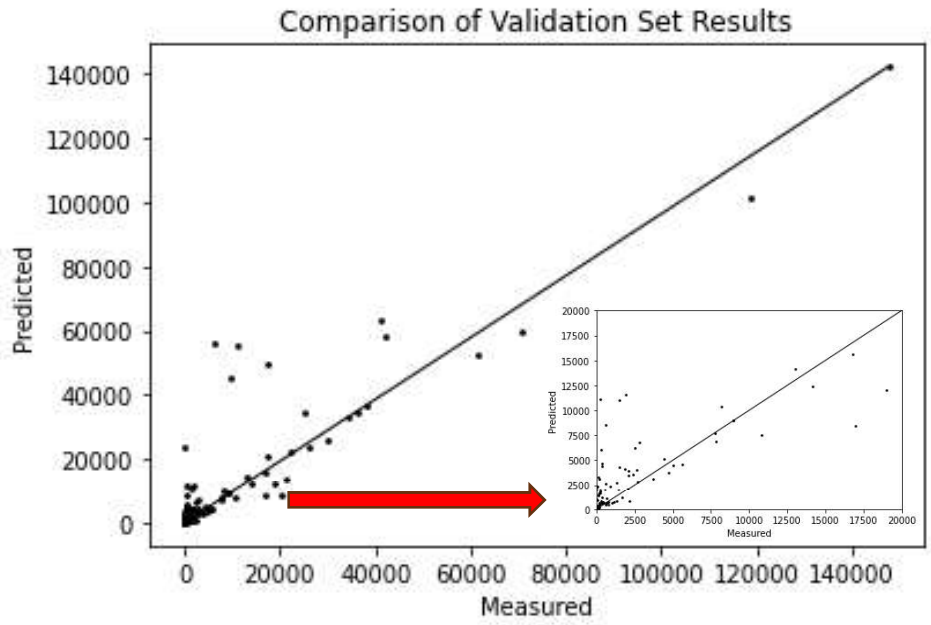

Figure 4.16. The comparison of validation set results for random forest model of hydrogen production rate with the most effective variables for 81 trees and 2 samples in the leaf node.

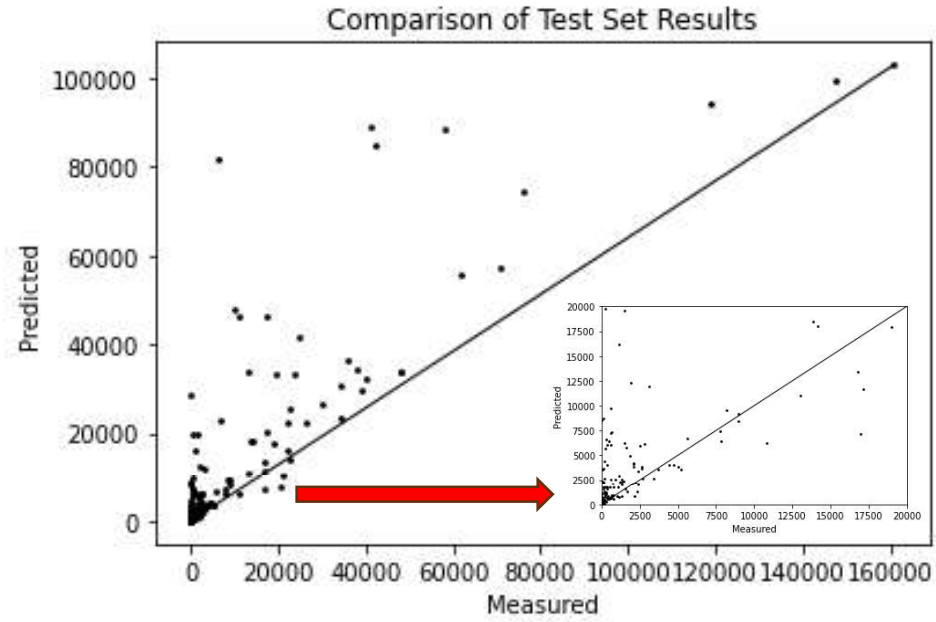

Figure 4.17. The comparison of test set results for random forest model of hydrogen production rate with the most effective variables for 81 trees and 2 samples in the leaf node.

Table 4.12. The RMSE results of random forest model with 81 trees and 2 samples in the leaf node.

Model	**R^2**	**RMSE**
Validation set results for all variables (Figure 4.14)	0.76	1.07×10^4
Test set results for all variables (Figure 4.15)	0.68	1.27×10^4
Validation set results for significant variables (Figure 4.16)	0.80	9.57×10^4
Test set results for significant variables (Figure 4.17)	0.71	1.09×10^4

In hydrogen production rate model, RMSE value range is different from band gap model. The reason behind this difference is band gap values are changing between 1.5 and 4.6 eV. However, hydrogen production rate values are changing between 0.1 and 300,000 μmol/g.h. These different ranges cause different RMSE values.

4.3.2. Artificial Neural Network Model

In artificial neural network model, the neuron number and activation function were found in a Python loop to determine the best model with the lowest RMSE and highest R-squared value. The best model was obtained from ReLU function with 63 neuron numbers by using 0.25 test/train split ratio and 4-fold cross validation.

Like in the previous Figures, validation and test set results were compared in artificial neural network model of hydrogen production rate. The results were quite similar to each other.

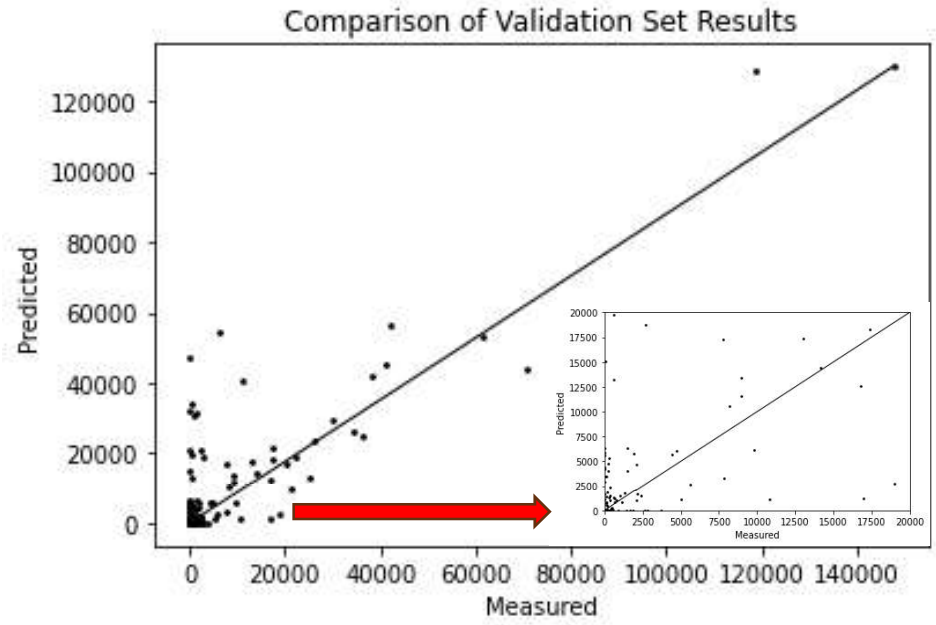

Figure 4.18. The comparison of validation set results for ANN model of hydrogen production rate with the most effective variables for 63 neurons and ReLU function.

While constructing ANN model for hydrogen production rate, more variables were used than ANN model of band gap. It affected the number of neurons for the models. The reason for using higher number of neurons in ANN model of hydrogen production rate is higher number of variables that are used during modeling.

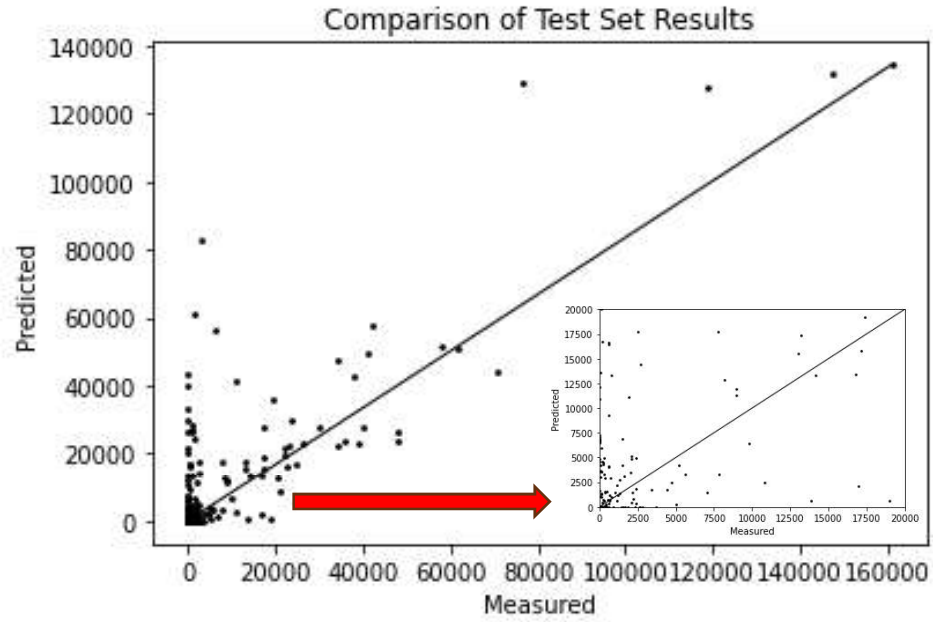

Figure 4.19. The comparison of test set results for ANN model of hydrogen production rate with the most effective variables for 63 neurons and ReLU function.

Table 4.13. The RMSE result of ANN model with 63 neurons and ReLU activation function.

Model	**R^2**	**RMSE**
Validation set results for all variables (Figure 4.18)	0.70	1.18×10^4
Test set results for all variables (Figure 4.19)	0.60	1.47×10^4

In conclusion, there were 4 models that were constructed, random forest model for band gap, ANN model for band gap, random forest model for hydrogen production rate and ANN model for hydrogen production rate. For each model, the best option was found by using a Python loop. For random forest models, the optimum tree number and sample number in a leaf node were determined. For ANN models, the optimum neuron number and activation function were found. By using these values, the best models with low Root Mean Square Errors were calculated. The calculated values were given in Table 4.15, as a summary.

Table 4.14. The R^2 and RMSE results of all models.

Model	**R^2**	**RMSE**
Results of validation set with all variables for RF band gap model (Figure 4.7)	0.77	0.216
Results of test set with all variables for RF band gap model (Figure 4.8)	0.68	0.305
Results of validation set with significant variables for RF band gap model (Figure 4.9)	0.83	0.206
Results of test set with significant variables for RF band gap model (Figure 4.10)	0.73	0.234
Validation set results for significant variables for ANN band gap model (Figure 4.11)	0.80	0.223
Test set results for significant variables for ANN band gap model (Figure 4.12)	0.70	0.282
Validation set results for all variables for RF hydrogen production rate model (Figure 4.14)	0.76	1.07×10^4
Test set results for all variables for RF hydrogen production rate model (Figure 4.15)	0.68	1.27×10^4
Validation set results for significant variables for RF hydrogen production rate model (Figure 4.16)	0.80	9.57×10^4
Test set results for significant variables for RF hydrogen production rate model (Figure 4.17)	0.71	1.09×10^4
Validation set results for all variables for ANN hydrogen production rate model (Figure 4.18)	0.70	1.18×10^4
Test set results for all variables for ANN hydrogen production rate model (Figure 4.19)	0.60	1.47×10^4

5. CONCLUSION AND RECOMMENDATIONS

5.1. Conclusion

The aim of this thesis is to extract knowledge from published literature about photocatalytic reforming of glycerol. Approximately 150 articles, which were published between 2007 and 2022, were reviewed on Web of Science. While constructing the database, 93 of 150 articles were used since the remaining part was not appropriate for the aim of this thesis. In these 93 articles, 791 data points were collected. In dataset, semiconductor, cocatalyst, weight percent of cocatalyst, structure and form of the photocatalyst, preparation methods, calcination temperature and duration, light type, experimental conditions such as solution volume, reaction temperature, band gap and hydrogen production rate were given. Python was used for analyzing and modeling the dataset. Band gap and hydrogen production rate were analyzed in separate parts. Random forest and artificial neural network models were used with k-fold cross validation. R-square and Root Mean Square Error were used as the measure of model performance.

In band gap model, Random Forest method was used to construct the model. In a Python loop, the number of trees were changed between 1 to 300 by 5. Also, the minimum sample in a leaf node was changed between 1 to 5. For each combination, the results were collected by changing test/train split ratio into the range of 20-30%, and k value for k-fold cross validation between 3, 4, 5, 7, and 10. The model with the best performance was determined. 0.25 test/train split ratio, 4-fold cross validation, 41 trees with 1 sample in a leaf node gave the best result with the lowest RMSE and highest R-squared value. Feature importance was applied to the dataset and the results were given in Figure 4.6. The most important variables were determined as weight percent of cocatalyst, percent of semiconductor and calcination temperature and duration. The least important variables were eliminated which were the second cocatalyst and the weight percent of cocatalyst in cocatalyst composite.

In ANN model of band gap, the hidden layer number is chosen between 1 to 5. To find the best model, different neuron numbers and different activation functions were analyzed by changing test/train split ratio into the range of 20-30%, and k value for k-fold cross validation between 3, 4, 5, 7, and 10. The number of neurons was changed between 2 to 100. Linear, sigmoid, tanh and rectified linear unit (ReLU) functions were used as activation function. For each combination, RMSE values were calculated. The model with the lowest Root Mean Square Error and highest R-squared value was chosen as the best model. 0.3 test/train split ratio, 4-fold cross validation, 52 neuron number and ReLU activation function gave the best result.

The missing values in band gap was predicted by using artificial neural network model. These predicted values were used in hydrogen production rate model. The other missing values were filled with the average mean of the training set of the corresponding variable like the previous analyses.

In random forest model of hydrogen production rate, the best model was obtained in a Python loop by changing the number of trees and number of samples in a leaf node. The range for the number of trees was taken between 1 to 300. The range of the number of samples in a leaf node was taken as 1 to 5. For each combination, R-squared value and Root Mean Square Error were calculated by changing test/train split ratio into the range of 20-30%, and k value for k-fold cross validation between 3, 4, 5, 7, and 10. The best model was constructed with 0.25 test/train split ratio, 5-fold cross validation, 81 trees and 2 samples in a leaf node.

In hydrogen production rate model, the significant variables were determined by using feature importance as well. In Figure 4.13, photocatalyst load, band gap, glycerol concentration, weight percent of cocatalyst, pH, calcination temperature were the most effective variables on the model. The least important variables such as the second cocatalyst, the percent of cocatalyst in a cocatalyst composite and the form of the photocatalyst were eliminated.

For both band gap and hydrogen production rate models, the model performance increases with the use of effective variables only and the model with the most significant variables becomes more successful to predict the output.

In ANN model of hydrogen production, different numbers of neuron and different activation functions were analyzed as well to find the best model. The number of neurons was changed between 2 to 100. Linear, sigmoid, tanh and rectified linear unit (ReLU) functions were used as activation function. For each combination, R-squared and RMSE values were calculated by changing test/train split ratio into the range of 20-30%, and k value for k-fold cross validation between 3, 4, 5, 7, and 10. The model with the lowest Root Mean Square Error and highest R-squared value was chosen as the best model. 63 neuron number and ReLU activation function gave the best result with 0.25 test/train split ratio and 4-fold cross validation.

5.2. Recommendations

The better models can be obtained with the following recommendations:

- More data can be used to construct the models. The model performance will increase with the increasing amount of data.

- Different machine learning methods can be used to compare the results. The prediction of output can be developed with other machine learning techniques.

REFERENCES

1. He, Q., J. McNutt and J. Yang, "Utilization of the Residual Glycerol from Biodiesel Production for Renewable Energy Generation", *Renewable and Sustainable Energy Reviews*, Vol. 71, pp. 63-76, 2017.

2. Sun, D., Y. Yamada, S. Sato and W. Ueda, "Glycerol Hydrogenolysis into Useful C3 Chemicals", *Applied Catalysis B: Environmental*, Vol. 193, pp. 75-92, 2016.

3. Veluturla, S., N. Archna, D. S. Rao, N. Hezil, I. S. Indraja and S. Spoorthi, "Catalytic Valorization of Raw Glycerol Derived from Biodiesel: A Review", *Biofuels*, Vol. 9, No. 3, pp. 305-314, 2016.

4. Ayodele, B. V., T. R. Abdullah, M. A. Alsaffar, S. I. Mustapa and S. F. Salleh, "Recent Advances in Renewable Hydrogen Production by Thermo-Catalytic Conversion of Biomass-Derived Glycerol: Overview of Prospects and Challenges", *International Journal of Hydrogen Energy*, Vol. 45, No. 36, pp. 18160-18185, 2020.

5. Estahbanati, M. R. K., M. Feilizadeh, F. Attar and M. C. Iliuta, "Current Developments and Future Trends in Photocatalytic Glycerol Valorization: Photocatalyst Development", *Industrial & Engineering Chemistry Research*, Vol. 59, No. 52, pp. 22330-22352, 2020.

6. Navarro, R. M., M. A. Peña and J. L. G. Fierro, "Hydrogen Production Reactions from Carbon Feedstocks: Fossil Fuels and Biomass", *Chemical Reviews*, Vol. 107, No. 10, pp. 3952-3991, 2007.

7. Tabassum, N., R. Pothu, A. Pattnaik, R. Boddula, P. Balla, R. Gundeboyina, P. Challa, R. Rajesh, V. Perugopu and N. Mameda, "Heterogeneous Catalysts for Conversion of Biodiesel-Waste Glycerol into High-Added-Value Chemicals", *Catalysts*, Vol. 12, No. 7, p. 767, 2022.

8. Altintas, C., O. F. Altundal, S. Keskin and R. Yildirim, "Machine Learning Meets with Metal Organic Frameworks for Gas Storage and Separation", *Journal of Chemical Information and Modeling*, Vol. 61, No. 5, pp. 2131-2146, 2021.

9. Gunay, M. E. and R. Yildirim, "Recent Advances in Knowledge Discovery for Heterogeneous Catalysis Using Machine Learning", *Catalysis Reviews*, Vol. 63, No. 1, pp. 120-164, 2020.

10. Larose, D. T. and C. D. Larose, *Discovering Knowledge in Data*, Second Edition, John Wiley & Sons, New Jersey, 2014.

11. Schonlau, M. and R. Y. Zou, "The Random Forest Algorithm for Statistical Learning", *the Stata Journal*, Vol. 20, No. 1, pp. 3-29, 2020.

12. Alpaydın, E. *Introduction to Machine Learning*, Third Edition, The MIT Press, Massachusetts, 2014.

13. Puga, A. V., "Photocatalytic Production of Hydrogen from Biomass-Derived Feedstocks", *Coordination Chemistry Reviews*, Vol. 315, pp. 1-66, 2016.

14. Kumar, M. K., B. S. Kwak, A. K. R. Police, M. Kang, "In-Situ Photo-Reduction of Silver Particles and Their SPR Effect in Enhancing the Photocatalytic Water Splitting of Ag_2O/TiO_2 Photocatalysts under Solar Light Irradiation: A Case Study", *Materials Research Bulletin*, Vol. 95, pp. 515-524, 2017.

15. Tahir, M. and N. Fajrina, "Monolithic Ag-Mt Dispersed Z-Scheme pCN-TiO_2 Heterojunction for Dynamic Photocatalytic H_2 Evolution Using Liquid and Gas Phase Photoreactors", *International Journal of Hydrogen Energy*, Vol. 45, No. 7, pp. 4355-4375, 2020.

16. Beltram, A., I. R. Ocaña, J. J. D. Jaen and T. Montini, "Photocatalytic Valorization of Ethanol and Glycerol over TiO_2 Polymorphs for Sustainable Hydrogen Production", *Applied Catalysis A: General*, Vol. 518, pp. 167-175, 2016.

17. Gholipour, M. R., C. T. Dinh, F. Béland and T. O. Do, "Nanocomposite Heterojunctions as Sunlight-Driven Photocatalysts for Hydrogen Production from Water Splitting", *Royal Society of Chemistry*, Vol 7, No. 18, pp. 8187-8208, 2015.

18. Rudd, O. P., *Data Mining Cookbook*, John Wiley & Sons, New York, 2000.

19. Oral B., E. Can and R. Yıldırım, "Analysis of Photoelectrochemical Water Splitting Using Machine Learning", *International Journal of Hydrogen Energy*, Vol. 47, No. 45, pp. 19633-19654, 2022.

20. Yılmaz, B. and R. Yıldırım, "Critical Review of Machine Learning Applications in Perovskite Solar Research", *Nano Energy*, Vol. 80, p. 105546, 2021.

21. Odabaşı, Ç. and R. Yıldırım, "Machine Learning Analysis on Stability of Perovskite Solar Cells", *Solar Energy Materials and Solar Cells*, Vol. 205, p. 110284, 2020.

22. Liu, Y., Y. Wang and J. Zhang, "New Machine Learning Algorithm: Random Forest", in *International Conference on Information Computing and Applications*, Berlin, Germany, 2012.

23. Shiruru, K., "An Introduction to Artificial Neural Network", *International Journal of Advance Research and Innovative Ideas in Education*, Vol. 1, No. 5, pp. 27-30, 2016.

24. López, O. A. M., A. M. López and J. Crossa, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, Springer, Mexico, 2022.

25. Siregar, S. P. and A. Wanto, "Analysis Accuracy of Artificial Neural Network Using Backpropagation Algorithm in Predicting Process (Forecasting)", *International Journal of Information System & Technology*, Vol. 1, No. 1, pp. 34-42, 2017.

26. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss and V. Dubourg, "Scikit-learn: Machine Learning

in Python", *Journal of Machine Learning Research*, Vol. 12, No. 85, pp. 2825-2830, 2011.

27. McKinney, W., "Data Structures for Statistical Computing in Python", in *Proceedings of the 9th Python in Science Conference*, Texas, Vol. 445, pp. 51-56, 2010.

28. Willmott, C. J. and K. Matsuura, "Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance", *Climate Research*, Vol. 30, No. 1, pp. 79-82, 2005.

29. Kasuya, E., "On the Use of R and R Squared in Correlation and Regression", *Ecological Research*, Vol. 34, No. 1, pp. 235-236, 2018.

30. Harris, C. R., K. J. Millman, S. J. Vanderwalt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg and N. J. Smith, "Array Programming with NumPy", *Nature*, Vol. 585, pp. 357-362, 2020.

31. Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean and M. Devin, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems", in *12th USENIX Conference on Operating Systems Design and Implementation Proceedings*, Savannah, pp. 265-283, 2016.

32. Chu, S., E. Wang, F. Feng, C. Zhang, J. Jiang, Q. Zhang, F. Wang, L. Bing, G. Wang and D. Han, "A Review of Noble Metal Catalysts for Catalytic Removal of VOCs", *Catalysts*, Vol. 12, p. 1543, 2022.

33. Chen, W., Y. Wang, S. Liu, L. Gao, L. Mao, Z. Fan, W. Shangguan and Z. Jiang, "Non-Noble Metal Cu as a Cocatalyst on TiO_2 Nanorod for Highly Efficient Photocatalytic Hydrogen Production", *Applied Surface Science*, Vol. 445, pp. 527-534, 2018.

34. Reddy, N. L., D. P. Kumar and M. V. Shankar, "Co-Catalyst Free Titanate Nanorods for Improved Hydrogen Production under Solar Light Irradiation", *Journal of Chemical Sciences*, Vol. 128, No. 4, pp. 649-656, 2016.

35. Manukumar, K. N., G. Nagaraju, D. P. Kumar and M. V. Shankar, "Facile Ionothermal Synthesis of TiO_2 Nanorods for Photocatalytic H_2 Generation", *Journal of Materials Science: Materials in Electronics*, Vol. 30, No. 2, pp. 1076-1083, 2018.

36. Kumar, D. P., V. D. Kumari, M. Karthik, M. Sathish and M. V. Shankar, "Shape Dependence Structural, Optical and Photocatalytic Properties of TiO_2 Nanocrystals for Enhanced Hydrogen Production via Glycerol Reforming", *Solar Energy Materials and Solar Cells*, Vol. 163, pp. 113-119, 2017.

37. Daskalaki, V. M. and D. I. Kondarides, "Efficient Production of Hydrogen by Photo-Induced Reforming of Glycerol at Ambient Conditions", *Catalysis Today*, Vol. 144, No. 1-2, pp. 75-80, 2009.

38. Majrik, K., Z. Pászti, L. Korecz, L. Trif, A. Domján, G. Bonura, C. Cannilla, F. Frusteri, A. Tompos and E. Tálas, "Study of PtO_x/TiO_2 Photocatalysts in the Photocatalytic Reforming of Glycerol: The Role of Co-Catalyst Formation", *Materials*, Vol. 11, No. 10, p. 1927, 2018.

39. Mizukoshi, Y., "Effects of Sonication on Photocatalytic Reforming of Aqueous Glycerol Solutio", *Ultrasonics Sonochemistry*, Vol. 51, pp. 182-185, 2019.

40. Kondarides, D. I., V. M. Daskalaki, A. Patsoura and X. E. Verykios, "Hydrogen Production by Photo-Induced Reforming of Biomass Components and Derivatives at Ambient Conditions", *Catalysis Letters*, Vol. 122, No. 1-2, pp. 26-32, 2007.

41. Fu, N. and G. Lu, "Photo-Catalytic H_2 Evolution over a Series of Keggin-Structure Heteropoly Blue Sensitized Pt/TiO_2 under Visible Light Irradiation", *Applied Surface Science*, Vol. 255, No. 8, pp. 4378-4383, 2009.

42. Li, M., Y. Li, S. Peng, G. Lu and S. Li, "Photocatalytic Hydrogen Generation Using Glycerol Wastewater over Pt/TiO_2", *Frontiers of Chemistry in China*, Vol. 4, No. 1, pp. 32-38, 2009.

43. Daskalaki, V. M., P. Panagiotopoulou and D. I. Kondarides, "Production of Peroxide Species in Pt/TiO_2 Suspensions under Conditions of Photocatalytic Water Splitting and Glycerol Photoreforming", *Chemical Engineering Journal*, Vol. 170, No. 2-3, pp. 433-439, 2011.

44. Fu, X., X. Wang, D. Y. C. Leung, Q. Gu, S. Chen and H. Huang, "Photocatalytic Reforming of C3-Polyols for H_2 Production: Part (I). Role of Their OH Groups", *Applied Catalysis B: Environmental*, Vol. 106, No. 3-4, pp. 681-688, 2011.

45. Beltram, A., I. Romero-Ocaña, J. Josè Delgado Jaen, T. Montini and P. Fornasiero, "Photocatalytic Valorization of Ethanol and Glycerol over TiO_2 Polymorphs for Sustainable Hydrogen Production", *Applied Catalysis A: General*, Vol. 518, pp. 167-175, 2016.

46. Melián, E. P., C. R. López, D. E. Santiago, R. Quesada-Cabrera, J. A. O. Méndez, J. M. D. Rodríguez and O. G. Díaz, "Study of the Photocatalytic Activity of Pt-Modified Commercial TiO_2 for Hydrogen Production in the Presence of Common Organic Sacrificial Agents", *Applied Catalysis A: General*, Vol. 518, pp. 189-197, 2016.

47. Jiang, X., X. Fu, L. Zhang, S. Meng and S. Chen, "Photocatalytic Reforming of Glycerol for H_2 Evolution on Pt/TiO_2: Fundamental Understanding the Effect of Co-Catalyst Pt and the Pt Deposition Route", *Journal of Materials Chemistry A*, Vol. 3, No. 5, pp. 2271-2282, 2015.

48. Estahbanati, M. R. K., N. Mahinpey, M. Feilizadeh, F. Attar and M. C. Iliuta, "Kinetic Study of the Effects of pH on the Photocatalytic Hydrogen Production from Alcohols", *International Journal of Hydrogen Energy*, Vol. 44, No. 60, pp. 32030-32041, 2019.

49. Estahbanati, M. R. K., M. Feilizadeh and M. C. Iliuta, "Photocatalytic Valorization of Glycerol to Hydrogen: Optimization of Operating Parameters by Artificial Neural Network", *Applied Catalysis B: Environmental*, Vol. 209, pp. 483-492, 2017.

50. Estahbanati, M. R. K., M. Feilizadeh and M. C. Iliuta, "An Intrinsic Kinetic Model for Liquid-Phase Photocatalytic Hydrogen Production", *American Institute of Chemical Engineers Journal*, Vol. 65, No. 11, p. 16724, 2019.

51. Chen, W. T., A. Chan, Z. H. N. Al-Azri, A. G. Dosado, M. A. Nadeem, D. Sun-Waterhouse, H. Idriss and G. I. N. Waterhouse, "Effect of TiO_2 Polymorph and Alcohol Sacrificial Agent on the Activity of Au/TiO_2 Photocatalysts for H_2 Production in Alcohol–Water Mixtures", *Journal of Catalysis*, Vol. 329, pp. 419-513, 2015.

52. Dosado, A. G., W. T. Chen, A. Chan, D. Sun-Waterhouse and G. I. N. Waterhouse, "Novel Au/ TiO_2 Photocatalysts for Hydrogen Production in Alcohol–Water Mixtures based on Hydrogen Titanate Nanotube Precursors", *Journal of Catalysis*, Vol. 330, pp. 238-254, 2015.

53. Xu, D., Y. Hai, X. Zhang, S. Zhang and R. He, "Bi_2O_3 Cocatalyst Improving Photocatalytic Hydrogen Evolution Performance of TiO_2", *Applied Surface Science*, Vol. 400, pp. 530-536, 2017.

54. Sadanandam, G., K. Lalitha, V. D. Kumari, M. V. Shankar and M. Subrahmanyam, "Cobalt doped TiO_2: A Stable and Efficient Photocatalyst for Continuous Hydrogen Production from Glycerol: Water Mixtures under Solar Light Irradiation", *International Journal of Hydrogen Energy*, Vol. 38, No. 23, pp. 9655-9664, 2013.

55. Reddy, N. L., K. K. Cheralathan, V. D. Kumari, B. Neppolian and S. M. Venkatakrishnan, "Photocatalytic Reforming of Biomass Derived Crude Glycerol in Water: a Sustainable Approach for Improved Hydrogen Generation Using $Ni(OH)_2$ Decorated TiO_2 Nanotubes under Solar Light Irradiation", *ACS Sustainable Chemistry & Engineering*, Vol. 6, No. 3, pp. 3754-3764, 2018.

56. Liu, R., H. Yoshida, S. Fujita and M. Arai, "Photocatalytic Hydrogen Production from Glycerol and Water with NiO_x/TiO_2 Catalysts", *Applied Catalysis B: Environmental*, Vol. 144, pp. 41-45, 2014.

57. Lalitha, K., G. Sadanandam, V. D. Kumari, M. Subrahmanyam, B. Sreedhar and N. Y. Hebalkar, "Highly Stabilized and Finely Dispersed Cu_2O/TiO_2: a Promising Visible Sensitive Photocatalyst for Continuous Production of Hydrogen from Glycerol: Water Mixtures", *The Journal of Physical Chemistry C*, Vol. 114, No. 50, pp. 22181-22189, 2010.

58. Montini, T., V. Gombac, L. Sordelli, J. J. Delgado, X. Chen, G. Adami and P. Fornasiero, "Nanostructured Cu/TiO_2 Photocatalysts for H_2 Production from Ethanol and Glycerol Aqueous Solutions", *Chemcatchem: The European Society of Journal for Catalysis*, Vol. 3, No. 3, pp. 574-577, 2011.

59. Bashiri, R., N. M. Mohamed, C. F. Kait and S. Sufian, "Hydrogen Production from Water Photosplitting using Cu/TiO_2 Nanoparticles: Effect of Hydrolysis Rate and Reaction Medium", *International Journal of Hydrogen Energy*, Vol. 40, No. 18, pp. 6021-6037, 2015.

60. Kumar, D. P, N. L. Reddy, M. M. Kumari, B. Srinivas, V. D. Kumari, B. Sreedhar, V. Roddatis, O. Bondarchuk, M. Karthik and B. Neppolian, "Cu_2O-Sensitized TiO_2 Nanorods with Nanocavities for Highly Efficient Photocatalytic Hydrogen Production under Solar Irradiation", *Solar Energy Materials and Solar Cells*, Vol. 136, pp. 157-166, 2015.

61. Zhang, M., R. Sun, Y. Li, Q. Shi, L. Xie, J. Chen, X. Xu, H. Shi and W. Zhao, "High H_2 Evolution from Quantum Cu (II) Nanodot-Doped Two-Dimensional Ultrathin TiO_2 Nanosheets with Dominant Exposed {001} Facets for Reforming Glycerol with Multiple Electron Transport Pathways", *The Journal of Physical Chemistry C*, Vol. 120, No. 20, pp. 10746-10756, 2016.

62. Kumar, D. P., M. V. Shankar, M. M. Kumari, G. Sadanandam, B. Srinivas and V. Durgakumari, "Nano-Size Effects on CuO/TiO_2 Catalysts for Highly Efficient H_2 Production under Solar Light Irradiation", *Chemical Communications*, Vol. 49, No. 82, pp. 9443-9445, 2013.

63. Lin, Y., S. Yang, Y. Liu, S. Zhang, H. Wang, H. Yu and F. Peng, "In-Situ Photo-Deposition CuO_{1-x} Cluster on TiO_2 for Enhanced Photocatalytic H_2-Production Activity", *International Journal of Hydrogen Energy*, Vol. 42, No. 31, pp. 19942-19950, 2017.

64. Ibrahim, S., Y. Cheng, D. Zhao and M. A. Nadeem, "A New Insight for Photocatalytic Hydrogen Production by a Cu/Ni Based Cyanide Bridged Polymer as a Cocatalyst on Titania Support in Glycerol Water Mixture", *International Journal of Hydrogen Energy*, Vol. 44, No. 5, pp. 2508-2518, 2019.

65. Ibrahim, S., I. Majeed, Y. Qian, A. Iqbal, D. Zhao, D. R. Turner and M. A. Nadeem, "Novel Hetero-Bimetallic Coordination Polymer as a Single Source of Highly Dispersed Cu/Ni Nanoparticles for Efficient Photocatalytic Water Splitting", *Inorganic Chemistry Frontiers*, Vol. 5, No. 8, pp. 1816-1827, 2018.

66. López-Tenllado, F. J., J. Hidalgo-Carrillo, V. Montes, A. Marinas, F. J. Urbano, J. M. Marinas, L. Ilieva, T. Tabakova and F. Reid, "A Comparative Study of Hydrogen Photocatalytic Production from Glycerol and Propan-2-Ol on M/TiO_2 Systems (M=Au, Pt, Pd)", *Catalysis Today*, Vol. 280, pp. 58-64, 2017.

67. Al-Azri, Z. H. N., W. T. Chen, A. Chan, V. Jovic, T. Ina, H. Idriss and G. I. N. Waterhouse, "The Roles of Metal Co-Catalysts and Reaction Media in Photocatalytic Hydrogen Production: Performance Evaluation of M/TiO_2 Photocatalysts (M = Pd, Pt, Au) in Different Alcohol–Water Mixtures", *Journal of Catalysis*, Vol. 329, pp. 355-367, 2015.

68. Bednarczyk, K., M. Stelmachowski and M. Gmurek, "The Influence of Process Parameters on Photocatalytic Hydrogen Production", *Environmental Progress & Sustainable Energy*, Vol. 38, No. 2, pp. 680-687, 2018.

69. Wang, C., X. Cai, Y. Chen, Z. Cheng, X. Luo, S. Mo, L. Jia, P. Lin and Z. Yang, "Improved Hydrogen Production from Glycerol Photoreforming over Sol-Gel Derived

TiO_2 Coupled with Metal Oxides", *Chemical Engineering Journal*, Vol. 317, pp. 522-532, 2017.

70. Seadira, T. W. P., G. Sadanandam, T. Ntho, C. M. Masuku and M. S. Scurrell, "Preparation and Characterization of Metals Supported on Nanostructured TiO_2 Hollow Spheres for Production of Hydrogen via Photocatalytic Reforming of Glycerol", *Applied Catalysis B: Environmental*, Vol. 222, pp. 133-145, 2018.

71. Chen, W. T., A. G. Dosado, A. Chan, D. S. Waterhouse and G. I. N. Waterhouse, "Highly Reactive Anatase Nanorod Photocatalysts Synthesized by Calcination of Hydrogen Titanate Nanotubes: Effect of Calcination Conditions on Photocatalytic Performance for Aqueous Dye Degradation and H_2 Production in Alcohol-Water Mixtures", *Applied Catalysis A: General*, Vol. 565, pp. 98-118, 2018.

72. Taylor, S., M. Mehta and A. Samokhvalov, "Production of Hydrogen by Glycerol Photoreforming Using Binary Nitrogen–Metal-Promoted N-M-TiO_2 Photocatalysts", *Chemphyschem: A European Journal of Chemical Physics and Physical Chemistry*, Vol. 15, No. 5, pp. 942-949, 2014.

73. Mohamed, R. M., M. W. Kadi and A. A. Ismail, "A Facile Synthesis of Mesoporous α- Fe_2O_3/TiO_2 Nanocomposites for Hydrogen Evolution under Visible Light", *Ceramics International*, Vol. 46, No. 10, pp. 15604-15612, 2020.

74. Gullapelli, S. M. S. Scurrell and D. K. Valluri, "Photocatalytic H_2 Production from Glycerol–Water Mixtures over Ni/γ-Al_2O_3 and TiO_2 Composite Systems", *International Journal of Hydrogen Energy*, Vol. 42, No. 22, pp. 15031-15043, 2017.

75. Kumar, M. K., J. Y. Do, P. A. K. Reddy and M. Kang, "Natural Solar Light-Driven Preparation of Plasmonic Resonance-based Alloy and Core-Shell Catalyst for Sustainable Enhanced Hydrogen Production: Green Approach and Characterization", *Applied Catalysis B: Environmental*, Vol. 231, pp. 137-150, 2018.

76. Ravi, P., V. N. Rao, M. V. Shankar and M. Sathish, "CuO@NiO Core-Shell Nanoparticles Decorated Anatase TiO_2 Nanospheres for Enhanced Photocatalytic Hydrogen Production", *International Journal of Hydrogen Energy*, Vol. 45, No. 3, pp. 7517-7529, 2020.

77. Reddy, N. R., M. M. Kumari, K. K. Cheralathan and M. V. Shankar, "Enhanced Photocatalytic Hydrogen Production Activity of Noble Metal Free MWCNT-TiO_2 Nanocomposites", *International Journal of Hydrogen Energy*, Vol. 43, No. 8, pp. 4036-4043, 2018.

78. Estahbanati, M. R. K., M. Feilizadeh, M. S. Yancheshmeh and M. C. Iliuta, "Effects of Carbon Nanotube and Carbon Sphere Templates in TiO_2 Composites for Photocatalytic Hydrogen Production", *Industrial & Engineering Chemistry Research*, Vol. 58, No. 8, pp. 2770-2783, 2019.

79. Tahir, M., "La-Modified TiO_2/Carbon Nanotubes Assembly Nanocomposite for Efficient Photocatalytic Hydrogen Evolution from Glycerol-Water Mixture", *International Journal of Hydrogen Energy*, Vol. 44, No. 7, pp. 3711-3725, 2019.

80. Ribao, P., M. A. Esteves, V. R. Fernandes, M. J. Rivero, C. M. Rangel and I. Ortiz, "Challenges Arising from the Use of TiO_2/rGO/Pt Photocatalysts to Produce Hydrogen from Crude Glycerol Compared to Synthetic Glycerol", *International Journal of Hydrogen Energy*, Vol. 44, No. 53, pp. 28494-28506, 2019.

81. Seadira, T. W. P, S. J. Baloyi, C. M. Masuku and M. S. Scurrell, "Solar Photocatalytic Hydrogen Production from Glycerol Reforming Using Ternary Cu/THS/Graphene", in *Conference of the South African Advanced Materials Initiative*, Vanderbijlpark, South Africa, 2020.

82. Nguyen, N. T., D. D. Zheng, S. S. Chen, C. T. Chang, C. M. Ma, N. C. Nguyen, H. T. Nguyen, S. S. Hsiao and S. Y. Kang, "Effects of Different Sacrificial Agents and Hydrogen Production from Wastewater by Pt-Graphene/TiO_2", *Journal of Nanoscience and Nanotechnology*, Vol. 18, No. 5, pp. 3563-3570, 2018.

83. Hafeez, H. Y., S. K. Lakhera, P. Karthik, M. Anpo and B. Neppolian, "Facile Construction of Ternary $CuFe_2O_4$-TiO_2 Nanocomposite Supported Reduced Graphene Oxide (rGO) Photocatalysts for the Efficient Hydrogen Production", *Applied Surface Science*, Vol. 449, pp. 772-779, 2018.

84. Hafeez, H. Y., S. K. Lakhera, S. Bellamkonda, G. R. Rao, M. V. Shankar, D. W. Bahnemann and B. Neppolian, "Construction of Ternary Hybrid Layered Reduced Graphene Oxide Supported g-C_3N_4-TiO_2 Nanocomposite and Its Photocatalytic Hydrogen Production Activity", *International Journal of Hydrogen Energy*, Vol. 43, No. 8, pp. 3892-3904, 2018.

85. Sadanandam, G., L. Zhang and M. S. Scurrell, "Enhanced Photocatalytic Hydrogen Formation over Fe-Loaded TiO_2 and g-C3N4 Composites from Mixed Glycerol and Water by Solar Irradiation", *Journal of Renewable and Sustainable Energy*, Vol. 10, No. 3, p. 034703, 2018.

86. Babu, S. G., R. Vinoth, D. P. Kumar, M. V. Shankar, H. L. Chou, K. Vinodgopal and B. Neppolian, "Influence of Electron Storing, Transferring and Shuttling Assets of Reduced Graphene Oxide at the Interfacial Copper Doped TiO_2 P-N Heterojunction for Increased Hydrogen Production", *Nanoscale*, Vol. 7, No. 17, pp. 7849-7875, 2015.

87. Sadanandam, G., D. K. Valluri and M. S. Scurrell, "Highly Stabilized Ag_2O-Loaded Nano TiO_2 for Hydrogen Production from Glycerol: Water Mixtures under Solar Light Irradiation", *International Journal of Hydrogen Energy*, Vol. 42, No. 2, pp. 807-820, 2017.

88. de Oliveira Melo, M. and L. A. Silva, "Visible Light-Induced Hydrogen Production from Glycerol Aqueous Solution on Hybrid Pt–CdS– TiO_2 Photocatalysts", *Journal of Photochemistry and Photobiology A: Chemistry*, Vol. 226, No. 1, pp. 36-41, 2011.

89. Bastos, S. A. L., P. A. L. Lopes, F. N. Santos and L. A. Silva, "Experimental Design as a Tool to Study the Reaction Parameters in Hydrogen Production from

Photoinduced Reforming of Glycerol over CdS Photocatalyst", *International Journal of Hydrogen Energy*, Vol. 39, No. 27, pp. 14588-14595, 2014.

90. Wang, J. J., Z. J. Li, X. B. Li, X. B. Fan, Q. Y. Meng, S. Yu, C. B. Li, J. X. Li, C. H. Tung and L. Z. Wu, "Photocatalytic Hydrogen Evolution from Glycerol and Water over Nickel-Hybrid Cadmium Sulfide Quantum Dots under Visible-Light Irradiation", *Chemsuschem: Chemistry and Sustainability, Energy and Materials*, Vol, 7, No. 5, pp. 1468-1475, 2014.

91. Bao, D., P. Gao, X. Zhu, S. Sun, Y. Wang, X. Li, Y. Chen, H. Zhou, Y. Wang and P. Yang, "ZnO/ZnS Heterostructured Nanorod Arrays and Their Efficient Photocatalytic Hydrogen Evolution", *Chemistry A European Journal*, Vol. 21, No. 36, pp. 12728-12734, 2015.

92. Sang, H. X., X. T. Wang, C. C. Fan and F. Wang, "Enhanced Photocatalytic H_2 Production from Glycerol Solution over ZnO/ZnS Core/Shell Nanorods Prepared by a Low Temperature Route", *International Journal of Hydrogen Energy*, Vol. 37, No. 2, pp. 1348-1355, 2012.

93. Chang, C. J., Y. G. Lin, H. T. Weng and Y. H. Wei, "Photocatalytic Hydrogen Production from Glycerol Solution at Room Temperature by ZnO-ZnS/Graphene Photocatalysts", *Applied Surface Science*, Vol. 451, No. 1, pp. 198-206, 2018.

94. Liu, S. X. Wang, K. Wang, R. Lv and Y. Xu, "ZnO/ZnS–PdS Core/Shell Nanorods: Synthesis, Characterization and Application for Photocatalytic Hydrogen Production from a Glycerol/Water Solution", *Applied Surface Science*, Vol. 283, pp. 732-739, 2013.

95. Manchala, S., L. R. Nagappagari, S. M. Venkatakrishnan and C. Shanker, "Facile Synthesis of Noble-Metal Free Polygonal Zn_2TiO_4 Nanostructures for Highly Efficient Photocatalytic Hydrogen Evolution under Solar Light Irradiation", *International Journal of Hydrogen Energy*, Vol. 43, No. 29, pp. 13145-13157, 2018.

96. Kadi, M. W. and R. M. Mohamed, "One-Step Sol-Gel Synthesis of $PbTiO_3$ Nanosheets and Photocatalytic Enhancement through Decoration by Platinum", *Journal of Nanoparticle Research*, Vol. 22, No. 9, p. 266, 2020.

97. Xiang, X., B. Zhu, B. Cheng, J. Yu and H. Lv, "Enhanced Photocatalytic H_2-Production Activity of CdS Quantum Dots Using Sn^{2+} as Cocatalyst under Visible Light Irradiation", *Small,* Vol. 16, No. 26, p. 2001024, 2020.

98. Alhaddad, M., R. M. Navarro, M. A. Hussein and R. M. Mohamed, "Bi_2O_3/g-C_3N_4 Nanocomposites as Proficient Photocatalysts for Hydrogen Generation from Aqueous Glycerol Solutions Beneath Visible Light", *Ceramics International*, Vol. 46, No. 16, pp. 24873-24881, 2020.

99. Kadi, M. W. and R. M. Mohamed, "Pt-Decorated CuO Nanosheets and Their Application in the Visible Light Photocatalytic Water Splitting Reaction", *Applied Nanoscience*, Vol. 10, No. 11, pp. 4291-4298, 2020.

100. Rawool, S. A., M. R. Rai, A. M. Banerjee, R. D. Bapat, C. Nayak and A. K. Tripathi, "Lab Scale Optimization of Various Factors for Photocatalytic Hydrogen Generation over Low Cost $Cu_{0.02}Ti_{0.98}O_{2-\delta}$ Photocatalyst under UV/Visible Irradiation and Sunlight", *International Journal of Hydrogen Energy*, Vol. 43, No. 3, pp. 1271-1284, 2018.

101. Lopes, P. A. L., A. J. S. Mascarenhas and L. A. Silva, "Sonochemical Synthesis of $Cd_{1-x}Zn_xS$ Solid Solutions for Application in Photocatalytic Reforming of Glycerol to Produce Hydrogen", *Journal of Alloys and Compounds*, Vol. 649, pp. 332-336, 2015.

102. Mohamed, R. M. and A. A. Ismail, "Mesoporous Pt/$La_{0.02}Na_{0.98}TaO_3$ Nanocomposites as Efficient Photocatalyst for Hydrogen Evolution", *Molecular Catalysis*, Vol. 486, p. 110885, 2020.

103. Mohamed, R. M., A. A. Ismail, A. S. Basaleh and H. A. Bawazir, "Photodeposition of Ag Nanoparticles on Mesoporous $LaNaTaO_3$ Nanocomposites for Promotion H_2 Evolution", *Materials Research Bulletin*, Vol. 131, p. 110962, 2020.

104. Mohamed, R. M., A. A. Ismail, A. S. Basaleh and H. A. Bawazir, "Construction of Highly Dispersed Nd_2O_3 Nanoparticles onto Mesoporous $LaNaTaO_3$ Nanocomposites for H_2 Evolution", *Journal of Photochemistry and Photobiology A: Chemistry*, Vol. 400, p. 112723, 2020.

105. Mohamed, R. M., A. A. Ismail, A. S. Basaleh and H. A. Bawazir, "Facile Fabrication of Mesoporous In_2O_3/ $LaNaTaO_3$ Nanocomposites for Photocatalytic H_2 Evolution", *International Journal of Hydrogen Energy*, Vol. 45, No. 38, pp. 19214-19225, 2020.

106. Kadi, M. W., R. M. Mohamed, A. A. Ismail and D. Q. Bahnemann, "H_2 Production Using CuS/ g-C_3N_4 Nanocomposites under Visible Light", *Applied Nanoscience*, Vol. 10, No. 1, pp. 223-232, 2019.

107. Vaiano, V. and G. Iervolino, "Photocatalytic Hydrogen Production from Glycerol Aqueous Solution Using Cu-Doped ZnO under Visible Light Irradiation", *Applied Sciences*, Vol. 9, No. 13, p. 2741, 2019.

108. Vempuluru, N. R., C. K. Krishnan, R. Parnapalli, J. Velusamy, S. Marappan, S. Pitchaimuthu, M. Murikinati and S. M. Venkatakrishnan, "Solar Hydrogen Generation from Organic Substance Using Earth Abundant CuS–NiO Heterojunction Semiconductor Photocatalyst", *Ceramics International*, Vol. 47, No. 7, pp. 10206-10215, 2021.

109. Osman, A. I., N. C. Skillen, P. K. J. Robertson, D. W. Rooney and K. Morgan, "Exploring the Photocatalytic Hydrogen Production Potential of Titania Doped with Alumina Derived from Foil Waste", *International Journal of Hydrogen Energy*, Vol. 45, No. 59, pp. 34494-34502, 2020.

110. López, C. R., E. P. Melián, J. A. O. Méndez, D. E. Santiago, J. M. D. Rodríguez and O. G. Díaz, "Comparative Study of Alcohols as Sacrificial Agents in H_2 Production by Heterogeneous Photocatalysis Using Pt/ TiO_2 Catalysts", *Journal of Photochemistry and Photobiology A: Chemistry*, Vol. 312, pp. 45-54, 2015.

111. Vaiano, V., M. A. Lara, G. Iervolino, M. Matarangolo, J. A. Navio and M. C. Hidalgo, "Photocatalytic H_2 Production from Glycerol Aqueous Solutions over Fluorinated Pt-TiO_2 with High {001} Facet Exposure", *Journal of Photochemistry and Photobiology A: Chemistry*, Vol. 365, pp. 52-59, 2018.

112. Kumar, D. P., M. V. Shankar, N. L. Reddy, D. Parasuramudu, D. Rajarajeswari and V. D. Kumari, "Solar Light Active CuO/ TiO_2 Nanobelt Photocatalyst for Enhanced H_2 Production", in *International Conference on Advanced Nanomaterials & Emerging Engineering Technologies*, Chennai, India, 2013.

113. Naffati, N., M. J. Sampaio, E. S. Da Silva, M. F. Nsib, Y. Arfaoui, A. Houas, J. L. Faria and C. G. Silva, "Carbon-Nanotube/ TiO_2 Materials Synthesized by a One-Pot Oxidation/Hydrothermal Route for the Photocatalytic Production of Hydrogen from Biomass Derivatives", *Materials Science in Semiconductor Processing*, Vol. 115, p. 105098, 2020.

114. Reddy, N. R., U. Bharagav, M. M. Kumari, K. K. Cheralathan, P. K. Ojha, M. V. Shankar and S. W. Joo, "Inclusion of Low Cost Activated Carbon for Improving Hydrogen Production Performance of TiO_2 Nanoparticles under Natural Solar Light Irradiation", *Ceramics International*, Vol. 47, No. 7, pp. 10216-10225, 2021.

115. Rao, V. N., V. Preethi, U. Bhargav, P. Ravi, A. Kumar, M. Sathish, V. Krishnan, V. Venkatramu, M. M. Kumari and K. R. Reddy, "Gram-Scale Synthesis of ZnS/NiO Core-Shell Hierarchical Nanostructures and Their Enhanced H_2 Production in Crude Glycerol and Sulphide Wastewater", *Environmental Research*, Vol. 199, p. 111323, 2021.

116. Reddy, N. R., U. Bhargav, C. C. Mohan, M. M. Kumari and M. V. Shankar, "Multiwalled Carbon Nanotubes in Titania based Nanocomposite as Trap for Photoexcitons for Enhanced Photocatalytic Hydrogen Production under Solar Light Irradiation", *Materials Research Bulletin*, Vol. 106, pp. 271-275, 2018.

117. Reddy, N. L., S. Emin, V. Kumari and S. M. Venkatakrishnan, "CuO Quantum Dots Decorated TiO_2 Nanocomposite Photocatalyst for Stable Hydrogen Generation", *Industrial & Engineering Chemistry Research*, Vol. 57, No. 2, pp. 568-577, 2018.

118. Zhang, H., Y. Dong, D. Li, G. Wang, Y. Leng, P. Zhang, H. Miao, Z. Wu, P. Jiang and Y. Zhu, "Photochemical Synthesis of Ni-$Ni(OH)_2$ Synergistic Cocatalysts Hybridized with CdS Nanorods for Efficient Photocatalytic Hydrogen Evolution", *Chemistry of Flat Materials*, Vol. 26, p. 100232, 2021.

119. Shaheer, A. R. M., V. Vinesh, S. K. Lakhera and B. Neppolian, "Reduced Graphene Oxide as a Solid-State Mediator in $TiO_2/In_{0.5}WO_3$ S-Scheme Photocatalyst for Hydrogen Production", *Solar Energy*, Vol. 213, pp. 260-270, 2021.

120. Chen, W. T., A. Chan, D. S. Waterhouse, J. Llorca, H. Idriss and G. I. N. Waterhouse, "Performance Comparison of Ni/TiO_2 and Au/TiO_2 Photocatalysts for H_2 Production in Different Alcohol-Water Mixtures", *Journal of Catalysis*, Vol. 367, pp. 27-42, 2018.

121. Mkhalid, I. A., "Hydrogen Evolution over Sol-Gel Prepared Visible-Light-Responsive $Ag_2O/SrAl_2O_4$/CNT Ternary Photocatalyst", *Ceramics International*, Vol. 48, No. 2, pp. 1542-1549, 2022.

122. Irfan, M., S. Shukrullah, M. Y. Naz, I. Ahmad, B. Shoukat, S. Legutko, J. Petrů, S. Rahman and M. A. Alsaiari, "$Si/SiO_2/Al_2O_3$ Supported Growth of CNT Forest for the Production of La/ZnO/CNT Photocatalyst for Hydrogen Production", *Materials*, Vol. 15, No. 9, p. 3226, 2022.

123. Gonuguntla, S., A. Tiwari, J. Gopinath, S. Yarasi, A. V. S. Sainath and U. Pal, "Rational Design of Ru(II)-Phenanthroline Complex Embedded Porous TiO_2

Photocatalyst for Efficient Hydrogen Production", *Renewable Energy*, Vol. 159, pp. 1-9, 2020.

124. Stelmachowski, M., K. Bednarczyk and M. Gmurek, "An Experimental Study of the Photocatalytic Hydrogen Production in the Laboratory-Scale Set-Up Depending on the Process Conditions", *WIT Transactions on Ecology and the Environment*, Vol. 224, pp. 339-348, 2017.

APPENDIX A: ARTICLES IN THE DATASET

Table A.1. Article references.

Article	Reference
1	Reddy et al., 2016 [34]
2	Manukumar et al., 2018 [35]
3	Kumar et al., 2017 [36]
4	Daskalaki and Kondarides, 2009 [37]
5	Majrik et al., 2018 [38]
6	Mizukoshi, 2019 [39]
7	Kondarides et al., 2007 [40]
8	Fu and Lu, 2009 [41]
9	Li et al., 2009 [42]
10	Daskalaki et al., 2011 [43]
11	Fu et al., 2011 [44]
12	Beltram et al., 2016 [45]
13	Melian et al., 2016 [46]
14	Jiang et al., 2015 [47]
15	Estahbanati et al., 2019 [48]
16	Estahbanati et al., 2017 [49]
17	Estahbanati et al., 2019 [50]
18	Chen et al., 2015 [51]
19	Dosado et al., 2015 [52]
20	Xu et al., 2017 [53]
21	Sadanandam et al., 2013 [54]
22	Reddy et al., 2018 [55]
23	Liu et al., 2014 [56]
24	Lalitha et al., 2010 [57]
25	Montini et al., 2011 [58]
26	Bashini et al., 2015 [59]
27	Kumar et al., 2015 [60]

Table A.1. Article references (cont.).

Article	Reference
28	Zhang et al., 2016 [61]
29	Kumar et al., 2013 [62]
30	Lin et al., 2017 [63]
31	Ibrahim et al., 2019 [64]
32	Ibrahim et al., 2018 [65]
33	Lopes-Tenllado et al., 2017 [66]
34	Al-Azri et al., 2015 [67]
35	Bednarczyk et al., 2018 [68]
36	Wang et al., 2017 [69]
37	Seadira et al., 2018 [70]
38	Chen et al., 2018 [71]
39	Kumar et al., 2017 [14]
40	Taylor et al., 2014 [72]
41	Mohamed et al., 2020 [73]
42	Gullapelli et al., 2017 [74]
43	Kumar et al., 2018 [75]
44	Ravi et al., 2020 [76]
45	Reddy et al., 2018 [77]
46	Estahbanati et al., 2019 [78]
47	Tahir, 2019 [79]
48	Tahir and Fajrina, 2020 [15]
49	Ribao et al., 2019 [80]
50	Seadira et al., 2020 [81]
51	Nguyen et al., 2018 [82]
52	Hafeez et al., 2018 [83]
53	Hafeez et al., 2018 [84]
54	Sadanandam et al., 2018 [85]
55	Babu et al., 2015 [86]
56	Sadanandam et al., 2017 [87]

Table A.1. Article references (cont.).

Article	Reference
57	de Oliveira Melo and Silva, 2011 [88]
58	Bastos et al., 2014 [89]
59	Wang et al., 2014 [90]
60	Bao et al., 2015 [91]
61	Sang et al., 2012 [92]
62	Chang et al., 2018 [93]
63	Liu et al., 2013 [94]
64	Manchala et al., 2018 [95]
65	Kadi and Mohamed, 2020 [96]
66	Xiang et al., 2020 [97]
67	Alhaddad et al., 2020 [98]
68	Kadi and Mohamed, 2020 [99]
69	Rawool et al., 2018 [100]
70	Lopes et al., 2015 [101]
71	Mohamed and Ismail, 2020 [102]
72	Mohamed et al., 2020 [103]
73	Mohamed et al., 2020 [104]
74	Mohamed et al., 2020 [105]
75	Kadi et al., 2019 [106]
76	Vaiano and Iervolino, 2019 [107]
77	Vempuluru et al., 2021 [108]
78	Osman et al., 2020 [109]
79	Lopez et al., 2015 [110]
80	Vaiano et al., 2018 [111]
81	Kumar et al., 2013 [112]
82	Naffati et al., 2020 [113]
83	Reddy et al., 2021 [114]
84	Rao et al., 2021 [115]
85	Reddy et al., 2018 [116]

Table A.1. Article references (cont.).

Article	Reference
86	Reddy et al., 2018 [117]
87	Zhang et al., 2021 [118]
88	Shaheer et al., 2021 [119]
89	Chen et al., 2018 [120]
90	Mkhalid, 2022 [121]
91	Irfan et al., 2022 [122]
92	Gonuguntla et al., 2020 [123]
93	Stelmachowski et al., 2017 [124]

APPENDIX B: COPYRIGHT PERMISSION FOR FIGURES

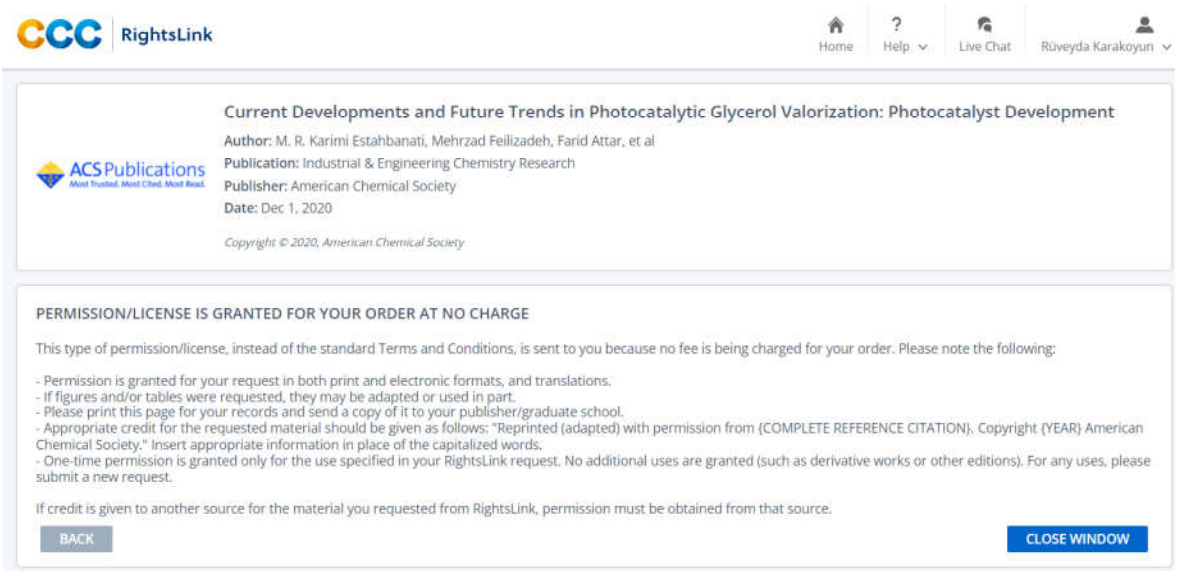

Figure B.1. Copyright permission for Figure 2.1.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 06, 2023

This Agreement between Mrs. Rüveyda Karakoyun ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5603081051299
License date	Aug 06, 2023
Licensed Content Publisher	Elsevier
Licensed Content Publication	Coordination Chemistry Reviews
Licensed Content Title	Photocatalytic production of hydrogen from biomass-derived feedstocks
Licensed Content Author	Alberto V. Puga
Licensed Content Date	May 15, 2016
Licensed Content Volume	315
Licensed Content Issue	n/a
Licensed Content Pages	66
Start Page	1
End Page	66
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations

Figure B.2. Copyright permission for Figure 2.3.

Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	MACHINE LEARNING ANALYSIS OF DATA COLLECTED FROM PUBLISHED LITERATURE ON PHOTOCATALYTIC REFORMING OF GLYCEROL
Institution name	Boğaziçi University
Expected presentation date	Aug 2023
Portions	Fig. 1
Requestor Location	Mrs. Rüveyda Karakoyun Pendik İstanbul Istanbul, 34896 Turkey Attn: Mrs. Rüveyda Karakoyun
Publisher Tax ID	GB 494 6272 12
Total	0.00 USD
Terms and Conditions	

Figure B.3. Copyright permission for Figure 2.3.

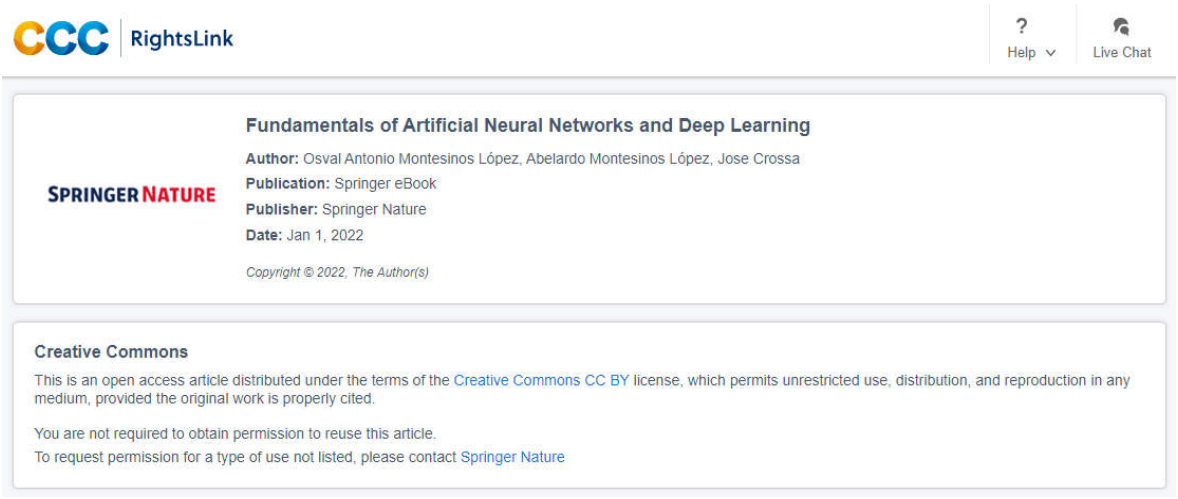

Figure B.4. Copyright permission for Figure 2.4.