DATA MINING FOR METHANE ADSORPTION OVER METAL ORGANIC FRAMEWORKS IN LITERATURE

by

Kadriye Zeynep Gülsoy Şerif

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

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

Graduate Program in Chemical Engineering

Boğaziçi University

2019

to my very precious niece Elif Serra
and nephew Ertuğrul Said…

ACKNOWLEDGEMENTS

I would like to acknowledge and thank to following valuable individuals who helped and supported me on this journey.

Foremost, I owe a deep gratitude to my precious extended “Gülsoy” family, especially my mother Ferhan Gülsoy, my father Hamdi Gülsoy and my elder sister Beyzanur Altıntaş for their love and support throughout my life. I am also thankful to my dear husband Muhammed Şerif for giving the right advice at the right time and being a constant source of motivation. He has never lost his faith in me.

Next, I would like to express my deepest thanks to my supervisor Prof. Dr. Ramazan Yıldırım for his excellent guidance, continuous support and endless patience during my graduate study. He always encouraged me to do better. I am also grateful to my co-advisor Prof. Dr. Seda Keskin for her insightful comments and kind advices when I consult her expertise. It was a pleasure to work with such great academicians.

I would also like to thank my committee members; Prof. Dr. Ahmet Erhan Aksoylu, Assoc. Prof. Burak Alakent and Assoc. Prof. Tuğba Davran Candan, for their contribution to my thesis by sparing their valuable time.

I would like to offer my sincere thanks to my very lovely “Jets” group; Feyza Sarı, Berrak Konca, Zeynep Taşkıran Yalçınkaya, Nurdan Erdoğan, Merve Karahan and Kübra Aras Türkcan for keeping me company with their endless joy. I am also thankful to my dear friend Ayşe Hilal Kaymak for always believing in me.

I am grateful to my colleagues at Gas Metrology Laboratory, TÜBİTAK; Kemal Özcan, Dr. Erinç Engin, Dr. Tanıl Tarhan, Dr. Aylin Boztepe and Gürol Koni who willingly helped me out with their valuable abilities.

Last of all, I would like to thank my ChE friends; Elif Can, Gizem Yumru, Ahmet Coşgun, Beyza Yılmaz, Çağla Özer, Neslihan Şener and Barış Demirci for their generous help when I need them.

ABSTRACT

DATA MINING FOR METHANE ADSORPTION OVER METAL ORGANIC FRAMEWORKS IN LITERATURE

A database containing 3000 data points for both gravimetric and volumetric CH_4 storage/delivery in metal organic frameworks (MOFs) was analysed using machine learning tools via *RStudio* functions to extract knowledge for generalization. First, the database was passed through the reprocessing stage to deal with the missing values and inappropriate input variables. The cross-correlation analysis was also performed during this stage as well and the correlated variables were determined. Then the optimization process took place to find out the optimum conditions of the best models. First, the database was reviewed to observe the basic trends and patterns. It was then analysed using decision trees and artificial neural networks (ANN) to extract hidden information and develop rules and heuristics for the future studies. Five-fold cross validations were used in each analysis to test the validity of the models with data not seen before. Decision tree analyses were carried out using six user defined descriptors and two structural properties, separately. The crystal structure and the total degree of unsaturation were found to be the effective user defined descriptors, whereas the pore volume and maximum pore diameter, as structural properties, were sufficient to determine the MOFs having high CH_4 storage/delivery capacity. Moreover, a high pore volume is always required as expected. In ANN analyses, models were performed by using the user defined descriptors and the structural properties separately. It was observed that the user defined descriptors were not sufficient to describe the CH_4 storage/delivery capacity of MOFs, whereas the structural properties, especially pore volume, provided accurate CH_4 storage/delivery prediction with low root mean square error (RMSE) and high coefficient of determination (R^2) values.

ÖZET

YAYINLANMIŞ MAKALELERDEN METAL ORGANİK YAPILARIN METAN ADSORPSİYONU İÇİN VERİ MADENCİLİĞİ

Metal organik yapılar (MOF) üzerinde hem kütlesel hem de hacimsel CH_4 depolama ve taşıma için 3000 veri noktası içeren bir veri tabanı, genellenebilir bilgiler çıkarmak için RStudio ortamında makine öğrenmesi araçları kullanılarak analiz edilmiştir. Veri tabanı, öncelikle eksik değerler ve uygun olmayan giriş değişkenleriyle başa çıkmak için yeniden işleme aşamasını geçmiştir. Bu aşamada çapraz korelasyon analizi de yapılmış ve ilişkili değişkenler belirlenmiştir. Sonra, en iyi modellerin koşullarını bulmak için optimizasyon süreci gerçekleştirilmiştir. İlk olarak, veri tabanı, temel eğilimleri ve modelleri gözlemlemek için gözden geçirilmiştir. Daha sonra saklı bilgileri ortaya çıkarmak ve gelecekteki çalışmalar için kurallar ve modeller geliştirmek için veri tabanı, karar ağaçları ve yapay sinir ağları kullanılarak analiz edilmiştir. Her analizde, daha önce görülmeyen verilerle modellerin geçerliliğini test etmek için beş kat çapraz doğrulama kullanılmıştır. Karar ağacı analizleri, altı kullanıcı tanımlı tanımlayıcı ve iki yapısal özellik kullanılarak ayrı ayrı gerçekleştirilmiştir. Kristal yapı ve MOF molekülünün toplam doymamışlık derecesi, etkin kullanıcı tanımlı tanımlayıcılar olarak bulunurken, yapısal özellikler olarak gözenek hacmi ve maksimum gözenek çapı, yüksek CH_4 depolama ve dağıtma kapasitesine sahip MOFları belirlemek için yeterli olmuştur. Beklendiği gibi her zaman yüksek bir gözenek hacminin gerekli olduğu görülmüştür. Yapay sinir ağları analizleri, kullanıcı tanımlı tanımlayıcılar ve yapısal özellikler ayrı ayrı kullanılarak yapılmıştır. Kullanıcı tanımlı tanımlayıcıların MOFların CH_4 depolama ve dağıtma kapasitesini tanımlamak için yeterli olmadığı, yapısal özelliklerin, özellikle gözenek hacminin, düşük RMSE ve yüksek R^2 değerleriyle doğru CH_4 depolama ve dağıtma tahmini sağladığı gözlemlenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

R^2	Coefficient of determination
R	Gas constant
S_g	Gravimetric surface area
Q_{st}	Isosteric heat of adsorption
M	Molar mass
P	Pressure
V_p	Pore volume
T	Temperature
ρ_{cryst}	Crystal density
ρ_{gas}	Density of gas

LIST OF ACRONYMS/ABBREVIATIONS

ANG	Adsorbed Natural Gas
ANN	Artificial Neural Network
CCDC	Cambridge Crystallographic Data Centre
DPD	Dominant Pore Diameter
FT–IR	Fourier-Transform Infrared Spectroscopy
FCV	Fuel Cell Vehicle
MPD	Maximum Pore Diameter
MAPE	Mean Absolute Percentage Error
ME	Mean Error
MOF	Metallic Organic Framework
MMM	Mixed Matrix Membrane
PLD	Pore Limiting Diameter
PSA	Pressure Swing Adsorption
QSPR	Quantitative Structure–Property Relationship
RMSE	Root Mean Square Error
SEM	Scanning Electron Microscope
SSE	Sum of Square Error
TSA	Temperature or Thermal Swing Adsorption
TGA	Thermal Gravimetric Analysis
UCV	Unit Cell Volume
DOE	United States Department of Energy
VSA	Vacuum Swing Adsorption
XRD	X-ray Powder Diffraction

1. INTRODUCTION

Natural gas is one of the most abundant source of energy, and it is the cleanest among the non-renewable energy sources [1,2]. Adsorbed natural gas (ANG) systems are designed to store and deliver natural gas for vehicular applications. Because the main component of natural gas is CH_4, which has the highest combustion energy, high methane delivery can facilitate the applications of natural gas fuelled vehicles. To transport CH_4, it should be first adsorbed at a high pressure (generally at 35 or 65 bar) and then desorbed at a low pressure (generally at 5 bar) after transportation. Hence, not only the *storage capacity* but also the difference between the amount of adsorbed and desorbed CH_4 which is considered as the *deliverable capacity* of CH_4 [3], are important criterions to compare the potential of different adsorbent materials.

Metal organic frameworks (MOFs) have been considered as promising alternatives to more traditional adsorbent materials such as zeolites and activated carbons. [3–5]. MOFs are relatively new crystalline porous materials that consist of metal ions as inorganic part and organic ligands [6]. Because of their large surface areas (up to 6000 m^2/g) and high pore volumes (1−4 cm^3/g), MOFs are in the forefront of porous materials [7,8]. Not only these features but also their ability to be tailor-made for the specific applications make them great candidates for gas separations, gas storage and delivery, catalytic applications [1,6]. MOFs have enormous number of possible structural variations to be synthesized. To date, several thousands of MOFs have been synthesized and yet infinite number of structures can be produced theoretically depending on the combination of organic linkers and metals. Although unlimited options to produce MOFs are counted as an opportunity, it is also a big challenge to find the best performing candidates for a target application. Synthesis and testing of a single MOF generally takes several weeks; hence, it is not practical to screen a large number of candidates for a target application.

At this point, the importance of data mining comes into prominence. Data mining is a branch of science that aims to discover the unknown pattern, trends, and correlations in large pre-existing databases. The machine learning tools such as decision tree classification, artificial neural network, principal component analysis and support vector

machine are applied to the database to find out the useful hidden information [9]. In fact, there are few studies that machine-learning tools were used for CH_4 uptake in MOFs. For instance, Fernandez *et al.* used decision trees and support vector machine to analyze a database of approximately 130,000 hypothetical MOFs (which are computer generated MOFs, not experimentally synthesized) to predict the CH_4 storage capacity from the structural descriptors such as void fraction and pore volume [10]. The same hypothetical MOF database was also modelled by Pardakhti *et al.*; the descriptors were divided into two, as chemical and structural, to predict the CH_4 uptake of MOFs via machine learning tools [2]. Ohno et at. have also applied machine learning tools to a dataset consists of 499 hypothetical MOF structures to predict their CH_4 uptake using structural descriptors [11].

The aim of this thesis is to focus on a considerably larger database containing simulated CH_4 uptake data over 3000 real, synthesized MOFs and use machine learning tools to identify the trends and patterns, extract hidden information in the network of data to develop rules and heuristics for the future investigations. The database first pre-analyzed using simple descriptive statistics to identify simple trends and to determine correlations among the variables to prevent multicollinearity in the input variables to be used in predictive models. Then, decision trees were constructed to develop heuristics for the MOF properties required for achieving a high CH_4 delivery capacity, while the artificial neural network was used to predict the CH_4 uptake of MOFs as the function of user defined descriptors representing structural properties of MOFs.

In Chapter 2, MOFs are introduced in details together with their application areas, production processes and determinative parameters; then the data mining tools utilized in this study and the published articles about MOFs in which data mining tools were used are summarized. Database reprocessing and the computational details about decision tree and artificial neural network algorithms are given in Chapter 3. In Chapter 4, all results obtained from the models were explained in detail. The important findings of this study for possible future studies are presented in the last chapter.

2. LITERATURE SURVEY

2.1. Metal Organic Frameworks as Adsorbents

2.1.1. Late History in Adsorbents

Metal Organic Frameworks (MOFs, also known as porous coordination polymers or PCPs) have been investigated as promising porous materials in the past two decades [12]. The main idea behind the porous materials is to utilize their capabilities to behave as hosts for particular guest molecules [13]. MOFs have been considered for a wide range of applications such as gas storage, adsorption-based gas/vapor separation, shape/size-selective catalysis, drug storage and delivery, and as templates in the preparation of low dimensional materials [14].

Traditional porous materials have been consisted of either organic or inorganic molecules. Before the metal-organic frameworks have been introduced, activated carbons and zeolites have been in the forefront of porous materials; most common organic porous materials are still activated carbons with high surface areas and high gas adsorption capacities [14]. The separation and storage of gases, the purification of water, and solvent removal and recovery are some of the usages of activated carbons [15]. Since activated carbons suffer from lack of ordered structure, the pursuit of adsorbent has continued with zeolites, which are highly ordered inorganic adsorbents used in separation and catalysis application. Inorganic frameworks such as zeolites have seldom diverged from Al, Si and chalcogens; so they have limited structures [16].

Basically, MOFs are relatively new kinds of crystalline porous materials that consist of both metal ions as inorganic part and organic linkers as organic part. Compared with zeolites and carbon materials, there is not any non-accessible bulk volume in MOFs. In other words, lack of dead volume in MOFs provides them the highest porosities and world record surface area that was reported as 4500 m^2/g [17]. Due the fact that MOFs are compromised of both organic and inorganic components, MOFs have enormous number of

possible options to be synthesized. As it is illustrated in Figure 2.1, metal ions or clusters connected by organic linkers compose 3D structures with unique physical and chemical properties. MOFs are flexible materials; their geometry, size and functionality can be altered easily [18,19]. Their tunable physical geometry and designed chemical functionality have made them scientifically searchable [20]. According to data taken from [21], 70,000 different MOFs being reported and studied in the past decade.

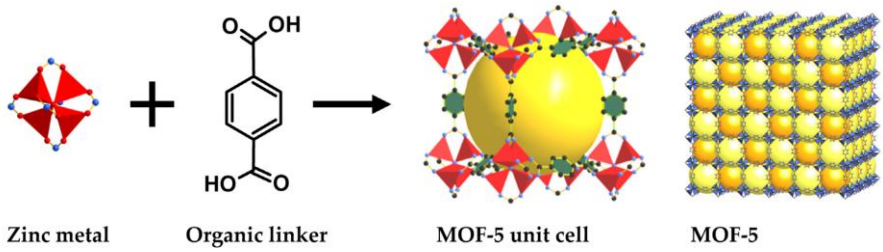

Figure 2.1. Basic construction units of MOF-5 [22].

2.1.2. Potential applications of MOFs

According to O'Keeffe *et al.* [23], the first 3D metal-organic framework that is known as MOF-4 was examined in 1997 by Yaghi and coworkers [24] whose 2D MOF research (namely MOF-2) in 1998 was also accepted as an early example of its kind [25]. Since that time, different MOFs had been identified to address various applications in chemical industry. It seems that the gas adsorption processes may use them significantly in the future because of their tailor-made features [26].

Adsorption process can be actualized in two ways as shown in Figure 2.2 physical adsorption (physisorption) and chemical adsorption (chemisorption). An adsorptive gas comes into contact with the solid surface called the adsorbent and then the physisorption occurs with the help of van-der Waals forces, the long-range London dispersion forces and the short-range intermolecular repulsion. There is no chemical reaction involved in this process and it is reversible [27,28]. On the other hand, chemical reaction occurs in chemisorption and it cannot be reversed. The heat of adsorption in chemisorptions is much greater than those in physisorption. As it is shown in Figure 2.2 the target molecules are attached to the surface by van-der Waals forces in physisorption whereas the molecules

reacts with certain binding sites on adsorbent surface in chemisorption. Unless otherwise stated, adsorption is referred to as physisorption for further sections.

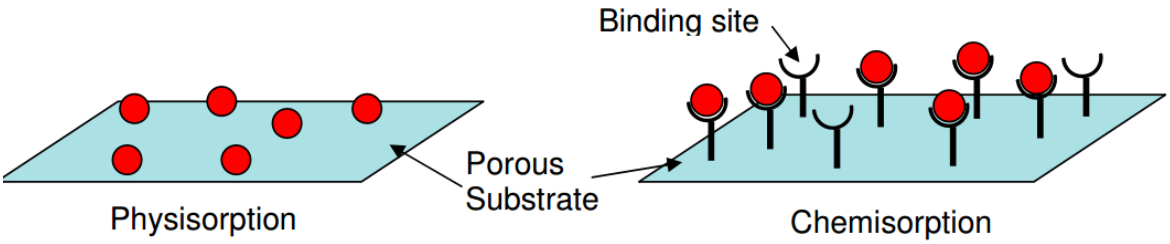

Figure 2.2. Schematic representation of physisorption and chemisorption [29].

Gas separation and purification processes seem to be the main application areas of MOFs considering their high surface areas, high porosities and well-defined structures [14]. Although distillation, extraction, adsorption, stripping, solvent extraction, crystallization and membranes are the important methods of gas separation and purification, MOFs are utilized in adsorption and membrane techniques [30,31]. A packed bed column filled with specific MOF might be used in an adsorption process as shown in Figure 2.3. After the adsorption process is done in the packed bed column, the compositions of the gas cylinder can be determined by mass spectrometry. Any kinds of pure gases such as N_2, CH_4, H_2, and O_2 can be purified by absorbing CO_2, which is one of the main contributors to global warming [30]. Also removal of impurities such as CO_2 or sulphur compounds from natural gas or flue gas might be the purpose of use of MOFs [12,26].

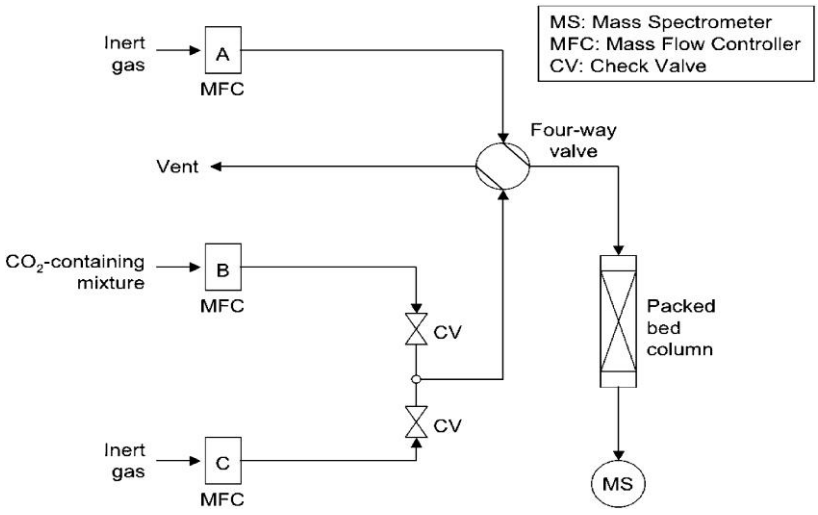

Figure 2.3. Setup for column measurements [30].

Adsorbents such as MOFs filled with gas need to be discharged for the further use in continuous adsorption processes. To do this, pressure swing adsorption (PSA), vacuum swing adsorption (VSA), and temperature swing adsorption (TSA) methods are used as the most common cyclic processes. Desorption of the desired gas is provided by either decreasing the pressure (in PSA or VSA) or increasing the temperature (in TSA). In Figure 2.4a, the most common version of PSA and VSA and their pressure profiles were shown. The feed flow enters to the column in both first and second step and desired gas is adsorbed in second step. Then the pressure is reduced to a certain level to blow down the undesired gas. In the final step, desired gas is absorbed by further decreasing the pressure [32]. Temperature or thermal swing adsorption cycle generally has four steps as shown in Figure 2.4b; the process in this figure is designed in order to adsorb CO_2 in flue gas. In the first step, flue gas enters to the column and the CO_2 is adsorbed while undesired gas (N_2 in Figure 2.4a leaves the column. As the temperature of the column is increased, adsorbed gas starts to be desorbed from the surface and then leaves the column with the help of increased pressure. When there is no gas left to be desorbed from the surface, the column is purged and then cooled to be ready for further adsorption cycles [12].

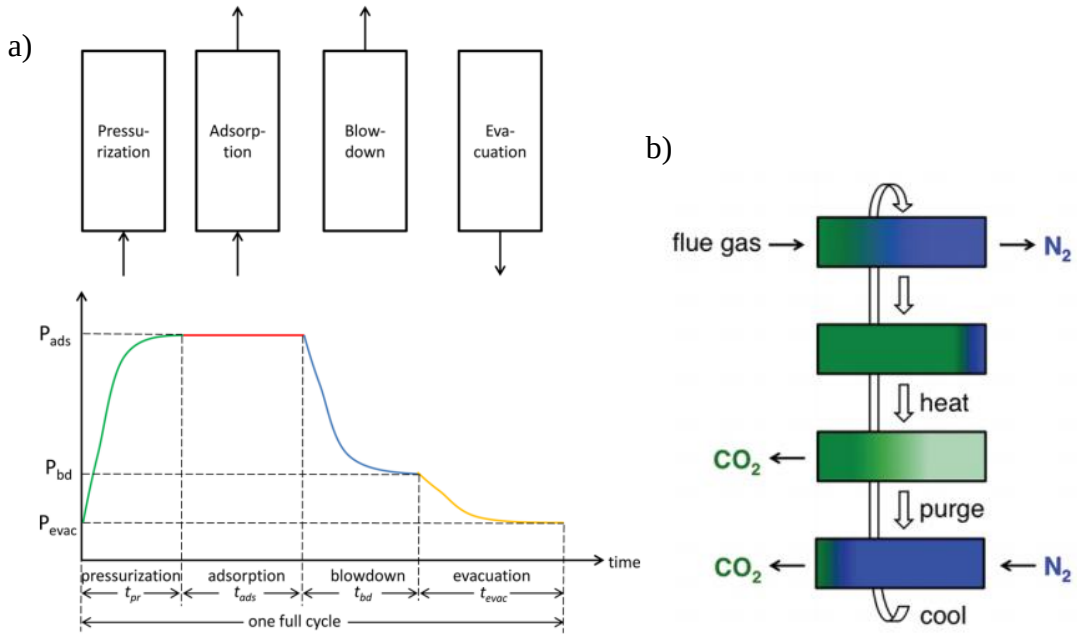

Figure 2.4. a) PSA / VSA steps and pressure profiles [32], b) TSA cycle [12].

Membrane-based separations especially made with polymers are also preferred in the gas field due to their low cost, their ease of manufacture and being versatile [31,33]. In

membrane-based separations, the feed gas, which consists of two bulk phases are separated in a physical way by a membrane as it is illustrated in Figure 2.5. The phase which goes through the membrane is called permeate whereas retentate represents the remained portion of the feed that is retained by the membrane. The material of the membrane and the operating conditions play key roles in controlling the transport of these two phases [34]. Such separation processes can be utilized in many industry related to energy and environment such as natural gas purification by removal of CO_2, purification of H_2 from syngas derived from coal or biomass gasification, air separation to enable oxyfuel firing of coal/gas and separation CO_2 from N_2 [31,33]. However, the polymeric membranes for separation applications suffer from the trade-off between permeability and selectivity. In other words, finding the optimum relation between permeability and selectivity is the main issue in the polymeric membranes [33,35,36]. In order to overcome that challenge, polymers were combined with non-polymeric material and heterogeneous or hybrid membranes were obtained; those membranes are called mixed matrix membranes (MMMs). In those membranes, organic or inorganic fillers are put into a polymer matrix; zeolites, silicas and activated carbons are examples of such MMM fillers. The first MOF based MMMs was explored by Yehia *et al.* [37]; from that day forward studies regarding MOF based membranes have showed a great progress because of the MOFs' high surface are pore volume and ease of tuning [33,35,36,38].

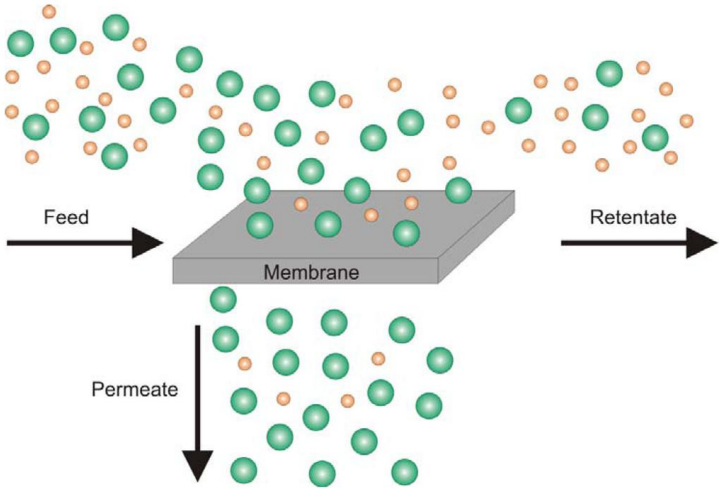

Figure 2.5. Principle of membrane-based separation [39].

In addition to aforementioned practice areas of MOFs, the new interpretations have also emerged through interdisciplinary research. The idea here is that inherent features of

MOFs are examined to modify existing systems in a more efficient way [40]. MOF-based materials have been developed to be used in chemical sensing [41,42], optoelectronics (a sub-field of photonics which is related with electronic devices and systems that source, detect and control light) [43,44] and biomedical applications [45,46].

2.1.3. Volumetric and Gravimetric Uptakes

The amount of gas adsorbed can be expressed in gravimetric and volumetric basis; their units are cm^3 gas adsorbed per unit mass of adsorbent and cm^3 gas adsorbed per unit volume of adsorbent respectively. Generally, gas uptakes have been measured gravimetrically and then crystal densities of MOFs have been used to convert gravimetric uptakes to volumetric ones [47].

Adsorption-based storage and transportation have been recently widespread for especially H_2 and CH_4 gases as energy containing materials; the fuel-cell vehicles (FCV) and adsorbed natural gas (ANG) systems have been used for storage and transportation of such gases [48]. Generally, CH_4 functions as a test gas in order to quantify the gas stored in ANG systems [47]. In those systems, various types of adsorbents are placed in vessels to enable the storage and then transportation of gas; MOFs have been utilized in that area as well. In order to determine the size and weight of those MOF-filled tanks in ANG systems, both volumetric and gravimetric CH_4 capacities have to be determined optimally so that the tank would not be too heavy or too large. In other words, the purpose of taking both volumetric and gravimetric adsorption capacities into consideration is to incorporate ideal characteristics within a single MOF as it is illustrated in Figure 2.6. This means that, if an increase in surface area leads to a decrease in volumetric capacity, further improvements in surface area would not be sufficient. At this point, being easily tunable for a specific process makes MOFs promising candidates as adsorbents [37,38,39]. U.S. Department of Energy (DOE) set targets for this challenge in order to find out the ideal MOF with both high gravimetric and volumetric adsorption capacity. For storage capacity, limits were set as 350 (cm^3 STP)/(cm^3) and 700 (cm^3 STP)/(g) for volumetric and gravimetric respectively whereas 315 (cm^3 STP)/(cm^3) for deliverable volumetric capacity [51,52].

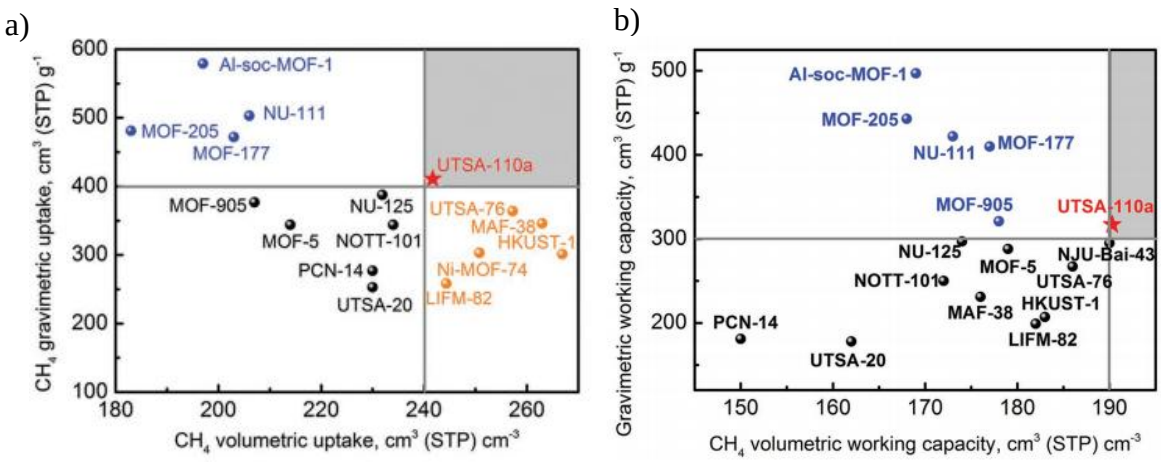

Figure 2.6. a) CH_4 uptakes at 65 bar b) CH_4 working capacities (between 5.8 and 65 bar) in both gravimetric and volumetric basis for the best robust MOF materials [52].

2.1.4. Deliverable Capacity and Storage Capacity

Having high adsorption capacity is not adequate by itself for the MOF to be used in transportation of gases; high delivery capacity is also an important criterion for adsorbents. The deliverable capacity, also known as working capacity, is the amount of gas desorbed by adsorbents which previously adsorb gas when the pressure decreases. The operating conditions in both adsorption and desorption processes affect the amount of delivered gas. For instance, if the heat released during adsorption process is not removed from the system, less CH_4 would be adsorbed whereas if the heat is not resupplied during the desorption process, the amount of delivered gas would be low [47].

Deliverable capacity is determined by subtracting the amount of methane adsorbed at lower limiting working pressure from that of adsorbed at upper limiting working pressure at isothermal conditions. Since the pressure level of 35 bar or 65 bar can be achieved by single-stage or two stage, they are taken as the upper limiting working pressure [47]. On the other hand, 5 bar has been considered as the lower limiting working pressure in several studies [53–55]. The aim of this thesis is to study the delivery capacity from 65 bar to 5 bar and uptake capacity at 35 bar both in gravimetric and volumetric basis. In Figure 2.7, the comparisons of such databases have been done for certain MOFs [52].

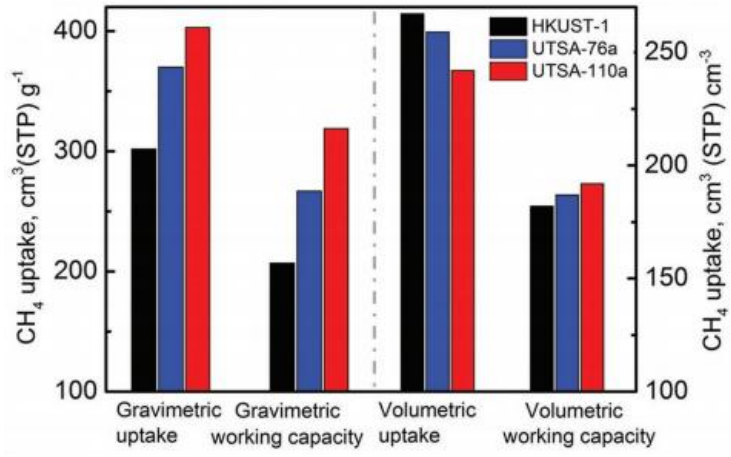

Figure 2.7. CH_4 storage capacities of three best-performing MOFs [52].

2.2. Synthesis Routes

Conventional solvothermal/hydrothermal, microwave-assisted, sonochemical, electrochemical, mechanochemical processes have been followed generally as synthesis routes of MOFs. In this section, these synthesis routes and comparative studies regarding them are briefly summarized.

2.2.1. Conventional Hydrothermal/Solvothermal Synthesis

Hydrothermal and solvothermal methods have been commonly used for more than 100 years and they are efficient synthetic routes to produce the nanomaterial with a diversity of morphologies. As it is illustrated in Figure 2.8, the reactants are put into a sealed vessel, which is filled with solvent (water or organic compound) and the reaction takes place at high temperature and pressure. If the solvent is water and the reaction occurs in aqueous solution, it is termed as hydrothermal method; whereas nonaqueous solution is utilized as solvent in solvothermal method [56–58]. The solid obtained by filtration is washed with solvent mixture and then dried to isolate the solvent (guest molecules). This activation step corresponds to the second part in Figure 2.8 (solvent exchange & washing and heat/vacuum treatment) and it can be applied in all synthesis to improve the properties of MOF. Hydrothermal/solvothermal conditions can also be utilized in other synthesis methods such as microwave-assisted and sonochemical by changing the source of heating.

For instance, electrical oven is utilized in conventional hydrothermal/solvothermal synthesis whereas microwave is used in microwave-assisted hydrothermal/solvothermal synthesis [59].

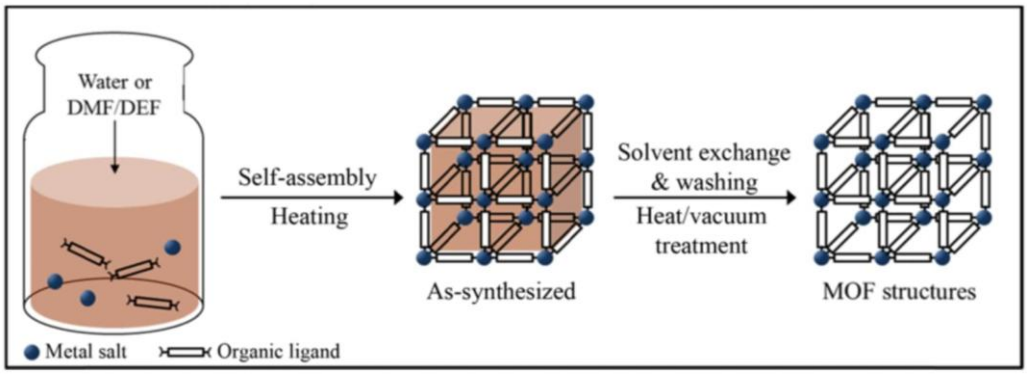

Figure 2.8. Conventional solvothermal/hydrothermal synthesis of MOF [60].

2.2.2. Microwave-assisted Synthesis

In microwave-assisted synthesis, the microwave irradiation was applied to supply required energy for reaction taken place on solvothermal or hydrothermal conditions. Local heating is achieved by dipole rotation and ionic conduction in this method, which is shown in Figure 2.9. This synthesis route has the potential advantages of high efficiency, phase selectivity, particle size reduction, and ease of morphology control [60–62]. Additionally, when it is compared with the conventional synthesis route, this method takes less time to obtain a MOF [63].

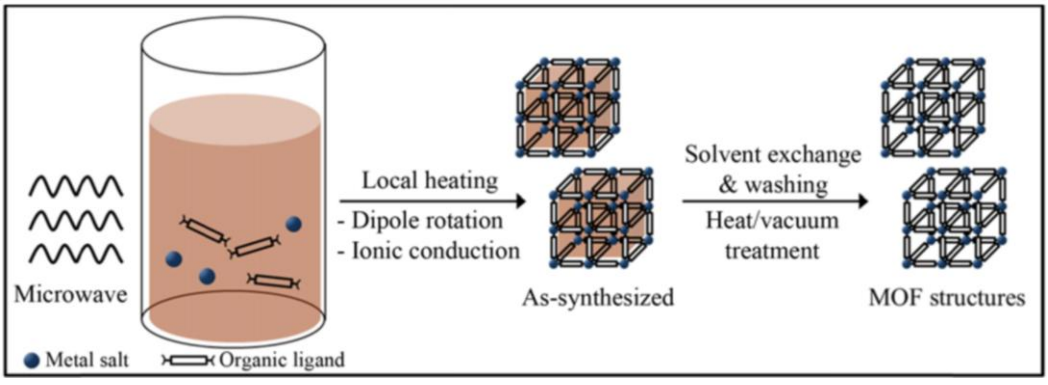

Figure 2.9. Microwave-assisted solvothermal/hydrothermal synthesis of MOF [60].

2.2.3. Sonochemical Synthesis

Sonochemical method also called as ultrasonic-assisted synthesis involves ultrasonic generator to heat the mixture. As it is shown in Figure 2.10, the formation and collapse of bubbles via sonication create turbulence and thus high local temperatures and pressures are produced. As a result, fine crystallites with a reduction in crystallization time have been observed in this method [59,60,63–65].

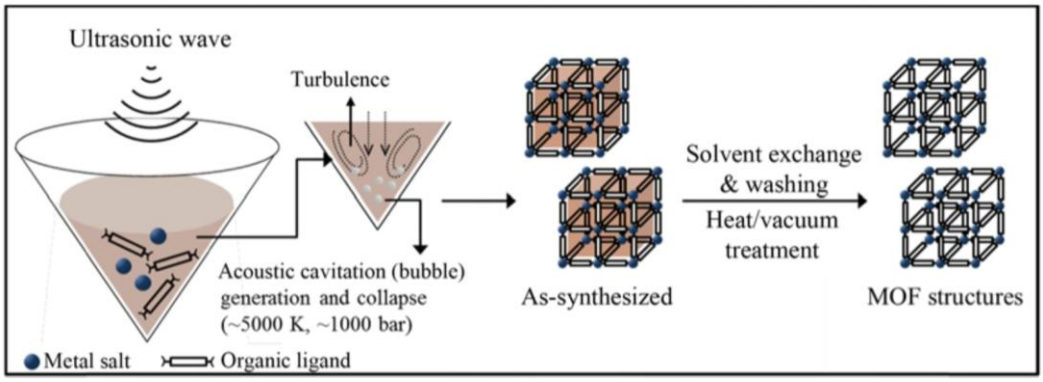

Figure 2.10. Ultrasonic-assisted solvothermal/hydrothermal synthesis of MOF [60].

2.2.4. Electrochemical Synthesis.

Two electrodes are placed in electrolysis cell, which contain dissolved organic molecules and solvent mixture as it is shown in Figure 2.11. The need of metal atoms is supplied from electrodes through anodic dissolution instead of adding metal salts. This process continues until a specific reaction time under constant voltage. Faster synthesis at lower temperatures is achieved compared to conventional methods [60,63].

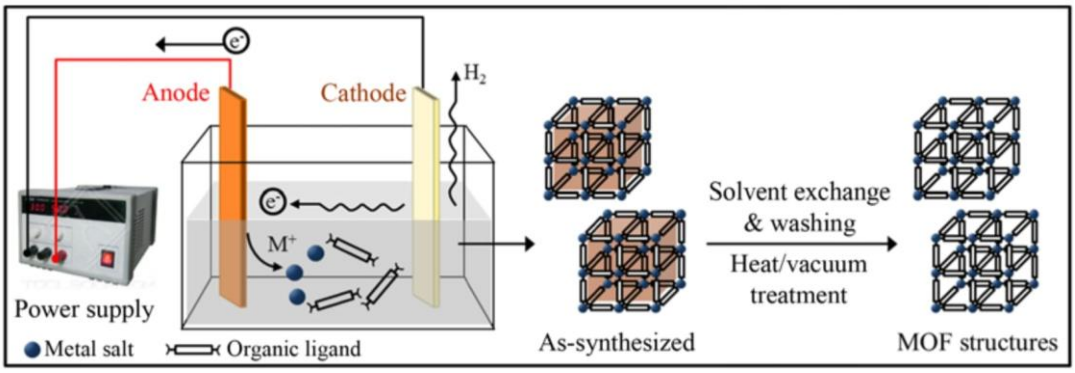

Figure 2.11. Electrochemical synthesis of MOF [60].

2.2.5. Mechanochemical Synthesis

Intermolecular bonds are broken mechanically in a grinding jar, which is presented in Figure 2.12. After breakage of bonds, chemical transformation occurs in mechanochemical synthesis. This synthesis route can be performed either under solvent-free condition or with solvent mixture. The advantage of first type is the non-necessity of the solvent removal whereas the latter one lead to increase in mechanochemical reactions [60,63]. To compare, the solvent-free HKUST-1 produced by Pichon *et al.* had 1084 m^2/g of surface area whereas the one with methanol as solvent had the surface area of 1364 m^2/g [66].

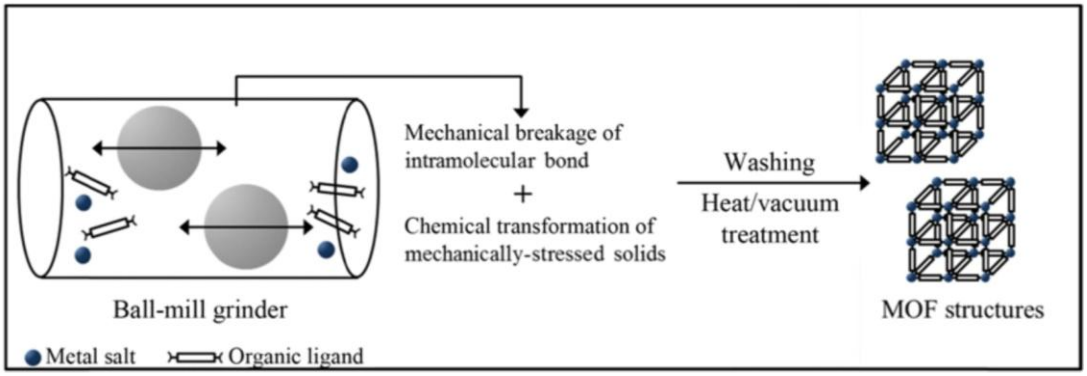

Figure 2.12. Mechanochemical synthesis of MOF [60].

2.2.6. Comparative Studies on Synthesis Routes

Sabouni *et al.* have investigated the difference between conventional solvothermal method using electrical oven and the two-step technique using ultrasonic followed by microwave method. Although, IRMOF-1 was tried to be synthesized using conventional oven technique for three times, none of them were successful. Hence, the results of the MOF obtained by two-step technique were compared with literature. Particle size of samples obtained by two-step technique was smaller than those of conventional solvothermal synthesis method. They have also observed that, particle size of samples obtained by two-step technique decreased as the ultrasonic bath temperature and sonication time increased [59].

In 2010, Schlesinger *et al.* have conducted a study in which six different synthesis methods (solvothermal, microwave-assisted, under atmospheric pressure at reflux, ultrasonic, electrochemical and mechanochemical) were compared. Based on this study, it was revealed that samples produced by using microwave-assisted solvothermal had high BET-surface area of 1499 m^2/g and specific pore volume of 0.79 cm^3/g. Moreover, this method is the fastest one among them to obtain pure HKUST-1 in high yield. Also, they have shown that better results were achieved with DMF solvent rather than ethanol/water mixture. Moreover, they have demonstrated that the samples produced by the solvent-assisted mechanochemical reaction gave comparable pore volumes (0.74 cm^3/g) with those obtained by microwave-assisted synthesis [67].

Loera-Serna and her coworkers have investigated the effect of synthesis methods on HKUST-1. Three synthesis routes as conventional solvothermal synthesis, ultrasonic-assisted synthesis and room temperature stirring synthesis have been examined, and the HKUST-1 was obtained in all these methods as the physicochemical structures determined by several characterization techniques such as XRD, FT–IR, SEM and TGA. However, the different coordinative and bulk molecular guests were observed among those distinctly synthesized MOFs. It was also seen that MOFs obtained with room temperature stirring method have high purity compared to those obtained with other syntheses routes [65].

Haque *et al.* have studied three paths to synthesize the CPO-27 under same conditions but different heating mode. These are namely, conventional solvothermal method, ultrasound method and microwave method. They have found that the crystal sizes decrease in the order of conventional > microwave > ultrasound whereas the adsorbed nitrogen, surface area, and pore volume are in the order of conventional < microwave < ultrasound [64].

In 2013, Wu *et al.* and coworkers have worked on a comparative M-MOF-74 synthesis study. They synthesized M-MOF-74 (M = Ni, Mg) with both hydrothermal and microwave-assisted method. Samples produced by using microwave-assisted method had larger surface area and pore volume than those produced with hydrothermal method. Also, the reaction time reduces from 1 day to 100 min in microwave-assisted synthesis. They have also tested the adsorption behavior of several gases (CO_2, CH_4, N_2, C_2H_4, C_2H_6, C_3H_6

and C_3H_8) on produced MOF-74 an observed that gas adsorption capacity and adsorption selectivity were improved with the microwave-assisted method [68].

2.3. Data Mining Analyses

Data mining, as a term, refers to 'knowledge mining or transferring from existed data'. From the perspective of data mining, all kinds of choices of people (for instance in market, in finance etc.) and events (i.e. weather forecast, natural disaster etc.) embodies data within itself so it is worth to store them [69]. Hence, there is significant amount of data present in the world and yet it is increasing day by day. At this point, getting the benefit from large existed databases is the ultimate target. The aim of the data mining is to extract the meaningful information and to identify the hidden patterns in large data. Data is analyzed from various perspectives and then all findings are summed up into useful and understandable information. In this sense, data mining is used for not only the characterization of the data but also prediction for the future [70].

Rows and columns compose datasets; the points or objects are indicated as rows whereas columns are mostly referred to the descriptors, variables, parameters or properties. The idea behind those attributes is that the entire columns except one are included to data to define the remaining one, which is called as output. Two kinds of attributes, namely categorical and continuous can be present in the database. For instance, type of gender as male and female can be considered as categorical whereas ages of people from 0 to 70 can be taken as continuous attribute. They are able to convert into each other for target application; this procedure was explained in sections that the implementation was presented. Before starting to apply data mining tools, getting to know about the related data is crucial; the data reprocessing is carried out to deal with the missing values and incorrect data points.

Data mining is the promising field in recent years; it combines the database, machine learning, statistics and pattern recognition with the help of several tools. The mostly utilized data mining methods for both classification of the existed data and the prediction of unseen data are linear regression, logistic regression, decision trees, artificial neural network, k-nearest neighboring and support vector machine. After constructing the model,

evaluation of model performance is fundamental in data mining; hold-out, k-fold cross validation, leave-one-out, and bootstrapping are the widely used methods for that purpose [9].

2.3.1. Linear and Logistic Regression

The mostly used statistical method in order to identify the relationship between two variables is the regression model in which output variable (also called as response or dependent variable) is dependent to the input variable (independent variable). If the output variable is any one of an infinite number of possible values (termed as continuous) the model is considered as linear regression whereas logistic regression is used when the response variable is any one of only a limited number of possible values (named as categorical).

The simple regression formula is given in Equation 3.1 where $\hat{y}$ is the predicted output variable, x is the input variable, b_0 and b_1 are called the regression coefficients. Those regression coefficients are found based on the least square analysis in which the overall error between actual (y_i) and predicted ($\hat{y}_i$) output variable is minimized. Overall error is called as sum of square errors (SSE), which is determined by the formula given in Equation 2.2 [9].

$$\hat{y} = b_0 + b_1 x \tag{2.1}$$

$$\text{SSE} = \sum_{i=1}^{n} e_1^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{2.2}$$

In case of the more than one input variables are included in data, multiple regression model is preferred whose formula is shown in Equation 2.3. Same procedure is carried out with simple regression to determine the regression coefficients (b_0, b_1, b_2 and b_3) [9].

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \tag{2.3}$$

Regression models are tailor-made for describing the linear relationship between dependent and independent variables. Hence, this method is insufficient for the examining the non-linearity of the systems.

2.3.2. Decision Trees

Decision tree is a classification method, in which the sequences of decisions are collected and then connected by branches. In a decision tree, *root node* represents the very first *decision node* and it is placed at the top of the decision tree. Then all attributes are tested at the decision nodes starting from the root node, and each of the instances are put into branches. There are two options present for branches; each branch either moves on to another decision node or terminates the algorithm as a *leaf node*. A simple decision tree for the weather data is illustrated in Figure 2.13.

a)

Number of instances	Outlook	Temperature	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rainy	mild	high	false	yes
5	rainy	cool	normal	false	yes
6	rainy	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rainy	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rainy	mild	high	true	no

b)

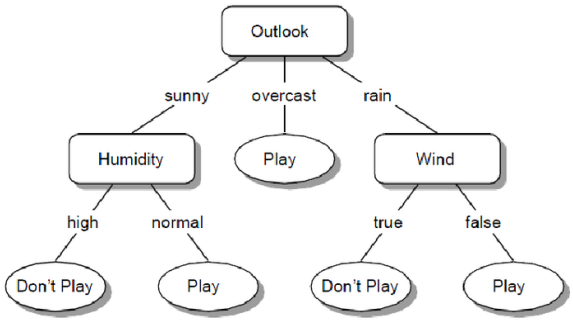

Figure 2.13. a) The weather data and b) the corresponding simple decision tree [69].

In Figure 2.13, there are 14 number of data points which have four attributes as outlook, temperature, humidity and windy. Those attributes have symbolic values; outlook can be sunny, overcast, or rainy; temperature can be hot, mild, or cool; humidity can be high or normal; and windy can be true or false. The output variable is whether to play golf at outside or not. What we obtained from this decision tree is that the decision to play or not can be made based on the weather conditions. For instance, if the outlook is sunny and the humidity is high, not to play golf is recommend by this decision tree.

Optimization is important in data mining tools in order to decide how far the model will grow. *Overlearning* is occurred when the repeated training iterations improve the performance of the model on training dataset by memorizing the training samples whereas the model shows poor performance on new data (test dataset). On the other hand, *underlearning* occurs in small trees where both errors are high because of the insufficient decision nodes [71]. Overlearning corresponds to the right hand side of the Figure 2.14, whereas the term underlearning is occurred at the left hand side. In the computational part of data mining, the optimum point to stop the tree further splitting is sought.

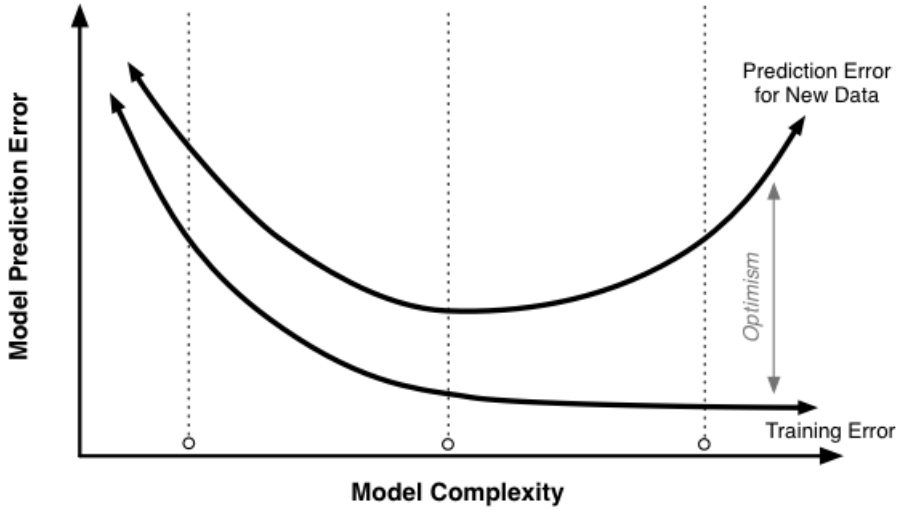

Figure 2.14. Model complexity vs. model prediction error [72].

2.3.3. Artificial Neural Network

Artificial neural network (ANN) is a non-linear mechanism that has been inspired by the learning mechanism of animal brain. Simply, it takes stimulus comes in to brain, process them through the combination function and generates the output [9]. To illustrate the similar idea behind the artificial neural network algorithm and animal brain process, Figure 2.15 was given below.

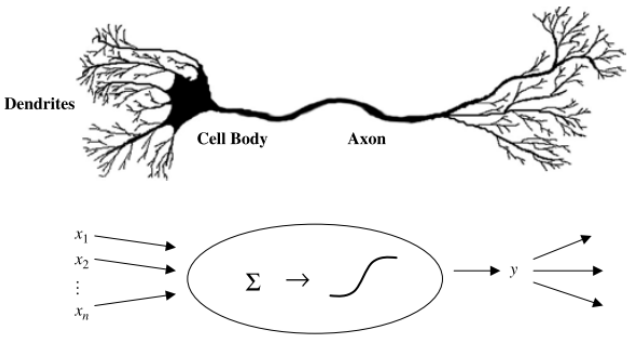

Figure 2.15. Real neuron vs. artificial neuron [9].

A neural network consists of input layer, hidden layer and output layer; however in some cases there may be more than one hidden layer depending on data. The user determines the number of nodes in each layer as well as the number of hidden layer. As it was mentioned before, optimization process to find out such optimum values is fundamental in machine learning.

A simple neural network algorithm is shown in Figure 2.16. Each node input has link to connect another node and inputs proportioned with their connection weights are summed up in the output node. The algorithm sets a threshold value for the difference between predicted and actual value; connection weights are varied until the threshold value is achieved. The linear combination of node inputs (x) and connection weights (w) are provided with the combination function and it is called as *net*. The formula is given in Equation 2.4. for a given node j, where x_{ij} represents the i^{th} input to node j, w_{ij} represents the weight related to the i^{th} input to node j, and there are n+1 inputs to node j. After the calculation of net_A, this value is converted to output value (y) via a non-linear activation function in order to be passed along the final node (i.e. Node Z). Same process is followed for each node and finally the output value of the Node Z is calculated by using its net_Z value via activation function [9].

$$net_j = \sum_i w_{ij}x_{ij} = w_{0j}x_{0j} + w_{1j}jx_{1j} + \ldots + w_{nj}x_{nj} \tag{2.4}$$

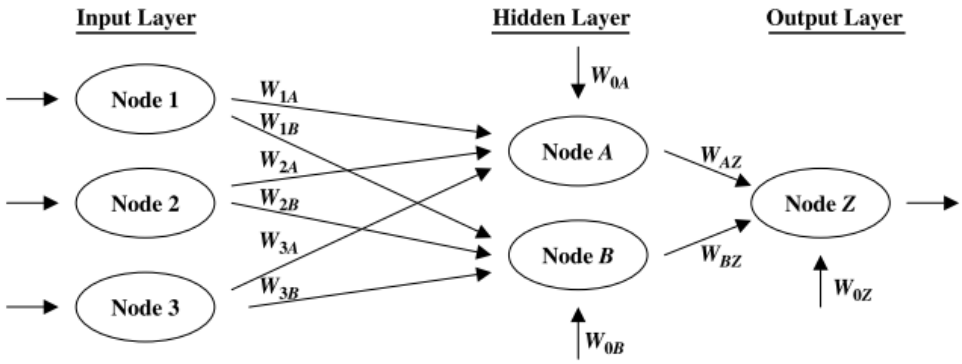

Figure 2.16. Simple neural network algorithm [9].

Activation function plays a significant role in neural network model and some common activation functions are given in Table 2.1.

Table 2.1. Activation functions.

Activation Function	Step	Linear	Log-sigmoid	Hyperbolic Tangent Sigmoid
Mathematical Notation	$y_j=\begin{cases}A, & net\geq 0\\ B, & net<0\end{cases}$	$y_j=net$	$y_j=\frac{1}{(1+e^{-net})}$	$y_j=\frac{(e^{net}+e^{-net})}{(e^{net}+e^{-net})}$

2.4. Data Mining Studies on Adsorption of CH_4 in MOFs

Wilmer and his group performed large-scale quantitative structure–property relationship (QSPR) modeling for CH_4 storage in MOFs in 2013. They investigated the effect of structural parameters such as pore size and void fraction on CH_4 storage of 130,000 hypothetical MOFs. Their aim was to develop models that make the prediction of CH_4 storage with high accuracy based only on the knowledge of geometric features of materials. Multi-linear regression models, decision trees and nonlinear support vector machines were developed. According to their study, the void fraction and the dominant pore diameter were the parameters that strongly correlated with CH_4 storage at high pressure [10].

In 2016, Ohno and Mukae conducted a study to predict the amount of CH_4 uptake over 499 hypothetical MOFs by using support vector regression, neural network, Gaussian

process regression and linear regression. Density, surface area and void fraction were found as the dominant parameters for CH_4 adsorption among the quantitative descriptors. The performance of the models were evaluated by comparing the coefficient of determination (R^2) and root mean square error (RMSE); it was observed that all nonlinear models were superior to the linear regression model [11].

Pardakhti *et al.* investigated predictive capability of decision tree, Poisson regression, support vector machine and random forest in terms of the CH_4 adsorption over 130,398 hypothetical MOFs. Input variables were arranged in a way that structural variables, chemical properties and both structural and chemical properties; all three classes of inputs were tested for volumetric and gravimetric based uptake data. Mean absolute percentage error (MAPE), mean error (ME) and root mean square error (RMSE) values of both training and test data set were calculated to evaluate the performance of the each model. They concluded that predictability of data mining models developed by combining the structural properties with chemical variables was best among all models [2].

3. COMPUTATIONAL WORKS

3.1. Data Preparation

A database containing 3000 real MOFs [73] was developed by taking additional parameters into consideration as it will be explained below; the development of the database is shown in Figure 3.1.

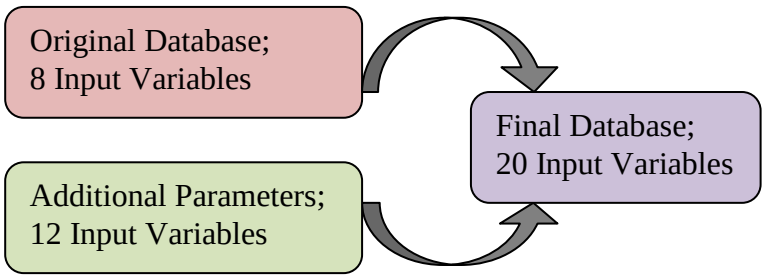

Figure 3.1. The development of the database.

3.1.1. The Original Form of the Database

The database, which contains CH_4 uptake values for 3000 real MOFs includes eight variables; these are crystal density (ρ_{cryst}), pore volume (V_p), gravimetric surface area (S_g), pore limiting diameter (PLD), maximum pore diameter (MPD), dominant pore diameter (DPD), isosteric heat of adsorption (Q_{st}) and unit cell volume (UCV) of MOFs [73]. GCMC simulations were performed at adsorption pressures of 1, 5, 10, 15, 20, 25, 30, 35, 50, and 65 bar and 298 K to calculate isosteric heats of adsorption values and CH_4 uptake of each MOF. In addition, unit cell volume (UCV) of each MOF was taken from the Cambridge Crystallographic Data Centre (CCDC) [74].

3.1.2. Additional Parameters

In order to obtain strong models constructed with data mining tools, MOFs should be well defined; in other words, all information that might have an effect on CH_4 uptake over MOFs should be included in the database. To do this, original database was improved and

the number of input variables increased. Cell lengths, cell angles, chemical formula, space group of MOFs and organic linker of the MOFs were taken from the CCDC [74]; based on this information, eleven new input variables were created in order to take all aspects of the MOFs into consideration.

First, the atoms contained in each MOF were counted and separated according to their elements; during this calculation, guest molecules and solvents were excluded. Then, group number and the atomic radius of the metal atoms for each MOF were taken as the input variables to test whether they have an effect on the uptake values or not. At this stage, it should be mentioned that some MOFs contains more than one type of metal atoms; in these cases, the dominant metal (having the larger number atom) was selected as the variable. If the metal atoms are even, then the first atom appeared in the formula was chosen as the dominant metal atom. The group number and the atomic radius were determined based on dominant metal atom [75]. Metallic percentage, electronegativity ratio, total degree of unsaturation and nitrogen to oxygen ratio were also calculated as described by Pardakhti *et al.* [2] using the information available in the CCDC [74]; the definition of these variables are given in Table 3.1. The cell lengths, cell angles, space groups and organic linkers were also kept in the database.

Table 3.1. Calculated input variables.

Variable	**Formula**
Metallic percentage	$M / T \times 100$
Electronegativity ratio	E / T
Total degree of unsaturation	$[(C \times 2) + 2 - H] / 2$
N / O	N / O

E: Total number of Nitrogen, Oxygen, Fluorine, Chlorine, Bromine and Sulfur atoms
M: Number of metal atoms
T: Number of total atoms
C: Number of Carbon atoms
H: Number of Hydrogen atoms
N: Number of Nitrogen atoms
O: Number of Oxygen atoms

3.1.3. The Final Database

The original form of the database was enhanced with additional parameters in order to make a complete analysis of CH_4 uptake over MOFs, and then the final database was modeled by using data mining tools. Since CH_4 uptake values were simulated at ten different adsorption pressures, ten different CH_4 uptake values were obtained; The number of gas molecules adsorbed per unit cell volume of MOF is given as the CH_4 adsorption unit. Isosteric heat of adsorption defines the interaction between CH_4 molecules and MOF atoms hence it can be considered as a chemical property rather than structural property. Therefore, it was also excluded from the database to build models based on easily definable descriptors and easily measurable structural properties. The unprocessed list of input and output variables was given Table 3.2; their ranges were calculated after the removal of missing data in each variable separately.

Table 3.2. The unprocessed input and output variables of the database.

Input Variables	**Ranges**	**Number of Missing Data**
Original Database		
ρ_{cryst}, g/cm^3	0.17 – 7.20	42
V_p, cm^3/g	0 – 5.55	42
S_g, m^2/g	0 – 6451.81	801
PLD, Å	1.72 – 71.07	42
MPD, Å	2.10 – 71.19	42
DPD, Å	0.05 – 71.05	763
UCV, Å^3	237.78 – 701,860.60	309
Additional Parameters		
Cell lengths, Å	2899 kinds	4
Cell angles, α β γ	987 kinds	5
Space group	191 kinds	7
Organic linker	66 kinds	2817

Table 3.2. The unprocessed input and output variables of the database. (cont.)

Input Variables	**Ranges**	**Number of Missing Data**
Additional Parameters		
Metal type	57 kinds	3
Second Metal type	55 kinds	2551
Group number*	1A, 2A, 3A, 4A, 5A, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, Lanthanide, Actinide	3
Atomic number*	1.53 – 2.43	3
Metallic percentage, %	0.45 – 23.81	313
Electronegativity ratio	0.03 – 0.67	313
Total degree of unsaturation	(-101) – 313	0
N / O	0 - 34	541
Output Variables (Number of adsorbed molecules/Unit cell)	**Ranges**	**Number of Missing Data**
CH_4 uptake at 1 bar	1.42×10^{-6} – 396.05	308
CH_4 uptake at 5 bar	7.04×10^{-6} – 834.01	308
CH_4 uptake at 10 bar	1.36×10^{-5} – 1116.58	308
CH_4 uptake at 15 bar	1.94×10^{-5} – 1495.08	308
CH_4 uptake at 20 bar	2.46×10^{-5} – 1847.52	308
CH_4 uptake at 25 bar	3.21×10^{-5} – 2163.86	308
CH_4 uptake at 30 bar	4.05×10^{-5} – 2062.54	308
CH_4 uptake at 35 bar	5.08×10^{-5} – 2236.24	308
CH_4 uptake at 50 bar	6.27×10^{-5} – 1875.76	308
CH_4 uptake at 65 bar	8.82×10^{-5} – 1982.07	308

*Belonging to dominant metal atom in the MOF

3.1.4. Reprocessing the Final Database

In this section, the steps that were followed to prepare the database for modeling were explained in detail. First, categorical variables containing a large number of types were arranged so that they can be used as input variables. Then the unit conversion of the output variable was carried out. Finally, the best sets of input variables were selected according to results of the cross-correlation analysis, and some minor restrictions were made for those input variables.

3.1.4.1. Categorical Variables. There are both categorical (i.e. metal type) and continuous (i.e. V_p) variables in the final database as it was seen in Table 3.2. To prepare the database for modeling, first, the categorical variables containing large number of species should be reduced to a relatively small number of species; otherwise models will not run. For this reason, cell length, cell angle, space group, organic linker and metal type variables were reexamined.

The cell length includes three sub-lengths a, b, c in the unit of Å; 2899 out of 2996 data points (2996 data points are left after subtracting missing data) are unique and the number of data points having the same cell lengths is at most two. Since the number of species cannot be reduced, it has been tried to transform the cell length column into three sub-columns as a, b and c; however, this also could not be done because the cell length would be meaningless on their own as they separated. Therefore, the cell length was excluded from the database. As in the case of cell length, cell angle also contains three sub-angles as α, β and γ. Since 987 data points out of 2995 are different from each other, the cell angle was also excluded from the database.

In general, there are 230 different space groups in 3D, and each space group is a subset of one of the seven crystal systems (monoclinic, triclinic, trigonal, tetragonal, cubic, hexagonal, orthorhombic); hence, without losing any data points, the space groups were represented by converting to crystal systems that they belonged [76]; for instance, if the space group of a MOF is one of the following space groups *P2, $P2_1$, C2, Pm, Pc, Cm, Cc, P2/m, $P2_1$/m, C2/m, P2/c, $P2_1$/c, C2/c,* its crystal structure will be *"monoclinic"*.

Since the missing data points in organic linker variable are 94 % of the entire database (2817 out of 3000), this variable was directly removed from the database. In the case of metal type, the MOFs containing only the most frequently (more than 50 data points) used seven metals (Zn, Cu, Cd, Co, Fe, Ni, and Mn) were considered. The type of the second metal was not used as an input variable in models because only 449 MOFs out of 3000 had two types of metal; however, the contribution of a second metal to the metallic percentage of the MOFs was considered.

3.1.4.2. Determination of Output Variable. The CH_4 delivery from 65 bar to 5 bar and CH_4 uptake at 35 bar were used as the measure for the CH_4 adsorption performance of MOFs. Because the computed CH_4 uptake values were in the units of number of adsorbed gas molecules/UCV of MOF, they were converted to both volumetric and gravimetric values; the conversion steps were given in Table 3.3. Then, to calculate CH_4 delivery as output variable, the amount CH_4 uptake at 5 bar was subtracted from the amount of CH_4 uptake at 65 bar; this procedure was followed for both volumetric and gravimetric CH_4 delivery.

Table 3.3. Conversion steps of CH_4 uptake values.

Obtained value	**Conversion steps**
Volumetric CH_4 uptake $\left(\frac{cm^3\ CH_4}{cm^3\ MOF}\right)$	$= \left(\frac{\text{Number of adsorbed gas molecules}}{\text{UCV}}\right)^{a} / (Å^3)^{b} \times \frac{1\ Å^3}{10^{-24} cm^3} \times \frac{22400\ cm^3\ CH_4}{1\ \text{mole}} \times \frac{1\ \text{mole}}{6.02\times10^{23}\ \text{molecules}}$
Gravimetric CH_4 uptake $\left(\frac{cm^3\ CH_4}{g\ MOF}\right)$	$=$ Volumetric CH_4 uptake $\left(\frac{cm^3\ CH_4}{cm^3\ MOF}\right) / \left(\frac{g}{cm^3}\right)^{c}$

[a] CH_4 uptake value of MOF

[b] UCV of MOF

[c] Density of MOF

The variables, which were already used to convert adsorbed CH_4 molecules per UCV of MOF to gravimetric and volumetric CH_4 uptake in Table 3.3, were not included in the

models as input variables; UCV and the density of MOFs were excluded in the analyses of gravimetric CH_4 whereas UCV of MOFs was removed from the volumetric based analyses.

3.1.4.3. Cross-correlation Analysis. Input variables were analysed so that one of the strongly correlated variables (based on the strength of correlation and the type of data mining tool) could be excluded from the database. Correlation coefficient ranges from -1 to +1; -1 means the perfect negative correlation whereas +1 indicates the perfect positive correlation [77]. In this analysis, values with a correlation coefficient of 0.9 or higher were identified as strong correlations, while values between 0.6-0.9 were considered as mild correlation. After removing all missing points, cross-correlation analysis was carried out to a database consisting 1792 data points using *cor* function in RStudio (Version 1.0.153); the results are given in the Table 3.4.

The best set of input variables to develop the ANN regression models were selected based on the outcomes of the cross-correlation analysis considering that the input variables should not be highly correlated; both strong and mildly correlated were excluded. Although the multicollinearity is also undesirable for the decision tree analysis, it is not as critical as in the case of neural network models [71,77]; hence we eliminated strong correlations in the decision tree as well, but kept the mildly correlated variables.

Table 3.4. Cross-correlation analysis.

	UCV	Atomic Radius	Uptake at 5 bar	Uptake at 35 bar	Uptake at 65 bar	ρ_{cryst}	V_p	S_g	PLD	MPD	DPD	Electronegativity Ratio	Metallic %	N/O	Total Degree of Unsaturation
UCV	1.00	-0.15	0.83	0.95	0.97	-0.38	0.57	0.47	0.45	0.65	0.63	-0.22	-0.13	0.04	0.47
Atomic Radius	-0.15	1.00	-0.16	-0.17	-0.17	0.34	-0.28	-0.26	-0.21	-0.25	-0.25	0.17	-0.09	-0.07	-0.12
Uptake at 5 bar	0.83	-0.16	1.00	0.92	0.87	-0.37	0.38	0.35	0.23	0.47	0.46	-0.27	-0.18	0.03	0.58
Uptake at 35 bar	0.95	-0.17	0.92	1.00	0.99	-0.43	0.57	0.53	0.38	0.64	0.62	-0.24	-0.13	0.02	0.55
Uptake at 65 bar	0.97	-0.17	0.87	0.99	1.00	-0.43	0.61	0.55	0.42	0.66	0.65	-0.22	-0.13	0.01	0.52
ρ_{cryst}	-0.38	0.34	-0.37	-0.43	-0.43	1.00	-0.69	-0.70	-0.43	-0.55	-0.54	0.60	0.46	0.01	-0.36
V_p	0.57	-0.28	0.38	0.57	0.61	-0.69	1.00	0.92	0.68	0.79	0.79	-0.35	-0.19	-0.05	0.35
S_g	0.47	-0.26	0.35	0.53	0.55	-0.70	0.92	1.00	0.54	0.70	0.70	-0.32	-0.14	-0.05	0.34
PLD	0.45	-0.21	0.23	0.38	0.42	-0.43	0.68	0.54	1.00	0.85	0.85	-0.24	-0.08	-0.05	0.14
MPD	0.65	-0.25	0.47	0.64	0.66	-0.55	0.79	0.70	0.85	1.00	1.00	-0.24	-0.07	-0.02	0.34
DPD	0.63	-0.25	0.46	0.62	0.65	-0.54	0.79	0.70	0.85	1.00	1.00	-0.24	-0.07	-0.02	0.33
Electronegativity Ratio	-0.22	0.17	-0.27	-0.24	-0.22	0.60	-0.35	-0.32	-0.24	-0.24	-0.24	1.00	0.65	0.05	-0.36
Metallic %	-0.13	-0.09	-0.18	-0.13	-0.13	0.46	-0.19	-0.14	-0.08	-0.07	-0.07	0.65	1.00	-0.02	-0.27
N/O	0.04	-0.07	0.03	0.02	0.01	0.01	-0.05	-0.05	-0.05	-0.02	-0.02	0.05	-0.02	1.00	0.08
Total Degree of Unsaturation	0.47	-0.12	0.58	0.55	0.52	-0.36	0.35	0.34	0.14	0.34	0.33	-0.36	-0.27	0.08	1.00

Two strong correlations were detected in the input variables: (i) between pore volume and surface area, (ii) among maximum pore diameter, dominant pore diameter and pore limiting diameter. Hence, surface area, DPD and PLD were excluded from the database because of having large number of missing data points; and consequently pore volume and MPD was selected to use in the models.

The pore volume and MPD as well as the metallic percentage and electronegativity ratio were found to be also mildly correlated with the correlation coefficient of 0.79 and 0.65 respectively. Since density could be used as an input variable in the volumetric based analyses, the correlations between density and pore volume as well as density and surface area were found as -0.69 and -0.70, respectively. Although even the moderate correlations may overshadow the effects of each other in ANN analysis, their exclusion seemed to cause some information loss; hence, they were also included in models while the analyses in their absence were also provided in Appendix B for comparison. The multicollinearity is also undesirable for the decision tree analysis even though it is not as critical as the other predictive models [71,77]; hence, we eliminated strong correlations in the decision tree as well, but we also kept the moderately correlated variables.

The data points, in which the value of pore volume was lower than 0.1 cm^3/g were removed from the database because these materials do not have enough space to store guest molecules.

The final list of input variables chosen as the user-defined descriptors, structural properties and output variables with their ranges are given in Table 3.5. The general structure for the decision tree and artificial neural network analyses are given in Figure 3.2 and Figure 3.3, respectively.

Table 3.5. Final list of input and output variables of the database.

Variables	**Ranges**
User Defined Descriptors	
Space group	monoclinic, triclinic, trigonal, tetragonal, cubic, hexagonal, orthorhombic
Metal type	Zn, Cu, Cd, Co, Fe, Ni, Mn
Group number*	1A, 2A, 3A, 4A, 5A, 1B, 2B, 3B, 4B, 5B, 6B, 7B, 8B, lanthanide, actinide
Atomic number*	1.53 – 2.43
Metallic percentage, %	0.45 – 23.81
Electronegativity ratio	0.03 – 0.67
Total degree of unsaturation	(-101) – 313
N / O	0 - 34
Structural Properties	
V_p, cm^3/g	0.10 – 5.55
MPD, Å	2.10 – 71.19
ρ_{cryst}, g/cm^3	0.17 – 2.76
Output Variables	
Gravimetric CH_4 delivery, cm^3/g	0.01 – 818.87
Gravimetric CH_4 uptake at 35 bar, cm^3/g	0.01 – 524.38
Volumetric CH_4 delivery, cm^3/cm^3	0.02 – 231.65
Volumetric CH_4 uptake at 35 bar, cm^3/cm^3	0.01 – 336.32

*Belonging to dominant metal atom in the MOF

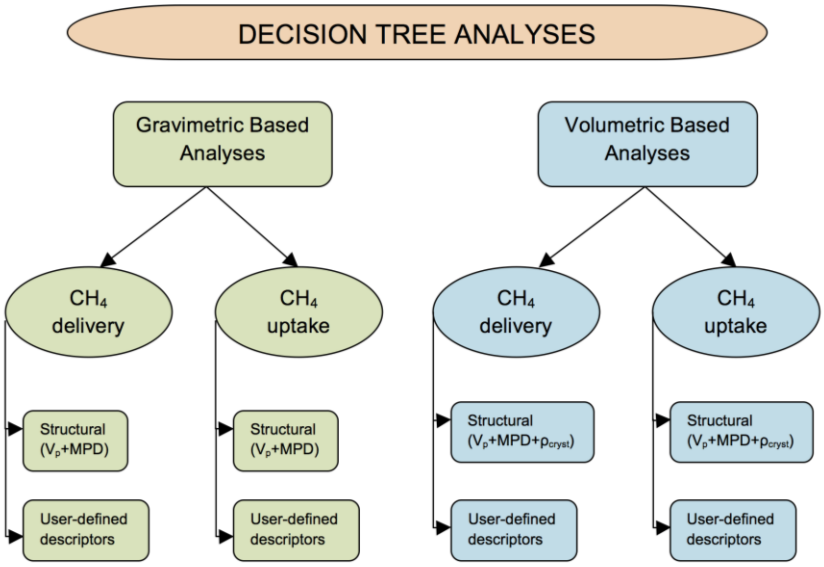

Figure 3.2. Plan for the decision tree analyses.

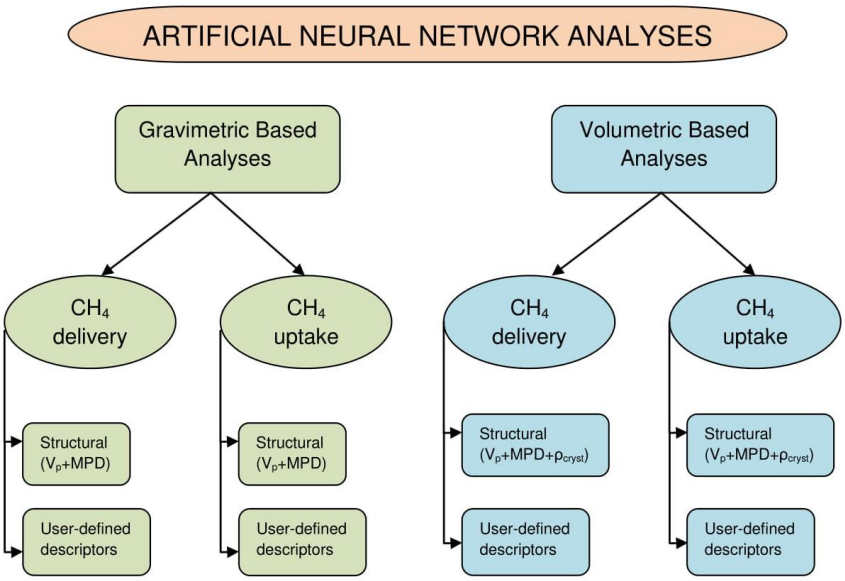

Figure 3.3. Plan for the artificial neural network analyses.

3.2. Computational Details

The computational analyses were performed using RStudio (Version 1.0.153), which is an open source for statistical computing and graphics. Several data mining algorithms

and functions were used; several codes were written in RStudio for this work to achieve the best results.

3.2.1. Decision Tree Analysis

Decision tree analysis was performed using *rpart* function of RStudio (Version 1.0.153) in classification mode. Gini index was used as the measure to select the optimal split; *minbucket* number (minimum number of observations in leaf / terminal node) and *maxdepth* number (depth of the tree) were tuned to obtain the optimal tree [78–80]. The five-fold cross validation was applied to determine the final tree structure; accordingly, the data was divided into five subsets; the decision tree is constructed from four sets and tested with the remaining one. The procedure was repeated for five times for all five combinations, and the average testing error (the fraction of the wrong classifications to the total observations) was calculated from all folds. To divide the database into five folds, first *rep_len* function is used to replicate the values of 1, 2, 3, 4 and 5 (number of folds) until reaching the number of the data points; then *sample* function is utilized to mix them randomly. This newly constructed index vector serves as the basis of the fold selection; in other words, elements having the value of 1 form the 1^{st} fold, the value of 2 forms the 2^{nd} fold and so on.

In order to determine the *minbucket* and *maxdepth* value, four nested loops were utilized; the outermost loop sets the number of run (in our case it was ten), the innermost loop controls the five-fold cross validation and the intermediate loops were utilized to check the *minbucket* and *maxdepth* values. Then testing and training accuracies were calculated by using Equation 3.1 and they were given in Table 3.6 and Table 3.7; the models with *maxdepth* values of 1 and 2 were indicated as “Nan” because data were not classified successfully. For all *maxdepth* values, it was detected that accuracies are decreasing as *minbucket* values increasing. When the number of observations in leaf node namely *minbucket* is very high, the decision tree will be too simple in order to classify these observations correctly and consequently the accuracy will be low. Hence, it is important to balance the accuracies with the purity of the terminal nodes. Since there were no significant differences in accuracies and the purity of the terminal nodes was another important criterion to determine the *minbucket* and *maxdepth* values, these values were

determined by checking the purity of the terminal nodes instead of calculating accuracies. As a result, the *minbucket* and *maxdepth* giving the relatively high average testing and training accuracies with sufficiently pure terminal nodes were determined of seven and eight, respectively for all models. Then the decision tree structures were constructed using these values to determine the rules and heuristics for the CH_4 delivery / CH_4 uptake.

$$\text{accuracy} = \frac{\text{correctly placed instances}}{\text{all instances}} \times 100 \tag{3.1}$$

Table 3.6. Testing accuracy (%) for gravimetric CH_4 delivery using user defined descriptors.

Minbucket Values	**Maxdepth Values**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	NaN	NaN	55.16	55.43	56.19	56.60	57.70	58.24	58.72	59.46
2	NaN	NaN	55.16	55.43	56.19	56.59	57.67	58.20	58.68	59.39
3	NaN	NaN	55.16	55.37	56.11	56.47	57.50	57.87	58.35	58.91
4	NaN	NaN	55.15	55.29	55.99	56.23	57.35	57.59	58.12	58.79
5	NaN	NaN	55.07	55.19	55.78	55.94	57.16	57.30	57.75	58.29
6	NaN	NaN	54.96	54.86	55.62	55.67	56.64	56.69	57.04	57.43
7	NaN	NaN	54.96	54.77	55.44	55.46	56.29	56.22	56.51	56.79
8	NaN	NaN	54.96	54.79	55.36	55.51	55.99	56.01	56.27	56.50
9	NaN	NaN	54.91	54.65	55.16	55.39	55.59	55.68	55.85	56.08
10	NaN	NaN	54.91	54.65	55.23	55.38	55.55	55.59	55.64	56.10
15	NaN	NaN	54.93	54.74	54.98	55.26	55.34	55.53	55.68	55.68
20	NaN	NaN	55.02	54.83	55.00	55.13	54.99	55.36	55.36	55.38
25	NaN	NaN	55.10	54.90	54.92	54.86	54.89	55.05	55.21	55.22
30	NaN	NaN	55.19	54.96	55.11	54.70	54.59	54.83	54.94	54.94
50	NaN	NaN	55.15	54.07	54.50	54.58	54.24	54.32	54.32	54.32

Table 3.7. Training accuracy (%) for gravimetric CH_4 delivery using user defined descriptors.

Minbucket Values	**Maxdepth Values**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	NaN	NaN	59.21	59.71	59.12	60.79	64.38	68.39	69.64	71.72
2	NaN	NaN	59.21	59.71	59.12	60.79	64.38	68.39	69.64	71.72
3	NaN	NaN	59.21	59.71	59.12	60.66	64.24	68.11	69.21	71.35
4	NaN	NaN	59.21	59.37	59.22	60.77	64.35	68.27	69.28	71.43

Table 3.7. Training accuracy (%) for gravimetric CH_4 delivery using user defined descriptors. (cont.)

Minbucket Values	Maxdepth Values									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
5	NaN	NaN	59.21	59.21	58.62	59.90	63.23	66.89	68.36	70.16
6	NaN	NaN	59.21	59.21	58.62	59.90	63.45	66.81	68.34	70.23
7	NaN	NaN	58.83	58.83	58.51	59.79	63.33	66.75	68.16	69.45
8	NaN	NaN	58.83	58.83	58.51	59.24	62.53	65.96	67.43	68.64
9	NaN	NaN	58.83	58.83	57.99	58.73	62.41	65.69	66.26	67.23
10	NaN	NaN	58.83	58.83	57.99	58.73	62.41	65.78	66.32	67.06
15	NaN	NaN	58.19	58.19	56.67	57.41	59.92	61.51	63.74	63.74
20	NaN	NaN	58.19	58.19	56.35	57.07	59.51	61.15	63.29	63.48
25	NaN	NaN	58.19	58.19	56.35	57.07	59.51	60.65	62.67	62.67
30	NaN	NaN	58.19	58.19	56.35	57.21	58.87	60.20	60.20	60.20
50	NaN	NaN	58.19	58.19	56.35	57.36	58.05	58.05	58.05	58.05

3.2.2. Artificial Neural Network Analysis

The artificial neural networks were created using *mlp* function of RStudio (Version 1.0.153). The continuous variables in the database were in various ranges; for instance total degree of unsaturation is in between -101 and 313 whereas N/O is in the range of 0-34. This kind of large gap between ranges may cause inappropriate results in ANN analysis. Therefore, normalization of both input and output variables is required so that all the variables are at a comparable range. To do that, Equation 3.2 was used; the mean and variance of data was transformed to zero and one, respectively. After the predictions were done by using ANN model, the output was converted to its real value using the same equation.

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{3.2}$$

After normalizing the data, the categorical variables were converted into continuous variable by using one-of-N encoding [81]. In this method, each category is represented as a separate variable. For example, the metal type is a categorical variable; hence metal types were converted into seven variables as Cu, Cd, Co, Fe, Mn, Zn, and Ni. Then the value of a metal used in a specific MOF was taken as one, while the other metals were taken as zero [9].

An ANN model is considered as successful when it as able to learn the patterns and then use this knowledge for generalization. If ANN learns too many patterns and loses its ability to generalize similar patterns, an issue called *overlearning* takes place [82]. In order to overcome the *overlearning* (memorizing the training samples by the model), validation of the model was performed on unseen data by applying *k*-fold cross validation analysis (with a "*k*" value of five in this work) [83–86]. Root mean square error (RMSE) was taken as the measure of the model fitness and calculated by using Equation 3.3 where y_i is the actual value, $\hat{y}_i$ is the predicted value and n is the number of data points. The entire database was randomly divided into five folds; and then data taken from four folds were used to construct the ANN model to predict the outcome of the remaining one-fold. This procedure was repeated for five times until all subsets were covered. Final RMSE testing error was calculated by taking average of testing errors obtained in all folds. RMSE of training was determined by using the entire database. Five-fold cross validation was carried out by using *rep_len* and *sample* functions in RStudio as well.

$$\text{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{3.3}$$

As it mentioned above, *mlp* function was used to create and run a multilayer perceptron. There are some elements in that function to be decided to optimize the model; these are namely *size* (number of units in the hidden layers), *hiddenActFunc* (the activation function of all hidden units), *learnFunc* (learning function), *linOut* (type of activation function).

Commonly used learning and activation functions have been chosen to find optimal functions [87–89] between all learning and activation functions [90]. Learning and activation functions given in Figure 3.4 were tried one by one by fixing one learning function while activation function changes. All combinations of functions were run ten times and RMSE of each of them were calculated; the average RMSE values were considered. The optimum functions were determined according to lowest RMSE values. This process was carried out with the help of four nested loops; outermost loop controls number of runs, the innermost loop controls the five-fold cross validation and the learning and activation functions were checked with two intermediate loops. The whole procedure

was carried out for all databases given in Figure 3.3 and a common result was obtained. It was seen from Table 3.8 - Table 3.15 that similar RMSE values were achieved except the value obtained by using the functions *"BackpropBatch"* and *"Act_StepFunc*". Consequently, conjugate gradient (gradient descent) algorithm namely *"SCG"* was used as the learning function to adapt the weights, whereas Sigmoid (or Logistic) function namely *"Act_Logistic"* was used as an activation function.

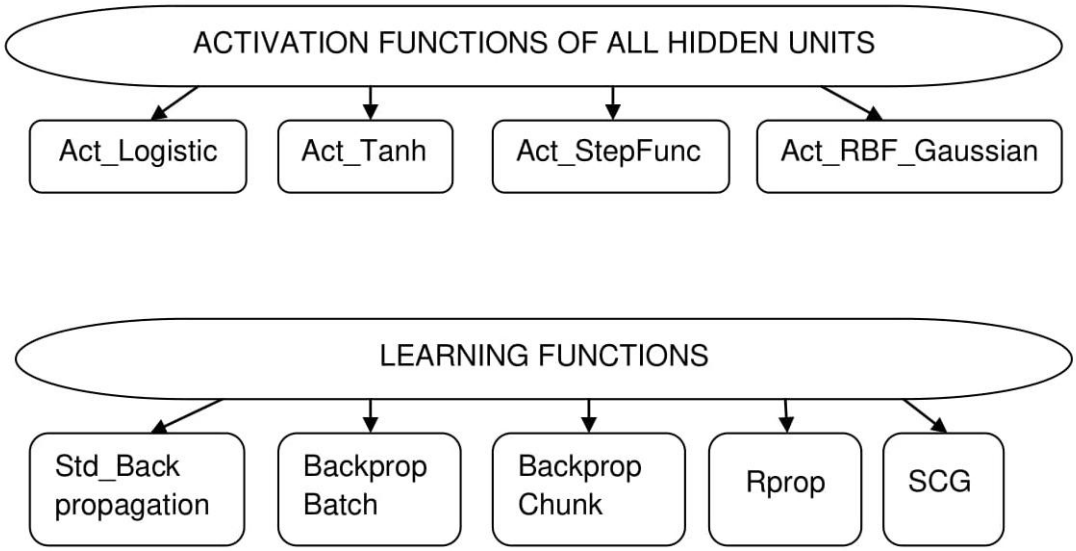

Figure 3.4. Activation and learning functions tested in the models.

Table 3.8. RMSE results of activation and learning function matrix for gravimetric CH_4 delivery using user defined descriptors.

RMSE		**ACTIVATION FUNCTION**			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	101.10	100.02	119.52	99.44
	BackpropBatch	104.93	104.49	105.76	104.89
	BackpropChunk	99.79	100.26	123.06	99.88
	Rprop	127.72	105.01	113.49	113.99
	SCG	100.68	101.74	109.60	102.73

Table 3.9. RMSE results of activation and learning function matrix for gravimetric CH_4 delivery using structural properties.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	30.13	29.84	103.36	30.00
	BackpropBatch	96.62	96.92	99.12	96.30
	BackpropChunk	30.40	30.32	105.62	30.32
	Rprop	30.06	30.17	108.85	30.12
	SCG	29.97	29.97	99.74	29.97

Table 3.10. RMSE results of activation and learning function matrix for gravimetric CH_4 uptake using user defined descriptors.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	85.64	85.60	98.16	85.66
	BackpropBatch	85.89	86.08	87.66	85.90
	BackpropChunk	85.88	85.50	100.85	85.88
	Rprop	84.26	83.59	98.43	83.92
	SCG	84.85	83.93	93.10	85.83

Table 3.11. RMSE results of activation and learning function matrix for gravimetric CH_4 delivery using structural properties.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	31.23	30.72	95.82	30.81
	BackpropBatch	84.72	84.58	87.11	84.60
	BackpropChunk	30.89	30.63	98.00	31.49
	Rprop	30.52	30.47	87.67	30.46
	SCG	29.97	29.97	99.74	29.97

Table 3.12. RMSE results of activation and learning function matrix for volumetric CH_4 delivery using user defined descriptors.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	50.32	50.48	59.24	50.06
	BackpropBatch	49.22	49.20	51.07	49.30
	BackpropChunk	49.96	50.39	60.06	50.45
	Rprop	50.02	50.42	56.57	50.84
	SCG	50.39	51.00	53.43	50.33

Table 3.13. RMSE results of activation and learning function matrix for volumetric CH_4 delivery using structural properties.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	35.43	34.89	55.41	35.78
	BackpropBatch	50.13	49.98	51.52	50.10
	BackpropChunk	35.56	35.61	57.64	35.54
	Rprop	34.83	34.76	53.22	34.65
	SCG	34.66	34.66	52.42	34.65

Table 3.14. RMSE results of activation and learning function matrix for volumetric CH_4 uptake using user defined descriptors.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	48.94	49.08	56.55	50.47
	BackpropBatch	47.87	47.82	51.07	47.89
	BackpropChunk	49.48	48.68	57.79	50.66
	Rprop	51.95	51.81	57.02	51.82
	SCG	47.78	48.03	53.00	48.52

Table 3.15. RMSE results of activation and learning function matrix for volumetric CH_4 delivery using structural properties.

RMSE		ACTIVATION FUNCTION			
		Act_Logistic	**Act_Tanh**	**Act_StepFunc**	**Act_RBF_ Gaussion**
LEARNING FUNCTION	**Std_Backpropagation**	38.56	39.04	60.48	39.08
	BackpropBatch	53.35	53.39	53.69	53.37
	BackpropChunk	38.60	39.31	59.60	38.92
	Rprop	38.21	38.32	55.28	38.37
	SCG	38.46	38.86	53.69	38.46

Optimum network topology (number of units in the hidden layers) was found by analysing performances of the several models according to their testing and training errors. Each network topology was run ten times and their average was considered. Seven network topologies having one hidden layer from one to seven neurons were tested. The training and testing errors for all topologies tested were given in Figure 3.5 to Figure 3.8 (in a-b-c representation, a, b, and c symbolize number of input variables, number of neuron in hidden layer, and number of output variable, respectively). The training error decreased when the network size increased (because of the use of more weights to represent the data) while the testing error decreased first and then increased again due to an increase in overlearning [83,84,91,92]. The topologies with the lowest testing errors, which indicate the ability to predict the unseen data, were selected for ANN models; based on these topologies, structures of ANN models for all databases were determined as in Table 3.16. The illustrations of the final ANN topologies belonging to gravimetric CH_4 delivery database were given in Figure 3.9.

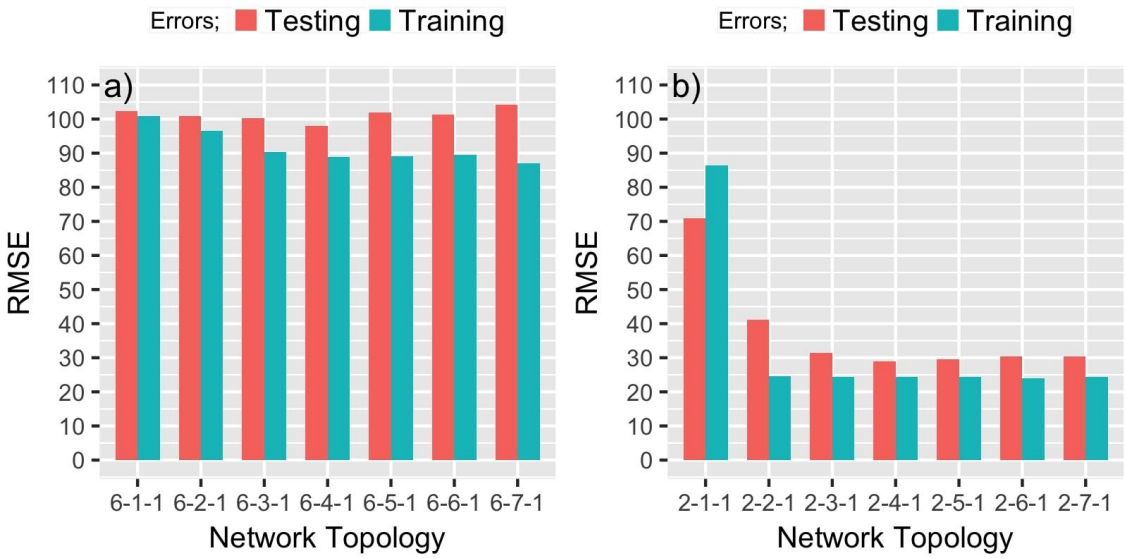

Figure 3.5. Network topologies for gravimetric CH_4 delivery a) using user defined descriptors b) using structural properties.

Figure 3.6. Network topologies for gravimetric CH_4 uptake a) using user defined descriptors b) using structural properties.

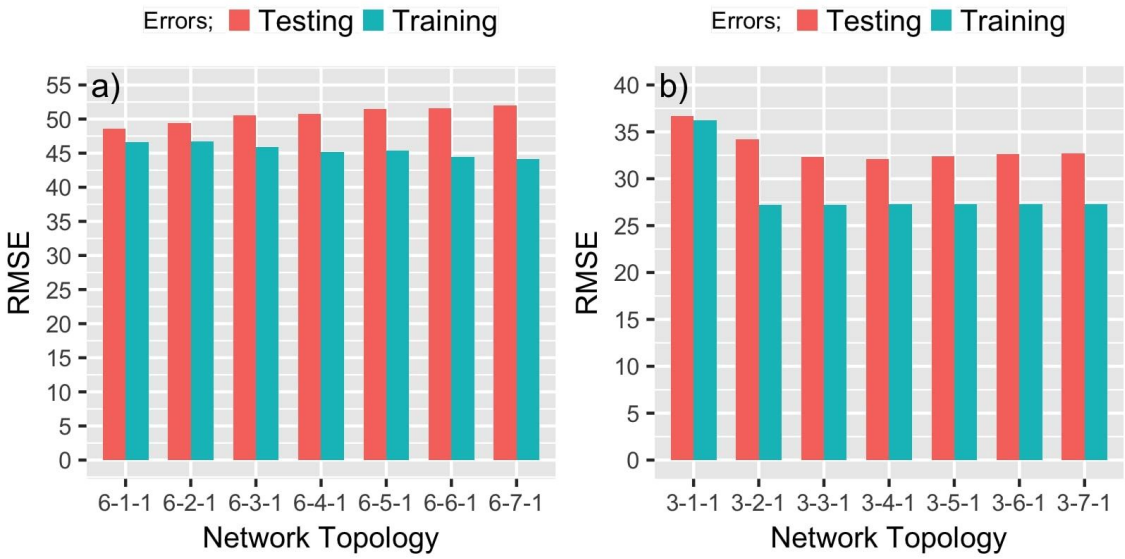

Figure 3.7. Network topologies for volumetric CH_4 delivery a) using user defined descriptors b) using structural properties.

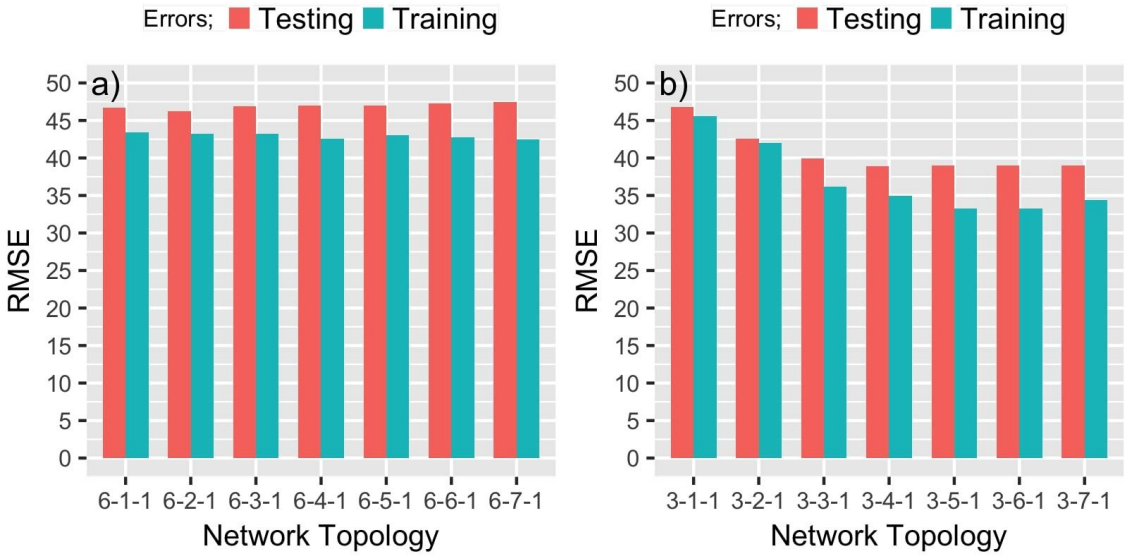

Figure 3.8. Network topologies for volumetric CH_4 uptake a) using user defined descriptors b) using structural properties.

Table 3.16. Determined ANN Structures.

Gravimetric Databases	**ANN Structures**	**Volumetric Databases**	**ANN Structures**
CH_4 delivery using user defined descriptors	6-4-1	CH_4 delivery using user defined descriptors	6-1-1
CH_4 delivery using structural properties	2-4-1	CH_4 delivery using structural properties	3-4-1
CH_4 uptake using user defined descriptors	6-2-1	CH_4 uptake using user defined descriptors	6-2-1
CH_4 uptake using structural properties	2-4-1	CH_4 uptake using structural properties	3-4-1

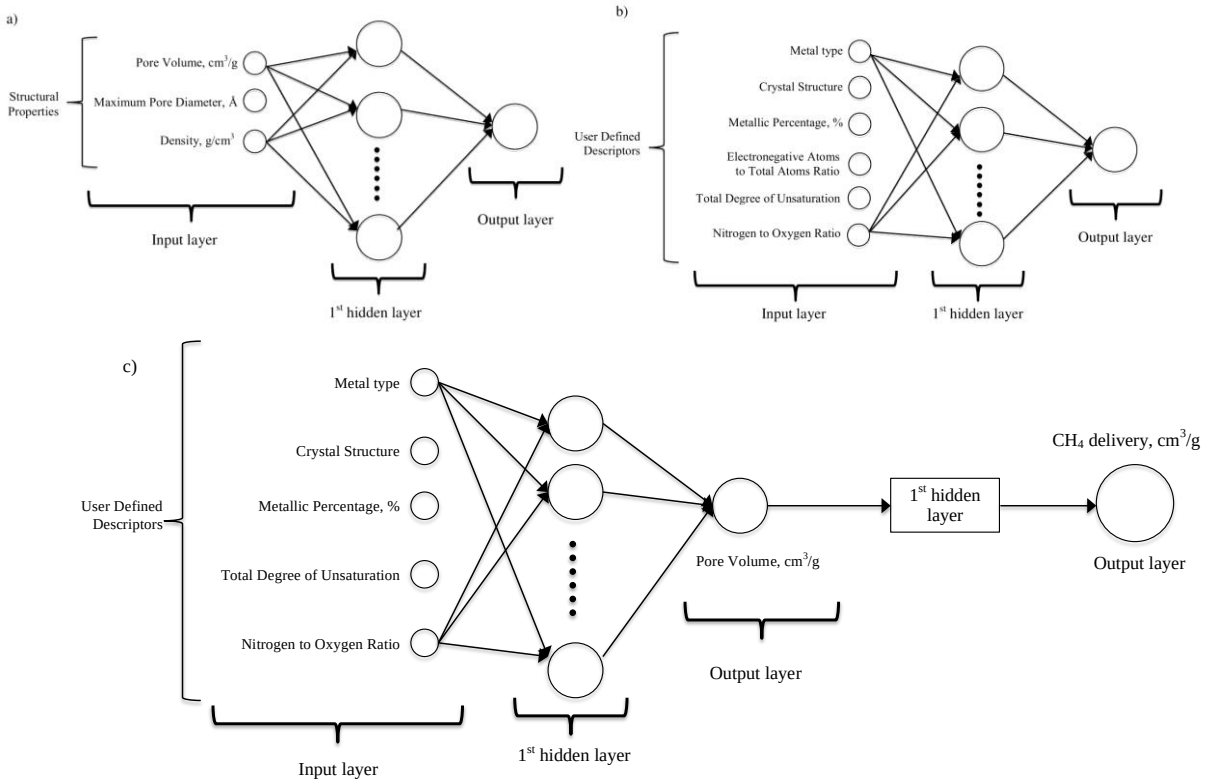

Figure 3.9. The ANN models for gravimetric CH_4 delivery a) using user defined descriptors b) using structural properties c) two-step model for CH_4 delivery (user defined descriptors ➔ pore volume ➔ CH_4 delivery).

4. RESULTS AND DISCUSSION

First, a general review and simple descriptive analyses of CH_4 delivery and uptake databases both on volumetric and gravimetric basis were performed to exhibit the basic trends and relations; the meaningful relationships were demonstrated. Then, the detailed analyses were carried out by constructing decision trees to deduce some general conclusions and heuristics that may be used for future studies. It was seen from the studies before splitting the input variables into two groups as user defined descriptors and structural properties that the structural properties may suppress the effects of user defined descriptors. Hence, decision tree analysis for the user defined descriptors and structural properties, which are easier to modify for a new MOF with a higher possibility of good performance, were performed separately. In ANN analyses for gravimetric CH_4 delivery, it was also tested whether the user defined descriptors can be used to predict CH_4 delivery through structural properties in two steps.

4.1. Review and Pre-analysis of Databases

Databases containing 1389 data points (remaining after removing the cases with correlations and missing data) showing the CH_4 delivery and CH_4 uptake both on gravimetric and volumetric basis against input variables listed in Table 3.5 were pre-analysed using *ggplot* function of RStudio (Version 1.0.153) and the main findings are presented below.

4.1.1. For Gravimetric Databases

Gravimetric CH_4 databases were analysed to illustrate the basic relations. Figure 4.1 shows the relation between the pore volume and MPD as well as their effects on CH_4 delivery and uptake. For gravimetric analyses shown in Figure 4.1a and Figure 4.1b, increase in both variables increases the adsorbed CH_4 amounts. The positive effects of high pore volume is understandable and it is one of the main reasons that MOFs have been studied as adsorbents [7,93–95]. However, there are also works that shows the positive [85,86] as well as the negative [96–98] effects of high MPD.

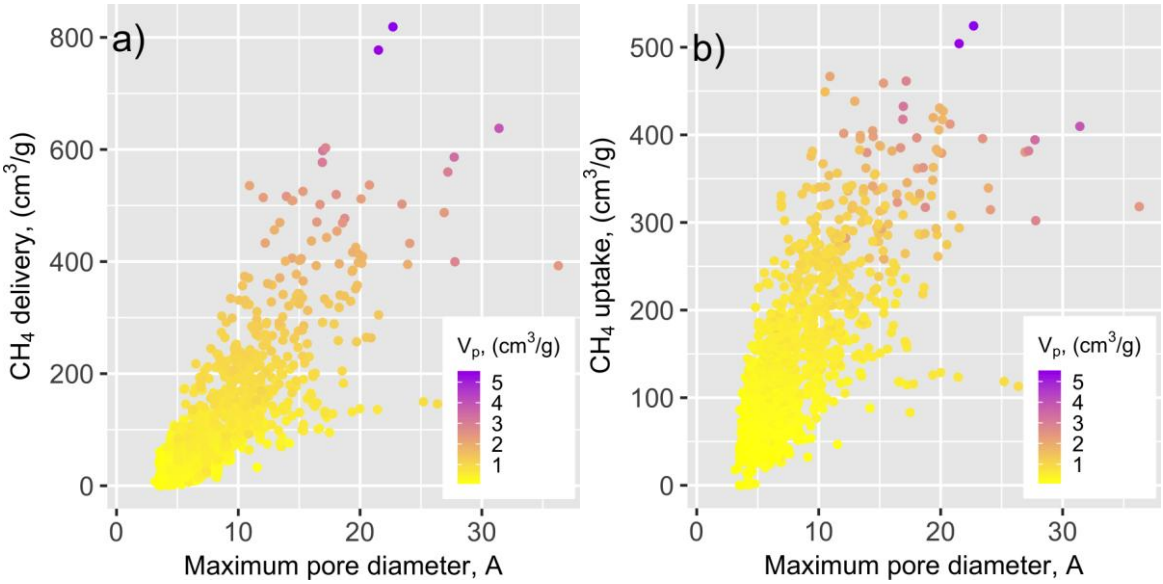

Figure 4.1. MPD vs. a) CH_4 delivery and b) CH_4 uptake with the change of pore volume.

Figure 4.2 illustrates the effect of pore volume on MOFs containing different metal types. In general, the pore volume should be as high as possible to obtain high CH_4 delivery or uptake. When the metal types are considered, MOFs containing Cu has the highest pore volume, which may be the main reason that the highest methane delivery or uptake values also belong to Cu containing MOFs. When Figure 4.2a and Figure 4.2b are compared; it is seen that they show similar trend in all metal atoms. However, the impact of metal types in MOFs on the gas uptake is still a challenge to be clarified. For example, Wu *et al.* investigated the effect of metal type on a same MOF (M_2 (dhtp)), and they claimed that MOF having Ni as the metal atom exhibits the highest CH_4 storage at 35 bar compared to the MOFs having Mg, Mn, Co, and Zn [99]. In another work, Koh *et al.* examined 17 metal atoms (M= Be, Ca, Co, Cr, Cu, Fe, Mn, Mg, Mo, Ni, Pb, Sc, Sn, Sr, Ti, V, W, and Zn) over the same framework (M-HKUST-1) and compared them with Cu-HKUST in terms of CH_4 delivery capacity. The MOFs containing Ni and Ca exceed the performance of Cu-HKUST [100]. Apparently, the effects of metal type also depend on the other components of the MOF.

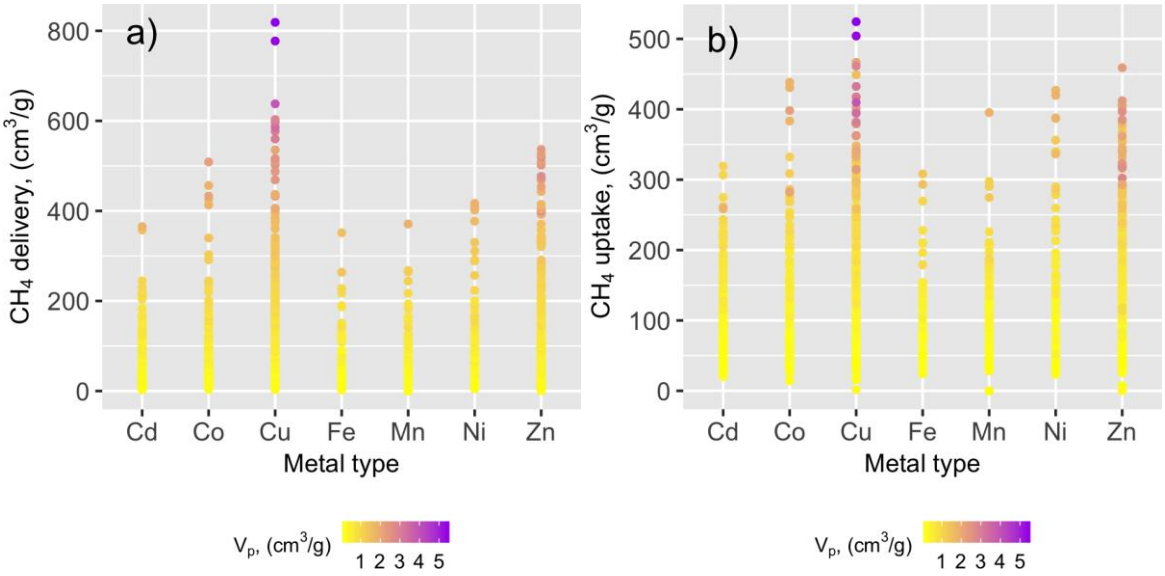

Figure 4.2. Metal type vs. a) CH_4 delivery and b) CH_4 uptake with the change of pore volume.

The effects of metallic percentage and nitrogen to oxygen ratio (N/O), on to CH_4 delivery are presented in Figure 4.3a and 4.3b, respectively. Although no clear trend was observed, Figure 4.3a.1 and Figure 4.3b.1 indicate that the higher amount of CH_4 delivery or uptake around the metallic percentage range of (0-2) % (based on the number of atoms). Figure 4.3a.2 and Figure 4.3b.2 show that MOFs with low N/O ratios in this database are superior in terms of CH_4 delivery or uptake; this could be partially because the database contains more MOFs having low N/O ratio. As far as we know, there is no study involving impacts of metallic percentage and N/O ratio on CH_4 uptake or delivery over MOFs except for the one reported by Pardakhti *et al.* [2]; they have investigated whether the prediction of CH_4 uptake is possible or not by using some variables containing metallic percentage and N/O ratio.

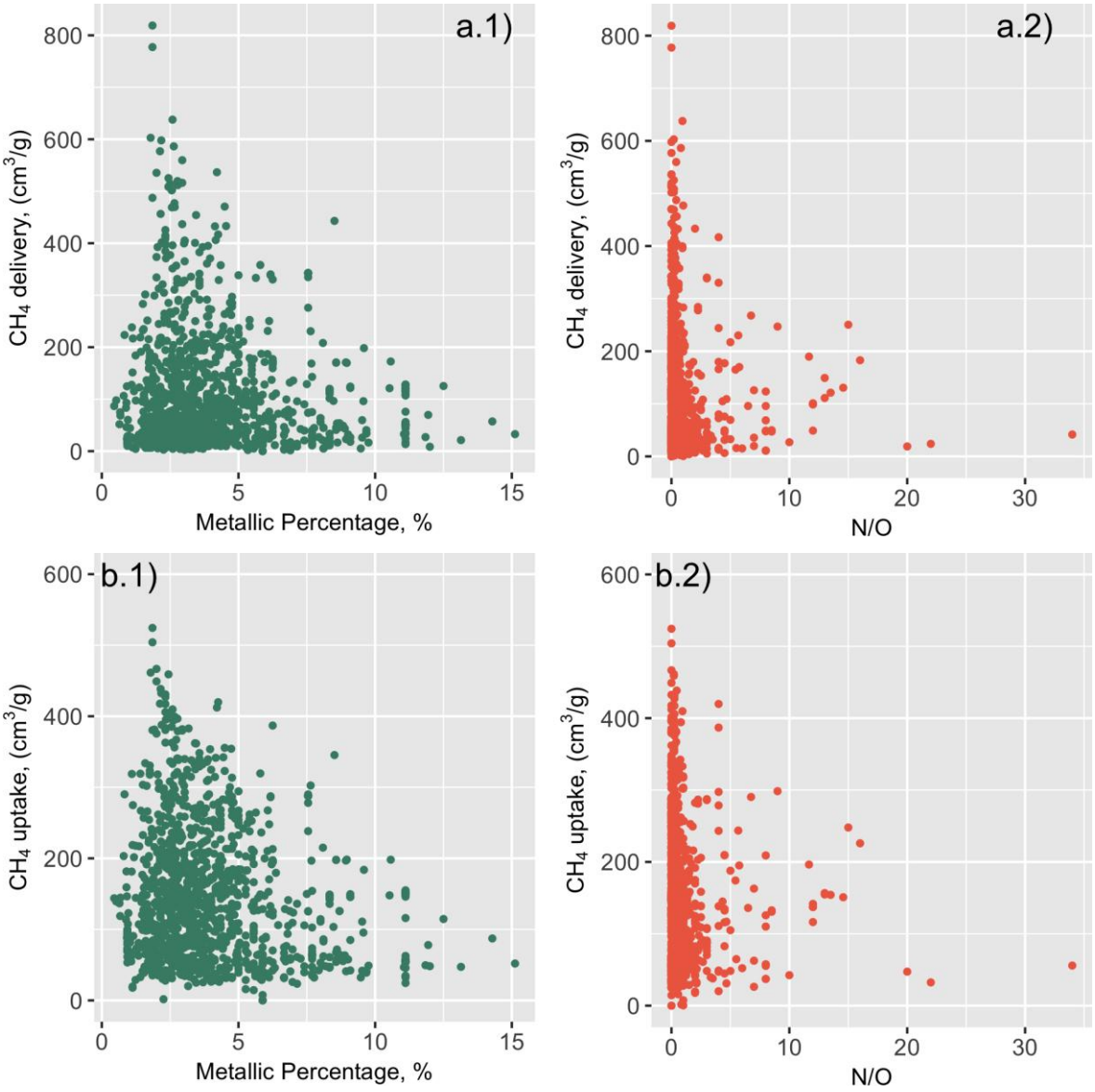

Figure 4.3. CH_4 delivery vs. a.1) metallic percentage and a.2) N/O ratio; CH_4 uptake vs. b.1) metallic percentage and b.2) N/O ratio.

Figure 4.4 and Figure 4.5 exhibit the relationship between the crystal structures of MOFs and the amount of CH_4 delivery and uptake for the entire database and seven metal types. The MOFs having tetragonal and cubic structure have higher CH_4 delivery and uptake, whereas the MOFs containing triclinic structure have lower delivery values in the entire databases. Similar trends were also observed for Cu and Zn based MOFs; this may also be true for the other metals as well, but number of data available for those cases is not sufficient. As far as we know, no works which we can use to verify these results were

reported presenting a relationship between the crystal structure and the CH_4 uptake or delivery over MOFs in the literature.

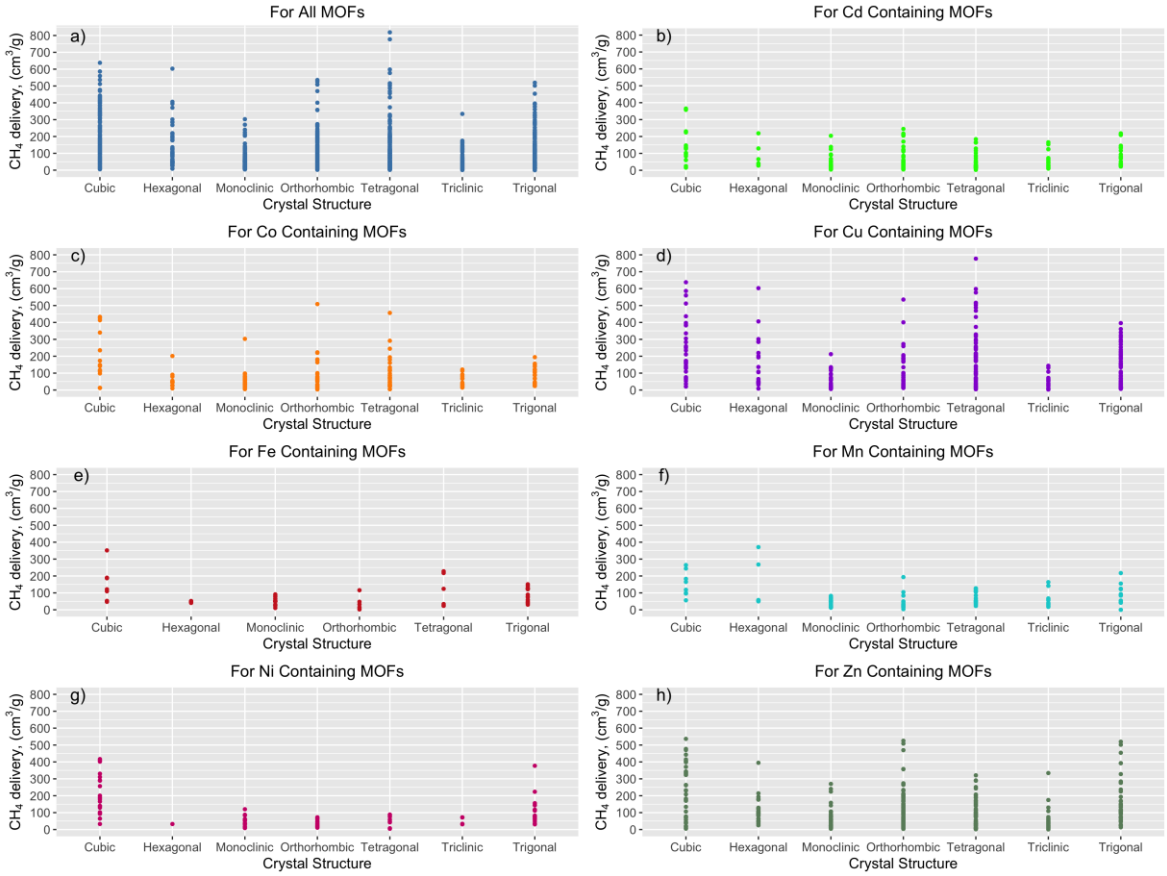

Figure 4.4. Crystal structure vs. CH_4 delivery for the entire database on the basis of each metal type.

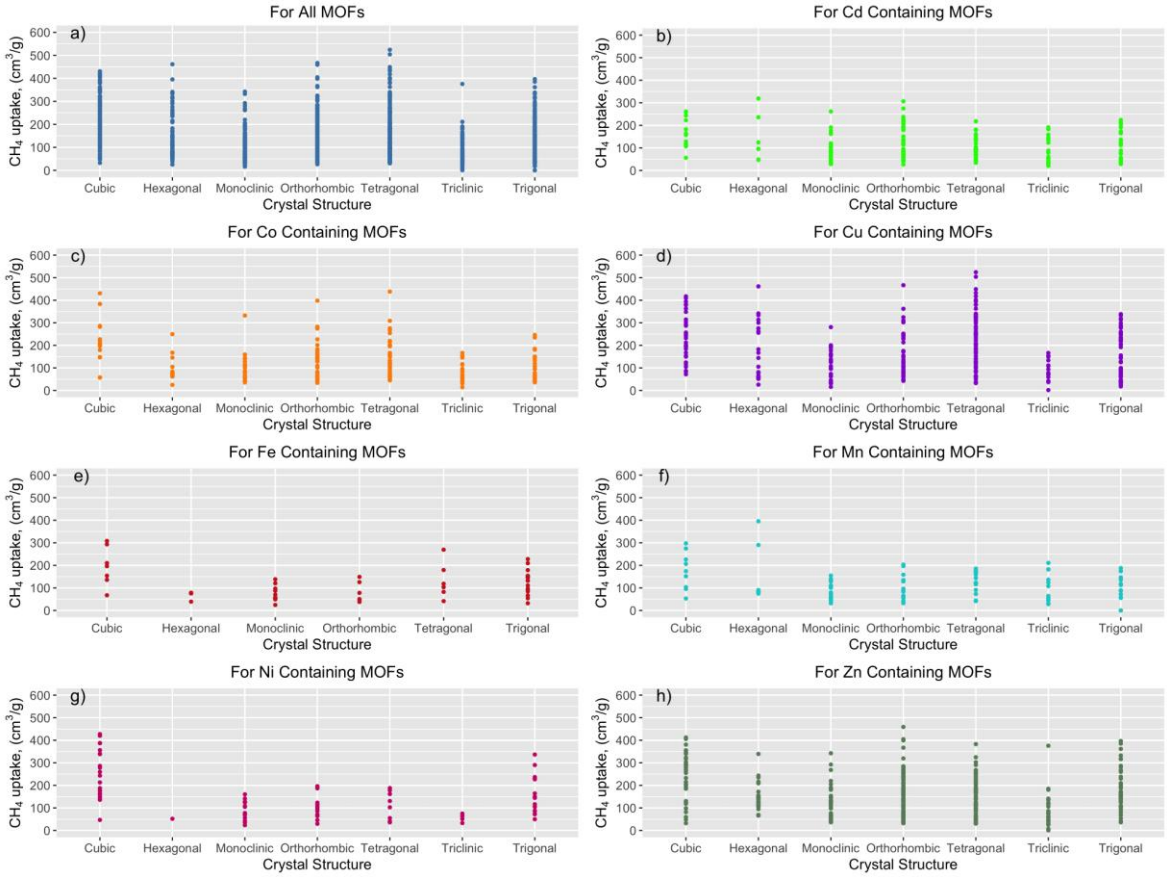

Figure 4.5. Crystal structure vs. CH_4 uptake for the entire database on the basis of each metal type.

The ratio of electronegative atoms to total number of atoms (electronegativity ratio) was also used as an input variable to examine its effect on CH_4 delivery and uptake roughly before the modelling stage (Figure 4.6 and Figure 4.7). The top graph (a) contains the entire database and is clustered with respect to metal atoms; this graph suggests that the MOFs containing Cu (shown with green dots) and Zn (shown with pink dots) may have a relation between the electronegativity ratio and the CH_4 delivery and uptake; the data points were located around the electronegativity ratio range of 0.1-0.2. Hence, MOFs containing Cu and Zn were re-plotted separately (b and c) to investigate this relationship further. It was observed that as the number of electronegative atoms increase in MOFs containing Cu or Zn as metal atoms, the CH_4 delivery decreases even though there are also MOFs with low electronegativity ratio but still has low CH_4. Although Pardakhti *et al.* used electronegativity ratio as one of the variable to predict the CH_4 uptake, they provided no information on its effect on the uptake or delivery capacity [2].

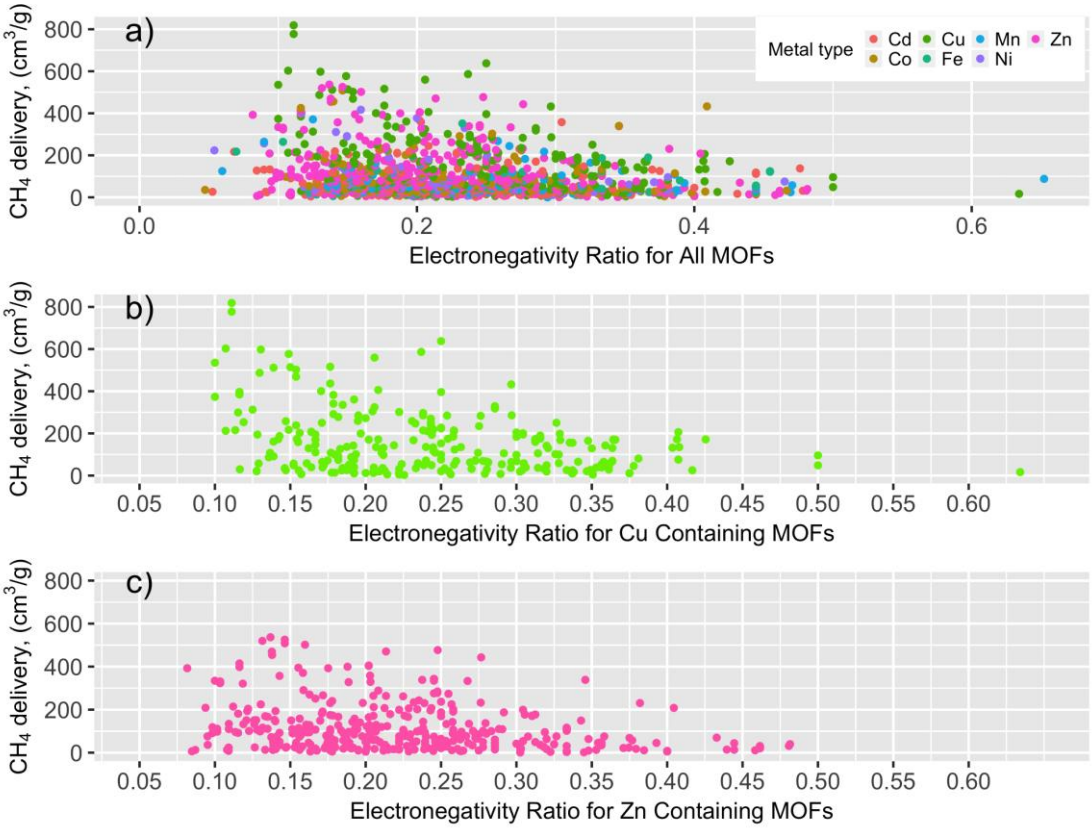

Figure 4.6. Electronegativity ratio vs. CH_4 delivery involving a) entire database b) Cu-based MOFs c) Zn-based MOFs.

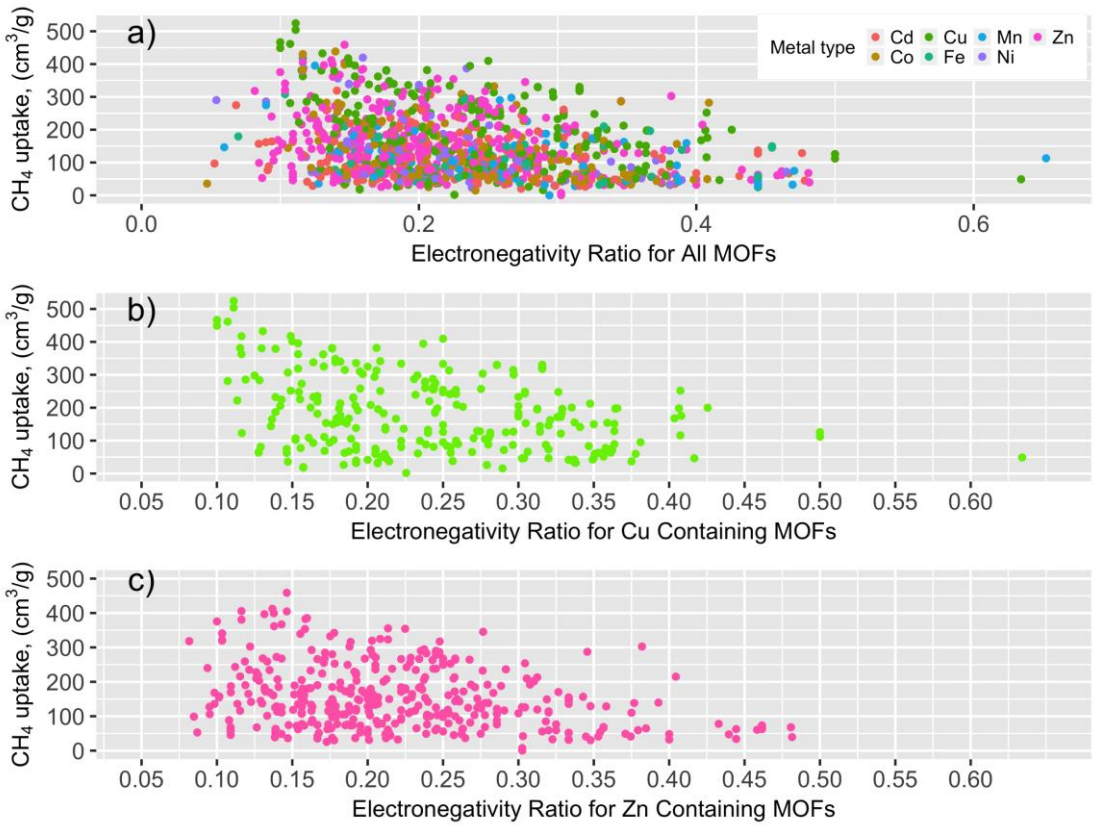

Figure 4.7. Electronegativity ratio vs. CH_4 uptake involving a) entire database b) Cu-based MOFs c) Zn-based MOFs.

4.1.2. For Volumetric Databases

In order to compare the effects input variables on volumetric CH_4 databases with gravimetric, the same analysis as in Section 4.1.1 was performed and the important relationships obtained are shown below. When Figure 4.8 and Figure 4.1 are compared, it is seen that the effects of MPD and pore volume on CH_4 delivery / uptake are similar but not identical; linearity is not very obvious in Figure 4.8. The data points in the MPD range of 10-20 Å, and the pore volume range of 3-4 cm^3/g have high CH_4 delivery and uptake.

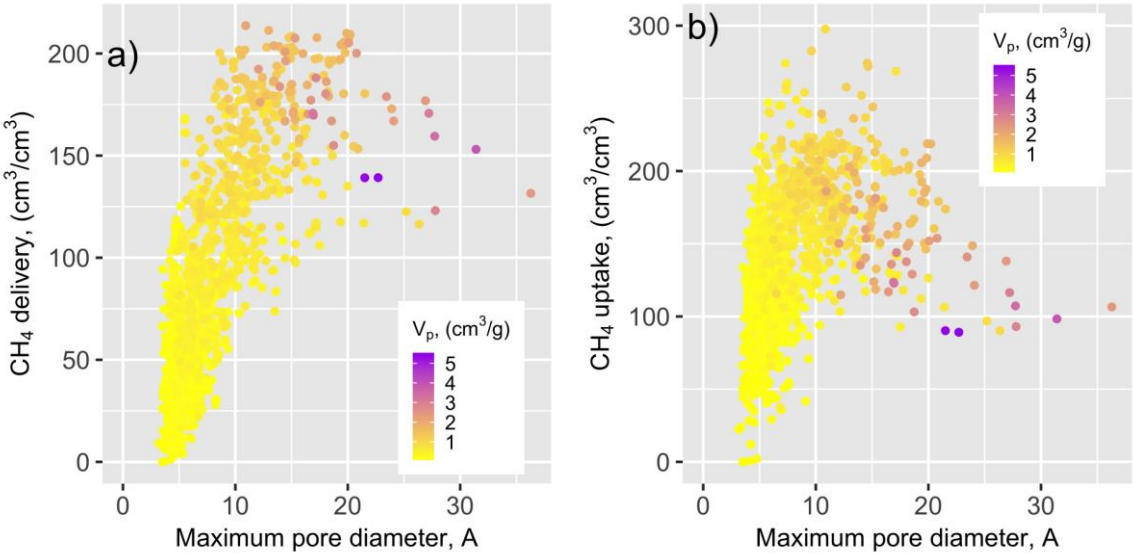

Figure 4.8. MPD vs. a) CH_4 delivery and b) CH_4 uptake with the change of pore volume.

As different from the gravimetric databases, the effect of density and pore volume can be examined on the volumetric amount of adsorbed CH_4 as in Figure 4.9. As the pore volume increases and the density decreases to a certain value (less than 1 g/cm^3), the amount of adsorbed CH_4 increases; with the further decreases in density after a certain value, the amount of CH_4 begins to fall. Consequently, as can be also seen from Figure 4.9, there is a limit on the density value for both CH_4 delivery and high uptake. In the work of Purewall *et al.*, the inverse proportion of pore volume and density was also detected [101]. When they compare MOFs having different densities (all of them are less than 1 g/cm^3) in terms of the volumetric amount of adsorbed H_2, they have found that when the density increases and get close to 1 g/cm^3, the amount of H_2 increases. Since the density values in Figure 4.9 were rounding to integers, decision tree analysis explained in further sections was useful to compare the results. Yuan *et al.* investigates the effect of density densities (all of them are less than 1 g/cm^3) on both volumetric and gravimetric amount of adsorbed CH_4, CO_2, N_2 and H_2 and they have found similar results with Purewall *et al.*; when the density increases and get close to 1 g/cm^3, the amount of CH_4 increases [101,102].

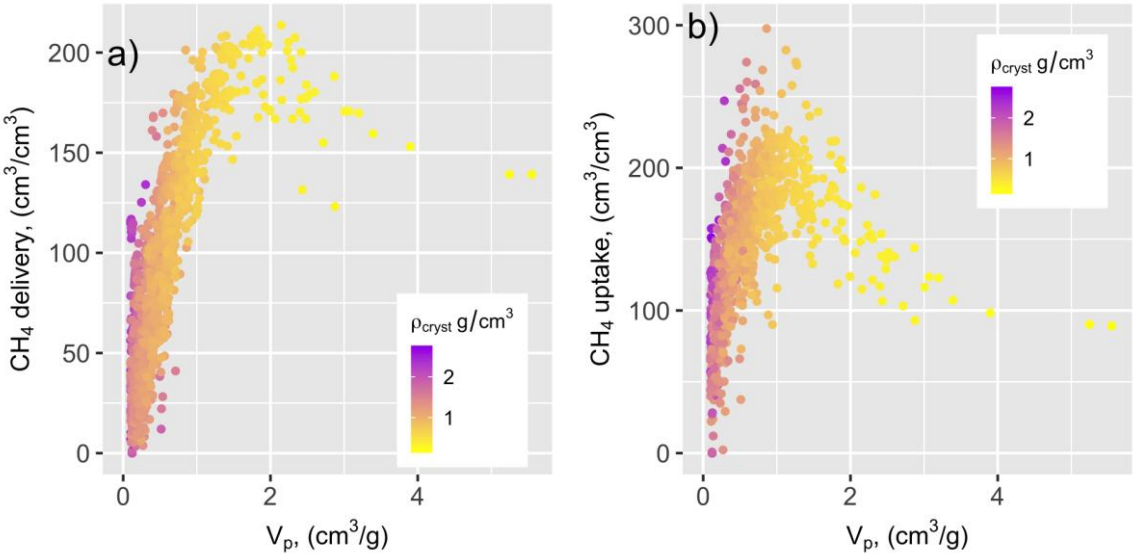

Figure 4.9. Pore volume vs. a) CH_4 delivery and b) CH_4 uptake with the change of density.

4.2. Data Mining Analyses

After reviewing the general relations of input variables between themselves and output variables, the decision tree and artificial neural network modelling were performed as explained in Figure 3.2 and Figure 3.3. First, the heuristics for high methane uptake were developed via decision trees analysis, and then the predictability of the output variable was tested by using artificial neural network; this procedure was followed for all databases (ie. gravimetric delivery, gravimetric uptake, volumetric delivery and volumetric uptake).

The decision tree analyses in classification mode were performed using user defined descriptors and structural properties given in Table 3.5. For the model developed using user defined descriptors (with 1389 instances), the delivery data divided into three classes as *low*, *medium*, and *high* (approximately 0-33rd, 34rd -66th and 67th-100th percentile, respectively); the number of instances were kept equal to avoid class imbalance problem [103]. The same delivery limits were also used in the model involving structural properties for comparison even though the total number of instances and their distribution among the classes were changed to 2224; however, the numbers of instances in all three classes were still close enough to avoid the class imbalance problem.

4.2.1. For Gravimetric Databases

In this section, gravimetric CH_4 delivery (from 65 bar to 5 bar) and CH_4 uptake (at 35 bar) were analysed using the decision tree and artificial neural network.

4.2.1.1. CH_4 Delivery. 453 instances having CH_4 delivery lower than 36 cm^3/g was classified as *low*, 454 instances having CH_4 delivery higher than 36 cm^3/g, and lower than 100 was classified as *medium*, and 482 instances with CH_4 delivery higher than 100 cm^3/g was classified as *high*. For structures properties (only pore volume and maximum pore diameter were remained after correlation analysis), the class limits were chosen the same as those used for user defined descriptors; with these limits, 818, 767, and 639 instances for CH_4 delivery were classified as *low*, *medium*, and *high*, respectively. Although the numbers of data points in each class were not as much balanced as in the tree for user defined descriptors given in Table 4.1, they were sufficiently close to avoid the imbalance problem.

Table 4.1. The range of the gravimetric amount of CH_4 delivery and corresponding number of instances.

Range	Classes	Number of instances	
		User defined descriptors	Structural properties
< 36 cm^3/g	Low performing	454	818
36 - 100 cm^3/g	Medium performing	453	767
> 100 cm^3/g	High performing	482	639

Decision Tree. After reprocessing database, the input variables given in Table 3.5 were analyzed separately as user defined descriptors and structural properties basis. First, all user defined descriptors were used and the tree shown in Figure A.1 was obtained. The atomic radius and the group number were not appeared as the split criteria in the tree as they were the features of the dominant metal atom rather than MOF; additionally, there are no such variables found in MOF modeling literature based on modelling as atomic radius and group number as well. Hence, the atomic radius and group number are excluded from the dataset (from the other set as well), another tree is constructed and its simplified

version presented in Figure 4.10 by excluding those two variables (full version of tree is presented in Appendix A, Figure A.2).

The accuracy rates were 66.3% and 58.1% for training and testing respectively; these values mean that the decision tree correctly placed 66.3 percent of training data points, which is already used in model building, and 58.1% of the data points that are not seen before. The training and testing accuracy rates of individual classes are shown in confusion matrices illustrated in Table 4.2. To calculate the testing accuracy, five-fold cross validation method was used as explained in Section 3.2.1. When the accuracy rates of each class are examined separately, it can be clearly seen that the accuracy of high and low delivery classes are much better; the low accuracy of medium class, which is usually the case because of the leaks from both sides, reduces the average classification accuracy entire tree. For example, the testing and training accuracy rates for high CH_4 delivery were 73% and 66.6% respectively. The accuracy of some of the individual branches, which are more important to develop rules and heuristics, are much higher.

The tree was first divided according to the MOFs' crystal structure, which is a categorical attribute with seven alternatives. If the crystal structure of the MOF was cubic, it resulted in higher CH_4 delivery depending on other limitations. The second decision point was the total degree of unsaturation of the MOF. The N/O ratio and metallic percentage were found to have small effects to determine if a MOF had high CH_4 delivery. The limits of branch divisions (for example metallic percent of 1.7) are usually the average of the highest value at one side and the lowest value at the other side; hence, there may not be such an exact data point. If that is the case, the first smaller and higher values should be considered as the limit of lower and upper classes, respectively.

Three branches (having more than 80% purity) of the tree were decided as heuristics for high CH_4 delivery and they were tabulated in Table 4.3 for the future studies. According to branch shown as H1 in Table 4.3 if the crystal structure of the MOF was cubic and the total degree of unsaturation was greater than or equal to 17, the CH_4 delivery was high only if the metallic percentage of the MOF was higher than or equal to 1.7. There were 125 MOFs satisfying these conditions and 110 of them were classified as having high

CH_4 delivery; (110/125)×100 = 88% is a strong statistical result that can be considered as heuristics for the future studies.

Table 4.2. Confusion matrices of decision tree constructed using user defined descriptors for gravimetric CH_4 delivery.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	454	339*	87	28	74.8
	Medium	453	127	230*	96	50.7
	High	482	67	63	352*	73.0
	Total	1389				66.3
Testing	Low	454	294*	115	45	64.9
	Medium	453	163	192*	98	42.3
	High	482	80	81	321*	66.6
	Total	1389				58.1

Three branches (having more than 80% purity) of the tree were decided as heuristics for high CH_4 delivery and they were tabulated in Table 4.3 for the future studies. According to branch shown as H1 in Table 4.3 if the crystal structure of the MOF was cubic and the total degree of unsaturation was greater than or equal to 17, the CH_4 delivery was high only if the metallic percentage of the MOF was higher than or equal to 1.7. There were 125 MOFs satisfying these conditions and 110 of them were classified as having high CH_4 delivery; (110/125)×100 = 88% is a strong statistical result that can be considered as heuristics for the future studies.

If the crystal structure was not cubic, there was still a chance of having a high CH_4 delivery capacity as apparent in Heuristics 2 and 3. According to the second heuristic route, the second decision point was N/O ratio. If this ratio was lower than 0.24, the algorithm questioned the crystal structure of the MOF again. If the crystal structure was orthorhombic, tetragonal, triclinic or trigonal and the total degree of unsaturation was greater than or equal to 37, these conditions yield high amount of CH_4 delivery; 68 out of 83 MOFs were classified correctly. In the third route, the metal type of the MOF became

crucial. MOFs containing iron or copper as metal atoms were identified as having high CH_4 delivery (47 out of 55 MOFs classified correctly); it should be remembered that the importance of copper atom was also apparent in the review and pre-analysis part of the work. Although the number of data points in these branches was lower than the branch for H1, they are still significant and their accuracies of 82% for H2 and 85% for H3 are quite high.

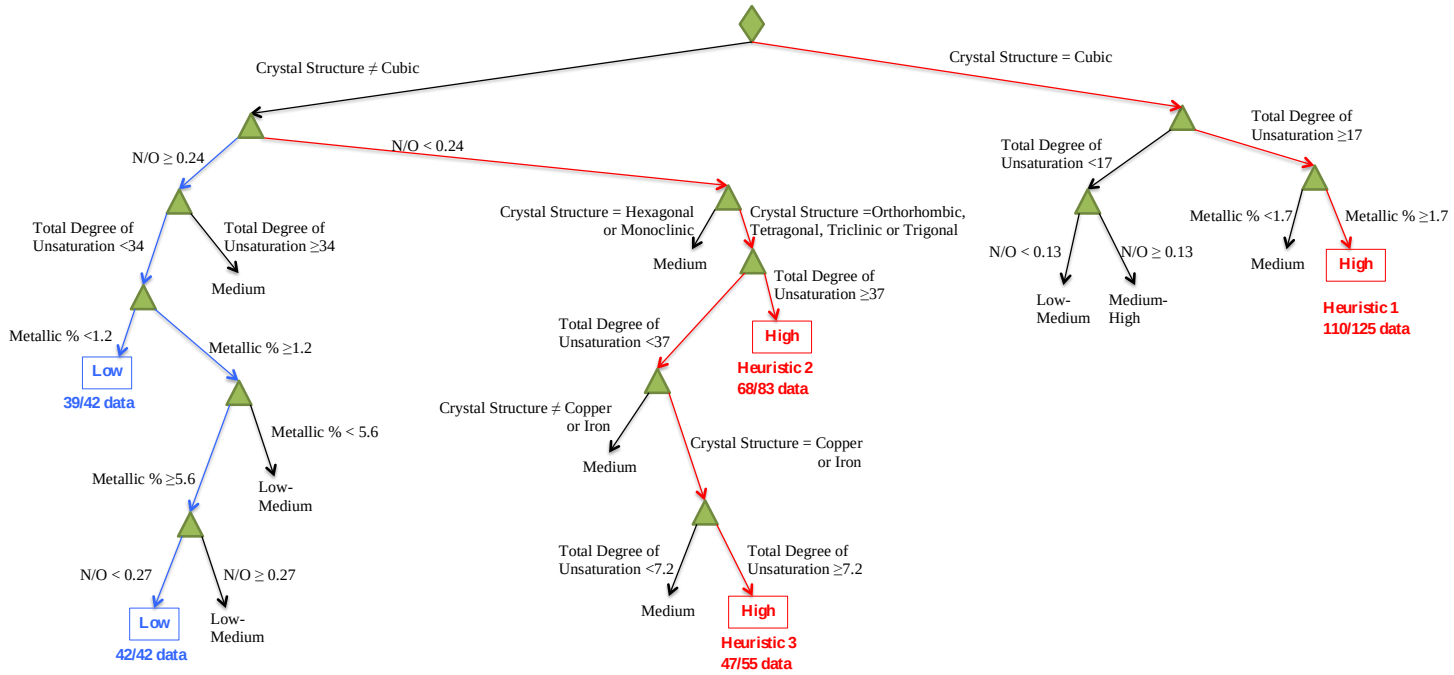

Figure 4.10. Simplified form of gravimetric CH_4 delivery classification tree developed by using user defined descriptors.

Table 4.3. Heuristics for high gravimetric CH_4 delivery for decision tree developed using user defined descriptors.

Variables and their limitations to have a high CH_4 delivery				**Accuracy of classification**	**Heuristics**
If crystal structure = cubic	Total degree of unsaturation > 17	Metallic percentage > 1.7		110/125	H1
If crystal structure = orthorhombic, tetragonal, triclinic or trigonal	N/O < 0.24	Total degree of unsaturation > 37		68/83	H2
If crystal structure = tetragonal or trigonal	N/O < 0.24	37 > Total degree of unsaturation > 7.2	Metal type = Cu or Fe	47/55	H3

The decision tree classification was also performed using structures properties (only pore volume and maximum pore diameter were remained after correlation analysis); the class limits and the number of corresponding instances were given in Table 4.1. The final decision tree was shown in Figure 4.11.

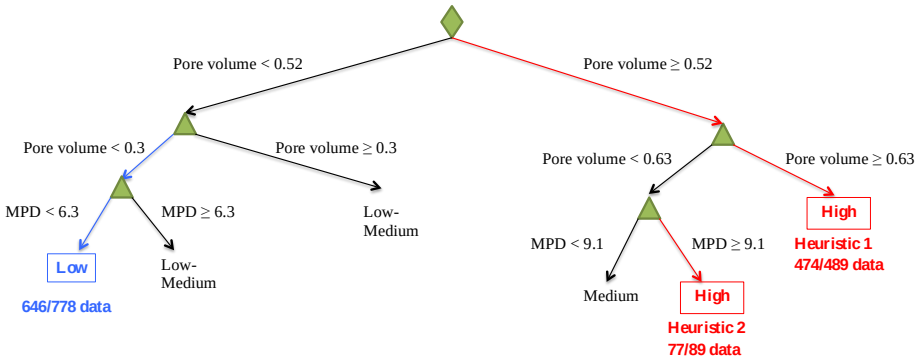

Figure 4.11. Simplified form of gravimetric CH_4 delivery classification tree developed by using structural properties.

The accuracy rates were found as 84.0% and 82.4% for training and testing, respectively. To calculate the testing accuracy, five-fold cross validation method was used. The training and testing accuracy rates of classification for individual classes were shown in confusion matrices illustrated in Table 4.4. The results for high CH_4 delivery were quite successful (more than 95% correctly classified MOFs); it was much better than the results of the decision tree constructed from user defined descriptors.

The decision tree had one major (Heuristic 1) and one minor (Heuristic 2) routes to obtain high CH_4 delivery; for both, the tree first divided according to the pore volume (greater than or equal to 0.52 cm^3/g). In Heuristics 1, the tree was divided again based on pore volume (greater than or equal to 0.63 cm^3/g) and reached to the final terminal node having 489 data points, of which 474 were correctly classified as high (the accuracy rate of 97 %). For Heuristic 2, on the other hand, MPD functioned as a final decision point (lower than 9.1 Å) for MOFs having pore volume between the range of 0.52 – 0.63 cm^3/g; the 77 of the 89 data points in the terminal node was correctly predicted (with accuracy rate of 87%).

Table 4.4. Confusion matrices of decision tree constructed using structural properties for gravimetric CH_4 delivery.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	818	712*	102	4	87.0
	Medium	767	139	594*	34	77.4
	High	639	0	77	562*	87.9
	Total	2224				84.0
Testing	Low	818	702*	10	6	85.8
	Medium	767	146	580*	41	75.6
	High	639	0	88	551*	86.2
	Total	2224				82.4

Considering that the major route for high CH_4 delivery were decided based on only the pore volume, a decision tree model involving only this structural property was also constructed and presented in Figure A.4. As expected, it generated the same branch having 489 data points with the same level of accuracy (474 correctly classified points). This indicates that only the pore volume can be considered as the structural properties in decision making since not only the 97% prediction accuracy (474 out of 489) of this branch is quite high, but also it covers 74% of all high-class cases (474 out of 639).

Artificial Neural Network. The artificial neural network analyses to predict the CH_4 delivery were performed using user defined descriptors (with 1389 instances) and structural properties (with 2224 instances), separately. The ANN analysis using pore volume only were also tried and given in the same section. Because the successful CH_4 delivery capacity predictions were obtained from the that model, a two-step approach was also tested to investigate whether the pore volume only can be expressed as a function of user defined descriptors and then be used to predict the methane delivery. The pore volume was first predicted from a separate ANN model using user defined descriptors, and then those predictions were used to construct a new model for CH_4 delivery prediction. The topologies of ANN models were given in Table 3.16; 5-fold cross validation method was applied as explained in Section 3.2.2. RMSE of both testing and training were used as the measure of fitness of model.

The results of ANN models to predict CH_4 delivery directly from the structural properties were given in Figure 4.12. The predictions obtained from ANN model were shown in *y*-axis, whereas the actual data was placed in *x*-axis. The RMSE of training was found as 85.0 cm^3/g, while that of testing was 100.9 cm^3/g (approximately corresponds to ± 101 cm^3/g error in CH_4 predictions), which is rather high. The model seems to over predict the low CH_4 delivery values, whereas it underpredicts the high delivery capacities; hence it cannot be considered as successful. The reason for the overestimation of the low deliveries may be originated from the fact that MOFs were assumed to have perfect crystal structures in GCMC, although they may have pinholes or disorders; consequently, the delivery may be overestimated in simulations. In the case of underestimation, there may be some other factors (like the presence of amine or extra framework action), which are not considered (only LJ interactions are used) in simulations even though they may affect the delivery capacity through chemisorption. Hence, it is not possible to simply correlate MOFs' CH_4 uptakes with a few properties; other factors, such as the presence of specific functional groups may also strongly affect the gas uptakes of MOFs.

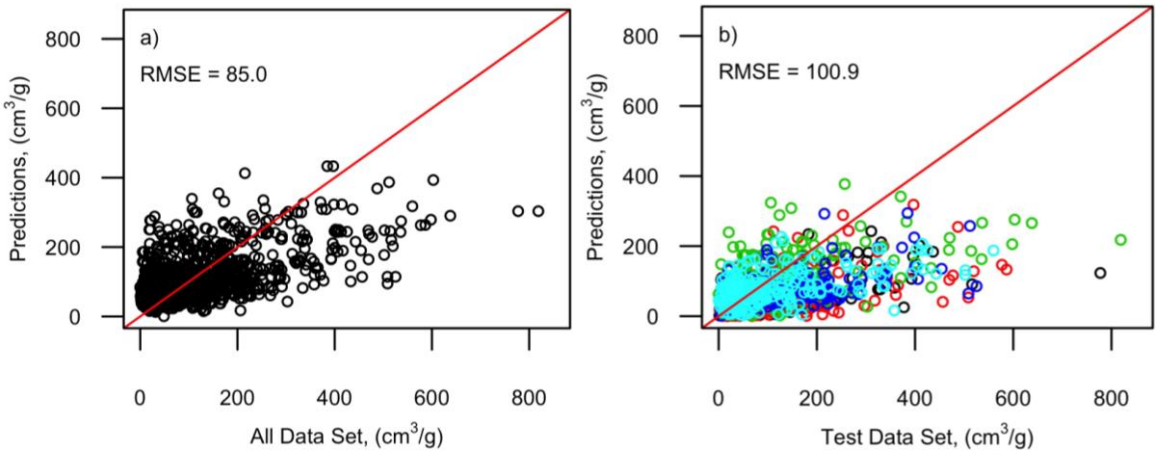

Figure 4.12. Actual versus predicted gravimetric CH_4 delivery values from ANN model constructed with user defined descriptors a) for training dataset and b) for testing dataset.

Then, the second ANN model to predict CH_4 delivery was constituted using structural properties The results obtained for testing and training are given in Figure 4.13; this time, the model is quite successful with the approximate RMSE values of 22.2 cm^3/g and 26.8 cm^3/g, respectively. The success of the model is also evident from the Figure 4.13a; the computed and model predicted CH_4 delivery plots were almost match with the 45°-line.

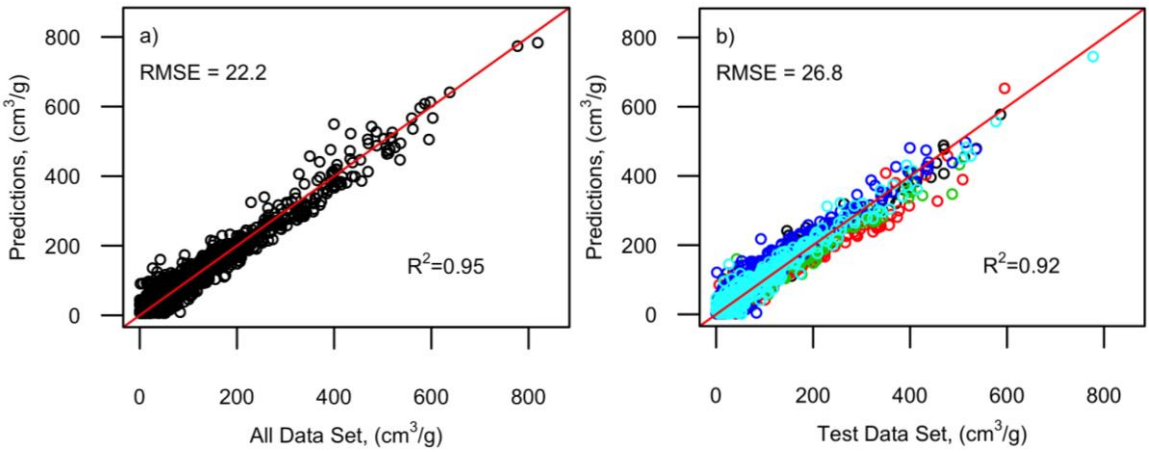

Figure 4.13. Actual versus predicted gravimetric CH_4 delivery values from ANN model constructed with structural properties a) for training dataset and b) for testing dataset.

Another ANN model to predict CH_4 delivery was constituted using pore volume only and given in Figure 4.14. To investigate the relationship between the pore volume and the CH_4 delivery capacity, the linear regression model was also fitted and given in Figure 4.15. Although the coefficient of determination (R^2), which is a measure of how well the model fits, was quite high (0.92), there seems to be some nonlinear effects because the ANN results are better especially at high delivery values.

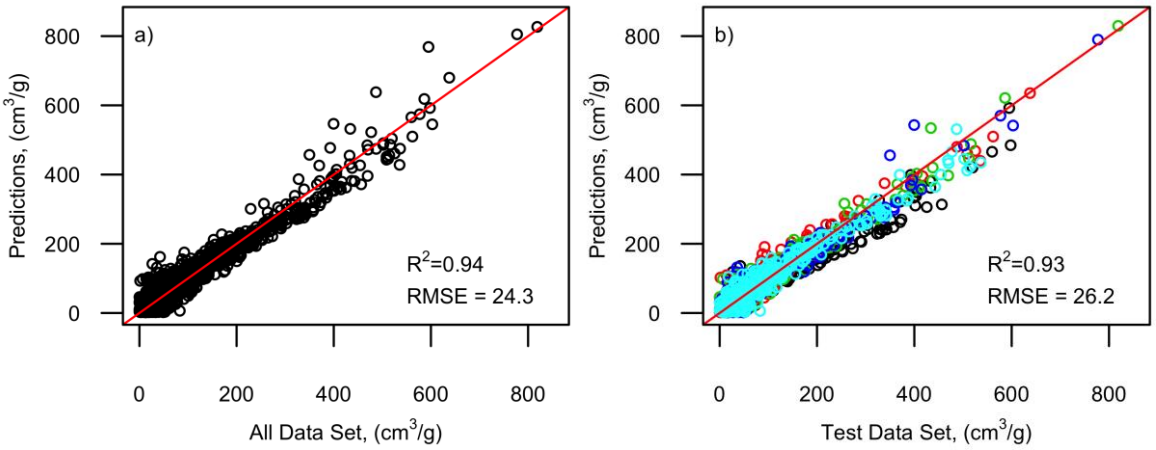

Figure 4.14. Actual versus predicted gravimetric CH_4 delivery values from ANN model constructed with pore volume only a) for training dataset and b) for testing dataset.

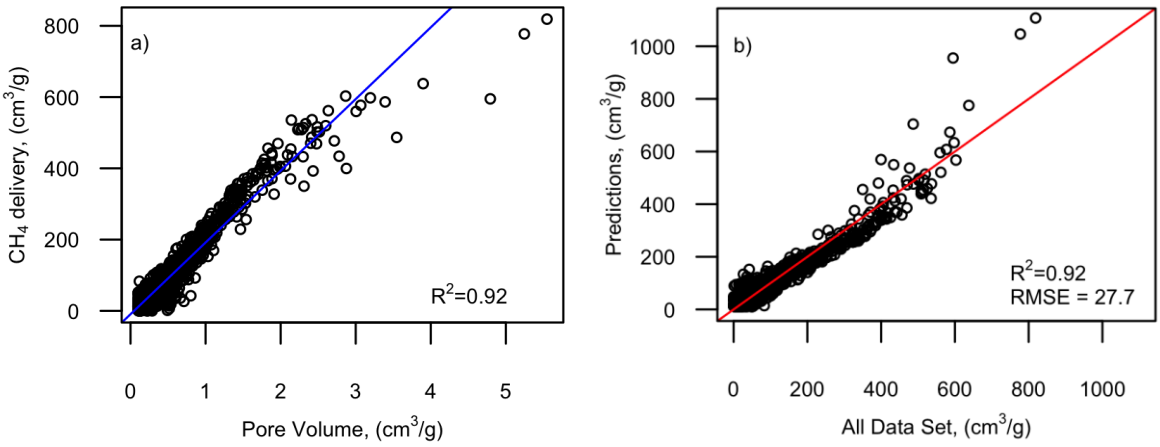

Figure 4.15. a) Linear regression model and corresponding best fit b) Actual versus predicted gravimetric CH_4 delivery values for entire database.

The results achieved with the model including pore volume only led us to investigate whether it is possible to go from user defined descriptors to pore volume and then to CH_4 delivery. Figure 4.16a and Figure 4.16b show the plots of actual versus predicted pore volume using user defined descriptors for testing and training. The results of ANN model for CH_4 delivery from predicted pore volume were given in Figure 4.16c and Figure 4.16d. When the RMSE of training and testing in Figure 4.12a and Figure 4.16c were compared, there was no improvement in the prediction of the CH_4 delivery capacity because of the fact that the pore volume predictions from the user defined descriptors were also poor. As a result, while the structural properties seems to define the CH_4 delivery mechanism well, neither pore volume nor the CH_4 delivery can be predicted from the user defined descriptors; this may verify the above discussion that there may be some missing variables/information in user defined descriptors.

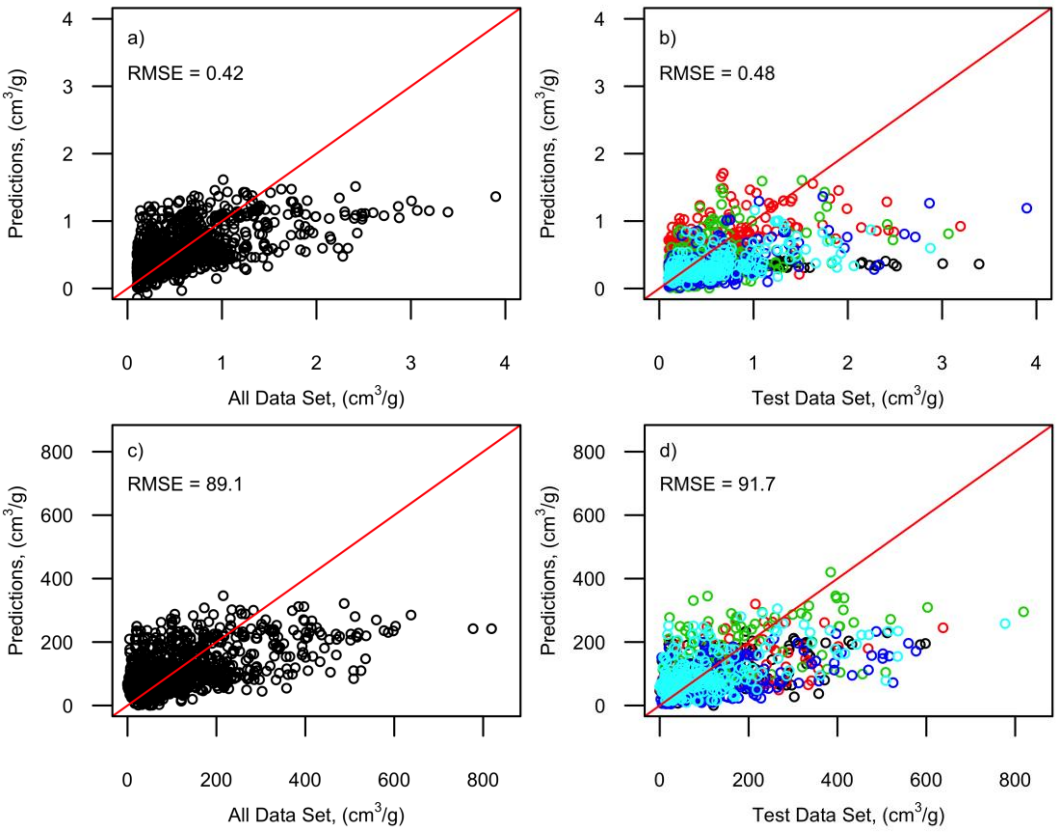

Figure 4.16. The results of two-step ANN model as actual versus predicted values; for predicting pore volumes a) training dataset b) testing dataset; for predicting CH_4 delivery values c) training dataset and d) testing dataset.

4.2.1.2. CH_4 Uptake. The range of the gravimetric amount of CH_4 uptake at 35 bar and corresponding number of instances for user defined descriptors and structural properties are given in Table 4.5.

Table 4.5. The range of the gravimetric amount of CH_4 uptake at 35 bar and corresponding number of instances.

Range	Classes	Number of instances	
		User defined descriptors	Structural properties
< 75 cm^3/g	Low performing	423	784
75 - 155 cm^3/g	Medium performing	491	797
> 155 cm^3/g	High performing	475	643

Decision Tree. Decision trees and the corresponding confusion matrices for gravimetric CH_4 uptake were given below. In Figure 4.17, user defined descriptors were taken as inputs; and the accuracies for testing and training from that tree were given in Table 4.6 as 58.4 % and 68.0 %, respectively.

Two branches having 80 % purity were selected as heuristics giving high CH_4 uptake amounts. In Heuristic 1, MOFs with less than 0.34 electronegativity, cubic structure and higher total degree of unsaturation than 9.5 were classified as having high gravimetric CH_4 uptake with 84 % accuracy. If the crystal structure was not cubic but the electronegativity ratio was still less than 0.34, the model questioned the value of total degree of unsaturation again to have high gravimetric CH_4 uptake (should be higher than 9.8). However, these split points were not sufficient to complete the Heuristic 2; MOFs' having N/O less than 0.16, also not having monoclinic and triclinic structure, metallic percentage less than 5.8 and the electronegativity ratio less than 0.27 were found as having high CH_4 uptake. There were 167 MOFs satisfying these conditions and 135 of them were predicted correctly; (135/167)×100 = 81% is a strong statistical result.

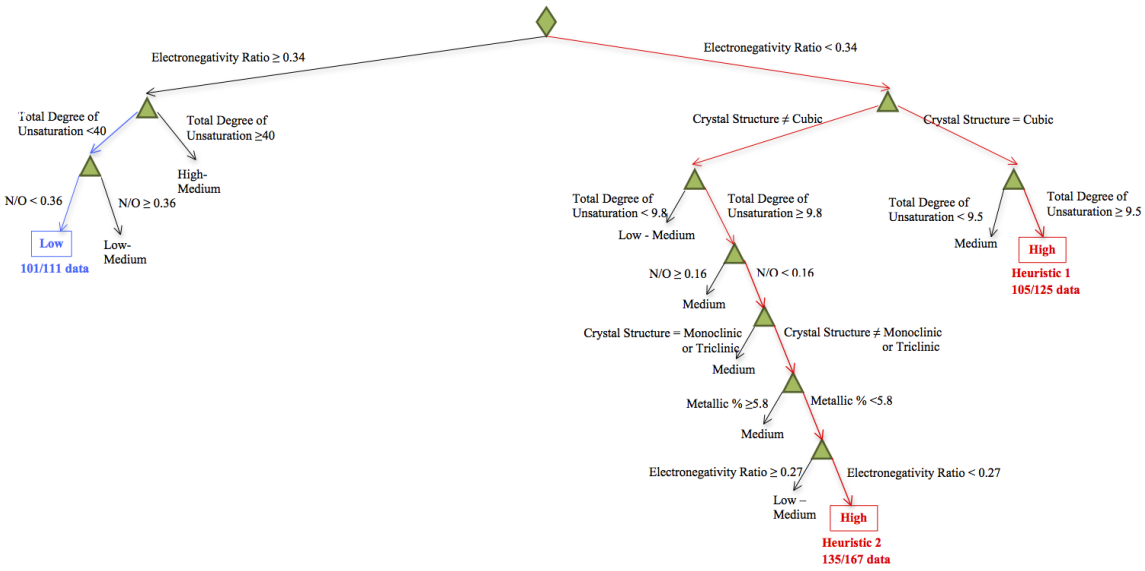

Figure 4.17. Simplified form of gravimetric CH_4 uptake classification tree developed by using user defined descriptors.

Table 4.6. Confusion matrices of decision tree constructed using user defined descriptors for gravimetric CH_4 uptake.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	423	244*	122	57	57.7
	Medium	491	59	347*	85	70.7
	High	475	25	96	354*	74.5
	Total	1389				68.0
Testing	Low	423	209*	151	63	49.4
	Medium	491	104	278*	109	56.6
	High	475	47	104	324*	68.2
	Total	1389				58.4

On the other hand, a relatively small and straightforward tree was obtained by using pore volume and MPD as structural properties. One major route for high gravimetric CH_4 uptake was indicated in Figure 4.18; classification successes of high performing class for both training and testing accuracy were greater than 90 %. The overall and the class-based accuracy rates were given in Table 4.7. There were 600 MOFs having pore volume greater than 0.56 cm^3/g in the database and 546 of them were correctly placed by using decision tree algorithm; corresponds to 91 % purity.

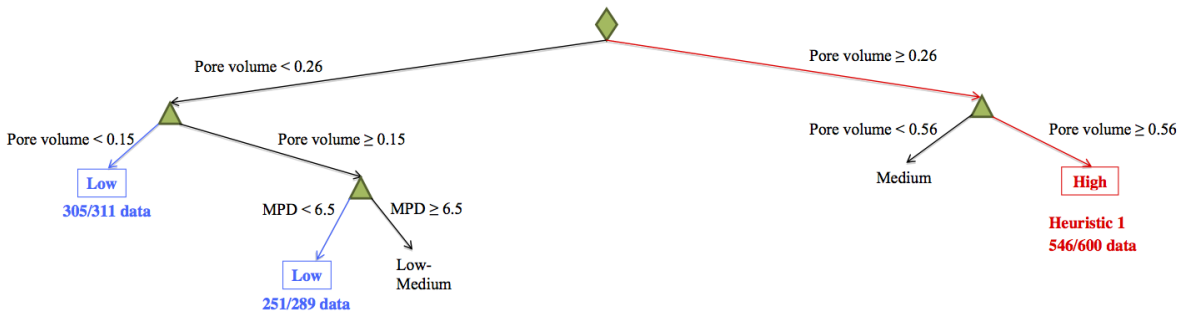

Figure 4.18. Simplified form of gravimetric CH_4 uptake classification tree developed by using structural properties.

Table 4.7. Confusion matrices of decision tree constructed using structural properties for gravimetric CH_4 uptake.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	784	703*	79	2	89.7
	Medium	797	85	629*	83	78.9
	High	643	0	33	610*	94.9
	Total	2224				87.3
Testing	Low	784	693*	89	2	88.4
	Medium	797	104	612*	81	76.8
	High	643	0	59	584*	90.8
	Total	2224				84.9

Artificial Neural Network. The gravimetric CH_4 uptake were predicted via ANN tool by using both user defined descriptors (with 1389 instances) and structural properties (with 2224 instances), separately. Determined ANN structures given in Table 3.16 were followed to decide the number of hidden layers; 5-fold cross validation was also carried out.

The predictions obtained from ANN models were plotted against actual values and the corresponding RMSE values were calculated. In Figure 4.19, better predictions were obtained by user defined descriptors when it was compared with Figure 4.12**;** but still the RMSE values were high (72.8 cm^3/g for training and 82.5 cm^3/g for testing).

The model achieved by using only structural properties was illustrated in Figure 4.20; the training and testing RMSE were 23.3 cm^3/g and 29.1 cm^3/g, respectively. Comparing them with the RMSE values of gravimetric CH_4 delivery, it was seen that better results were obtained with models shown in Figure 4.13.

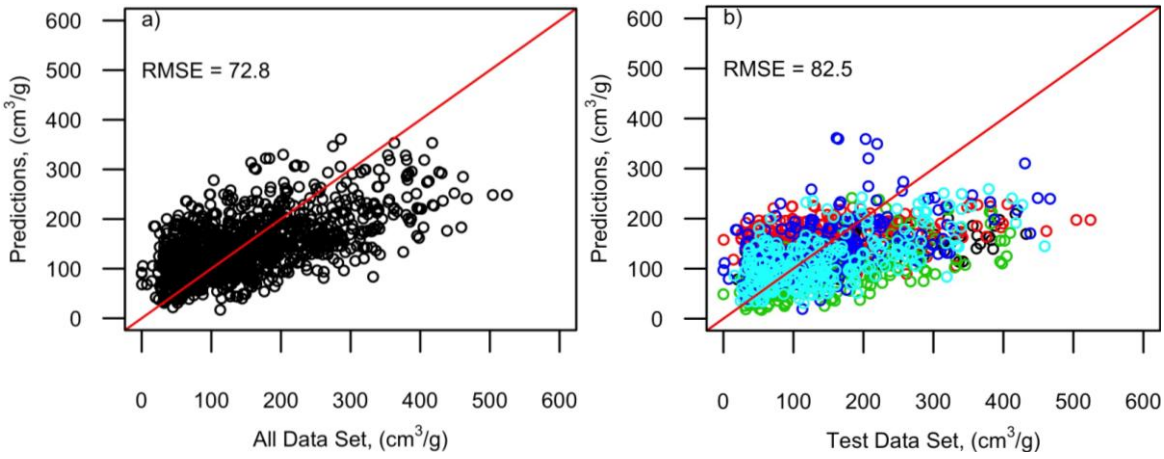

Figure 4.19. Actual versus predicted gravimetric CH_4 uptake values from ANN model constructed with user defined descriptors a) for training dataset and b) for testing dataset.

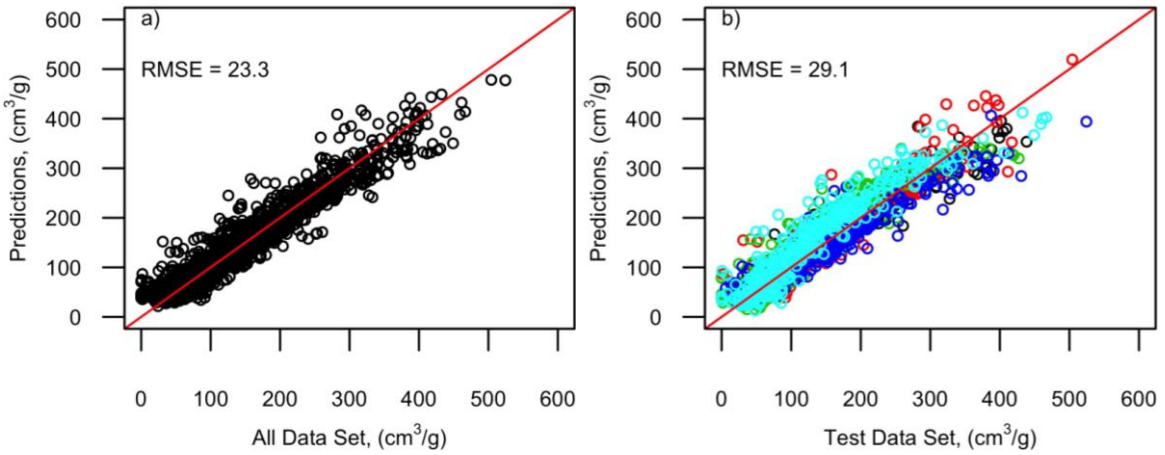

Figure 4.20. Actual versus predicted gravimetric CH_4 uptake values from ANN model constructed with structural properties a) for training dataset and b) for testing dataset.

4.2.2. For Volumetric Databases

In this section, decision tree and artificial neural network models were constructed for both volumetric CH_4 delivery (from 65 bar to 5 bar) and uptake (at 35 bar). User defined descriptors and structural properties were approached in separate analyses; 1389 and 2224 data points were used respectively.

4.2.2.1. CH_4 Delivery. The range of the volumetric amount of CH_4 delivery and corresponding number of instances for user defined descriptors and structural properties are given in Table 4.8.

Table 4.8. The range of the volumetric amount of CH_4 delivery and corresponding number of instances.

Range	Classes	Number of instances	
		User defined descriptors	Structural properties
< 50 cm^3/g	Low performing	433	773
50 - 105 cm^3/g	Medium performing	458	748
> 105 cm^3/g	High performing	498	703

Decision Tree. In order to determine the routes, which reach to the MOFs having high volumetric CH_4 delivery, the decision tree analyses were performed; the simplified version of tree constructed by using user defined descriptors was given in Figure 4.21 (the full tree in Figure A.7). It was first divided based on the type of crystal structure; if MOF has a cubic structure with metallic percentage higher than 1.7 and total degree of unsaturation greater than 17, this MOF will be included in high performing class with a probability of 88 % (110/125×100). As a second heuristic route, if the crystal structure is another type rather than cubic, there is still chance to have high volumetric CH_4 delivery. Additional four conditions were required to have high delivery amounts with a probability of 75 %. These conditions were listed as follows: having N/O, total degree of unsaturation and metallic percentage less than 0.24, greater than 17 and less than 4.8, respectively with a condition of having hexagonal, tetragonal or trigonal structure. Training and testing accuracy rates in terms of classes were given in Table 4.9; it is seen that they are similar with those of belonging to gravimetric database.

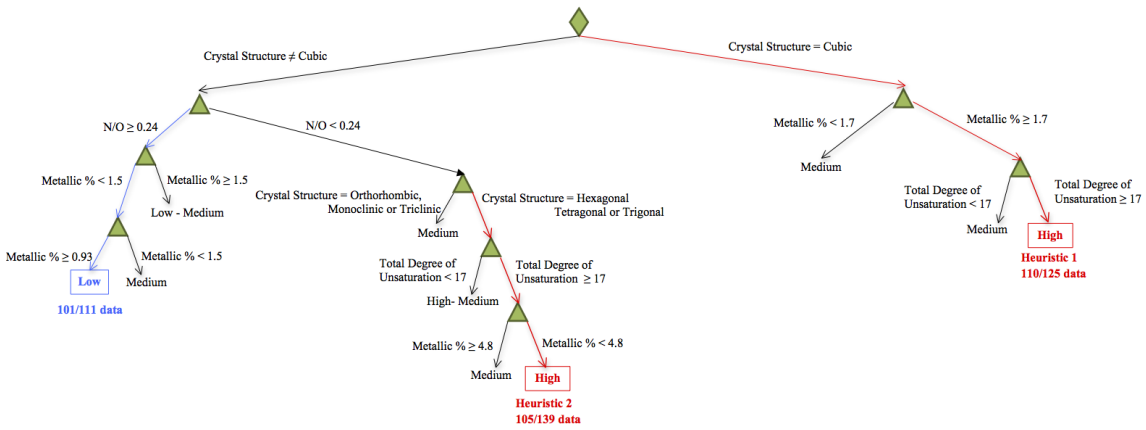

Figure 4.21. Simplified form of volumetric CH_4 delivery classification tree developed by using user defined descriptors.

Table 4.9. Confusion matrices of decision tree constructed using user defined descriptors for volumetric CH_4 delivery.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	433	350*	49	34	80.9
	Medium	458	162	228*	68	49.8
	High	498	69	46	383*	76.9
	Total	1389				69.2
Testing	Low	433	282*	101	50	65.1
	Medium	458	180	181*	97	39.5
	High	498	74	85	339*	68.1
	Total	1389				57.7

In volumetric based analysis, density can be considered as a structural property since the output variable is volumetric. It was seen from the Heuristic 1 in Figure 4.22, if the pore volume is higher than 0.68, 445 MOFs has high volumetric CH_4 delivery; 427 data points out of them were classified correctly as high performing class by using the algorithm of this model. When MPD (higher than 7.4 Å) and density (higher than 1.1 g/cm^3) get involved to route, which corresponds to Heuristic 2, relatively small amount of data points (76 out of 89) were classified correctly; however it has to be mention that the

purity was still statistically strong (85 %). Classification accuracy rates for training and testing sets were given in Table 4.10; it was seen that correctly placed data points were lower than that of the gravimetric database.

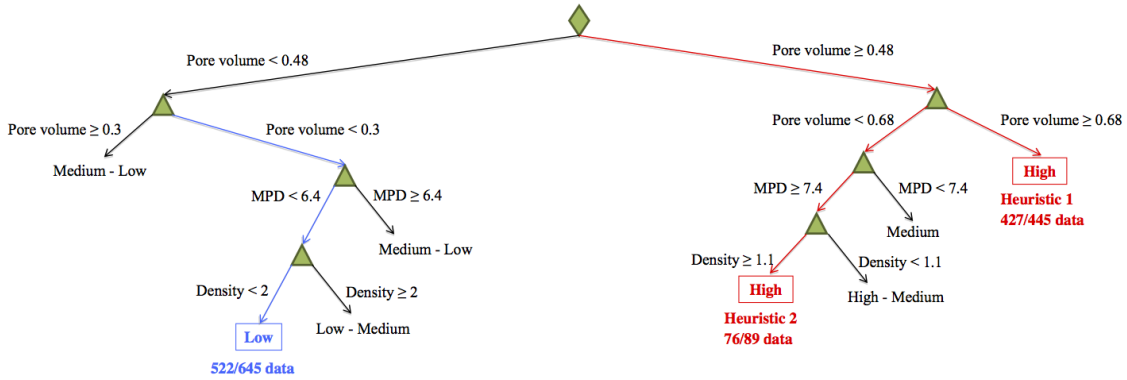

Figure 4.22. Simplified form of volumetric CH_4 delivery classification tree developed by using structural properties.

Table 4.10. Confusion matrices of decision tree constructed using structural properties for volumetric CH_4 delivery.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	773	641*	120	12	82.9
	Medium	748	149	554*	45	74.1
	High	703	21	97	585*	83.2
	Total	2224				80.0
Testing	Low	773	620*	138	15	80.2
	Medium	748	172	518*	58	69.2
	High	703	27	117	559*	79.5
	Total	2224				76.3

Artificial Neural Network. In order to predict the volumetric CH_4 delivery, ANN were performed by using two separate groups of input variable; with user defined descriptors (with 1389 instances) and structural properties (with 2224 instances). Number of hidden layers was taken from Table 3.16 and 5-fold cross validation was also applied for each analysis.

The results obtained from the models constructed with using user defied descriptors and with pore volume were shown in Figure 4.23 and Figure 4.24, respectively. Since the unit of RMSE in volumetric based analyses is cm^3/cm^3 whereas cm^3/g in gravimetric based analyses, the results of ANN models were compared within the same kind of database. By comparing the below figures, it was seen that user defined descriptors were far from defining the volumetric CH_4 delivery and structural properties was not able to sufficient to predict volumetric CH_4 delivery.

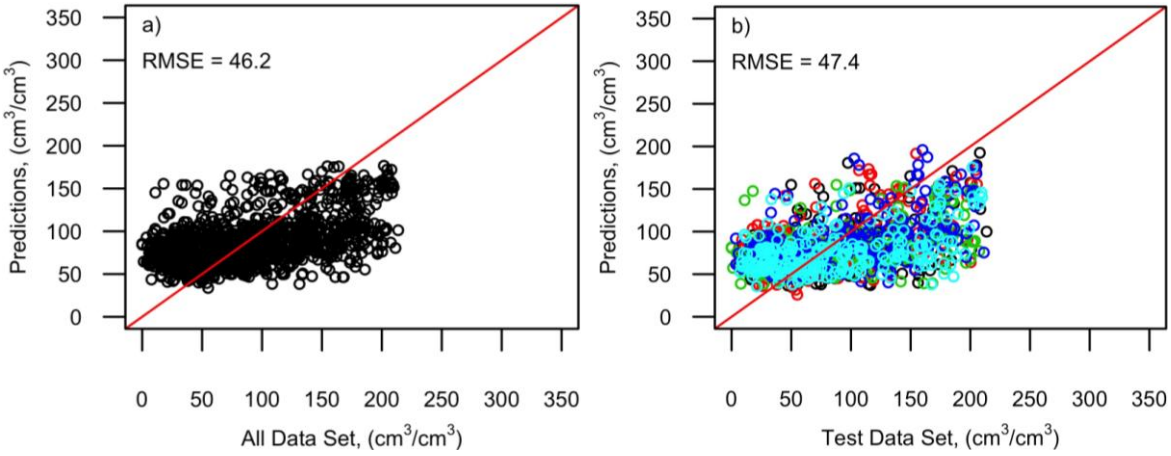

Figure 4.23. Actual versus predicted volumetric CH_4 delivery values from ANN model constructed with user defined descriptors a) for training dataset and b) for testing dataset.

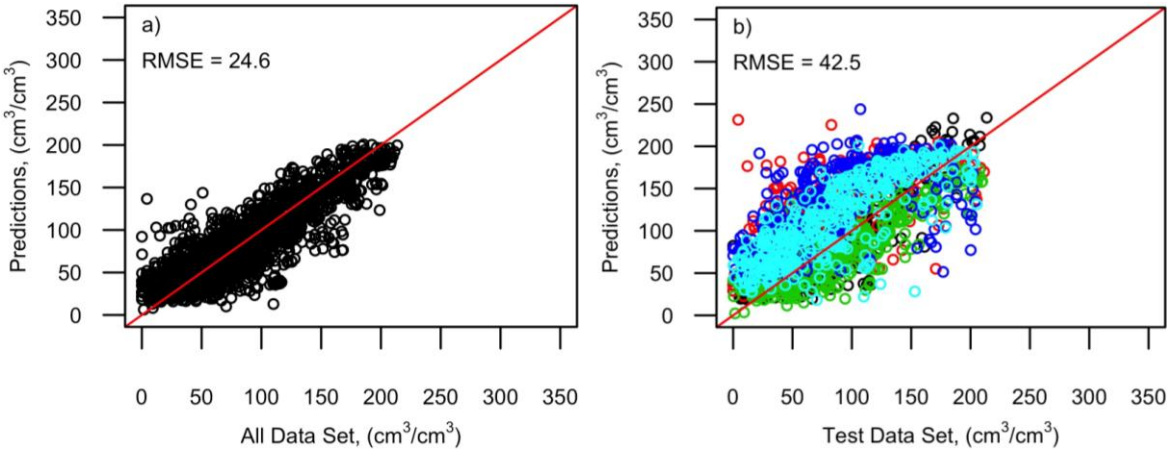

Figure 4.24. Actual versus predicted volumetric CH_4 delivery values from ANN model constructed with structural properties a) for training dataset and b) for testing dataset.

4.2.2.2. CH_4 Uptake. The range of the volumetric amount of CH_4 uptake at 35 bar and corresponding number of instances for user defined descriptors and structural properties are given in Table 4.11.

Table 4.11. The range of the volumetric amount of CH_4 uptake at 35 bar and corresponding number of instances.

Range	Classes	Number of instances	
		User defined descriptors	Structural properties
< 108 cm^3/g	Low performing	410	742
108 - 157 cm^3/g	Medium performing	484	724
> 157 cm^3/g	High performing	495	758

Decision Tree. For the decision tree of volumetric CH_4 uptake constructed by using user defined descriptors, one major and two minor heuristic routes were obtained. The first decision point of the tree is the type of crystal structure; there was no route generated to reach the high performing class for MOFs having hexagonal, monoclinic or triclinic structure. The details of the all heuristic routes were given in Figure 4.25; their classification accuracy rates were 75 %, 65 % and 100 %, respectively. The heuristic 3 showed a path in which terminal node of that path includes only high performing MOFs. In other words, if the heuristic 3 is followed, MOFs having high volumetric CH_4 uptake in 100 % are obtained. According to this route, MOFs must have Cu metal and be in tetragonal or cubic form; at the same time, they should have a degree of unsaturation in between 34 – 45, and N/O higher than 0.24.

Comparing the confusion matrix of this decision tree (Table 4.12) with the one belonging to the tree constructed for volumetric CH_4 delivery (Table 4.9), it was seen that the overall classification accuracy and more importantly the accuracy of the high performing class decreased.

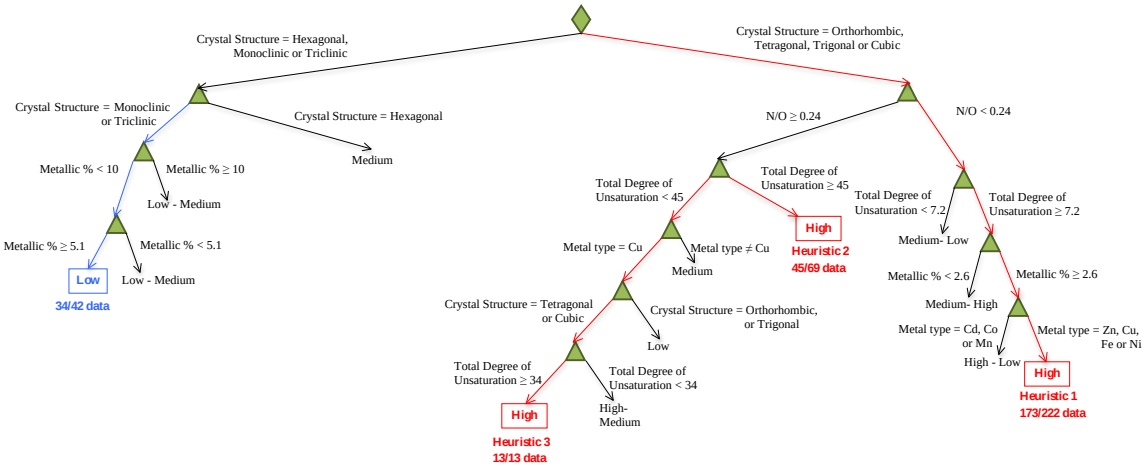

Figure 4.25. Simplified form of volumetric CH_4 uptake classification tree developed by using user defined descriptors.

Table 4.12. Confusion matrices of decision tree constructed using user defined descriptors for volumetric CH_4 uptake.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	410	259*	104	47	63.2
	Medium	484	80	300*	104	62.0
	High	495	73	62	360*	72.7
	Total	1389				66.2
Testing	Low	410	186*	149	77	45.4
	Medium	484	114	256*	114	52.9
	High	495	79	114	302*	61.0
	Total	1389				53.6

In case of modeling the volumetric CH_4 uptake database, structural properties tend to have higher accuracy rates than user defined descriptors as in all databases. Two major routes shown in Figure 4.26 had 87 % and 89 % purity to obtain high performing MOFs. Being pore volume higher than 0.7 cm3/g was the only condition for heuristic 1 whereas the density of the MOFs /with pore volume in between 0.46 - 0.7 cm3/g) must be higher than 1.1 g/cm3 according to heuristic 2. Confusion matrix given in Table 4.13 shows the

all accuracy rates of all classes. When Table 4.10 and Table 4.13 are compared, the average classification accuracies were lower in Table 4.13.

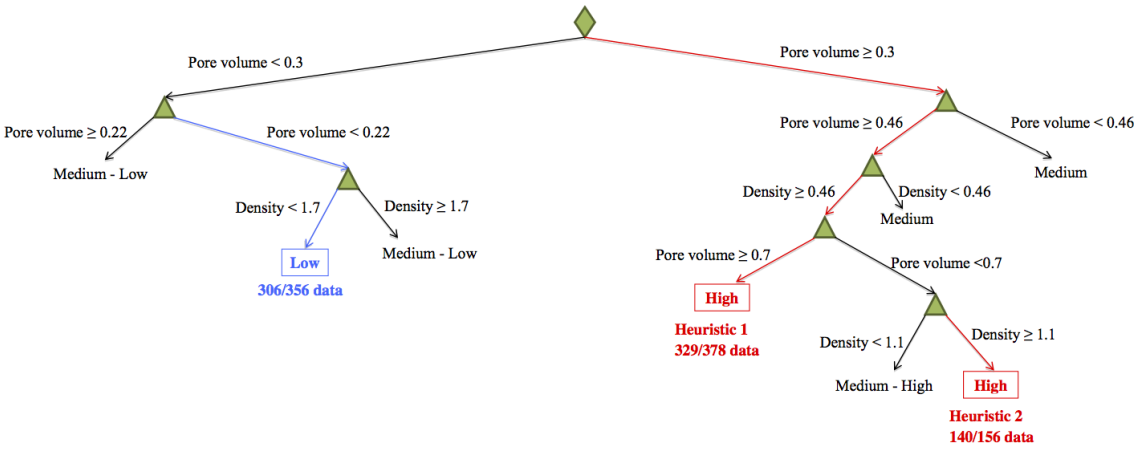

Figure 4.26. Simplified form of volumetric CH_4 uptake classification tree developed by using structural properties.

Table 4.13. Confusion matrices of decision tree constructed using structural properties for volumetric CH_4 uptake.

	Database		Predictions			Classification accuracy, %
	Class	Number of Data	Low	Medium	High	
Training	Low	742	630*	92	20	84.9
	Medium	724	137	462*	125	63.8
	High	758	10	84	664*	87.6
	Total	2224				79.0
Testing	Low	742	577*	145	20	77.8
	Medium	724	130	432*	162	60.0
	High	758	10	109	639*	84.3
	Total	2224				74.1

Artificial Neural Network. Predictions of the volumetric CH_4 uptake values were made with two separate ANN analyses as the other databases; in the first model, user defined descriptors were used whereas structural properties was utilized for second ANN analysis. Optimum number of hidden layers was given in Section 3.2.2 and the corresponding structures in Table 3.16 were utilized; 5-fold cross validation was also applied for each analysis.

The results of ANN model constructed by using user defined descriptors were given in Figure 4.27; the training and the testing RMSE values were calculated as 42.9 cm^3/cm^3 and 47.6 cm^3/cm^3, respectively. From the comparison of this model with the one shown in Figure 4.23, it can be said that user defined descriptors described volumetric CH_4 uptake better than volumetric CH_4 delivery.

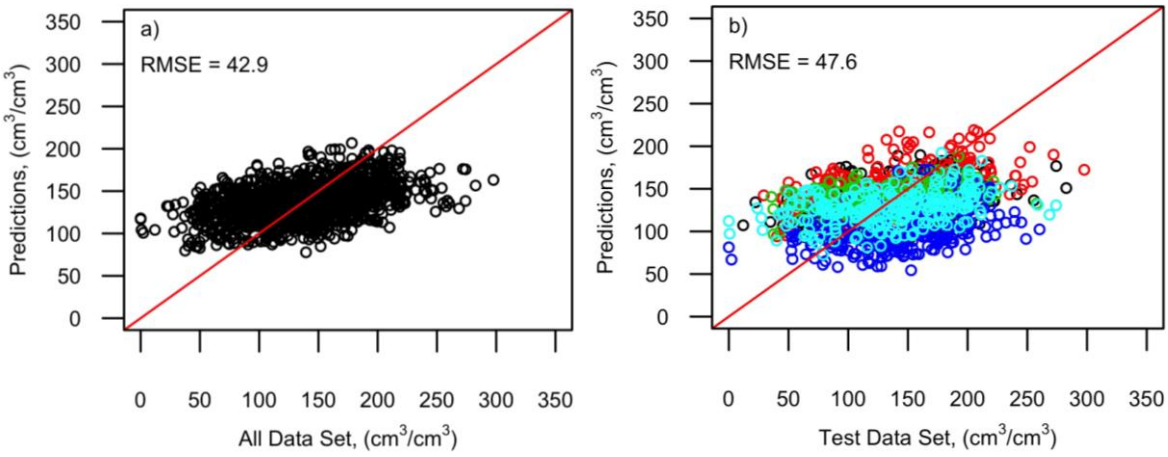

Figure 4.27. Actual versus predicted volumetric CH_4 uptake values from ANN model constructed with user defined descriptors a) for training dataset and b) for testing dataset.

After that, the second ANN model to predict volumetric CH_4 uptake was built by using structural properties; relatively higher RMSE values than those of the model constructed to predict volumetric CH_4 delivery, which was illustrated in Figure 4.24, were obtained.

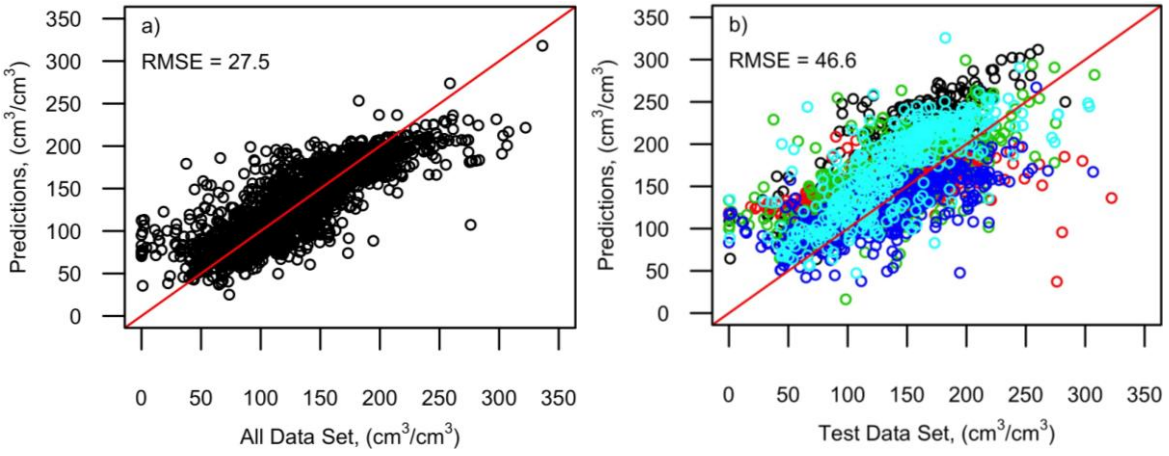

Figure 4.28. Actual versus predicted volumetric CH_4 uptake values from ANN model constructed with structural properties a) for training dataset and b) for testing dataset.

4.3. Validating Models

In this section, validation of our successful models (decision tree from user defined descriptors and structural properties and neural network from structural properties) using the experimental results reported in the literature was performed. Since the experimental uptake values were generally reported at 35 bar gravimetrically, we compare the results of gravimetric CH_4 uptake models with the experimental findings. Due to the fact that GCMC simulations generate absolute-based CH_4 uptake values, our database and models were constructed on the absolute basis. On the other hand, experimental uptake values are generally on the excess basis. In order to make reliable comparison, excess-based experimental values were converted to absolute-based values by using Equation 4.1. Density of adsorbed gas, ρ_{gas}, at 35 bar and 298 K is calculated by the ideal gas density formula given in Equation 4.2, whereas V_p is the pore volume of corresponding MOF and taken from the database.

$$\text{Absolute uptake} = \text{Excess uptake} + V_p \times \rho_{gas} \tag{4.1}$$

$$\rho_{gas} = \frac{P \times M}{R \times T} \tag{4.2}$$

The experimental results are compared with decision tree predictions in Table 4.14 and Table 4.15; whether the experimental results comply the heuristics suggested by decision tree for high CH_4 uptake (>155 cm^3/g) was checked. The model developed from user defined descriptors successfully predicts the exact class of five experimental results out of 10. Two of the of remaining five was also successfully predicted if the user defined variables we computed used for prediction; the discrepancies for these two points are not between experimental results and model prediction but between the amount of uptake reported in the papers and those we computed from GCMC. The decision tree model constructed from the structural properties, however, classifies eight out of 10 correctly, which is a considerable success; again the remaining two are misclassified due to the uptake computed, not due to decision tree model itself.

Table 4.14. Comparison of heuristics obtained from decision trees developed by using user defined descriptors with experimental data.

MOF name	User defined descriptors						CH_4 uptake at 35 bar (cm^3 CH_4/g MOF)		Classification accuracy
	Metal	Crystal Structure	Elec. ratio	Metal %	N/O	Total degree of unsaturation	Experimental	Class in DT	
UTSA-20[104]	Cu	Hexagonal	0.21	5.26	0	97	222(2)	High	+
Cu-TDPAT[105]	Cu	Tetragonal	0.30	4.35	0.4	37	217(1)	High	-
CuBTC[54]	Cu	Cubic	0.18	3.57	0	31	255(1)	High	+
PCN-14[54]	Cu	Trigonal	0.17	3.33	0	22	244(1)	High	+
PCN-11[106]	Cu	Trigonal	0.24	4.76	0	13	257(2)	High	+
PCN-46[107]	Cu	Trigonal	0.24	4.76	0	16	258(2)	High	+
PCN-80[108]	Cu	Tetragonal	0.18	3.03	0.2	48	230(2)	High	-
NU-125[109]	Cu	Tetragonal	0.29	3.57	0.6	52	311(1)	High	-
CPO-27-Ni[99]	Ni	Trigonal	0.33	11.11	0	8	171(2)	High	-*
CPO-27-Co[99]	Co	Trigonal	0.33	11.11	0	8	164(2)	High	-*

(1) Absolute-based experimental values (at 298 K and 35 bar) are taken from corresponding reference.

(2) Absolute-based experimental values (at 298 K and 35 bar) are calculated from excess-based experimental values taken from corresponding reference.

*Our simulated uptake is in medium performing class (as tree predicted) while the experimental values are in high performing class.

Table 4.15. Comparison of heuristics obtained from decision trees developed by using structural properties with experimental data.

MOF name	Structural properties		CH_4 uptake at 35 bar (cm^3 CH_4/g MOF)		Classification accuracy
	V_p, (cm^3/g)	MPD, Å	Experimental	Class in DT	
UTSA-20[104]	0.69	9.43	222[(2)]	High	+
Cu-TDPAT[105]	0.95	16.31	217[(1)]	High	+
CuBTC[54]	1.70	19.35	255[(1)]	High	+
PCN-14[54]	0.85	10.79	244[(1)]	High	+
PCN-11[106]	1.07	10.65	257[(2)]	High	+
PCN-46[107]	1.31	11.6	258[(2)]	High	+
PCN-80[108]	1.38	15.26	230[(2)]	High	+
NU-125[109]	1.40	19.23	311[(1)]	High	+
CPO-27-Ni[99]	0.53	11.32	171[(2)]	High	-*
CPO-27-Co[99]	0.54	11.52	164[(2)]	High	-*

[(1)] Absolute-based experimental values (at 298 K and 35 bar) are taken from corresponding reference.

[(2)] Absolute-based experimental values (at 298 K and 35 bar) are calculated from excess-based experimental values taken from corresponding reference.

*Our simulated uptake is in medium performing class (as tree predicted) while the experimental values are in high performing class.

Validation of artificial neural network model developed from structural properties using the same experimental CH_4 uptake at 35 bar was also done; we did not compare the model constructed from user defined variables because it was already found to be unsuccessful. The results are presented in Table 4.16; they were quite successful. The model predicted five of 10 MOFs with the errors less than 10 % while it predicts the three of the remaining MOFs with errors within 10-20 %; only two MOFs were predicted with higher errors.

Table 4.16. Comparison of predicted CH_4 uptakes with experimental data.

MOF name	Experimental CH_4 uptake (cm^3 CH_4/g MOF)	Predicted CH_4 uptake by ANN (cm^3 CH_4/g MOF)	Error, %
UTSA-20[104]	222[(2)]	192	14
Cu-TDPAT[105]	217[(1)]	230	6
CuBTC[54]	255[(1)]	334	31
PCN-14[54]	244[(1)]	225	8
PCN-11[106]	257[(2)]	269	5
PCN-46[107]	258[(2)]	307	19
PCN-80[108]	230[(2)]	306	33
NU-125[109]	311[(1)]	297	5
CPO-27-Ni[99]	171[(2)]	145	15
CPO-27-Co[99]	164[(2)]	149	9

[(1)] Absolute-based experimental values (at 298 K and 35 bar) are taken from corresponding reference.

[(2)] Absolute-based experimental values (at 298 K and 35 bar) are calculated from excess-based experimental values taken from corresponding reference.

5. CONCLUSION

The aim of this thesis was to extract knowledge for both gravimetric and volumetric CH_4 delivery capacity (from 65 to 5 bar) and uptake capacity (at 35 bar) of various real MOFs by using decision tree analysis and artificial neural network. Before modeling, a review and pre-analysis of the database were carried out to observe the simple trends.

Decision trees using both user defined descriptors and structural properties were constructed for four types of CH_4 databases. The obtained accuracy rates for testing and training sets were given in Table 5.1. According to these results, the classification trees for gravimetric databases by using structural properties show great success. However, the aim for decision tree analysis is to find out the routes of high performing class of CH_4 delivery or uptake, the overall accuracy rates may be misleading. For that reason, the heuristics, which give high CH_4 delivery or uptake in all trees, were obtained and shown in corresponding section for future studies. In general, the crystal structure and the total degree of unsaturation were found to be the effective among user defined descriptors, whereas the pore volume was sufficient among structural properties to obtain MOFs having high CH_4 delivery and uptake.

Table 5.1. Accuracy rates of constructed decision trees for all databases.

Databases	**By using user defined descriptors**		**By using structural properties**	
	Testing accuracy, %	**Training accuracy, %**	**Testing accuracy, %**	**Training accuracy, %**
Gravimetric CH_4 delivery	58.1	66.3	82.4	84.0
Gravimetric CH_4 uptake	58.4	68.0	84.9	87.3
Volumetric CH_4 delivery	57.7	69.2	76.3	80.0
Volumetric CH_4 uptake	53.6	66.2	74.1	79.0

To investigate the predictability of the CH_4 delivery, ANN models were constructed as well for all four databases; the RMSE values obtained were given in Table 5.2. Since

the unit of RMSE changes with respect to kind of database (gravimetric or volumetric), the comparisons within the same kind of database were done. Gravimetric CH_4 delivery was predicted more properly than the gravimetric CH_4 uptake by using structural properties whereas gravimetric CH_4 uptake was predicted more properly than the gravimetric CH_4 delivery by using user defined descriptors. From the volumetric based analysis, it was observed that RMSE values were close to each other. In overall, it can be said that gravimetric CH_4 delivery values, which were predicted by using structural properties, overlapped significantly with the actual values. Linear model between pore volume and gravimetric CH_4 delivery was also tested and found to be quite well. However, ANN predictions were better indicating that there may be some nonlinear relations between pore volume and CH_4 delivery. Moreover, two-step ANN model was constructed to investigate whether it is possible to go from user defined descriptors to pore volume and then to gravimetric CH_4 delivery; however, no significant improvement was achieved. Apparently, there are some missing variables in user defined descriptors that the CH_4 delivery capacity mechanism cannot be fully defined.

Table 5.2. RMSE values of ANN models for all databases.

Databases	**By using user defined descriptors**		**By using structural properties**	
Gravimetric	**Testing RMSE, cm^3/g**	**Training RMSE, cm^3/g**	**Testing RMSE, cm^3/g**	**Training RMSE, cm^3/g**
CH_4 delivery	100.9	85.0	26.8	22.2
CH_4 uptake	82.5	72.8	29.1	23.3
Volumetric	**Testing RMSE, cm^3/cm^3**	**Training RMSE, cm^3/cm^3**	**Testing RMSE, cm^3/cm^3**	**Training RMSE, cm^3/cm^3**
CH_4 delivery	47.4	46.2	42.5	24.6
CH_4 uptake	42.9	47.6	46.6	27.5

Both decision tree and artificial neural network models were validated using the experimental results reported in literature and both models based on the structural properties were predicted the experimental results successfully.

REFERENCES

1. Zhou, H.C., J.R. Long, O.M. Yaghi, "Introduction to Metal-Organic Frameworks", *Chem. Rev.*, Vol. 112, pp. 673–674, 2012.

2. Pardakhti, M., E. Moharreri, D. Wanik, S.L. Suib, R. Srivastava, "Machine Learning Using Combined Structural and Chemical Descriptors for Prediction of Methane Adsorption Performance of Metal Organic Frameworks (MOFs)", *ACS Comb. Sci.*, Vol. 19, pp. 640–645, 2017.

3. Mason, J.A., J. Oktawiec, M.K. Taylor, M.R. Hudson, J. Rodriguez, J.E. Bachman, M.I. Gonzalez, A. Cervellino, A. Guagliardi, C.M. Brown, P.L. Llewellyn, N. Masciocchi, J.R. Long, "Methane Storage in Flexible Metal-Organic Frameworks with Intrinsic Thermal Management", *Nature*, Vol. 527, pp. 357–361, 2015.

4. Tsivion, E., M. Head-Gordon, "Methane Storage: Molecular Mechanisms Underlying Room-Temperature Adsorption in $Zn_4O(BDC)_3$ (MOF-5)", *J. Phys. Chem. C.*, Vol. 121, pp. 12091–12100, 2017.

5. Gómez-Gualdrón, D.A., T.C. Wang, P. García-Holley, R.M. Sawelewa, E. Argueta, R.Q. Snurr, J.T. Hupp, T. Yildirim, O.K. Farha, "Understanding Volumetric and Gravimetric Hydrogen Adsorption Trade-off in Metal-Organic Frameworks", *ACS Appl. Mater. Interfaces.*, Vol. 9, pp. 33419–33428, 2017.

6. Haldoupis, E., S. Keskin, S. Nair, D.S. S., United States Patent No: US 8,725,482 B2, 2010.

7. Altintas, C., I. Erucar, S. Keskin, "High-Throughput Computational Screening of the Metal Organic Framework Database for CH_4/H_2 Separations", *ACS Appl. Mater. Interfaces.* Vol. 10, pp. 3668–3679, 2018.

8. Uzun, A., S. Keskin, "Site Characteristics in Metal Organic Frameworks for Gas

Adsorption", *Prog. Surf. Sci.*, Vol. 89, pp. 56–79, 2014.

9. Larose, D.T., *Discovering Knowledge in Data. An Introduction to Data Mining*, John Wiley & Sons, New York, 2005.

10. Fernandez, M., T.K. Woo, C.E. Wilmer, R.Q. Snurr, "Large-Scale Quantitative Structure-Property Relationship (QSPR) Analysis of Methane Storage in Metal-Organic Frameworks", *J. Phys. Chem. C.*, Vol. 117, pp. 7681–7689, 2013.

11. Ohno, H., Y. Mukae, "Machine Learning Approach for Prediction and Search: Application to Methane Storage in a Metal-Organic Framework", *J. Phys. Chem. C.*, Vol. 120, pp. 23963–23968, 2016.

12. Sumida, K., D.L. Rogow, J.A. Mason, T.M. McDonald, E.D. Bloch, Z.R. Herm, T.H. Bae, J.R. Long, "Carbon Dioxide Capture in Metal-Organic Frameworks", *Chem. Rev.*, Vol. 112, pp. 724–781, 2012.

13 Qiu, S., G. Zhu, "Molecular Engineering for Synthesizing Novel Structures of Metal-Organic Frameworks with Multifunctional Properties", *Coord. Chem. Rev.*, Vol. 253, pp. 2891–2911, 2009.

14. Kuppler, R.J., D.J. Timmons, Q.R. Fang, J.R. Li, T.A. Makal, M.D. Young, D. Yuan, D. Zhao, W. Zhuang, H.C. Zhou, "Potential Applications of Metal-Organic Frameworks", *Coord. Chem. Rev.*, Vol. 253, pp. 3042–3066, 2009.

15. Manocha, S.M., "Porous Carbons", *Sadhana*, Vol. 28, pp. 335–348, 2003.

16. Férey, G., "Hybrid Porous Solids: Past, Present, Future", *Chem. Soc. Rev.*, Vol. 37, pp. 191–214, 2008.

17. Mueller, U., M. Schubert, F. Teich, H. Puetter, K. Schierle-Arndt, J. Pastré, "Metal–Organic Frameworks—Prospective Industrial Applications", *J. Mater. Chem.*, Vol. 16, pp. 626–636, 2006.

18. Bon, V., "Metal-Organic Frameworks for Energy-Related Applications", *Curr. Opin. Green Sustain. Chem.*, Vol. 4, pp. 44–49, 2017.

19. Furukawa, H., K.E. Cordova, M. O'Keeffe, O.M. Yaghi, "The Chemistry and Applications of Metal-Organic Frameworks", *Science*, Vol. 341, pp. 974-986, 2013.

20. Bichoutskaia, E., M. Suyetin, M. Bound, Y. Yan, M. Schröder, "Enhancement of Methane Adsorption in Metal-Organic Frameworks Containing Nanographene Linkers : A Multiscale Computational Study", *J. Phys. Chem. C.*, Vol. 118, pp. 15573−15580, 2014.

21. Moghadam P.Z., A. Li, S.B. Wiggin, A. Tao, A.G.P. Maloney, P.A. Wood, S.C. Ward, D. Fairen-Jimenez, "The Development of a CSD Subset: A Collection of Metal-Organic Frameworks for Past, Present and Future", *Chem. Mater.*, Vol. 29 (7), pp. 2618-2625, 2017.

22. Perez E. V., C. Karunaweera, I. H. Musselman, K. J. Balkus, J. P. Ferraris, "Origins and Evolution of Inorganic-Based and MOF-Based Mixed-Matrix Membranes for Gas Separations", *Processes*, Vol. 4, 32, 2016.

23. O'Keeffe, M., "Design of MOFs ind Intellectual Content in Reticular Chemistry: A Personal View", *Chem. Soc. Rev.*, Vol. 38, pp. 1215-1217, 2009.

24. Yaghi, O.M., C.E. Davis, G. Li, H. Li, "Selective Guest Binding by Tailored Channels in a 3-D Porous Zinc(II)-Benzenetricarboxylate Network", *J. Am. Chem. Soc*., Vol. 119, pp. 2861–2868, 1997.

25. Li, H., M. Eddaoudi, T.L. Groy, O.M. Yaghi, "Establishing Microporosity in Open Metal-Organic Frameworks: Gas Sorption Isotherms for Zn(BDC) (BDC = 1,4-Benzenedicarboxylate)", *J. Am. Chem. Soc*., Vol. 120, pp. 8571–8572, 1998.

26. Czaja, A.U., N. Trukhan, U. Müller, "Industrial Applications of Metal–Organic Frameworks", *Chem. Soc. Rev*., Vol. 38, pp. 1284-1293, 2009.

27. Thommes, M., "Physical Adsorption Characterization of Nanoporous Materials", *Chemie-Ingenieur-Technik*, Vol. 82, pp. 1059–1073, 2010.

28. Rouquerol, J., F. Rouquerol, P. Llewellyn, G. Maurin, K.S.W. Sing, *Adsorption by Powders and Porous Solids: Principles, Methodology and Applications*, Second Edition, Elsevier, London, 2012.

29. Berger, A.H., A.S. Bhown, "Comparing Physisorption and Chemisorption Solid Sorbents for use Separating CO_2 from Flue Gas Using Temperature Swing Adsorption", *Energy Procedia*, Vol. 4, pp. 562–567, 2011.

30. Belmabkhout, Y., R. Serna-Guerrero, A. Sayari, " Adsorption of CO_2-Containing Gas Mixtures over Amine-Bearing Pore-Expanded MCM-41 Silica: Application for Gas Purification", *Adsorpt. J. Int. Adsorpt. Soc*., Vol. 49, pp. 359–365, 2010.

31. Keskin, S., D.S. Sholl, "Selecting Metal Organic Frameworks as Enabling Materials in Mixed Matrix Membranes for High Efficiency Natural Gas Purification", *Energy Environ. Sci*., Vol. 3, pp. 343–351, 2010.

32. Hasan, M.M.F., R.C. Baliban, J.A. Elia, C.A. Floudas, "Modeling, Simulation, and Optimization of Postcombustion CO_2 Capture for Variable Feed Concentration and Flow Rate. 2. Pressure Swing Adsorption and Vacuum Swing Adsorption Processes", *Ind. Eng. Chem. Res.*, Vol. 51, pp. 15665–15682, 2012.

33. Basu, S., A. Cano-Odena, I.F.J. Vankelecom, "MOF-Containing Mixed-Matrix Membranes for CO_2/CH_4 and CO_2/N_2 Binary Gas Mixture Separations", *Sep. Purif. Technol*., Vol. 81, pp. 31–40, 2011.

34. Adatoz, E., A.K. Avci, S. Keskin, "Opportunities and Challenges of MOF-Based Membranes in Gas Separations", *Sep. Purif. Technol.*, Vol. 152, pp. 207–237, 2015.

35. Zornoza, B., C. Tellez, J. Coronas, J. Gascon, F. Kapteijn, "Metal Organic Framework Based Mixed Matrix Membranes: An Increasingly Important Field of

Research with a Large Application Potential", *Microporous Mesoporous Mater*., Vol. 166, pp. 67–78, 2013.

36. Rezakazemi, M., A. Ebadi Amooghin, M.M. Montazer-Rahmati, A.F. Ismail, T. Matsuura, "State-Of-The-Art Membrane Based CO_2 Separation Using Mixed Matrix Membranes (MMMs): An Overview on Current Status and Future Directions", *Prog. Polym. Sci*., Vol. 39, pp. 817–861, 2014.

37. Yehia, H., T.J. Pisklak, J.P. Ferraris, K.J. Balkus, I.H. Musselman, "Methane Facilitated Transport Using Copper(II) Biphenyl Dicarboxylatetriethylenediamine /Poly(3-Acetoxyethylthiophene) Mixed Matrix Membranes", *Polym. Prepr.*, Vol. 45, pp. 35-36, 2004.

38. Perez, E. V., K.J. Balkus, J.P. Ferraris, I.H. Musselman, "Mixed-Matrix Membranes Containing MOF-5 for Gas Separations", *J. Memb. Sci.*, Vol. 328, pp. 165–173, 2009.

39. Hunger, K., N. Schmeling, H.B.T. Jeazet, C. Janiak, C. Staudt, K. Kleinermanns, "Investigation of Cross-Linked and Additive Containing Polymer Materials for Membranes with Improved Performance in Pervaporation and Gas Separation", *Membranes (Basel)*, Vol. 2, pp. 727–763, 2012.

40. Zhou, H.-C. “Joe”, S. Kitagawa, "Metal–Organic Frameworks (MOFs)", *Chem. Soc. Rev*., Vol. 43, pp. 5415–5418, 2014.

41. Dolgopolova, E.A., A.M. Rice, C.R. Martin, N.B. Shustova, "Photochemistry and Photophysics of MOFs: Steps Towards MOF-Based Sensing Enhancements", *Chem. Soc. Rev*., Vol. 47, pp. 4710-4728, 2018.

42. Hu, Z., B.J. Deibert, J. Li, "Luminescent Metal-Organic Frameworks for Chemical Sensing and Explosive Detection", *Chem. Soc. Rev*., Vol. 43, pp. 5815–5840, 2014.

43. Wang, C., T. Zhang, W. Lin, "Rational Synthesis of Noncentrosymmetric Metal-

Organic Frameworks for Second-Order Nonlinear Optics", Chem. Rev., Vol. 112, pp. 1084–1104, 2012.

44. Stavila, V., A.A. Talin, M.D. Allendorf, "MOF-Based Electronic and Opto-Electronic Devices", *Chem. Soc. Rev.*, Vol. 43, pp. 5994–6010, 2014.

45. Horcajada, P., R. Gref, T. Baati, P.K. Allan, G. Maurin, P. Couvreur, G. Férey, R.E. Morris, C. Serre, "Metal-Organic Frameworks in Biomedicine", *Chem. Rev.*, Vol. 112, pp. 1232–1268, 2012.

46. Rocca, J. D., D. Liu, W. Lin, "Nanoscale Metal-Organic Frameworks for Biomedical Imaging and Drug Delivery", *Acc. Chem. Res.*, Vol. 44, pp. 957–968, 2011.

47. He, Y., W. Zhou, G. Qian, B. Chen, "Methane Storage in Metal-Organic Frameworks", *Chem. Soc. Rev.*, Vol. 43, pp. 5657–5678, 2014.

48. Vasiliev, L.L., L.E. Kanonchik, D. A. Mishkinis, M.I. Rabetsky, "Adsorbed Natural Gas Storage and Transportation Vessels", *Int. J. Therm. Sci.*, Vol. 39, pp. 1047–1055, 2000.

49. Goldsmith, J., A.G. Wong-foy, M.J. Cafarella, D.J. Siegel, J. Goldsmith, A.G.W.-Foy, M.J. Cafarella, D.J. Siegel, "Theoretical Limits of Hydrogen Storage in Metal-Organic Frameworks : Opportunities and Trade-Offs Theoretical Limits of Hydrogen Storage in Metal-Organic Frameworks : Opportunities and Trade-Offs", *Chem. Mater*., Vol. 25, pp. 3373−3382 ,2013.

50. Ahmed, A., Y. Liu, J. Purewal, L.D. Tran, A.G. Wong-Foy, M. Veenstra, A.J. Matzger, D.J. Siegel, "Balancing Gravimetric and Volumetric Hydrogen Density in MOFs", *Energy Environ. Sci.*, Vol. 10, pp. 2459–2471, 2017.

51. Sharma, A., R. Babarao, N. V Medhekar, A. Malani, "Methane Adsorption and Separation in Slipped and Functionalized Covalent Organic Frameworks", *Ind. Eng.*

Chem. Res., Vol. 57, pp. 4767–4778, 2018.

52. Wen, H.M., B. Li, L. Li, R.B. Lin, W. Zhou, G. Qian, B. Chen, "A Metal–Organic Framework with Optimized Porosity and Functional Sites for High Gravimetric and Volumetric Methane Storage Working Capacities", *Adv. Mater.*, Vol. 30, pp. 1–6, 2018.

53. He, Y., W. Zhou, T. Yildirim, B. Chen, "A Series of Metal-Organic Frameworks with High Methane Uptake and an Empirical Equation for Predicting Methane Storage Capacity", *Energy Environ. Sci.*, Vol. 6, pp. 2735–2744, 2013.

54. Mason, J.A., M. Veenstra, J.R. Long, "Evaluating Metal–Organic Frameworks for Natural Gas Storage", *Chem. Sci*., Vol. 5, pp. 32–51, 2014.

55. Peng, Y., V. Krungleviciute, I. Eryazici, J.T. Hupp, O.K. Farha, T. Yildirim, "Methane Storage in Metal − Organic Frameworks: Current Records, Surprise Findings, and Challenges", *J. Am. Chem. Soc.*, Vol. 135, pp. 11887–11894, 2013.

56. Cushing, B.L., V.L. Kolesnichenko, C.J. O'Connor, "Recent Advances in the Liquid-Phase Syntheses of Inorganic Nanoparticles", *Chem. Rev.*, Vol. 104, pp. 3893–3946, 2004.

57. Feng, S.-H., G.-H. Li, "Hydrothermal and Solvothermal Syntheses", *Mod. Inorg. Synth. Chem*., pp. 73–104, 2017.

58. Rao, B.G., D. Mukherjee, B.M. Reddy, "Novel Approaches for Preparation of Nanoparticles", *Nanostructures Nov. Ther.*, pp. 1–36, 2017.

59. Sabouni, R., H. Kazemian, S. Rohani, "A Novel Combined Manufacturing Technique for Rapid Production of IRMOF-1 Using Ultrasound and Microwave Energies", *Chem. Eng. J.*, Vol. 165, pp. 966–973, 2010.

60. Lee, Y.R., J. Kim, W.S. Ahn, "Synthesis of Metal-Organic Frameworks: A Mini

Review", *Korean J. Chem. Eng.*, Vol. 30, pp. 1667–1680, 2013.

61. Choi, J.S., W.J. Son, J. Kim, W.S. Ahn, "Metal-Organic Framework MOF-5 Prepared by Microwave Heating: Factors to be Considered", *Microporous Mesoporous Mater*., Vol. 116, pp. 727–731, 2008.

62. Sun, Y., H.C. Zhou, "Recent Progress in The Synthesis of Metal-Organic Frameworks", *Sci. Technol. Adv. Mater.*, Vol. 16, pp. 1-11 ,2015.

63. Pirzadeh, K., A.A. Ghoreyshi, M. Rahimnejad, M. Mohammadi, "Electrochemical Synthesis, Characterization and Application of A Microstructure $Cu_3(BTC)_2$ Metal Organic Framework For CO_2 And CH_4 Separation", *Korean J. Chem. Eng.*, Vol. 35, pp. 974–983, 2018.

64. Haque, E., S.H. Jhung, "Synthesis of Isostructural Metal-Organic Frameworks, CPO-27s, with Ultrasound, Microwave, and Conventional Heating: Effect of Synthesis Methods and Metal Ions", *Chem. Eng. J.*, Vol. 173, pp. 866–872, 2011.

65. Loera-Serna, S., M.A. Oliver-Tolentino, M. De Lourdes López-Núñez, A. Santana-Cruz, A. Guzmán-Vargas, R. Cabrera-Sierra, H.I. Beltrán, J. Flores, "Electrochemical Behavior of $[Cu_3(BTC)_2]$ Metal-Organic Framework: The Effect of The Method of Synthesis", *J. Alloys Compd.*, Vol. 540, pp. 113–120, 2012.

66. Pichon, A., S.L. James, "An Array-Based Study of Reactivity Under Solvent-Free Mechanochemical Conditions—Insights And Trends", *CrystEngComm*., Vol. 10, pp. 1839-1847, 2008.

67. Schlesinger, M., S. Schulze, M. Hietschold, M. Mehring, "Evaluation of Synthetic Methods for Microporous Metal-Organic Frameworks Exemplified by The Competitive Formation of $[Cu_2(Btc)_3(H_2O)_3]$ and $[Cu_2(Btc)(OH)(H_2O)]$", *Microporous Mesoporous Mater*., Vol. 132, pp. 121–127, 2010.

68. Wu, X., Z. Bao, B. Yuan, J. Wang, Y. Sun, H. Luo, S. Deng, "Microwave Synthesis

and Characterization of MOF-74 (M=Ni, Mg) for Gas Separation", *Microporous Mesoporous Mater*., Vol. 180, pp. 114–122, 2013.

69. Witten, I.H., E. Frank, M.A. Hall, C.J. Pal, *Data Mining: Practical Machine Learning Tools and Techniqu*es, Fourth Edition, Elsevier, Cambridge, 2016.

70. Ali, S.M., R.R. Tuteja, "Data Mining Techniques", International Journal of Computer Science and Mobile Computing, Vol. 3 (2014) 879-883.

71. Lahiri, S.K., K.C. Ghanta, "Development of An Artificial Neural Network Correlation for Prediction of Hold-Up of Slurry Transport in Pipelines", *Chem. Eng. Sci*., Vol. 63, pp. 1497–1509, 2008.

72. Ballabio, C., "Spatial Prediction of Soil Properties in Temperate Mountain Regions Using Support Vector Regression", *Geoderma*, Vol. 151, pp. 338–350, 2009.

73. Sezginel, K.B., *Computational and Experimental Investigation of CH_4 Adsorption in Pure and Ionic Liquid Incorporated Metal Organic Frameworks,* Master of Science Thesis, Koç University, 2015.

74. Allen, F.H., "The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising", *Acta Crystallogr. Sect. B Struct. Sci*., Vol. 58, pp. 380–388, 2002.

75. R. Soc. Chem., *Periodic Table*, 2017, http://www.rsc.org/periodic-table, accessed at January 2019.

76. Birkbeck Coll. Univ. London, *High-Resolution Space Group Diagrams and Tables*, 1999, http://img.chem.ucl.ac.uk/sgp/large/sgp.htm, accessed at January 2019.

77. Mukaka, M.M., "A Guide To Appropriate Use of Correlation Coefficient in Medical Research", *Malawi Med. J.*, Vol. 24, pp. 69–71, 2012..

78. Song, Y.-Y., Y. Lu, "Decision Tree Methods: Applications for Classification and Prediction", *Shanghai Arch. Psychiatry*, Vol. 27, pp. 130–135, 2015.

79. James, G., D. Witten, T. Hastie, T. Robert, *An Introduction to Statistical Learning*, Springer, New York, 2013.

80. Kingsford, C., S.L. Salzberg, "What are Decision Trees?", *Nat. Biotechnol.*, Vol. 26, pp. 1011–1013, 2008.

81. Dietterich, T.G., "Machine Learning in Ecosystem Informatics and Sustainability" In: Artificial Intelligence, *Twenty-First International Joint Conference*, 2009.

82. Zhang, Q.B., R.J. Hebda, Q.J. Zhang, R.I. Alfaro, "Modeling Tree-Ring Growth Responses to Climatic Variables Using Artificial Neural Networks", *For. Sci.*, Vol. 46, pp. 229–239, 2000.

83. Gunay, M.E., R. Yildirim, "Neural network Analysis of Selective CO Oxidation over Copper-Based Catalysts for Knowledge Extraction from Published Data in the Literature", *Ind. Eng. Chem. Res.*, Vol. 50, pp. 12488–12500, 2011.

84. Khajeh-Hosseini-Dalasm, N., S. Ahadian, K. Fushinobu, K. Okazaki, Y. Kawazoe, "Prediction and Analysis of The Cathode Catalyst Layer Performance of Proton Exchange Membrane Fuel Cells Using Artificial Neural Network And Statistical Methods", *J. Power Sources*, Vol. 196, pp. 3750–3756, 2011.

85. Eddaoudi, M., J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, O.M. Yaghi, "Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage Published by: American Association for the Advancement of Science Linked references are available on JSTOR for this article: Systematic Design", *Science*, Vol. 295, pp. 469–472, 2002.

86. Huang, J.R,. Li, K. Wang, T. Han, M. Tong, L. Li, Y. Xie, Q. Yang, D. Liu, C. Zhong, "An in Situ Self-Assembly Template Strategy for the Preparation of

Hierarchical-Pore Metal-Organic Frameworks", *Nat. Commun.*, Vol. 6, pp. 1–8, 2015.

87. Shenouda, E.A.M.A., "A Quantitative Comparison of Different MLP Activation Functions in Classification", In: J. Wang, Z. Yi, J.M. Zurada, B.-L. Lu, H. Yin (Eds.), *Adv. Neural Networks - ISNN 2006*, Berlin, Springer Berlin Heidelberg , 2006.

88. Karlik, B., A. V Olgac, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks", *Int. J. Artif. Intell. Expert Syst.*, Vol. 1, pp. 111–122, 2010.

89. Duch, W., N. Jankowski, "Survey of Neural Transfer Functions", *Neural Comput. Surv.*, Vol. 2, pp. 163–212, 1999.

90. Zell, A., G. Mamier, M. Vogt, N. Mache, R. Hübner, S. Döring, K.-U. Herrmann, T. Soyez, M. Schmalzl, T. Sommer, A. Hatzigeorgiou, D. Posselt, T. Schreiner, B. Kett, G. Clemente, J. Wieland, J. Gatter, "SNNSv4.2.Manual", pp. 1–350, 2016.

91. Gunay, M.E., F. Akpinar, Z.I. Onsan, R. Yildirim, "Investigation of Water Gas-Shift Activity of Pt-MO_x-CeO_2/Al_2O_3(M = K, Ni, Co) Using Modular Artificial Neural Networks", *Int. J. Hydrogen Energy*, Vol. 37, pp. 2094–2102, 2012.

92. Gunay, M.E., R. Yildirim, "Neural Network Aided Design of Pt-Co-Ce/Al_2O_3 Catalyst for Selective CO Oxidation in Hydrogen-Rich Streams", *Chem. Eng. J.*, Vol. 140, pp. 324–331, 2008.

93. Erucar, I., S. Keskin, “High-Throughput Molecular Simulations of Metal Organic Frameworks for CO_2 Separation: Opportunities and Challenges”, *Front. Mater.*, Vol. 5, pp. 1–6, 2018.

94. Borboudakis, G., T. Stergiannakos, M. Frysali, E. Klontzas, I. Tsamardinos, G.E. Froudakis, "Chemically Intuited, Large-Scale Screening of MOFs by Machine

Learning Techniques", *Npj Comput. Mater*., Vol. 3, pp. 1–6, 2017.

95. J. Yu, L.-H. Xie, J.-R. Li, Y. Ma, J.M. Seminario, P.B. Balbuena, "CO_2 Capture and Separations Using MOFs: Computational and Experimental Studies", *Chem. Rev.*, Vol. 117, pp. 9674–9754, 2017.

96. Lin, X., I. Telepeni, A.J. Blake, A. Dailly, C.M. Brown, J.M. Simmons, M. Zoppi, G.S. Walker, K.M. Thomas, T.J. Mays, P. Hubberstey, N.R. Champness, M. Schroder, "High Capacity Hydrogen Adsorption in Cu (II) Tetracarboxylate Framework Materials : The Role of Pore Size, Ligand Functionalization, and Exposed Metal Sites", *J. Am. Chem. Soc*., Vol. 131, pp. 2159–2171, 2009.

97. Sun, T., X. Ren, J. Hu, S. Wang, "Expanding Pore Size of Al-BDC Metal-Organic Frameworks as a Way to Achieve High Adsorption Selectivity for CO_2/CH_4 Separation", *J. Phys. Chem. C*., Vol. 118, pp. 15639–15639, 2014.

98. Rowsell, J.L.C., O.M. Yaghi, "Strategies for Hydrogen Storage in Metal-Organic Frameworks", *Angew. Chemie Int. Ed.*, Vol. 44, pp. 4670–4679, 2005.

99. Wu, H., W. Zhou, T. Yildirim, "High-Capacity Methane Storage in Metal - Organic Frameworks M_2 (dhtp): The Important Role of Open Metal Sites", *J. Am. Chem. Soc.*, Vol. 131, pp. 4995–5000, 2009.

100. Koh, H.S., M.K. Rana, A.G. Wong-Foy, D.J. Siegel, "Predicting Methane Storage in Open-Metal-Site Metal-Organic Frameworks", *J. Phys. Chem. C*., Vol. 119, pp. 13451–13458, 2015.

101. Purewal, J.J., D. Liu, J. Yang, A. Sudik, D.J. Siegel, S. Maurer, U. Müller, "Increased Volumetric Hydrogen Uptake of MOF-5 by Powder Densification", *Int. J. Hydrogen Energy*., Vol. 37, pp. 2723–2727, 2012.

102. Yuan, D., D. Zhao, D. Sun, H.C. Zhou, "An Isoreticular Series of Metal-Organic Frameworks with Dendritic Hexacarboxylate Ligands and Exceptionally High Gas-

Uptake Capacity", *Angew. Chemie Int. Ed.*, Vol. 49, pp. 5357–5361, 2010.

103. Sener, A.N., M.E. Gunay, A. Leba, R. Yildirim, "Statistical Review of Dry Reforming of Methane Literature Using Decision Tree and Artificial Neural Network Analysis", Catal. Today, Vol. 299, pp. 289–302, 2018.

104. Guo Z., H. Wu, G. Srinivas, Y. Zhou, S. Xiang, Z. Chen, Y. Yang, W. Zhou, M. O'Keeffe, B. Chen, "A Metal-Organic Framework with Optimized Open Metal Sites and Pore Spaces for High Methane Storage at Room Temperature", *Angew. Chemie - Int. Ed.*, Vol. 50, pp. 3178–3181, 2011.

105. Li B., Z. Zhang, Y. Li, K. Yao, Y. Zhu, Z. Deng, F. Yang, X. Zhou, G. Li, H. Wu, N. Nijem, Y.J. Chabal, Z. Lai, Y. Han, Z. Shi, S. Feng, J. Li, "Enhanced Binding Affinity, Remarkable Selectivity, and High Capacity of CO_2 by Dual Functionalization of a Rht-Type Metal-Organic Framework", *Angew. Chemie - Int. Ed.*, Vol. 51, pp. 1412–1415, 2012.

106. Wang X.S, M. Shengqian, K. Rauch, J.M. Simmons, D. Yuan, X. Wang, T. Yildirim, W.C. Cole, J.J. López, A. De Meijere, H.C. Zhou, "Metal-Organic Frameworks Based on Double-Bond-Coupled Di-Isophthalate Linkers with High Hydrogen and Methane Uptakes", *Chem. Mater.*, Vol. 20, pp. 3145–3152, 2008.

107. Zhao D., D. Yuan, A. Yakovenko, H.C. Zhou, "A Nbo-Type Metal-Organic Framework Derived from a Polyyne-Coupled Di-Isophthalate Linker Formed in Situ", *Chem. Commun.*, Vol. 46, pp. 4196–4198, 2010.

108. Lu W., D. Yuan, T.A. Makal, J.R. Li, H.C. Zhou, "A Highly Porous and Robust (3,3,4)-Connected Metal-Organic Framework Assembled with A 90° Bridging-Angle Embedded Octacarboxylate Ligand", *Angew. Chemie - Int. Ed.*, Vol. 51, pp. 1580–1584, 2012.

109. Wilmer C.E., O.K. Farha, T. Yildirim, I. Eryazici, V. Krungleviciute, A.A. Sarjeant, R.Q. Snurr, J.T. Hupp, "Gram-Scale, High-Yield Synthesis of a Robust Metal-

Organic Framework for Storing Methane and Other Gases", *Energy Environ. Sci.*, Vol. 6, pp. 1158–1163, 2013.

APPENDIX A: FULL VERSION OF DECISION TREES

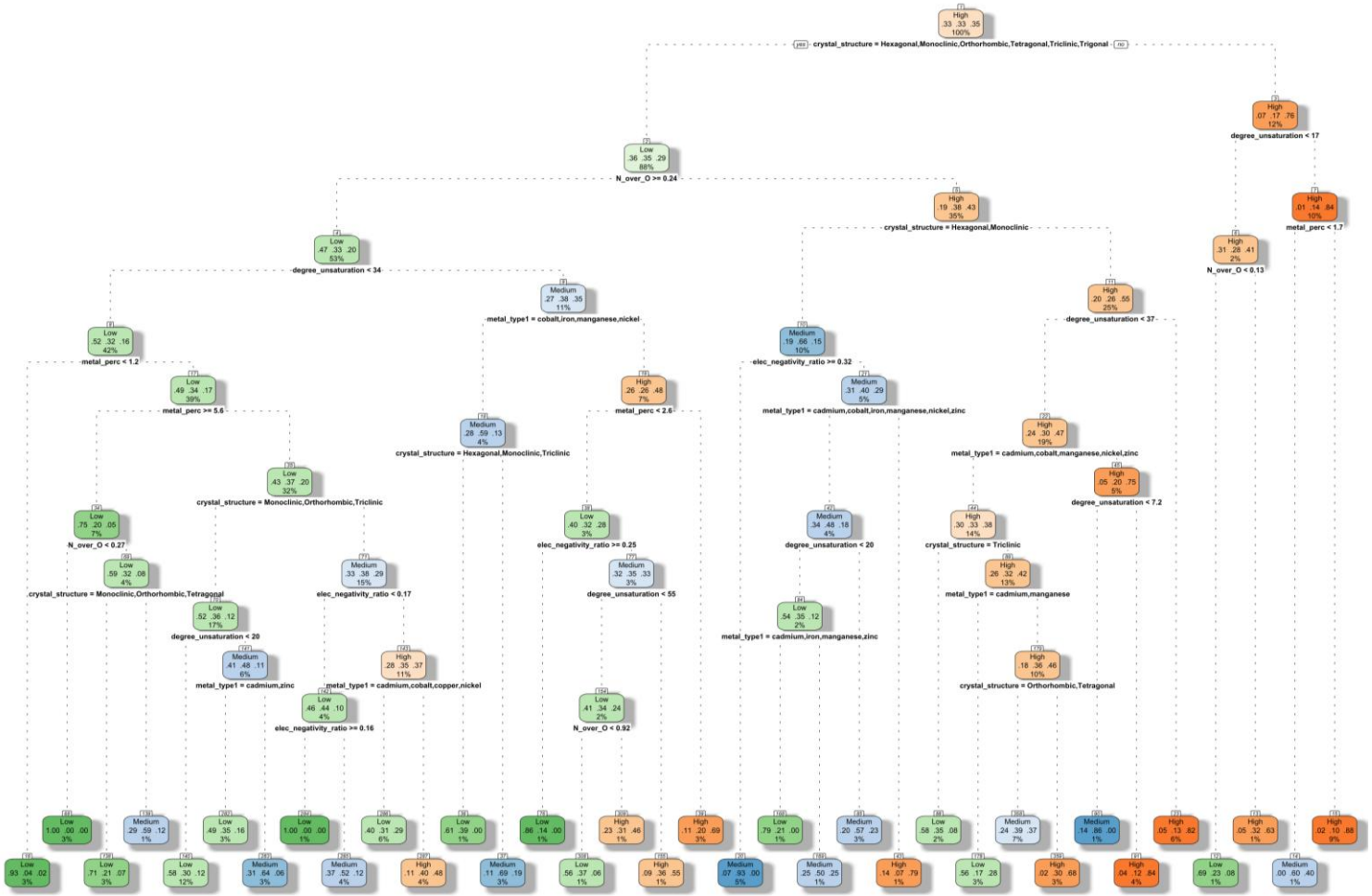

Figure A.1. Decision tree constructed with all user defined descriptors (including atomic radius and group number) for gravimetric CH_4 delivery.

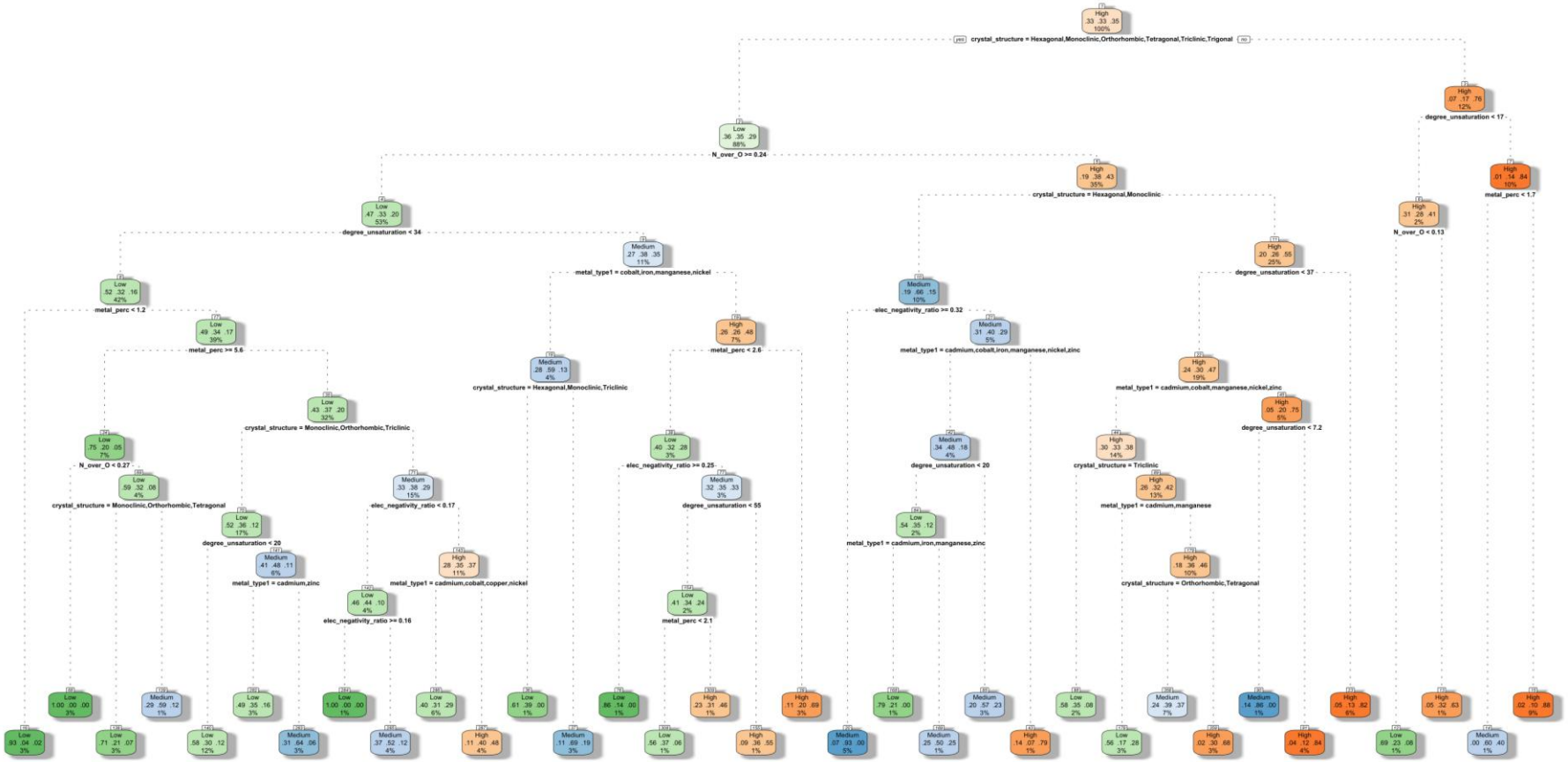

Figure A.2. Decision tree constructed by using user defined descriptors for gravimetric CH_4 delivery.

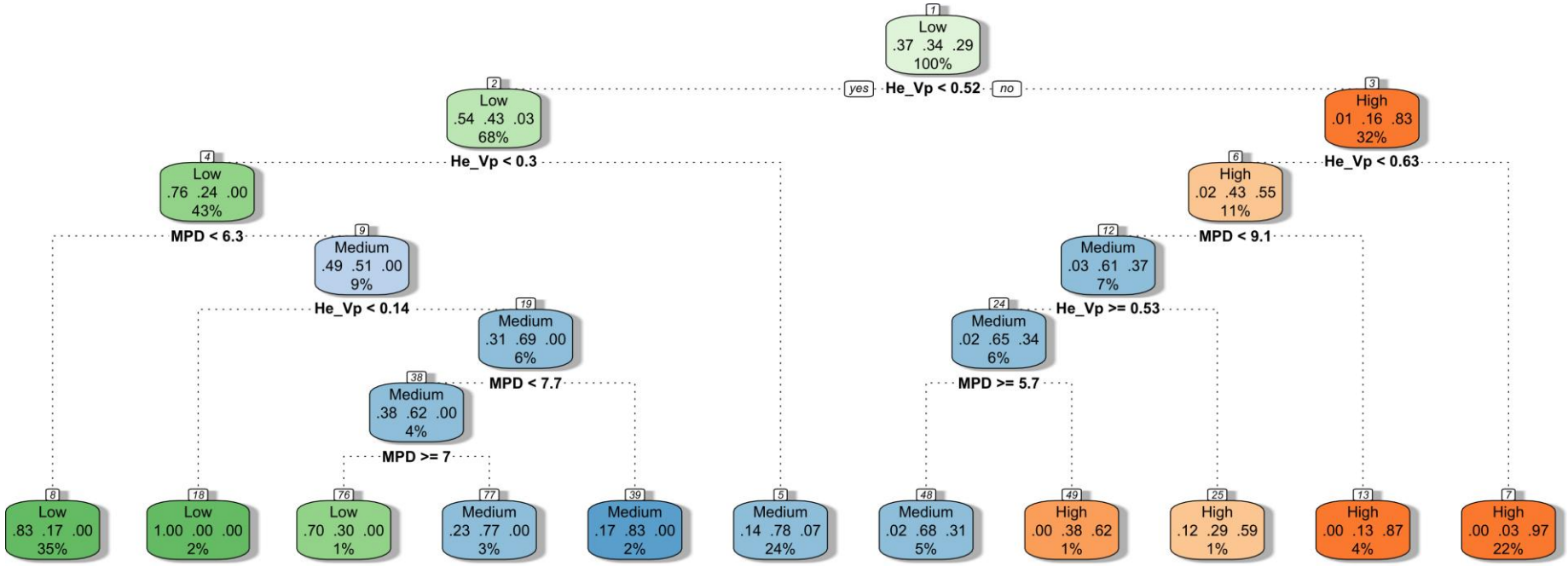

Figure A.3. Decision tree constructed by using structural properties gravimetric CH_4 delivery.

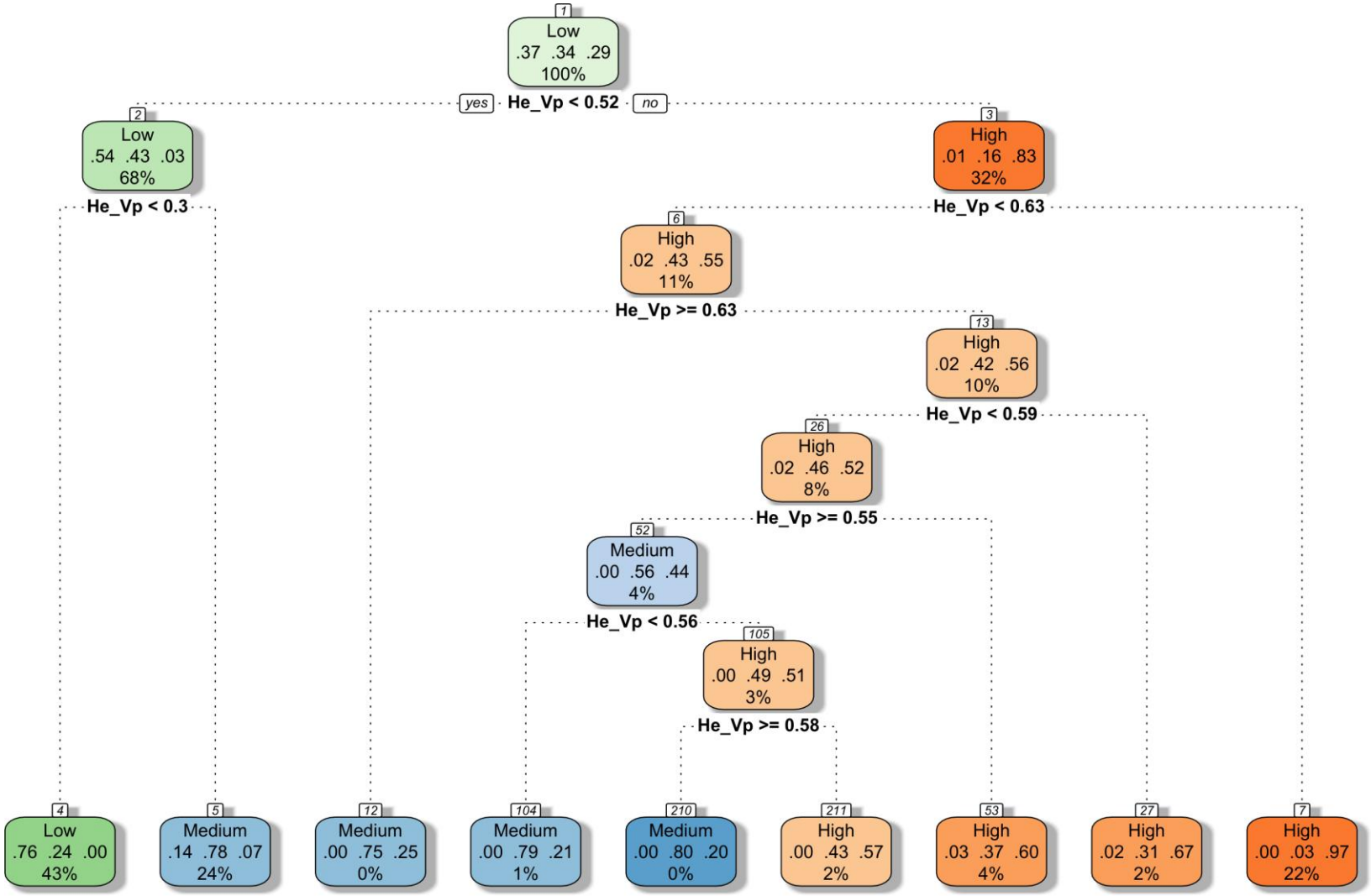

Figure A.4. Decision tree constructed by using pore volume only for gravimetric CH_4 delivery.

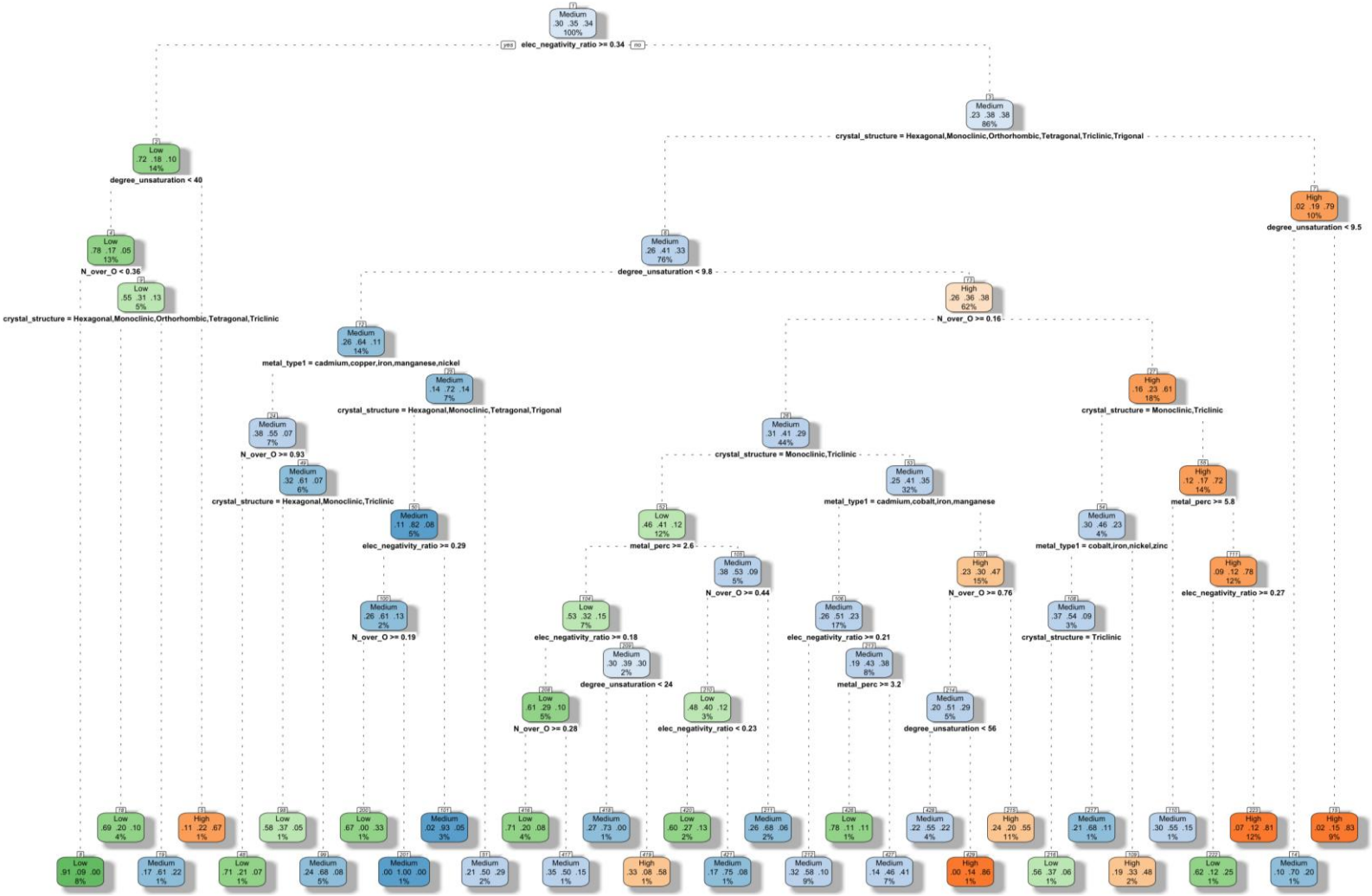

Figure A.5. Decision tree constructed by using user defined descriptors for gravimetric CH_4 uptake.

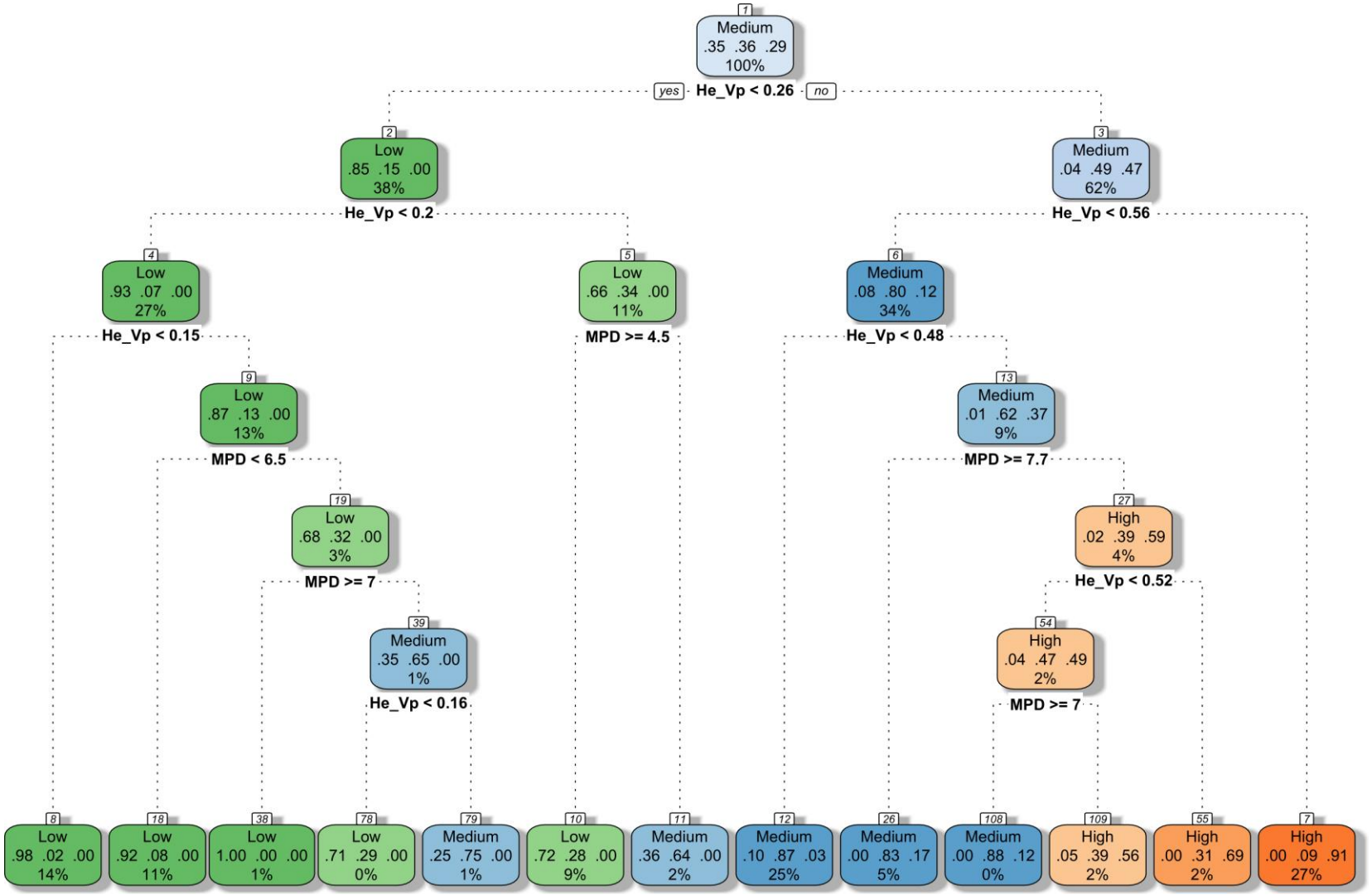

Figure A.6. Decision tree constructed by using structural properties gravimetric CH_4 uptake.

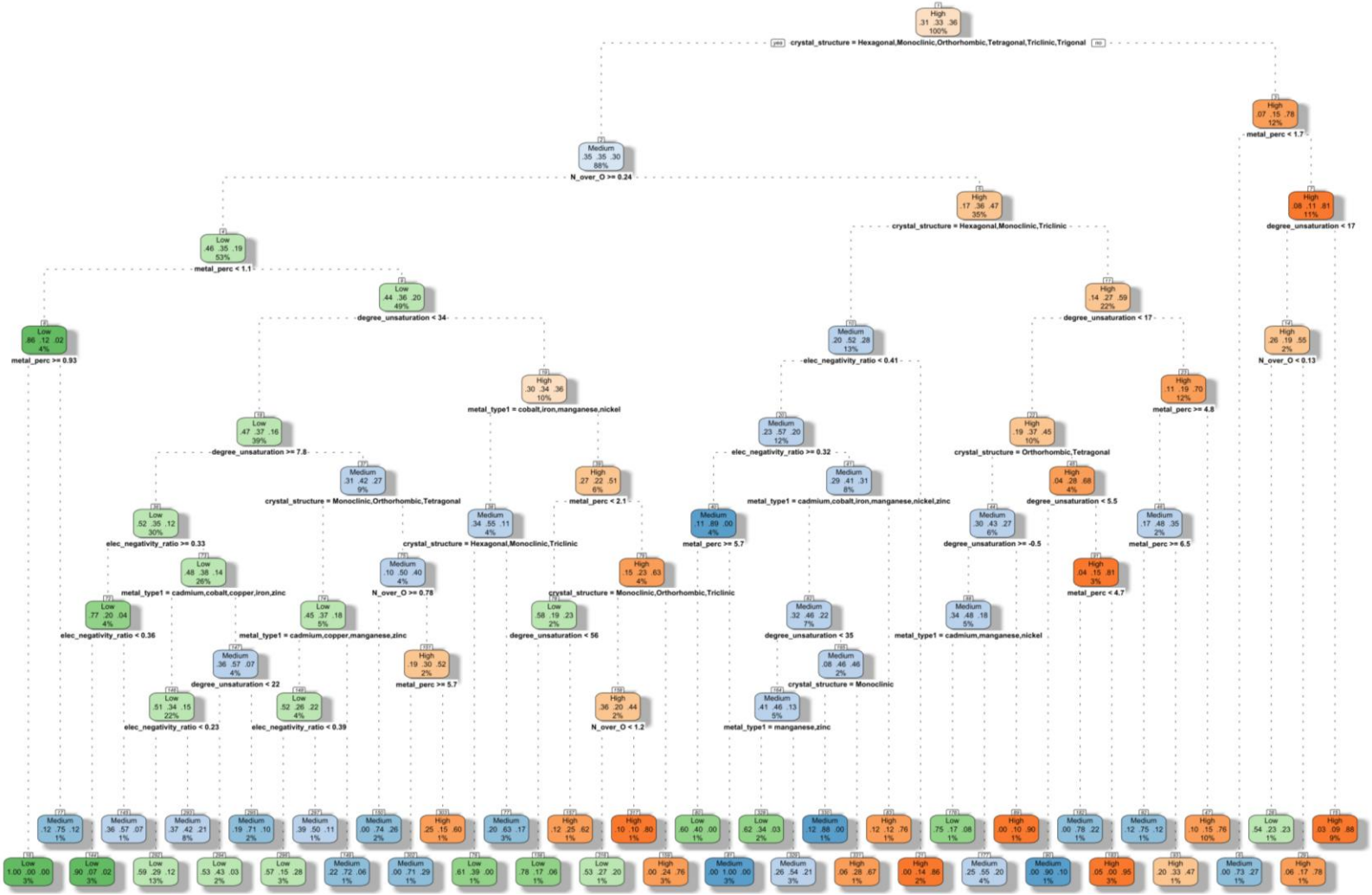

Figure A.7. Decision tree constructed by using user defined descriptors for volumetric CH_4 delivery.

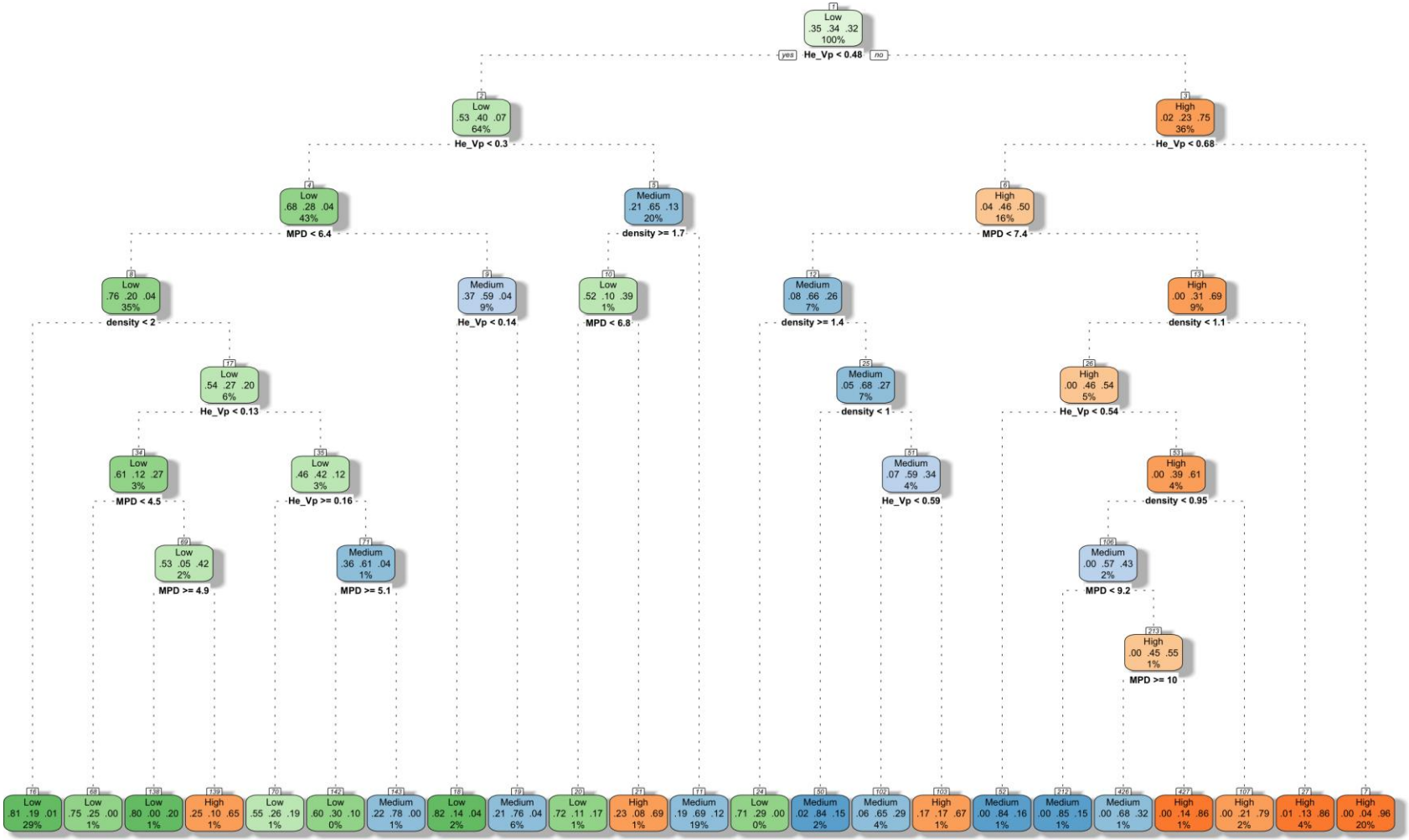

Figure A.8. Decision tree constructed by using structural properties volumetric CH_4 delivery.

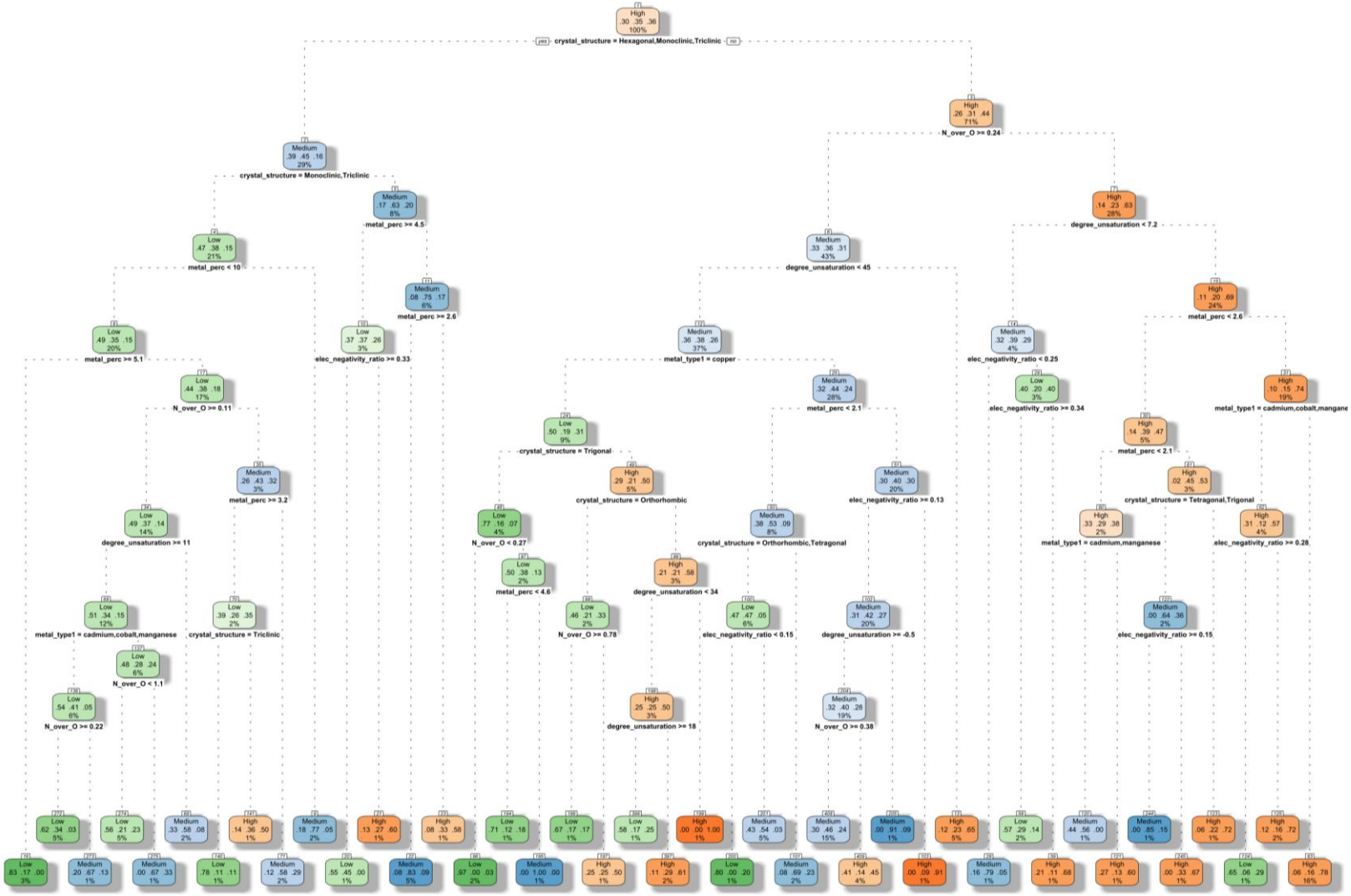

Figure A.9. Decision tree constructed by using user defined descriptors for volumetric CH_4 uptake.

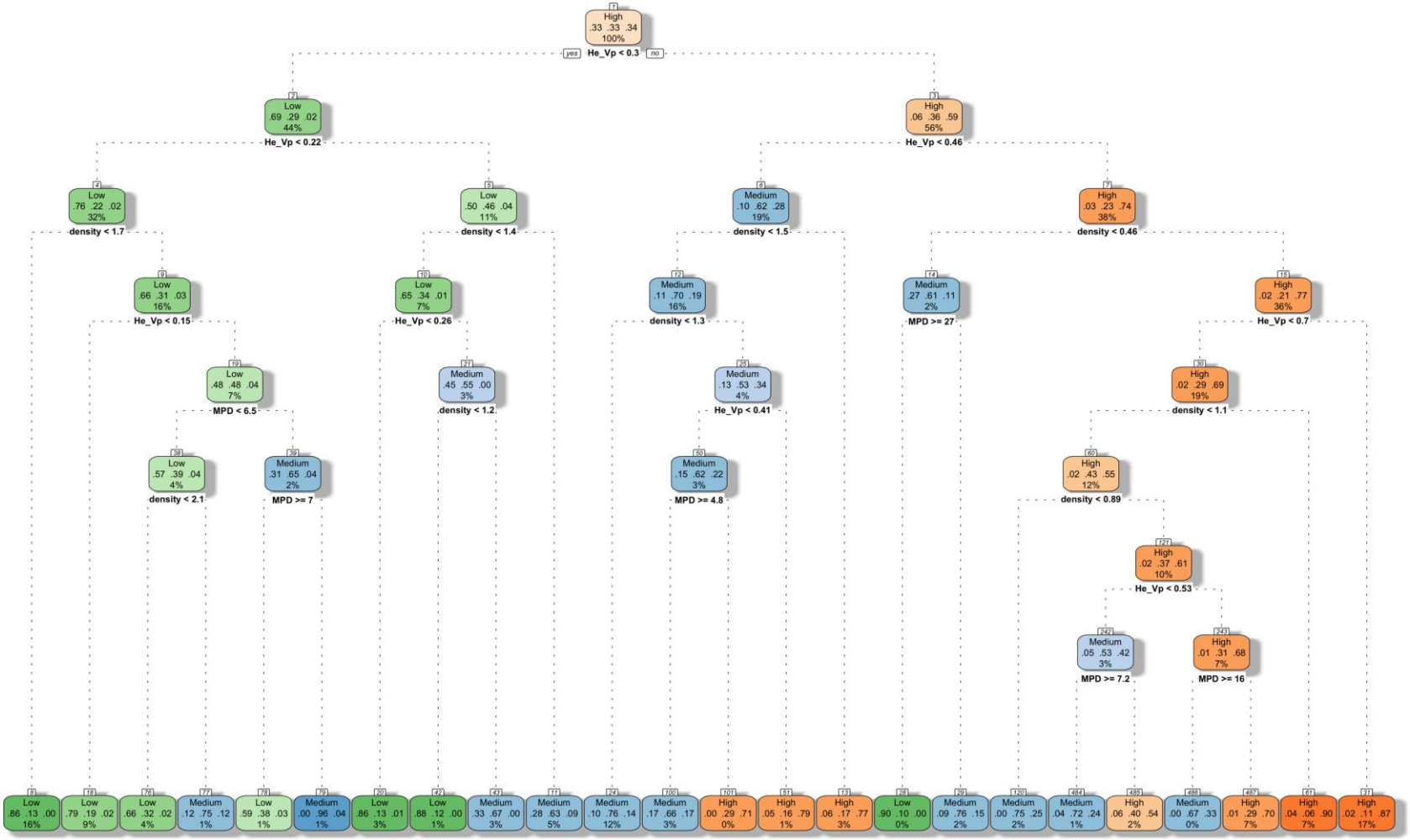

Figure A.10. Decision tree constructed by using structural properties volumetric CH_4 uptake.

APPENDIX B: ALTERNATIVE ANN MODELS

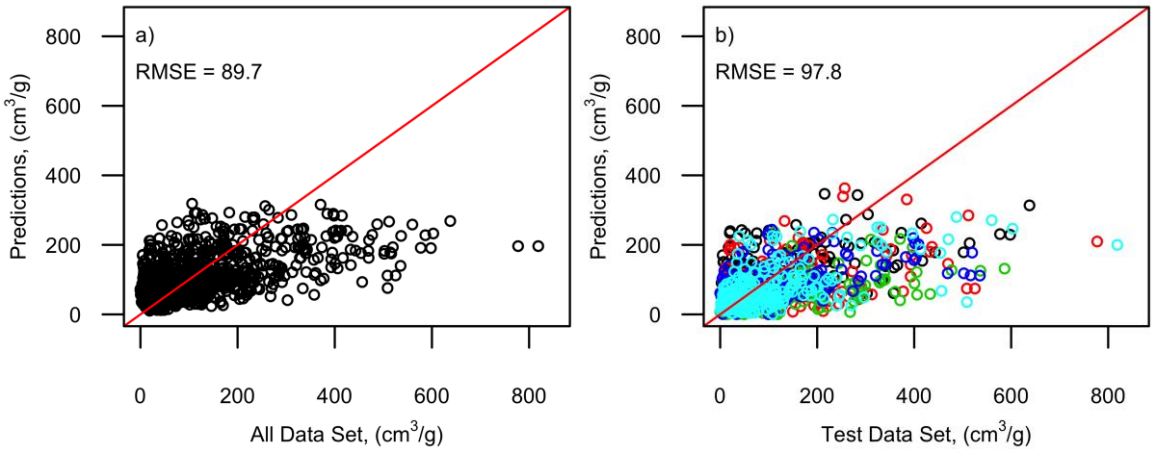

Figure B.1. Actual versus predicted gravimetric CH_4 delivery values from ANN model constructed with user defined descriptors (excluding electronegativity ratio) a) for training dataset and b) for testing dataset.

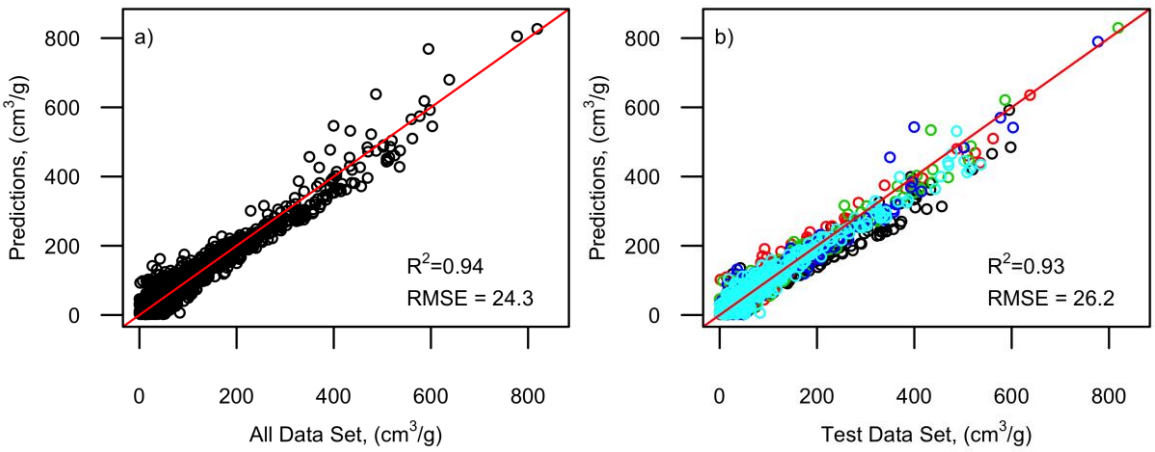

Figure B.2. Actual versus predicted gravimetric CH_4 delivery values from ANN model constructed with pore volume only a) for training dataset and b) for testing dataset.

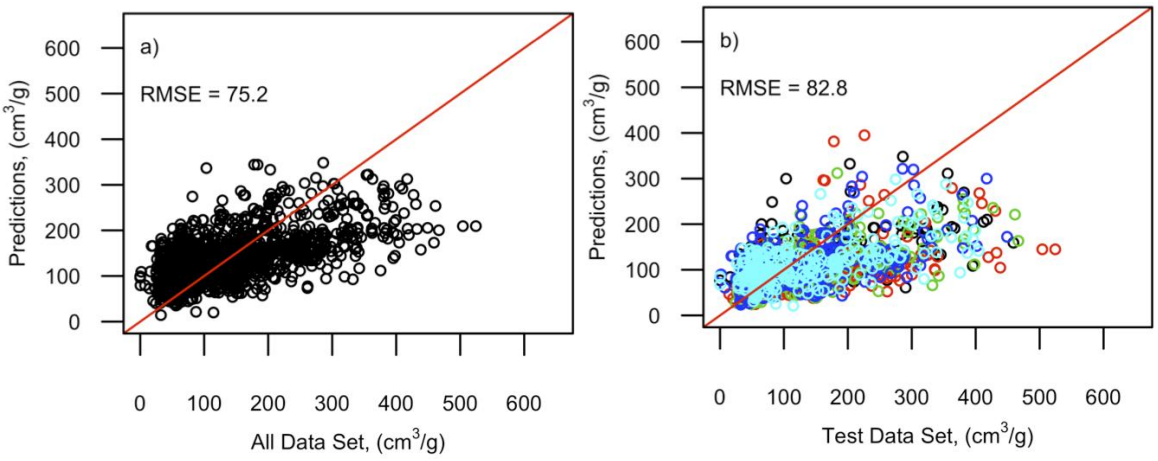

Figure B.3. Actual versus predicted gravimetric CH_4 uptake values from ANN model constructed with user defined descriptors (excluding electronegativity ratio) a) for training dataset and b) for testing dataset.

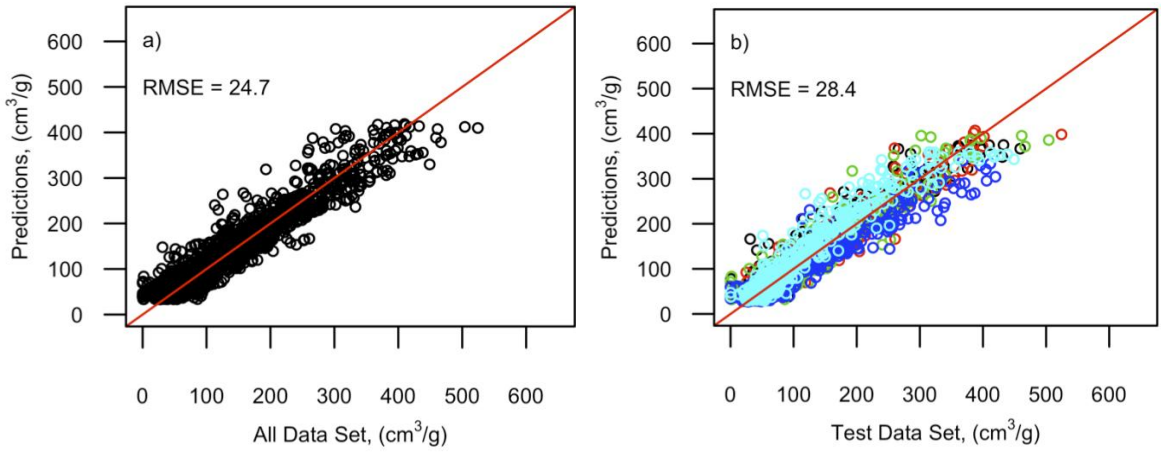

Figure B.4. Actual versus predicted gravimetric CH_4 uptake values from ANN model constructed with pore volume only a) for training dataset and b) for testing dataset.

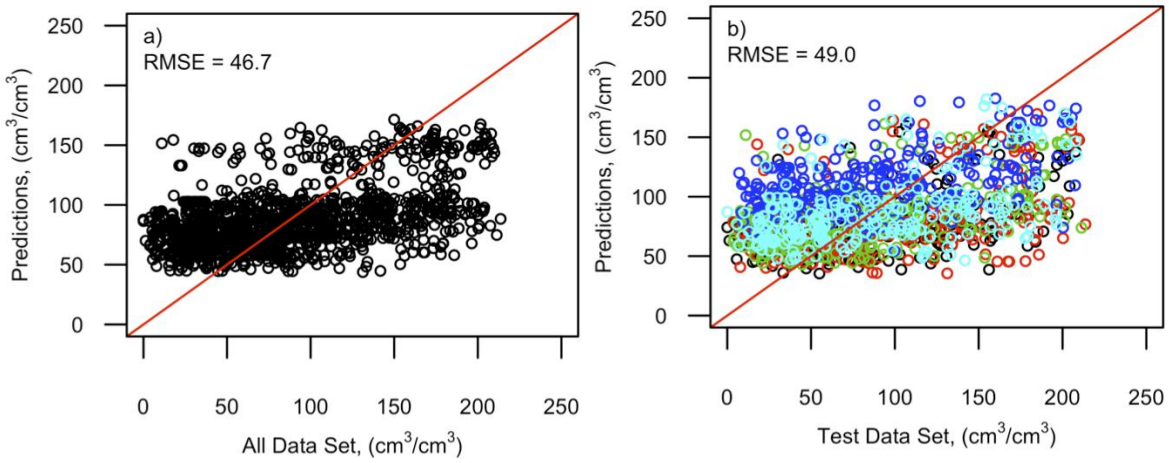

Figure B.5. Actual versus predicted volumetric CH_4 delivery values from ANN model constructed with user defined descriptors (excluding electronegativity ratio) a) for training dataset and b) for testing dataset.

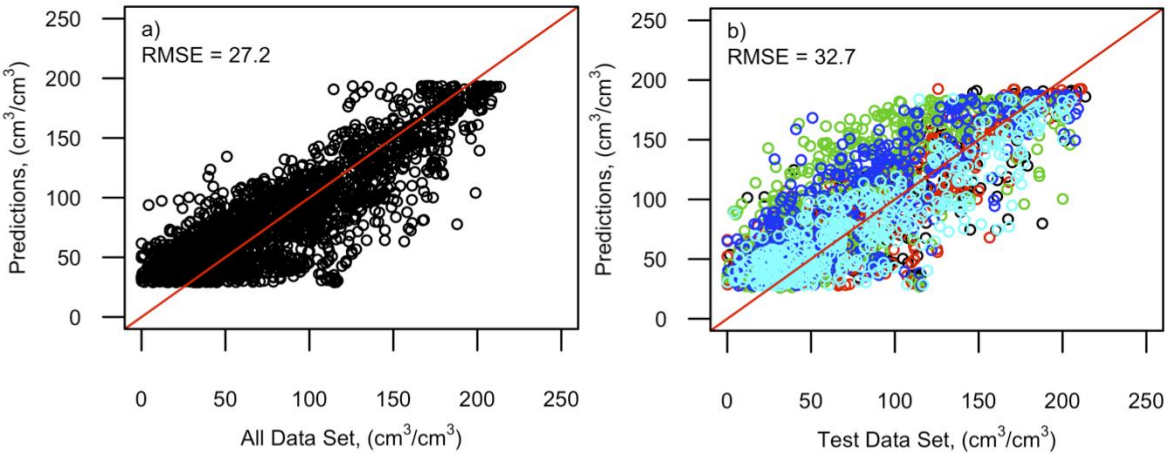

Figure B.6. Actual versus predicted volumetric CH_4 delivery values from ANN model constructed with pore volume only a) for training dataset and b) for testing dataset.

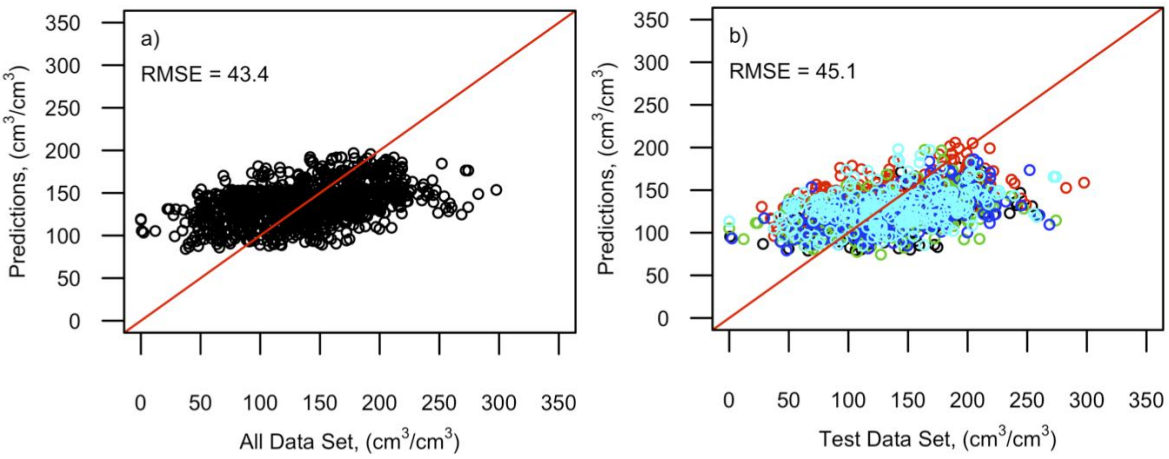

Figure B.7. Actual versus predicted volumetric CH_4 uptake values from ANN model constructed with user defined descriptors (excluding electronegativity ratio) a) for training dataset and b) for testing dataset.

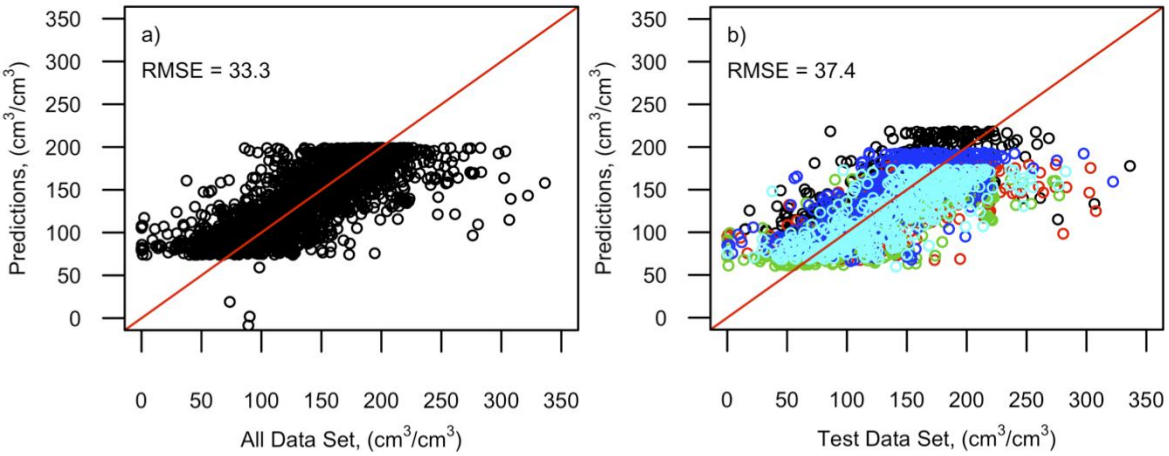

Figure B.8. Actual versus predicted volumetric CH_4 uptake values from ANN model constructed with pore volume only a) for training dataset and b) for testing dataset.