


Article

Design of Polymeric Membranes for Air Separation by Combining Machine Learning Tools with Computer Aided Molecular Design

Jie-Ying Cheun¹, Joshua-Yeh-Loong Liew¹, Qian-Ying Tan¹, Jia-Wen Chong¹, Jecksin Ooi² and Nishanth G. Chemmangattualappil^{1,*} 

¹ Department of Chemical & Environmental Engineering, University of Nottingham Malaysia, Jalan Broga, Semenyih 43500, Malaysia; ebyjc5@nottingham.edu.my (J.-Y.C.); ebyjl4@nottingham.edu.my (J.-Y.-L.L.); ebyqt1@nottingham.edu.my (Q.-Y.T.); jiawen960218@gmail.com (J.-W.C.)

² School of Engineering and Physical Sciences, Heriot-Watt University Malaysia, No. 1, Jalan Venna P5/2, Precinct 5, Putrajaya 62200, Malaysia; j.ooi@hw.ac.uk

* Correspondence: nishanth.c@nottingham.edu.my

Abstract: The growing importance of the membrane-based air separation processes results in an increasing demand for suitable polymeric membrane structures. This has spurred the interest in designing polymer structures for O₂/N₂ separation by employing a systematic approach. In this work, a computer-aided molecular design (CAMD)-based framework was developed to identify promising structures of polymers that can be used for air separation. To incorporate constraints in CAMD, the rough set-based machine learning (RSML) method was implemented to establish predictive models for the physical and transport properties of polymer owing to its interpretability. The deterministic rules generated from RSML would be interpreted scientifically reflecting the structure–property relationship to ensure that the molecules generated were feasible according to a scientific point of view. The most prominent rules selected were then integrated as constraints in CAMD. The relevant properties in this framework comprised of glass transition temperature (T_g), molar volume (V_m), cohesive energy (E_{coh}), O₂ permeability and O₂/N₂ selectivity. The solutions from CAMD optimisation were demonstrated in case studies. Results indicated the capability of a novel approach in identifying potential polymeric membrane candidates for air separation application that meet the permeability and selectivity requirements.

Keywords: polymer membrane; air separation; topological indices; rough set-based machine learning; computer-aided molecular design



Citation: Cheun, J.-Y.; Liew, J.-Y.-L.; Tan, Q.-Y.; Chong, J.-W.; Ooi, J.; Chemmangattualappil, N.G. Design of Polymeric Membranes for Air Separation by Combining Machine Learning Tools with Computer Aided Molecular Design. *Processes* **2023**, *11*, 2004. <https://doi.org/10.3390/pr11072004>

Academic Editor: Xiong Luo

Received: 18 May 2023

Revised: 29 June 2023

Accepted: 30 June 2023

Published: 4 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Polymeric membrane separation has been transitioning from a laboratory curiosity to a commercial reality for the separation of common gases, which is gaining popularity over the commercial process accomplished by adsorption, cryogenic distillation and amine absorption. The global gas separation membrane market size has expanded from USD 1.88 billion in 2022 to USD 2.09 billion in 2023 at a compound annual growth rate (CAGR) of 11.3% [1]. The market size is forecasted to further grow to USD 3.01 billion in 2027 with a CAGR of 9.5% [2]. This is because the membrane offers simplicity in operation, lower energy costs, a smaller footprint and viable economics as compared to distillation and adsorption; thus, it is extensively used in petrochemical industries, ammonia plants, natural gas processing units and air separation fields [2].

One of the main focuses for membrane-based application is the air separation process which is of great significance to the chemical industry to produce enriched oxygen and nitrogen. Air separation by membranes makes up approximately USD 155 million of the overall membrane gas separation business [3]. Membrane performance greatly depends on

its properties including tensile strength, selectivity and permeability. Therefore, efforts and developments are made to synthesise novel polymers with better separation properties and to understand the optimum properties for a polymeric membrane that functions to separate oxygen and nitrogen in the air. Ideally, commercially viable membranes shall possess superior selectivity and permeability in addition to mechanical stability for long operating life. However, there are still yet feasible membranes with such features to be applied in large-scale industrial air separation processes [4].

The design and screening of suitable polymeric membrane material is a resource-intensive process and depends heavily on the available polymer property databases in the market. Moreover, there are limited complete databases reported on structural properties, membrane permeability and selectivity towards O₂/N₂. Traditional methods of developing a new polymeric membrane structure consume a vast number of chemical compound assessments prior to the final selection, which increases the possibility of overlooking potential polymer structure. To resolve such challenges, a reverse formulation approach known as computer-aided molecular design (CAMD) was employed in this study. It is an effective tool for the design and screening of molecules by forecasting the molecular structure using a set of chemical, physical and structural properties. The pre-requisite for initiation of CAMD modelling is to develop property predictive models including topological, structural, physiochemical and electronic descriptors which are closely related to polymer molecules. Prediction models were therefore developed using machine learning algorithms to relate polymer properties, represented by topological indices with their molecular structures.

1.1. Rough Set Machine Learning (RSML)

Machine Learning (ML) is a discipline of artificial intelligence (AI) focusing on the use of computational algorithms that are designed to emulate human intelligence through exploring patterns in a series of data for future prediction [5]. Nevertheless, most of the modelling which employs ANN, support vector machine (SVM) and Gaussian process regression (GPR) methods require large datasets. In the case where these modelling methods are used even if lack of large data amount, it will lead to unreliable ML algorithms for designing polymers [6]. Furthermore, both ANN and SVM algorithms are black-box treatments by nature, making the learning of principles behind the models challenging. It is necessary to interpret the influence of a specific property on the prediction approach for a plausible gas permeability prediction [7]. Among the ML techniques, rough set ML (RSML) is reported to have the benefit of interpretability to support decisions based on scientific reasoning [8]. Therefore, RSML introduced by Pawlak in the early 1980s, was chosen to be applied in this study owing to its ability to deal with vagueness, inconsistency and uncertainty [9]. The main advantage of rough set theory is that none of the preliminary information about the data is required such as statistical probability distribution, grade of membership or fuzzy set possibility value.

Rough set theory is proven to be useful in real-life applications and it is applicable for the case when the data are limited. It has compelling applications in multiple fields such as medicine, engineering design, business, pharmacology, decision analysis, banking and others. It works in the form of rules which are induced by learning from training examples [9]. RSML produces a set of decision rules from the specified attributes covering all the training data examples with the derived certainty, strength and coverage. RSML has been implemented to determine secure geological reservoirs for the minimisation of CO₂ emissions by analysing data from storage sites. Predictive models generated from RSML showed similar results with site selection rules which were established according to proficient knowledge [10]. One of the recent works employed RSML for the prediction of energy consumption within a building, where it was used to eliminate redundant influencing factors prior to the identification of crucial component contributing to the energy consumption [11]. RSML was also used to develop models to predict the fragrance of molecules based on their chemical structure [12]. In another contribution, RSML has

been successfully used to identify the optimal operating conditions to produce bio-oil from biomass through fast pyrolysis [13].

Generally, RSML is useful to establish the concealed order in a dataset for the generation of decision algorithms, classification of data and discovering cause–effect relationships of the attributes and decision rules [9]. The concept of indiscernibility relation, which is correlated with a set of attributes, applies to RSML. The variables will be classified into conditional attributes (inputs) and decision attributes or objects (outputs). Suppose U is a finite set of objects, also called the universe, and A is a finite set of attributes. For each attribute $a \in A$, the value set is associated with V_a [14]. Every attribute a establishes a function as shown in Equation (1). With each subset of attributes B of A , the indiscernibility relation on U is defined and denoted as $I(B)$ in Equation (2).

$$f_a: U \rightarrow V_a \tag{1}$$

$$I(B) = \{(x, y) \in U \times U : f_a(x) = f_a(y), \forall a \in B\} \tag{2}$$

RSML can also be used for approximation concepts in the case of any vague or indefinite concept. It is categorised into lower and upper approximations where the former comprises all objects that surely belong to the concept while the latter contains all objects that are less certain to belong to the concept. Approximation ideas are presented in Equations (3) and (4), where U is the universe, X is a subset or concept of the universe and B is a subset of A .

$$B_*(X) = \{x \in U : B(x) \subseteq X\} \tag{3}$$

$$B^*(X) = \{x \in U : B(x) \cap X \neq \emptyset\} \tag{4}$$

where $B_*(X)$ is B-lower approximation and $B^*(X)$ is B-upper approximation of X .

For *B-boundary region* of X - $BN_B(X)$, referring to the difference between upper and lower approximations, it is defined in Equation (5).

$$BN_B(X) = B^*(X) - B_*(X) \tag{5}$$

The accuracy of the approximation is characterised numerically by coefficient α_B , in Equation (6).

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|} \tag{6}$$

Rough membership is another concept in RSML that considers the uncertainty of the elements in the universe, i.e., a vague concept with boundary line situations. Therefore, the uncertainty is coupled with the membership of a set of elements to form a rough membership function as described in Equations (7) and (8). $\mu_x(x)$ is the membership function depicted as a conditional probability, also interpreted as the degree of uncertainty to which x belongs to X [14].

$$\mu_x^B(x) = \frac{|X \cap B(x)|}{|B(x)|} \tag{7}$$

$$\mu_x^B(x) \in [0, 1] \tag{8}$$

In terms of approximation cases, rough membership function is shown in Equations (9)–(11). It is evident from the equations that a strict relation is present between uncertainty and vagueness in RSML. Vagueness is thus associated with sets while uncertainty is associated with the elements of sets.

$$B_*(X) = \{x \in U : \mu_x^B(x) = 1\} \tag{9}$$

$$B^*(X) = \{x \in U : \mu_x^B(x) > 0\} \tag{10}$$

$$BN_B(X) = \{x \in U : 0 < \mu_x^B(x) < 1\} \quad (11)$$

Reduct generation is introduced in RSML, which means a minimal subset of attributes are generated that allows same categorisation of elements in the universe as the entire set of attributes and preserves the indiscernibility relation in the system. Another significant property is the core, which represents the key attribute of all reducts, i.e., it is the intersection of every reduct and will be included in all reducts. None of the core elements can be removed without influencing the attributes classification [14]. Since there might be more than one reduct generated in a single dataset, further evaluation shall be conducted to select the appropriate reducts that fulfil certain requirements.

To further analyse the decisions and rules generated by RSML, certainty (cer_x), strength (σ_x) and coverage (cov_x) are computed. Certainty, which is equivalent to precision, means the frequency of objects having a decision, D fulfilling the conditions, C in a set of elements. High certainty deduces a more confident chance for a molecule or element to be classed in the right decision class. Strength indicates the percentage of the total number of data classified under certain rules out of the entire dataset. Coverage, which is also known as recall, defines the objects' frequency having conditions, C in the decision class, D . Both are evaluated as the generalisation power of a rule. The overall accuracy of the models, which is the number of correct predictions over the total number of predictions, was also estimated. These parameters will be implemented in this work to select the most eminent rules quantitatively [15]. Ideally, decision rules with both high certainty and coverage are desired to be utilised as predictive models [15].

1.2. Computer-Aided Molecular Design (CAMD)

CAMD is a reverse engineering methodology that is capable of predicting, estimating and designing molecules that match a predetermined set of target properties [16]. In CAMD, predetermined target properties and feasibility conditions are expressed as constraints in the problem formulation. An optimization model is then used to generate the best possible molecular structure, based on an objective function that seeks to either maximise or minimise some of the desired properties [17]. Usually, most CAMD approaches rely on property prediction models, such as group contribution (GC) methods to predict a molecule's properties, which are then used to assess how well the molecule meets the desired set of properties [18]. Other property prediction models employed in CAMD involve using topological indices (TIs) as molecular descriptors that rely on quantitative structure–property relationship (QSPR) for property estimation [19].

Numerous systematic frameworks, procedures, and algorithms based on the CAMD approach have been developed extensively in recent years. For instance, Wang and Cheng introduced a CAMD framework to identify a suitable bio-compatible solvent for the extractive fermentation and separation process [20]. This CAMD problem was formulated as a multi-objective optimisation (MOO) problem whereby the goal was to simultaneously optimise several targets, which included maximising the production rate, extraction efficiency, and limiting solvent usage. A MOO approach using CAMD had been used to identify effective solvents to extract palm oil from palm-pressed fibre [21]. The method optimized multiple targets such as performance targets, safety, health and environmental objectives using Fuzzy Analytic Hierarchy Process. A COSMO-CAMD framework was developed [22] to design solvents for liquid–liquid extraction processes of phenol and hydroxymethylfurfural from water. This framework includes quantum mechanical information incorporated into CAMD, which predicts properties more accurately and independently of experimental data.

Aside from the vast application of CAMD in solvent design, it is also widely applied in various chemical product design applications. Yee et al. created a systematic framework for designing personal care products that integrate safety, health, and performance considerations into the CAMD formulation [23]. By placing limitations on safety and health risks during CAMD, they were able to generate molecules that were less toxic while still

exhibiting outstanding product performance. CAMD approaches have also been used in the design of polymeric membranes based on group contribution methods [24] and molecular dynamics [25]. In these approaches, a set of desirable properties of the polymer has been targeted and the polymeric structures that meet those properties have been generated. There were a few recent studies in CAMD that focused on the development of fragrance products. For instance, a recent methodology involved the use of a series of MINLP models for screening and designing fragrances in shampoo using CAMD, which eliminated molecules that did not meet specific constraints and fragrance design properties [26]. In another study, an enhanced hyperbox ML approach was integrated with CAMD to generate rules that were used to create fragrance property prediction models [27]. Similarly, a CAMD framework was developed to facilitate the design of fragrance molecules using a rough set-based machine learning (RSML) model to generate constraints for the prediction of odour properties [28]. The hybrid CAMD-ML approach resulted in a diverse array of feasible compounds that met structural and physical property requirements. Both studies demonstrated the effectiveness of CAMD in generating potential fragrant molecules for consumer products. For more comprehensive information on the latest developments in this area, readers can refer to review articles by Chemmangattuvalappil et al. [29] and Zhang et al. [30].

One of the most important prerequisites for developing a polymeric membrane is the identification of structures that possess the desirable attributes needed to be used for this application. Although the CAMD-based polymer design methodology can effectively determine the polymer structures with desirable properties, there was a need to develop predictive modes for various polymer-based products to cater for wider applications [24]. Rough set machine learning (RSML) was employed to develop reliable predictive models for polymer properties, comprising of topological indices in this work. Since there are no existing models relating TIs with the O₂/N₂ membrane separation characteristics, they are used as structural descriptors for property correlations in predicting polymer properties for air separation. The prominent rules will then be integrated into CAMD as property constraints after screening the rules generated from RSML. Classification models were established for physical properties including glass transition temperature (T_g), molar volume (V_m), cohesive energy (E_{coh}) and transport properties such as permeability and selectivity in view of their impacts towards polymeric membrane functionality and O₂/N₂ separation feasibility. The polymeric membrane design using CAMD was formulated to optimise the high tensile strength of the polymeric membrane molecule structure.

2. Methodology

Figure 1 illustrates the methodology developed to design air separation polymeric membrane molecules by employing RSML and CAMD tools. The entire work is separated into 4 main steps, starting with the identification of significant attributes impacting polymeric membranes, followed by establishment of property prediction models and implementation of CAMD model for the membrane design and lastly design model verification.

Step 1: Polymer attributes/properties identification

For air separation applications, the role of polymeric membranes is to ensure effective separation between oxygen and nitrogen. Technical requirements for the polymer attributes were separated into physical and transport properties. The essential physical properties to be fulfilled are the fundamental polymer properties to function at normal operating conditions which are glass transition temperature, molar volume and cohesive energies. Glass transition temperature (T_g) indicates the temperature region where polymer changes from rigid “glassy” state to flexible “rubbery” state which is undesired in this case since change in polymer physical state will affect polymer chain flexibility and separation efficiency. Glassy state membrane generally has higher permeabilities to gases as compared to rubbery polymer with higher permeabilities for organic solvents [31]. On the other hand, molar volume (V_m) is related to the fractional free volume of a polymeric membrane representing the free space and accessible volume in membrane model affecting transport behaviour of small gas molecules, i.e., membrane diffusivity–separation properties [32]. Cohesive

energy (E_{coh}) is the energy required to break all intermolecular physical links per mole of polymer which also indicates dispersion, polar and hydrogen bonding interactions [33]. For transport properties, permeability and selectivity are important factors in the selection of polymeric membrane structures. Permeability defines the speed at which the gas molecules transport across the membrane whilst selectivity indicates the separation degree of the target gas molecules from other molecules [32].

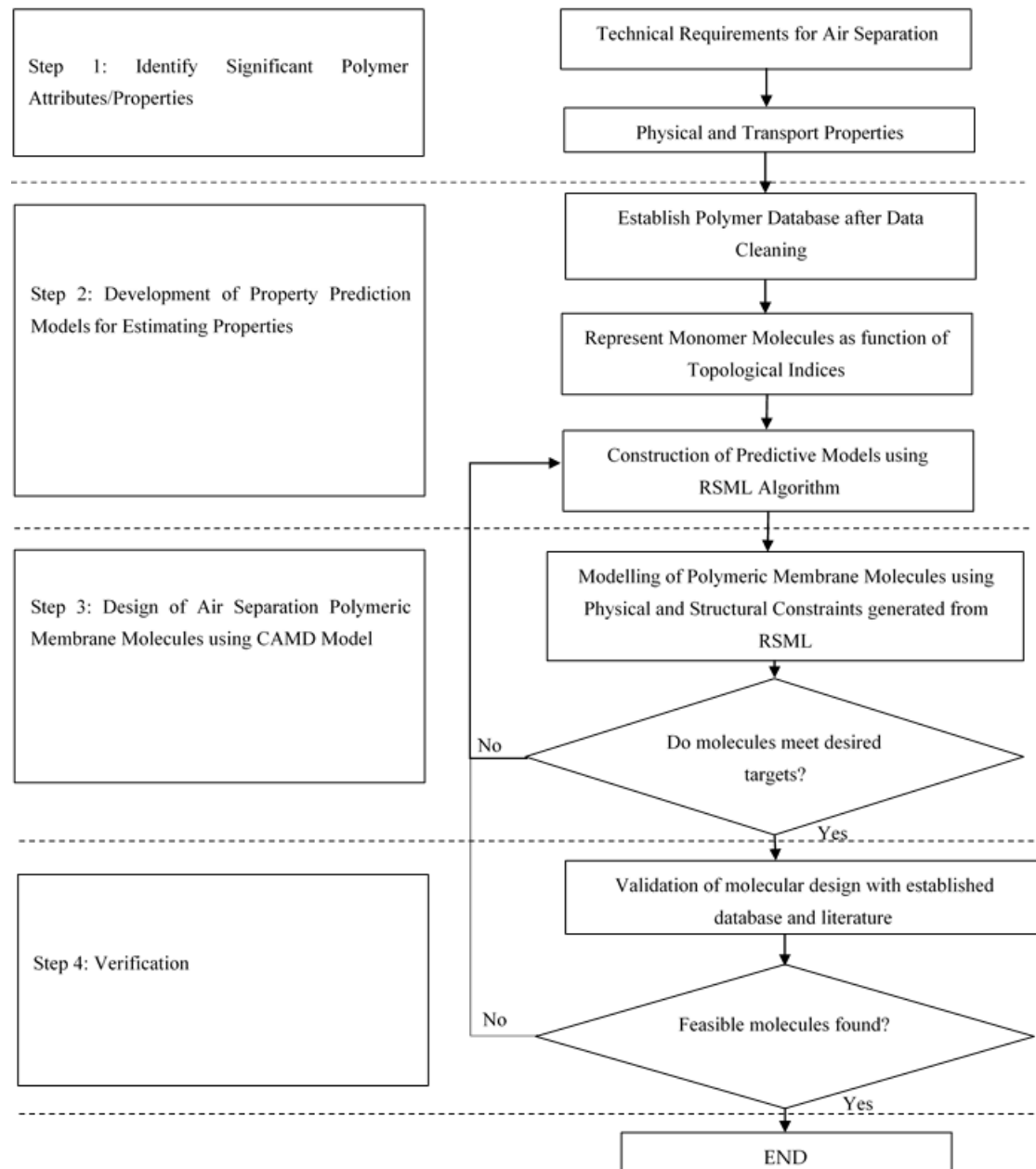


Figure 1. Methodology of CAMD development to design air separation polymeric membrane.

Step 2: Development of property prediction models for estimating properties

RSML was utilised to develop predictive models for physical and transport properties using topological indices. Though there are existing models for the physical properties, it has been reported that the accuracy for T_g and E_{coh} models is relatively low [34]. Moreover, complete polymer data available for both the properties and their respective topological indices are not abundant, making RSML suitable to be applied in addition to its interpretability. Furthermore, the classification predictive model for each property is more

relevant in this case as it determines the class to which a polymer's property belongs, thereby determining its suitability for air separation applications.

Step 2.1: Database and properties classification

A polymer database was required to develop ML models for data training. Database established by Van Krevelen and Te Nijenhuis [35] and Bicerano [33] containing polymer physical properties information were used in this work. Gas permeability and selectivity information were obtained from Jia and Xu [36] with approximately 60 entries. All the properties are quantitative and were directly input to build up the information system. Each physical and transport property was categorised into 2 classes in view of the data range and availability. Low boundary of T_g was set at 300 K, since from the original dataset, approximately 67% of the T_g falls above 300 K which is the typical temperature region to fulfil normal operating conditions. Based on a similar concept, boundary for V_m was set at 100 cm³/mol whereas E_{coh} was 35,000 J/mol. It was desired to have polymer with T_g higher than 300 K, V_m more than 100 cm³/mol and E_{coh} lower than 35,000 J/mol as higher cohesive energy increases the chain's density making it harder to allow molecule permeation [37].

Oxygen separation membranes are attractive if membrane's selectivity ranges from 4 to 6 while oxygen permeability is more than 10 Barrers [38]. In order to achieve more extensive penetration and obtain higher product purity, higher membrane selectivity and permeability values are desired. Therefore, the lower boundary for membrane permeability was set at 10 Barrers whilst O₂/N₂ selectivity was set at 4.

Step 2.2: Representation of monomer molecules using topological indices

Topological indices (TIs) used in this study were zeroth and first-order connectivity indices including its valency, electro-topological state (E-state) index and shape indices including Kappa Order 1 to 3 ($^1\kappa$, $^2\kappa$, $^3\kappa$) as well as Kappa Alpha Order 1 to 3 ($^1\kappa_\alpha$, $^2\kappa_\alpha$, $^3\kappa_\alpha$) and Kappa Flexibility Index. They were utilised to represent numerically the monomer's physical properties in terms of structural aspects such as cross-linking, bond types, branching, electronic information, etc. The connectivity indices were incorporated since they provide information about the number of non-hydrogen atoms and bonding type of the polymer molecules that will influence polymer physical properties such as T_g , where more non-hydrogen atoms result in higher T_g .

Kier and Hall's electro-topological state (E-state) index incorporates both electronic information and molecular topology to describe the chemical structure at atomic level which is useful in this case. It is the sum of intrinsic state of atom and perturbation factor depicting the influence of the remaining atoms in the molecule [39]. On the other hand, Kappa shape indices characterise molecular structure quantitatively and take into account spatial density. Kappa shape index also encodes information about size, degree of cyclicity and degree of separation in branching [40]. By incorporating the shape indices, it provides insights into polymer cross-linking, degree of branching and molecule sizes, which will affect the polymer permeability and selectivity. First order Kappa shape index, $^1\kappa$, is defined by one bond fragment counts, 1P where only linear graph is considered. Second order shape index, $^2\kappa$, is described by two-bond paths, 2P_i , with two shape extremes of $^2P_{max}$ and $^2P_{min}$ [41]. Likewise, count of paths of three adjacent bonds, 3P_i , configures the basis for the quantification of third order shape index, $^3\kappa$. Inclusion of alpha in shape indices demonstrates the influence of covalent radius on molecular shape, giving more information about the polymer structure. Lastly, incorporation of flexibility index is based on the role of molecular size, cycles, branching and heteroatom content. The flexibility index is made equivalent to the product of first and second Kappa order, normalised to the number of atoms in the graph [41].

All the topological indices information was extracted from Toxicity Estimation Software Tool (TEST). The software estimates the toxicity as well as topological indices of chemical structure using QSAR methods. Moreover, the tool eases information collection

where the toxicity values and other relevant molecular information will be presented once user inputs the respective molecular structure [42].

Step 2.3: Construction of predictive models using RSML

The identified physical properties and tabulated data for the respective properties accomplished in the previous steps were defined as the decision attributes. For topological indices which are the structural descriptors, they were determined as the condition attributes. The polymers were selected from the available database with completed TIs information, for instance, 194 polymers were utilised for generating T_g predictive model. A total of 70% of the entire dataset were used for training, while 15% was used for validation and testing, respectively. Out of the 194 polymers, 65 of them have T_g less than 300 K which were classified as Class 1 while the remaining 129 polymers were equal and more than 300 K, classified as Class 2. However, the completed dataset collected for permeability and selectivity was much lesser with only 62 and 53 data, respectively. Due to the limited data availability for permeability and selectivity, 70% of the data were used for training and the remaining 30% were for validation. There was insufficient data available for testing permeability and selectivity; therefore, only validation was performed. As mentioned above, polymers with permeability below 10 Barrers were classed as Class 1 and those above 10 Barrers were classed as Class 2. A similar approach was used for selectivity categorisation, with O_2/N_2 less than 4 classified as Class 1, and equal or more than 4 classified as Class 2. Likewise, the other physical properties— V_m and E_{coh} were also classified into 2 classes with the boundaries as stated in Step 2.1.

All the conditional attributes are composed of topological indices which are continuous attributes whereas decision attributes consisted of all the physical properties that have been classified into classes; hence, decision attributes are integer attributes. There were 5 information tables constructed since 5 physical properties predictive models were aimed to be developed. Table 1 tabulates the layout of a simplified information table where C1, C2 and C3 are continuous attributes.

Table 1. Simplified polymer information system.

Polymer	Conditional Attributes			Decision Attribute
	C1	C2	C3	D1
P1	5.9	6.5	3.1	1
P2	5.2	0	0	1
P3	7.7	7.2	4.9	2
P4	9.2	8.9	5.6	2
P5	5.5	5.3	3.4	1

Once the information system was input with complete data, reduction of attributes was executed to remove redundant attributes. As shown in Table 1, $U = \{P1, P2, \dots, P5\}$ is the finite non-empty set whilst $R = \{C1, C2, C3\}$ represents attribute set. Indiscernibility (I) indicates the polymers that have the same conditional attribute sets. Indiscernibility for the simplified table of complete relation R , $C1\&C2$, $C1\&C3$ and $C2\&C3$ are shown in Equations (12)–(15). From Equations (13) and (15), removal of either $C1$, $C2$ or $C3$ attributes from relation R shows no effect on the table where it still results in the same classification as original information table. Therefore, $C1$, $C2$ and $C3$ each are indispensable in this case.

$$(R) = \{P1\}, \{P2\}, \{P3\}, \{P4\}, \{P5\} \quad (12)$$

$$I(R - \{C3\}) = I(R) \quad (13)$$

$$I(R - \{C2\}) = I(R) \quad (14)$$

$$I(R - \{C1\}) = I(R) \quad (15)$$

In this context, classification generated by all the 3 conditional attributes $C1$, $C2$ and $C3$ is identical to the classification of $C1\&C2$, $C1\&C3$ and $C2\&C3$. In order to determine the reducts of R , pairs of attributes $C1\&C2$, $C1\&C3$ and $C2\&C3$ are to be checked if they are independent. Since $I(C1\&C2) \neq I(C1)$ and $I(C1\&C2) \neq I(C2)$, pairs of $C1\&C2$ are independent. Likewise, $C1\&C3$ and $C2\&C3$ are also determined to be independent. Therefore, the reducts of R are found to be $\{C1\}$, $\{C2\}$ and $\{C3\}$. The core is not present in this example since there is no attribute which is the intersection of all reduct sets. Moreover, there are no superfluous attributes which can be omitted in this example since each attribute is a standalone reduct, i.e., each reduct is capable of determining the classification of the system. Subsequently, each reduct was applied to derive a set of rules. For example, attributes $C2$ and $C3$ were omitted during rules generation for reduct $\{C1\}$. Similar approach was implemented for the other reducts. Three rules were generated in this example:

1. Rule 1: $(C1 < 6.8) \rightarrow (D1 = 1)$
2. Rule 2: $(C2 \geq 6.85) \rightarrow (D1 = 2)$
3. Rule 3: $(C2 \geq 4.15) \rightarrow (D1 = 2)$

In this work, there were 12 conditional attributes in the completed information table and 2 classes in each decision attribute to distinguish between normal or robust and unusual or less desired physical properties. A section of the permeability information table is attached in Appendix A—Table A1. Similar steps were employed to generate decision rules by deriving reducts from training data. Reducts and rules were generated using the software Rough Set Data Explorer-ROSE2 [43]. Rules generated were then validated using validation dataset by evaluating certainty (cer_x), strength (σ_x) and coverage (cov_x). These parameters are defined in Equations (16)–(18), by letting $S = (U, C, D)$. The final rules selected were after evaluation of testing dataset. The testing data were sourced from other references to avoid bias issues. Generally, the benchmark for coverage and certainty was set to be higher than 70% to select high coverage and certainty rules. Finalised rules were to be used as property constraints in CAMD. Results of ROSE2 and variation of rules selection are discussed in Section 3.1.2.

$$\sigma_x(C, D) = \frac{supp_x(C, D)}{card(U)} \quad (16)$$

$$cer_x(C, D) = \frac{card(C(x) \cap D(x))}{card(C(x))} = \frac{supp_x(C, D)}{card(C(x))} = \frac{\sigma_x(C, D)}{\pi(C(x))} \quad (17)$$

$$cov_x(C, D) = \frac{card(C(x) \cap D(x))}{card(D(x))} = \frac{supp_x(C, D)}{card(D(x))} = \frac{\sigma_x(C, D)}{\pi(D(x))} \quad (18)$$

where $\pi(C(x)) = \frac{card(C(x))}{card(U)}$ and $\pi(D(x)) = \frac{card(D(x))}{card(U)}$.

Step 3: Design of air separation polymeric membrane molecules using CAMD model

In the optimisation model, the objective function was to maximise tensile strength of the polymer. High tensile strength is essential so that polymeric membrane is able to withstand mechanical stresses during operation and maintain their integrity to allow for appropriate gas flux across the membrane [44]. Property constraints were included in CAMD model which was derived from RSML algorithms decision rules.

Step 3.1: Formulation of structural constraints

In order to generate a feasible molecular structure, structural constraints were included in CAMD model so that molecules do not violate basic feasibility criteria such as octet rule. The molecules should not contain any free bonds or have any unattached sites or multiple bonds attached to the same site. First, suitable first-order molecular groups were selected that may potentially form the building blocks for a monomer molecule design. The first-order molecular groups determined were as stated in Appendix B—Table A2. Linear structural constraints were developed using integer variables based on the algorithms developed by Churi and Achenie [16]. Let m be the number of structural groups, v_k be

the valence of the k th group while s_{max} indicates the maximum valence of all the groups in the basis set. Moreover, n is the number of groups in the designed structure and n_{max} is the maximum number of groups allowed in a molecule. m structural groups having v_k with maximal valency, s_{max} , were specified at a reasonable n_{max} . Lower limit of n_{max} is 2 as it is the minimum group to form a molecule. The actual number of groups, n , will then be obtained from the mathematical programming model. Therefore, in this study, the parameters were

$$m = 12$$

$$v_k = [1 \ 1 \ 1 \ 1 \ 3 \ 3 \ 1 \ 1 \ 3 \ 2 \ 2 \ 2]$$

$$s_{max} = \max \{v_k\} = 3$$

$$n_{max} = 14$$

The entire structural constraints consist of three binary and discrete variables— u , z and w , defined with indices i, j, k and p . Indices i and p define structural group's position in a designed molecule. Index k specifies the type of functional group while j implies the site of which i th group is attached to p th group. As shown in Equation (19), u_{ik} defines if i th position is occupied by k th group in the molecule and it restricts each position i with only one group k . Octet rule is defined in Equation (20) to ensure the number of bonds connected to a group that corresponds to its valency. z_{ijp} indicates if i th group is attached to p th group via j th site.

$$\sum_{i=1}^{n_{max}} \sum_{k=1}^m u_{ik} \leq n_{max} \quad (19)$$

$$\sum_{p=1}^{n_{max}} \sum_{j=1}^{s_{max}} z_{ijp} = \sum_{k=1}^m u_{ik} v_k ; i = 1 \dots n_{max} \quad (20)$$

Equation (21) constrains the i th group to be attached to one of the groups before it, defined by $(i - 1)$. w in the equation is also a binary vector which signifies valence site; thus, the first and second terms are zero since they will be occupied. The presence of first term is emphasised in Equation (23) and $(i + 1)$ th group is only present if i th group is present to assure that only one molecule is formed, which is defined in Equation (24).

$$\sum_{p=1}^{i-1} \sum_{j=1}^{s_{max}} z_{ijp} \geq -w_i ; i = 2 \dots n_{max} \quad (21)$$

$$\sum_{i=1}^{n_{max}} \sum_{k=1}^m u_{ik} + \sum_{i=1}^{n_{max}} w_i = n_{max} \quad (22)$$

$$w_i = 0 \quad (23)$$

$$w_i \leq w_{i+1} ; i = 1 \dots (n_{max} - 1) \quad (24)$$

To account for various group valences, Equation (25) is introduced in linear form, stating that for k th kind of i th group, the group should not have any attachments for its sites $(v_k + 1)$ to s_{max} which are non-existent. M is a significantly larger number than all other terms in the equation, specified as 50 in this model. Furthermore, Equation (26) denotes the symmetry constraints, for instance, the first group attached to second group is equivalent to the second group connected to first one. Since a group cannot be attached to itself, p is set to start from 2.

$$\sum_{j=v_k}^{s_{max}} \sum_{p=1}^{n_{max}} z_{ijp} - \sum_{p=1}^{n_{max}} z_{i v_k p} + M u_{ik} \leq M ; i = 1 \dots n_{max}, k = 1 \dots m \quad (25)$$

$$\sum_{j=1}^{s_{max}} z_{ijp} = \sum_{j=1}^{s_{max}} z_{pji} ; i = 1 \dots (n_{max} - 1), p = (i + 1) \dots n_{max} \quad (26)$$

Equation (27) ensures that a group's site can only be attached at most once to another group. Lastly, for any existence of i th group, $(i - 1)$ th group should also be present, as defined in Equation (28). Structural constraints from Equations (19)–(28) are all linear, forming a convex hull.

$$\sum_{p=1}^{n_{max}} z_{ijp} \leq 1 ; i = 1 \dots n_{max}, j = 1 \dots s_{max} \quad (27)$$

$$\sum_{k=1}^m u_{ik} - \sum_{k=1}^m u_{i-1,k} \leq 0 ; i = 2 \dots n_{max} \quad (28)$$

In addition, the prevention of free bonds number formed in generated molecule is described in Equation (29).

$$v_k \sum_{i=1}^{n_{max}} u_{ik} - 2(n_{max} - 1) = 0 ; k = 1 \dots m \quad (29)$$

Step 3.2: Modelling of Air Separation Polymeric Membrane Molecule

After formulating all the structural criteria, the objective function (Equation (30)) was encoded, where the predictive model for tensile strength, σ , was extracted from Eslick et al. [45]. The 1x in the predictive model means first order connectivity index whereas ${}^1x^V$ is the first order valence connectivity index. CD indicates crosslink density that is computed in Equation (31) where DC is degree of conversion determined empirically, w_i is the weight fraction of monomer i , nv_i is the number of vinyl groups in monomer i and MW_i is the monomer i molecular weight.

$$\sigma = 1406.6 - 7484.5{}^1x + 6611.6{}^1x^V + 78,231.7CD_{max} - 149,268.6CD \quad (30)$$

$$CD = DC \sum_i \frac{w_i(nv_i - 1)}{MW_i} \quad (31)$$

The framework was a single objective problem aiming to maximise polymer tensile strength. However, mathematical programming algorithms were to be developed to correlate binary terms in structural constraints with the connectivity indices in the predictive model. The correlation terms were derived for 1x and ${}^1x^V$ in Equations (32) and (33), respectively.

$${}^1x = \sum_{i=1}^{n_{max}} \sum_{p=i+1}^{n_{max}} \frac{\sum_{j=1}^{s_{max}} z_{ijp}}{\sqrt{\Delta_i \Delta_j}} \quad (32)$$

$${}^1x^V = \sum_{i=1}^{n_{max}} \sum_{p=i+1}^{n_{max}} \frac{\sum_{j=1}^{s_{max}} z_{ijp}}{\sqrt{\Delta_i^V \Delta_j^V}} \quad (33)$$

$$\Delta_i = \sum_{k=1}^{n_{max}} u_{ik} \delta_k ; k = 1 \dots m \quad (34)$$

$$\Delta_i^V = \sum_{k=1}^{n_{max}} u_{ik} \delta_k^V ; k = 1 \dots m \quad (35)$$

Equations (32)–(35) were utilised to determine the connectivity indices of the bond between the attachment of the groups at different positions. Nevertheless, connectivity

indices within the first-order group itself were also calculated using similar approach where edges of the groups are known [46], as presented in Equations (36) and (37).

$${}^1x = \sum_{i=1}^{n_{max}} \frac{u_{ik}}{\sqrt{\delta_i \delta_j}} ; k = 1 \dots m \quad (36)$$

$${}^1x^V = \sum_{i=1}^{n_{max}} \frac{u_{ik}}{\sqrt{\delta_i^V \delta_j^V}} ; k = 1 \dots m \quad (37)$$

Incorporation of Equations (30)–(37) make the optimisation formulation non-convex because of the trilinear terms. Therefore, it is necessary to linearise some of the equations and make it a convex problem to obtain feasible solutions. In this context, the trilinear terms in Equations (32) and (33) were modified to make the square root denominator terms known values, which will be demonstrated in Section 3.2.2.

Step 3.3: Incorporation of physical constraints in CAMD modelling

To ensure a viable molecule could be generated, property constraints determined from RSML were included. Since tensile strength is influenced by connectivity indices as observed in Equation (30), the constraints to be included in CAMD consist of only the selected predictive rules comprising connectivity indices. The remaining selected predictive rules containing other topological indices were cross-checked and verified again after the molecular structure was derived. The property constraints model will be either of upper and/or lower bound number range derived from RSML to ensure the molecule designed is under the desired category. At this stage, the CAMD formulation has been formulated by maximising σ , subjected to structural and property constraints. The CAMD problem was then solved using global solver in LINGO extended version 20.0 after transforming the non-linear terms in Equations (32) and (33) to be convex functions, which will be elaborated in Section 3.2.2.

Step 4: Verification

Once a molecule was generated from CAMD, it was first verified whether the molecule exists in the present polymer database. If the designed molecule was present in existing database, this proved the model's accuracy in discovering potential polymer structure candidates. In the context where the polymer generated was not suitable for air separation purposes, integer cuts constraints would be incorporated to generate different solutions. Otherwise, if it is not available in present database, a literature review could be performed to determine its separation characteristics. If the designed molecule could neither be found in the literature nor existing database, experimental verification should be conducted to validate the molecule's properties. Solutions from CAMD could be utilised to guide the focus of experimental analysis. However, if the generated molecules were not able to meet the desired properties, RSML shall be revisited and modified to improve the prediction reliability and accuracy.

3. Results and Discussions

The development of polymer properties predictive models is pre-requisite for CAMD formulation; therefore, the determined rules generated from RSML were discussed extensively prior to incorporating them as property constraints into CAMD problem. Approach to generate feasible polymer structure was demonstrated with case studies as well.

3.1. Development of Predictive Models Using RSML

Polymer predictive models from RSML were used as constraints in the generation of feasible polymeric membrane molecules. The generation of cores, reducts and rules as well as selection of most prominent rules were discussed in the following sections.

3.1.1. Cores and Reducts Generation

There were five information systems established for T_g , V_m , E_{coh} , O_2 permeability and O_2/N_2 selectivity, respectively. Each of the information systems consisted of 12 conditional attributes whilst the decision attribute was each classified into 2 classes—Class 1 being the less desired property ranges and Class 2 as the more favoured property ranges. There was no core generated from either of the information systems. Nevertheless, the number of reducts generated from T_g , V_m , E_{coh} , O_2 permeability and O_2/N_2 selectivity information systems were 19, 20, 20, 11 and 10, respectively, where repeated rules were identified in the subsequent analysis.

3.1.2. Rules Generated from Reducts

All the rules generated from all the reducts were evaluated based on their strength, certainty and coverage. In this study, a total of 602 rules were generated from T_g reducts, 305 rules for V_m , 206 rules for E_{coh} , 192 rules for permeability and 157 rules generated for selectivity. Table 2 presents the example rules from each decision class extracted from selectivity reduct 1.

Table 2. Example rules generated from selectivity reduct 1.

Rule	Kappa Order 3	Kappa Alpha Order 2	Decision	Strength	Coverage (Recall)	Certainty (Precision)	Accuracy
2	5.432 to 11.545	-	Class 1 (Selectivity < 4)	13.51%	29.41%	100%	83%
10	<3.828	0.671 to 1.773	Class 2 (Selectivity \geq 4)	10.81%	20%	100%	60%

Based on Table 2, rule 10 covers 4 data out of the entire training dataset consisting of 37 data, having a strength of 10.81%, which is considered as a feasible rule to be applied in CAMD by constraining Kappa Order 3 and Kappa Alpha Order 2 values according to the values derived from RSML. Rule 10 in this case also depicts that Kappa Order 3 is lesser than 3.83 and the effect of Kappa Alpha Order 2 potentially results in high certainty that the polymer has O_2/N_2 selectivity of more than 4. All the other rules were interpreted in a similar manner, however, since there were abundant rules generated that fall under the desired properties class, further interpretation and analysis were performed to select reasonable rules.

Generally, the rules generated for T_g , V_m and E_{coh} show high coverage and strength as compared to the rules derived for permeability and selectivity. This may be due to the lack of data available for both polymer permeability towards oxygen and O_2/N_2 selectivity, leading to low coverage and strength. All the rules were analysed using a validation dataset where the filtered rules with their respective strength, coverage and certainty were shown in Appendix C—Tables A3 and A4. These rules were filtered based on a certainty of more than 75% and were selected for the desired properties class. However, it was observed that about 90% of the rules for permeability and selectivity attributes could only fulfil one data in the validation set. On the contrary, there were more physical property rules satisfying certainty of more than 75% as well as having higher average coverage (\approx 30%) than the transport properties' rules. The accuracy of individual rules is lower because those rules are developed to classify molecules into one of the categories. However, since the certainty is high for all the rules, it can be confirmed that the chosen molecules have the potential to meet the property in the desired range. These rules will be further verified in the testing section and with respect to scientific findings.

Furthermore, it can be noticed that Kappa shape indices including first, second, and third order as well as the incorporation of alpha were present in approximately 65% of the total filtered rules. As a result, Kappa shape indices can be regarded as a significant parameter encoding molecular structure information that could potentially influence the polymeric membrane performance. Nevertheless, it shall be noted on overlapping cases where the exact same polymer might fall in more than one rule under the same decision

class. Though a huge number of rules were filtered from the validation dataset, there were still abundant remaining rules to be selected as the finalised constraints in CAMD.

3.1.3. Evaluation of Model Performance and Scientific Coherency of Rules Generated

The physical property rules were tested using a dataset retrieved from different reference sources to gauge the model performance when dealing with entirely new sets of molecules. Through the testing evaluation, rules with a certainty above 80% were further evaluated by analysing any overlapping molecules and the scientific coherency between the conditional and decision attributes. Since the focus is to design desired physical and transport properties of a polymer, only rules falling under the proper category were further analysed which are Class 2 for T_g , V_m , permeability and selectivity attributes whereas Class 1 for E_{coh} attribute.

From the analysis, it was found that there were large overlapping polymers between the rules, particularly for physical properties. Therefore, the rules were selected considering the largest coverage as summarised in Table 3. The strength, coverage and certainty were based on rule combinations tabulated from testing data for physical properties, while transport properties were according to validation data due to the lack of a database. A higher 1x indicates a higher number of vertices in the hydrogen-suppressed graph which means more non-hydrogen atoms that could lead to higher T_g . In addition, Kappa order 3 provides more detailed molecular shape information than the first and second order. A lower Kappa third-order value implies a more spherical and symmetrical molecular structure with more organised polymer chain packing resulting in higher T_g [47].

Table 3. Rules selected for CAMD modelling.

Rule	Decision	Strength	Coverage (Recall)	Certainty (Precision)	Accuracy
$^1x \geq 2.94$ and Kappa Order 3 < 2.98	$T_g = \text{Class 2}$	31%	44%	89%	83%
Kappa Alpha Order 2 ≥ 7.03 , or Kappa Alpha Order 3 from 5.16 to 6.31	$V_m = \text{Class 2}$	42.3%	50%	100%	85%
$^0x \geq 2.5$ and E-state Index < 13.81, or 1x from 1.404 to 3.59 and E-state Index < 15.08, or Kappa Alpha Order 1 ≥ 2.72 and E-state Index < 15.08	$E_{coh} = \text{Class 1}$	29.4%	100%	100%	86%
E-state Index < 18.25 and Kappa Order 3 ≥ 4.67 , or 0x from 4.63 to 5.08	Permeability = Class 2	11.1%	40%	100%	83%
Kappa Order 3 < 3.83 and Kappa Alpha Order 2 from 0.67 to 1.77, or Kappa Flexibility Index from 2.72 to 3.32	Selectivity = Class 2	12.5%	28.57%	100%	89%

For V_m , a higher Kappa order such as, in this case, Kappa order 3, tends to have a lower molar volume, thus, the constraint for the third order is lower than the second order as observed in Table 3. This is because branched or networked polymers occupy a smaller space than linear polymers of the same molecular weight [48]. Two rules were combined for V_m to increase the coverage. The higher value of the zeroth order connectivity index indicates a greater degree of connectivity within the polymer; simultaneously, the cohesive energy can be reduced by having a lower E-state index denoting lesser electronic delocalization [49]. Therefore, the polymer has weaker intermolecular interactions resulting in lower cohesive energy. A similar concept applies to the rule consisting of a first-order connectivity index with a narrower limit. The third rule in E_{coh} can be explained by having a higher Kappa alpha order 1 value which defines a less ordered structure due to higher branching [50], having lower cohesive energy.

In the case of permeability, 2 rules were combined to improve the coverage. A higher E-state index is considered as having stronger intermolecular interactions with gas molecules that reduce the permeation through the polymer. Moreover, a higher degree of polymer branching reflects by a higher value of Kappa order 3 brings about a more porous and open structure allowing gas molecules to diffuse through [50]. Therefore, it is reasonable from a scientific point of view for the first rule of permeability to be selected. Zeroth order connectivity index was also proven to affect polymer permeability by Bicerano [33]. A more porous structure leads to higher permeability performance but lower selectivity; therefore, Kappa order 3 in the selectivity rule is to be below 3.83 to prevent larger pore size than oxygen molecules. The range of Kappa alpha order 2 might be derived based on the optimum range for oxygen selectivity. The second rule for selectivity involving Kappa flexibility index ensures that the polymer formed is not too flexible nor rigid to selectively allow oxygen but no other gases to permeate through. All the constraints selected in Table 3 demonstrate a trend satisfying the rationale behind scientific reasoning with the numerical value derived from RSML programming based on training dataset pattern.

Referring to Table 3, it is also noted that the E-state index and Kappa shape indices would be widely incorporated in the CAMD modelling. Therefore, the reverse approach was used in the subsequent step to verify whether the molecules generated satisfy these RSML rules since only connectivity indices constraints were included in the optimisation framework.

3.2. Generated Air Separation Polymer Molecules

This sub-section presents the results obtained from solving the optimisation model. Various case studies were conducted to produce a set of solutions. Molecules that fulfil structural, physical and transport property constraints are identified as the potential candidates to be used as air separation membranes.

3.2.1. Non-Convexity in CAMD Modelling

As aforementioned, the typical formulation of this optimisation problem, as demonstrated in Section 2, would yield an MINLP problem. As a consequence of the formation of trilinear terms in Equations (32) and (33), the model became a non-convex problem that will be hard to solve. The non-linearities were contributed by the crosslinking term in the tensile strength equation (Equation (31)) and connectivity indices correlation terms (Equations (32)–(37)). Therefore, the problem was relaxed to form a convex problem through modifications of trilinear equations.

3.2.2. CAMD Model with Linearised Connectivity Index Terms

Since the non-linearities were contributed by the connectivity indices correlation terms, linear formulations were proposed. All the structural constraints from Equations (19)–(29) were still incorporated in this case study while assumptions were made to derive the correlation terms. In the first attempt, only one heteroatom group was included which was CF group with CH₃ and CH₂ groups, i.e., $k = 3$ in this case. To reduce the number of integer terms, m was set to be 6. Another assumption made was that with any presence of the CF group, it would be attached to three groups of CH₂. Moreover, CH₃ groups would only be attached to CH₂.

With these connection assumptions, first order connectivity index was formulated as Equation (38) where n_{CF} , n_{CH_3} and n_{CH_2} define the number of CF, CH₃ and CH₂ groups, respectively. These numbers were defined from U_{ik} terms by specifying k term. The first term in this equation refers to the connectivity index within CF in addition to the bond connections between CF with three CH₂. The second term defines 1x between the CH₃ and CH₂ group and the final term depicts the connectivity index between CH₂ groups only, where n_{CF} and n_{CH_3} are deducted to avoid duplication since two of the CH₂ groups from CF are attached to CH₃, leaving one CH₂ connected to CF which is not connected to CH₃. Equation (39) ensures the number of CH₂ groups is equal to or more than the other functional groups. The $^1x^V$ applies the same approach, as in Equation (40). Since the edges of groups are known, δ_i values will be specified making the denominator terms in the equations to be known values, i.e., δ_{CF} , δ_{CH_2} , δ_{CH_3} are not variables, resulting in linear equations. Hence, Equations (32)–(35) would be replaced with the following linear equations.

$${}^1x = nCF({}^1x_{CF} + \frac{3}{\sqrt{\delta_{CF}\delta_{CH_2}}}) + nCH_3\left(\frac{1}{\sqrt{\delta_{CH_2}\delta_{CH_3}}}\right) + (nCH_2 - nCF - nCH_3 + 1)\left(\frac{1}{\sqrt{(\delta_{CH_2})^2}}\right) \quad (38)$$

$$nCH_2 \geq (3nCF + nCH_3) \quad (39)$$

$${}^1x^V = nCF({}^1x^V_{CF} + \frac{3}{\sqrt{\delta_{CF}^V\delta_{CH_2}^V}}) + \left(\frac{nCH_3}{\sqrt{\delta_{CH_2}^V\delta_{CH_3}^V}}\right) + (nCH_2 - nCF - nCH_3 + 1)\left(\frac{1}{\sqrt{(\delta_{CH_2}^V)^2}}\right) \quad (40)$$

Following next, DC in the crosslinked density term was estimated to be 0.7 and w_i to be 1. This has now become a convex formulation as the terms in the denominator are known values in this case, instead of variables. Global optimum results were then able to be generated since there were values generated for 1x and ${}^1x^V$ to be substituted in the objective function (Equation (30)). For this combination of functional groups, it was determined that heteroatoms were not favoured to maximise polymer tensile strength subjected to the property constraint. The result obtained was a short hydrocarbon structure (butane).

As a result, it is evident that by linearising the correlation terms, the optimisation model becomes convex. This example only considers three first-order groups in the molecules—CF, CH₃ and CH₂; hence, with different first-order groups and specific group attachment assumptions, the structural constraints formulations from Equations (19)–(28) would need to be modified for each assumption. However, the formulation is linear and can generate reliable results for each class of polymer molecules.

3.3. Verification of Model

From Tables 4 and 5, the potential candidates were generated according to the structural assumptions made where six out of the seven candidates are available in the existing polymer database [51]. This proves the model's accuracy in identifying potential polymer structure candidates and the potential of RSML to generate new polymer molecules for effective air separation. The first assumption was to analyse the performance of straight-chain hydrocarbons. Results showed that monomer 1-hexene was the optimum structure fulfilling the permeability rules at the same time satisfying the O₂ permeability classification under Class 2. However, none of the selectivity rules were fulfilled which corresponds to its selectivity value that is less than 4. Since T_g and V_m of poly(1-hexene) fall under Class 1, none of the rules in Table 4 are satisfied. On the other hand, its E_{coh} is under Class 1, i.e., less than 35,000 J/mol, satisfying E_{coh} rule generated from RSML, where poly(1-hexene) ${}^0x \geq 2.5$ and E-state index < 13.81. This verifies the robustness and effectiveness of the predictive model from RSML.

Poly(4-methyl-1-pentene) is the isomer of poly(1-hexene) with branching. Despite altering the assumption to have branching, the optimisation model still yielded a six-carbon structure indicating that the six-carbon chain has the maximum tensile strength subjected to the constraints. Furthermore, the tensile strength of the branched chain is expected to be lower than the straight hydrocarbon chain in view of the less ordered chain packing leading to weaker intermolecular forces [52], which is also observed in Table 4. Though its tensile strength is lower, both O₂ permeability and selectivity fall within the desired classes, simultaneously fulfilling all the physical properties requirements as well. This makes poly(4-methyl-1-pentene) more attractive than poly(1-hexene) as the candidate for air separation membranes.

Benzene rings were also considered in this study on air separation performance. The aromatic rings positioned in the polymer backbone potentially form a rigid structure with stronger mechanical strength and thermal stability [53]. The hetero group incorporated was the carbonate functional group. CAMD showed polycarbonate as the optimum structure of high tensile strength. However, this result was obtained by relaxing the permeability constraint $-{}^0x$ between 4.63 and 5.08 rule, even so, polycarbonate structure still does not meet the other permeability rule where E-state index < 18.25 and Kappa Order 3 ≥ 4.67 . The actual value of polycarbonate permeability

towards oxygen is 1.5 barrers (Class 1) which again verifies the effectiveness and accuracy of the RSML predictive model since the permeability of polycarbonate does not fall in Class 2. Although polycarbonate portrays high tensile strength characteristics and meets the desired selectivity, it is not considered a potential candidate due to its low oxygen permeability.

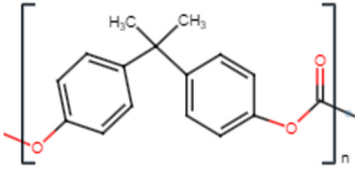
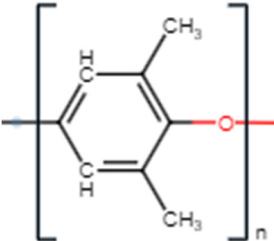
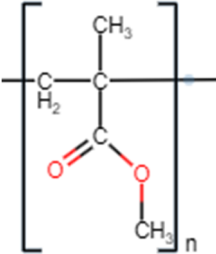
Table 4. CAMD results.

Polymer Name	Poly(1-Hexene)	Poly(4-Methyl-1-Pentene)	Poly(5-Methyl-Hexene-1)	Poly(3-Chlorohexene)
Monomer Molecular Structure				
Formula	C ₆ H ₁₂	C ₆ H ₁₂	C ₇ H ₁₄	C ₆ H ₁₁ Cl
CAS number	592-41-6	691-37-2	3524-73-0	53101-38-5
Structural Assumptions	<ul style="list-style-type: none"> • CH = CH₂ attach to one CH₂ • CH₃ only attach to CH₂ 	<ul style="list-style-type: none"> • CH = CH₂ attach to one CH₂ • (CH₃)₂CH only attach to CH₂ 	<ul style="list-style-type: none"> • CH = CH₂ attach to one CH₂ • (CH₃)₂CH only attach to CH₂ 	<ul style="list-style-type: none"> • CH = CH₂ attach to one CHCl • (CH₃)₂CH only attach to CH₂
TS [1]	3	4	5	2
O ₂ Permeability (Barrers)	10	32.3	20	Not available
O ₂ Selectivity	2.6	4.225	2.5	Not available
⁰ _x	4.406	4.992	5.698	5.698
¹ _x	2.932	2.770	3.270	3.3081
¹ _x ^V	2.932	2.379	2.879	3.011
E-state index	11.5	11.833	13.333	15.4444
¹ _κ	6	6	7	7
² _κ	5	3.2	4.167	4.167
³ _κ	5.333	5.333	6	3.840
¹ _{κ_α}	5.740	5.740	6.740	7.026
² _{κ_α}	4.740	2.951	3.915	4.192
³ _{κ_α}	5.105	5.105	5.740	3.867
Φ	4.535	2.824	3.769	4.208
T _g (K)	223	302	259	Not available
V _m (cm ³ /mol)	97.9	235	139.6	Not available
E _{coh} (J/mol)	13,000	26,160	7900	Not available
Literature TS (MPa)	39	28	40	Not available

Furthermore, polyphenylene oxide was the optimum structure from CAMD by considering branching and oxides from the benzene ring. Polyphenylene oxide shows relatively high tensile strength which also satisfies the desired permeability and selectivity class. This also illustrates that the predictive rules selected from RSML are fulfilled in this case. Polyphenylene oxide has the highest cohesive energy among the generated molecules, even so, it is still within the acceptable range and does not show any adverse effect towards oxygen molecules permeation. Therefore, polyphenylene oxide emerges to be one of the potential candidates for air separation application. The final generated molecule was polymethyl methacrylate incorporating consideration of the presence of carbonate groups in a straight chain. Though it

satisfies the permeability constraint, the molecule does not meet the desired selectivity. Its tensile strength generated from the CAMD model contradicted the trend from the literature which might be due to the inaccuracy of the objective function model extracted from the literature. Nevertheless, based on scientific reasoning, benzene ring structures have stronger mechanical strength as compared to straight chain structures [54].

Table 5. CAMD results (continued).

Polymer Name	Polycarbonate	Polyphenylene Oxide	Polymethyl Methacrylate
Monomer Molecular Structure			
Formula	$C_{15}H_{16}O_2$	C_8H_8O	$C_5O_2H_8$
CAS number	25037-45-0	25134-01-4	9011-14-7
Structural Assumptions	<ul style="list-style-type: none"> • C_6H_4O attach to one C • C only attach to CH_3 and C_6H_4COO 	<ul style="list-style-type: none"> • C_6H_3O attach to CH_3 	<ul style="list-style-type: none"> • C = CH_2 attach to one COO and CH_3
TS *	1	7	6
O ₂ Permeability (Barrers)	1.5	16.8	20
O ₂ Selectivity	5.769	4.421	3.71
0x	3.577	4.690	5.492
1x	1.732	3.450	3.189
$^1x^V$	1.354	2.230	2.274
E-state index	8.667	11.095	20.833
$^1\kappa$	4	3.938	7
$^2\kappa$	3.740	1.240	3.061
$^3\kappa$	1.333	0.490	2.667
$^1\kappa_\alpha$	1.105	3.218	6.377
$^2\kappa_\alpha$	0	0.874	2.533
$^3\kappa_\alpha$	0	0.302	2.121
Φ	1.033	0.402	2.307
T _g (K)	423	488	378
V _m (cm ³ /mol)	320	76.6	89.3
E _{coh} (J/mol)	14,400	33,300	27,700
Literature TS (MPa)	62.1	75	50

* Tensile strength is ranked based on CAMD results.

4. Conclusions

In this study, a computer-aided molecular design (CAMD) framework incorporating rough set-based machine learning (RSML) algorithms for the determination of polymeric structures that has the potential to be considered for air separation has been developed. Topological indices

were used to estimate both the physical and transport properties of polymer molecules where the deterministic rules were generated in RSML. The promising rules generated with the highest coverage and certainty were studied qualitatively from scientific standpoints to ensure that they were reliable to be included as property constraints in CAMD modelling. The original non-convex formulation of the CAMD model was transformed into a convex equivalent by transforming the equations into an alternative form. Results demonstrated that the rough set model was able to precisely predict the polymer characteristics of all molecules generated from the optimisation model, proving the reliability of RSML predictive models. After analysing the results, poly(4-methyl-1-pentene) (PMP) and polyphenylene oxide (PPO) emerge to be the most potential candidates for air separation since these two polymers fulfil both oxygen permeability and selectivity requirements as well as the desired physical properties in this study. The results depicted that the proposed methodology in this work could potentially be implemented for the systematic design of air separation polymeric membrane structure. To improve the quality of the models predicted by this method in the future, it is suggested to enhance the robustness and accuracy of the RSML model by incorporating more attributes that could potentially relate to the structure–property relationship. Furthermore, before utilizing a polymeric membrane for air separation applications, it is advisable to conduct economic analysis and feasibility studies to assess aspects such as scale-up feasibility.

Author Contributions: Conceptualization, J.O., N.G.C. and J.-W.C.; methodology, J.-Y.C., J.-Y.-L.L. and Q.-Y.T.; software, J.-W.C.; validation, J.-W.C., J.O. and N.G.C.; formal analysis, J.-Y.C.; investigation, J.O. and N.G.C.; resources, N.G.C. and J.-W.C.; data curation, J.-Y.C. and Q.-Y.T.; writing—original draft preparation, J.-Y.C. and J.-Y.-L.L.; writing—review and editing, N.G.C. and J.O.; visualization, J.-W.C.; supervision, N.G.C. and J.O.; project administration, N.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

Symbol	Description
${}^0\chi$	Zeroth Order Connectivity (Chi) Index
${}^0\chi^V$	Zeroth Order Valence Connectivity (Chi) Index
${}^1\chi$	First Order Connectivity (Chi) Index
${}^1\chi^V$	First Order Valence Connectivity (Chi) Index
δ_i	Number of sigma electrons in the hydrogen suppressed graph
δ_i^V	Number of valence electrons
$\delta_i^V \delta_j^V$	Number of edges in the molecules with bond s terminates on vertices i and j
${}^1\kappa$	First Order Kappa Shape Index
${}^2\kappa$	Second Order Kappa Shape Index
${}^3\kappa$	Third Order Kappa Shape Index
${}^1\kappa_\alpha$	First Order Kappa Alpha Shape Index
${}^2\kappa_\alpha$	Second Order Kappa Alpha Shape Index
${}^3\kappa_\alpha$	Third Order Kappa Alpha Shape Index
Φ	Kappa Flexibility Index
u_{ik}	Binary variable that indicates if i th position is occupied by k th group
z_{ijp}	Binary variable that indicates if i th group is attached to p th group via j th site

Appendix A. Example of Information Table

Table A1. Part of permeability information table with 12 conditional attributes.

Tag	Polymer	Decision Attribute	Condition Attributes											
		P _{O2} (Class)	θ_x	θ_x^V	l_x	l_x^V	E-States Index	Kappa Order 1	Kappa Order 2	Kappa Order 3	Kappa Alpha Order 1	Kappa Alpha Order 2	Kappa Alpha Order 3	Kappa Flexibility Index
1	Poly[1-(trimethylsilyl)-1-propyne]	2	5.91	6.50	3.06	6.00	13.69	7.00	2.34	6.00	7.08	2.40	6.08	2.43
2	Poly(tert-butylacetylene)	2	5.21	2.91	2.56	1.06	15.42	6.00	1.63	5.33	5.56	1.34	4.95	1.24
3	Poly(1-n-heptyl-propyne)	2	7.66	7.24	4.91	4.31	26.95	10.00	9.00	9.14	9.56	8.56	8.71	8.18
4	Poly[o-(trimethylsilyl)phenylacetylene]	2	9.19	8.89	5.55	7.62	24.96	10.08	3.81	2.49	9.39	3.35	2.14	2.62
5	Poly(1-chloro-2-n-butylacetylene)	2	5.54	5.26	3.41	2.88	15.61	7.00	6.00	6.00	6.84	5.84	5.84	5.71
6	Poly(1-chloro-2-n-hexylacetylene)	2	6.95	6.67	4.41	3.88	21.51	9.00	8.00	8.00	8.84	7.84	7.84	7.71
7	Poly(1-chloro-2-n-octylacetylene)	2	8.36	8.08	5.41	4.88	27.28	11.00	10.00	10.00	10.84	9.84	9.84	9.70
8	Poly[o-(trifluoromethyl)phenylacetylene]	2	9.19	6.02	5.55	3.18	18.07	10.08	3.81	2.49	8.68	2.91	1.81	2.10
9	Poly(1-n-hexyl-2-phenylacetylene)	2	10.06	8.92	6.93	5.47	30.95	12.07	8.32	6.19	10.86	7.21	5.20	5.59
10	Poly(1-ethyl-2-phenylacetylene)	2	7.23	6.09	4.93	3.47	19.42	8.10	4.76	3.11	6.89	3.74	2.29	2.58
11	Poly(1-phenyl-1-propyne)	1	6.53	5.39	4.43	2.91	16.53	7.11	3.92	2.38	5.91	2.94	1.62	1.93
12	Poly(1-chloro-2-phenylacetylene)	1	6.53	5.52	4.43	2.98	13.97	7.11	3.92	2.38	6.19	3.16	1.79	2.17

Table A1. Cont.

Tag	Polymer	Decision Attribute	Condition Attributes											
		P _{O2} (Class)	θ_x	θ_x^V	l_x	l_x^V	E-States Index	Kappa Order 1	Kappa Order 2	Kappa Order 3	Kappa Alpha Order 1	Kappa Alpha Order 2	Kappa Alpha Order 3	Kappa Flexibility Index
13	Poly(oxydimethylsilylene)	2	8.41	9.14	4.27	9.17	25.28	10.00	2.94	5.53	10.70	3.40	6.20	3.64
14	Hydrogenated Polybutadiene	2	3.41	2.57	1.91	1.15	9.65	4.00	3.00	4.00	3.48	2.48	4.56	2.16
15	Poly(1,3-butadiene)	2	3.41	2.57	1.91	1.15	9.65	4.00	3.00	4.00	3.48	2.48	4.56	2.16
16	Polyisoprene (NR)	2	5.00	3.70	3.49	2.39	11.67	5.00	2.25	4.00	4.48	1.77	3.48	1.58
17	Polychloroprene	1	3.70	3.63	2.39	2.12	13.78	5.00	2.25	4.00	4.77	4.77	3.77	1.94
18	Polystyrene	1	5.40	4.67	3.97	3.02	16.67	6.13	3.11	1.80	5.10	2.31	1.21	1.48

Appendix B. List of First-Order Groups

Table A2. Selected first-order groups.

First-Order Groups			
CH ₃	CH=CH ₂	COOH	CH=O
CF	CCl	CH ₂ OH	C=ONH ₂
CH ₃ Si	COO	-O-	CH ₂

Appendix C. Rules Filtered from Validation Dataset

Table A3. Physical properties rules filtered from validation dataset.

Rule	Glass Transition Temperature (T _g)		
	Decision	Strength	Coverage
$\theta_x \geq 4.49$ and Kappa Alpha Order 3 < 2.35	Class 2	24%	50%
$\theta_x \geq 4.49$ and Kappa Order 3 < 2.98	Class 2	24%	50%

Table A3. Cont.

Rule	Glass Transition Temperature (T_g)			
	Decision	Strength	Coverage	Certainty
${}^0x \geq 4.49$ and Kappa Alpha Order 3 < 2.35	Class 2	24%	50%	100%
${}^0x \geq 4.49$ and Kappa Order 3 < 2.98	Class 2	24%	50%	100%
${}^0x \geq 5.34$ and Kappa Alpha Order 2 < 3.43	Class 2	24%	50%	100%
${}^1x \geq 2.94$ and Kappa Alpha Order 3 < 2.35	Class 2	24%	50%	100%
${}^1x \geq 2.94$ and Kappa Order 2 < 3.861	Class 2	24%	50%	100%
${}^1x \geq 2.94$ and Kappa Order 3 < 2.98	Class 2	24%	50%	100%
E-state Index ≥ 10.17 and Kappa Flexibility Index from 1.08 to 2.32	Class 2	34%	57%	80%
Molar Volume (V_m)				
Kappa Alpha Order 2 ≥ 7.03	Class 2	10.71%	13.64%	100%
Kappa Alpha Order 2 from 5.27 to 6.4	Class 2	7.14%	9.09%	100%
Kappa Alpha Order 2 ≥ 4.89 and Kappa Alpha Order 3 from 3.96 to 6.31	Class 2	7.14%	9.09%	100%
Kappa Alpha Order 2 from 3.85 to 4.69	Class 2	17.86%	18.18%	80%
Kappa Alpha Order 2 ≥ 2.96 and Kappa Alpha Order 3 ≤ 3.19	Class 2	21.43%	27.27%	100%
Kappa Alpha Order 2 ≥ 2.40 and Kappa Alpha Order 3 ≤ 1.92	Class 2	42.86%	50%	91.67%
Kappa Alpha Order 3 from 5.16 to 6.31	Class 2	3.57%	4.55%	100%
Kappa Flexibility Index ≥ 6.65	Class 2	10.71%	13.64%	100%
Kappa Flexibility Index from 3.66 to 4.45	Class 2	17.86%	18.18%	80%
Kappa Alpha Order 3 from 3.96 to 6.31 and Kappa Flexibility Index ≥ 4.66	Class 2	3.57%	4.55%	100%
Kappa Alpha Order 3 ≤ 3.19 and Kappa Flexibility Index ≥ 2.54	Class 2	7.14%	9.09%	100%
Kappa Alpha Order 3 from 1.19 to 1.92 and Kappa Flexibility Index ≥ 1.67	Class 2	35.71%	45.45%	100%
${}^0x^V \geq 7.76$	Class 2	25%	31.82%	100%
${}^0x^V$ from 5.61 to 6.56	Class 2	28.56%	36.36%	100%

Table A3. Cont.

Rule	Glass Transition Temperature (T_g)			
	Decision	Strength	Coverage	Certainty
${}^0x^V$ from 5.16 to 5.32	Class 2	3.57%	4.55%	100%
${}^0x^V$ from 5.56 to 5.59	Class 2	7.14%	9.09%	100%
${}^1x \geq 3.98$ and Kappa Alpha Order 2 < 5.17	Class 2	39.29%	45.45%	90.9%
${}^1x \geq 5.68$	Class 2	32.14%	36.36%	88.89%
1x from 3.98 to 5.65 and Kappa Alpha Order 3 ≥ 3.96	Class 2	7.14%	9.09%	100%
${}^1x \geq 3.98$ and Kappa Alpha Order 3 < 3.93	Class 2	57.14%	63.64%	87.5%
1x from 3.24 to 3.68 and Kappa Alpha Order 3 ≥ 3.51	Class 2	3.57%	4.55%	100%
1x from 3.96 to 4.71 and Kappa Alpha Order 3 ≥ 1.32	Class 2	21.43%	22.72%	83.33%
${}^1x \geq 3.98$ and Kappa Flexibility Index < 4.85	Class 2	67.86%	77.27%	89.47%
1x from 3.98 to 5.65 and E-state Index ≥ 25.58	Class 2	17.86%	18.18%	80%
${}^1x \geq 3.98$ and E-state Index < 25.42	Class 2	32.14%	36.36%	88.89%
${}^1x \geq 3.24$ and E-state Index from 13.25 to 15.5	Class 2	3.57%	4.55%	100%
1x from 3.96 to 4.71 and Kappa Alpha Order 1 ≥ 5.85	Class 2	25%	27.27%	85.71%
1x from 4.76 to 5.65	Class 2	10.71%	13.64%	100%
${}^1x < 3.68$ and Kappa Alpha Order 1 ≥ 6.68	Class 2	3.57%	4.55%	100%
${}^1x^V \geq 4.09$	Class 2	35.71%	40.91%	90%
${}^1x^V$ from 3.05 to 3.54	Class 2	28.57%	31.82%	87.5%
${}^1x^V$ from 3.64 to 4.07	Class 2	7.14%	9.09%	100%
E-state Index ≥ 25.58 and Kappa Alpha Order 2 from 2.96 to 6.4	Class 2	32.14%	36.36%	88.89%
E-state Index from 14.42 to 25.42 and Kappa Alpha Order 2 from 2.76 to 3.46	Class 2	17.86%	18.18%	80%
Kappa Alpha Order 2 from 2.4 to 2.5	Class 2	3.57%	4.55%	100%
E-state Index ≥ 36.33	Class 2	14.29%	18.18%	100%

Table A3. Cont.

Rule	Glass Transition Temperature (T_g)			
	Decision	Strength	Coverage	Certainty
E-state Index from 20.92 to 30.01 and Kappa Alpha Order 3 \geq 3.96	Class 2	10.71%	13.64%	100%
E-state Index \geq 30.42 and Kappa Alpha Order 3 \geq 5.16	Class 2	3.57%	4.55%	100%
E-state Index from 14.42 to 15.08	Class 2	3.57%	4.55%	100%
Kappa Alpha Order 3 from 1.32 to 1.92	Class 2	35.71%	45.45%	100%
E-state Index \geq 27.39 and Kappa Flexibility Index from 2.06 to 6.36	Class 2	25%	27.27%	85.71%
E-state Index from 20.92 to 27.25 and Kappa Flexibility Index from 1.71 to 3.21	Class 2	7.14%	9.09%	100%
E-state Index \geq 13.25 and Kappa Flexibility Index from 3.66 to 4.85	Class 2	17.86%	18.18%	80%
Kappa Flexibility Index from 4.86 to 6.36	Class 2	3.57%	4.55%	100%
E-state Index from 13.81 to 18.63 and Kappa Flexibility Index from 1.67 to 2.17	Class 2	7.14%	9.09%	100%
E-state Index from 25.58 to 30.31 and Kappa Order 1 \geq 7.06	Class 2	21.43%	22.73%	83.33%
E-state Index \geq 30.42 and Kappa Order 1 \geq 6.06	Class 2	25%	31.82%	100%
E-state Index from 17.67 to 24.43 and Kappa Order 1 \geq 7.06	Class 2	32.14%	36.36%	88.89%
E-state Index $<$ 15.5 and Kappa Order 1 \geq 6.06	Class 2	7.14%	9.09%	100%
E-state Index $<$ 25.42 and Kappa Order 1 \geq 8.05	Class 2	17.86%	18.18%	80%
E-state Index from 25.58 to 30.31 and Kappa Order 2 \geq 4.23	Class 2	17.86%	18.18%	80%
E-state Index \geq 30.42 and Kappa Order 2 \geq 3.22	Class 2	25%	31.82%	100%
E-state Index from 15.5 to 22.64 and Kappa Order 2 from 3.22 to 4.15	Class 2	21.43%	27.27%	100%
Kappa Order 2 \geq 7.82	Class 2	7.14%	9.09%	100%
E-state Index \geq 27.39 and Kappa Order 2 from 3.09 to 4.15	Class 2	7.14%	9.09%	100%
E-state Index \geq 30.42 and Kappa Order 3 \geq 3.06	Class 2	17.86%	22.72%	100%
E-state Index $<$ 15.5 and Kappa Order 3 \geq 5.36	Class 2	3.57%	4.55%	100%
E-state Index from 13.81 to 15.5 and Kappa Order 3 $<$ 3.92	Class 2	3.57%	4.55%	100%
E-state Index from 25.58 to 30.31 and Kappa Alpha Order 1 \geq 8.04	Class 2	17.86%	18.18%	80%

Table A3. Cont.

Rule	Glass Transition Temperature (T_g)			
	Decision	Strength	Coverage	Certainty
E-state Index ≥ 30.42 and Kappa Alpha Order 1 ≥ 5.85	Class 2	25%	31.82%	100%
E-state Index from 18.63 to 24.43 and Kappa Alpha Order 1 ≥ 6.98	Class 2	17.86%	18.18%	80%
E-state Index < 18.63 and Kappa Alpha Order 1 from 5.85 to 7.52	Class 2	10.71%	13.64%	100%
Kappa Order 2 from 5.16 to 6.98 and Kappa Alpha Order 3 ≥ 3.96	Class 2	7.14%	9.09%	100%
Kappa Order 2 ≥ 3.97 and Kappa Alpha Order 3 < 3.19	Class 2	17.86%	22.72%	100%
Kappa Order 2 from 3.16 to 3.93 and Kappa Alpha Order 3 < 2.3	Class 2	28.57%	31.82%	87.5%
Kappa Order 2 ≥ 4.23 and Kappa Alpha Order 3 < 3.93	Class 2	17.86%	18.18%	80%
Kappa Order 2 ≥ 7.32 and Kappa Alpha Order 3 ≥ 6.82	Class 2	10.71%	13.64%	100%
Kappa Order 2 from 3.09 to 3.16 and Kappa Alpha Order 3 ≥ 1.32	Class 2	3.57%	4.55%	100%
Kappa Order 2 < 5.72 and Kappa Alpha Order 1 ≥ 8.04	Class 2	25%	27.27%	85.71%
Kappa Alpha Order 2 ≥ 11.36	Class 2	17.86%	22.72%	100%
Kappa Order 3 ≥ 3.06 and Kappa Alpha Order 1 from 6.68 to 7.52	Class 2	3.57%	4.55%	100%
Kappa Order 2 from 3.09 to 3.93 and Kappa Alpha Order 1 ≥ 5.85	Class 2	32.14%	36.36%	88.89%
Kappa Order 2 < 6.98 and Kappa Alpha Order 1 ≥ 9.42	Class 2	10.71%	13.64%	100%
Kappa Order 3 from 4.49 to 7.09 and Kappa Alpha Order 2 ≥ 4.89	Class 2	7.14%	9.09%	100%
Kappa Order 3 < 3.59 and Kappa Alpha Order 2 ≥ 2.96	Class 2	21.4%	27.27%	100%
Kappa Alpha Order 2 ≥ 7.03	Class 2	10.71%	13.64%	100%
Kappa Order 3 < 4.37 and Kappa Alpha Order 2 ≥ 4.89	Class 2	3.57%	4.55%	100%
Kappa Order 3 < 2.6 and Kappa Alpha Order 2 ≥ 2.4	Class 2	42.86%	50%	91.67%
Kappa Order 3 < 3.92 and Kappa Alpha Order 2 ≥ 3.85	Class 2	7.14%	9.09%	100%
Kappa Order 3 from 1.6 to 2.6 and Kappa Flexibility Index ≥ 1.67	Class 2	35.71%	45.45%	100%
Kappa Order 3 from 5.36 to 7.09	Class 2	7.14%	9.09%	100%
Kappa Order 3 from 3.06 to 3.59	Class 2	7.14%	9.09%	100%

Table A3. Cont.

Rule	Glass Transition Temperature (T_g)			
	Decision	Strength	Coverage	Certainty
	Cohesive Energy (E_{coh})			
${}^0x \geq 2.5$ and E-state Index < 13.81	Class 1	23.5%	57.14%	100%
E-state Index < 10.75	Class 1	23.5%	57.14%	100%
0x from 2.5 to 3.78	Class 1	5.88%	14.29%	100%
Kappa Alpha Order 1 < 2.69	Class 1	11.77%	28.57%	100%
${}^0x < 4.7$ and Kappa Alpha Order 2 ≥ 1.73	Class 1	11.77%	28.57%	100%
${}^0x < 4.7$ and Kappa Flexibility Index ≥ 1.57	Class 1	11.77%	28.57%	100%
Kappa Order 3 ≥ 3.25 and Kappa Flexibility Index from 1.52 to 2.33	Class 1	11.77%	28.57%	100%
1x from 1.4 to 3.59 and E-state Index < 15.08	Class 1	23.53%	57.14%	100%
1x from 1.4 to 2.13	Class 1	11.77%	28.57%	100%
1x from 2.35 to 2.6	Class 1	5.88%	14.29%	100%
${}^1x < 3.59$ and Kappa Alpha Order 2 from 1.73 to 2.57	Class 1	17.65%	42.86%	100%
E-state Index < 13.81 and Kappa Order 1 ≥ 3.1	Class 1	23.53%	57.14%	100%
${}^1x < 2.6$ and Kappa Flexibility Index ≥ 1.57	Class 1	11.77%	28.57%	100%
E-state Index from 11.33 to 13.81	Class 1	11.77%	28.57%	100%
E-state Index < 13.81 and Kappa Alpha Order 1 ≥ 2.72	Class 1	23.53%	57.14%	100%
E-state Index < 13.81 and Kappa Alpha Order 2 ≥ 1.73	Class 1	11.77%	28.57%	100%
E-state Index < 20.92 and Kappa Alpha Order 3 ≥ 3.42	Class 1	11.77%	28.57%	100%
E-state Index < 13.81 and Kappa Flexibility Index ≥ 1.57	Class 1	11.77%	28.57%	100%
E-state Index < 13.81 and Kappa Flexibility Index < 1.52	Class 1	23.53%	57.14%	100%
${}^0x^V$ from 1.85 to 2.73	Class 1	17.65%	42.86%	100%
${}^1x^V < 1.07$	Class 1	11.77%	28.57%	100%

Table A4. Transport properties rules filtered from validation dataset.

Rule	O ₂ Permeability Decision	Strength	Coverage	Certainty
0x from 4.63 to 5.08	Class 2	5.56%	20%	100%
${}^0x^V$ from 4.55 to 4.62	Class 2	5.56%	20%	100%
1x from 2.71 to 2.78	Class 2	5.56%	20%	100%
1x from 2.92 to 3.12	Class 2	5.56%	20%	100%
${}^1x^V$ from 5.84 to 6.13	Class 2	5.56%	20%	100%
E-state Index < 18.24 and Kappa Order 3 \geq 4.67	Class 2	11.11%	40%	100%
Kappa Order 3 from 5.43 to 6.28	Class 2	5.56%	20%	100%
Kappa Alpha Order 1 from 5.52 to 5.75	Class 2	5.56%	20%	100%
Kappa Alpha Order 2 from 2.94 to 3.02	Class 2	5.56%	20%	100%
Kappa Alpha Order 3 from 4.87 to 5.51	Class 2	5.56%	20%	100%
O₂/N₂ Selectivity				
${}^0x \geq 9.92$	Class 2	25%	42.86%	75%
0x from 6.26 to 6.61 and Kappa Alpha Order 2 ≥ 2.94	Class 2	6.25%	14.29%	100%
${}^0x^V \geq 9.99$	Class 2	6.25%	14.29%	100%
${}^1x \geq 6.16$	Class 2	25%	42.86%	75%
${}^1x^V$ from 6.08 to 7.58	Class 2	12.5%	28.57%	100%
E-state Index from 22.09 to 23.82	Class 2	12.5%	28.57%	100%
E-state Index ≥ 31.88	Class 2	12.5%	28.57%	100%
Kappa Order 1 ≥ 12.22	Class 2	12.5%	28.57%	100%
Kappa Order 2 from 6.12 to 7.84	Class 2	6.25%	14.29%	100%

Table A4. Cont.

Rule	O ₂ Permeability Decision	Strength	Coverage	Certainty
Kappa Order 2 from 0.67 to 1.48	Class 2	6.25%	14.29%	100%
Kappa Order 3 < 2.3 and Kappa Alpha Order 2 \geq 2.38	Class 2	6.25%	14.29%	100%
Kappa Order 3 < 3.83 and Kappa Alpha Order 2 from 0.67 to 1.77	Class 2	6.25%	14.29%	100%
Kappa Order 3 < 3.83 and Kappa Alpha Order 2 \geq 4.03	Class 2	6.25%	14.29%	100%
Kappa Alpha Order 1 \geq 11.34	Class 2	12.5%	28.57%	100%
Kappa flexibility Index from 3.8 to 5.55	Class 2	6.25%	14.29%	100%
Kappa flexibility Index from 1.7 to 1.85	Class 2	6.25%	14.29%	100%
Kappa flexibility Index from 2.72 to 3.32	Class 2	6.25%	14.29%	100%

References

1. Gas Separation Membrane Market (2023–2032). The Business Research Company. 2023. Available online: <https://www.openpr.com/news/3068812/gas-separation-membrane-market-2023-2032-top-companies> (accessed on 7 June 2023).
2. Lasseguette, E.; Comesaña-Gándara, B. Polymer Membranes for Gas Separation. *Membranes* **2022**, *12*, 207. [CrossRef] [PubMed]
3. Murali, R.S.; Sankarshana, T.; Sridhar, S. Air separation by polymer-based membrane technology. *Sep. Purif. Rev.* **2013**, *42*, 130–186. [CrossRef]
4. Chong, K.C.; Lai, S.O.; Thiam, H.S.; Teoh, H.C.; Heng, S.L. Recent progress of oxygen/nitrogen separation using membrane technology. *J. Eng. Sci. Technol.* **2016**, *11*, 1016–1030.
5. Bell, J. What Is Machine Learning? In *Machine Learning and the City*; Wiley: New York, NY, USA, 2022; pp. 207–216. [CrossRef]
6. El-Banbi, A.; Alzahabi, A.; El-Maraghi, A. Artificial Neural Network Models for PVT Properties. *PVT Prop. Correl.* **2018**, 225–247. [CrossRef]
7. Tayyebi, A.; Alshami, A.S.; Yu, X.; Kolodka, E. Can machine learning methods guide gas separation membranes fabrication? *J. Membr. Sci. Lett.* **2022**, *2*, 100033. [CrossRef]
8. Pedrycz, W.; Succi, G. Genetic granular classifiers in modeling software quality. *J. Syst. Softw.* **2005**, *76*, 277–285. [CrossRef]
9. Pawlak, L.; Grzvmala-Busse, L.; Slowinski, R.; Ziarko, W. Rough Sets. *Commun ACM* **1995**, *38*, 88–95. [CrossRef]
10. Aviso, K.B.; Janairo, J.I.B.; Promentilla, M.A.B.; Tan, R.R. Prediction of CO₂ storage site integrity with rough set-based machine learning. *Clean Technol. Environ. Policy* **2019**, *21*, 1655–1664. [CrossRef]
11. Lei, L.; Chen, W.; Wu, B.; Chen, C.; Liu, W. A building energy consumption prediction model based on rough set theory and deep learning algorithms. *Energy Build.* **2021**, *240*, 110886. [CrossRef]
12. Heng, Y.P.; Lee, H.Y.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattualappil, N.G. Incorporating Machine Learning in Computer-Aided Molecular Design for Fragrance Molecules. *Processes* **2022**, *10*, 1767. [CrossRef]
13. Chong, J.W.; Ng, L.Y.; Aboagwa, O.A.; Thangalazhy-Gopakumar, S.; Muthoosamy, K.; Chemmangattualappil, N.G. Computer-Aided Framework for the Design of Optimal Bio-Oil/Solvent Blend with Economic Considerations. *Processes* **2021**, *9*, 2159. [CrossRef]
14. Pawlak, Z. Rough set approach to knowledge-based decision support. *Eur. J. Oper. Res.* **1997**, *99*, 48–57. [CrossRef]
15. Pawlak, Z. Rough sets, decision algorithms and Bayes' theorem. *Eur. J. Oper. Res.* **2002**, *136*, 181–189. [CrossRef]
16. Churi, N.; Achenie, L.E.K. Novel Mathematical Programming Model for Computer Aided Molecular Design. *Ind. Eng. Chem. Res.* **1996**, *35*, 3788–3794. [CrossRef]
17. Zhou, T.; McBride, K.; Zhang, X.; Qi, Z.; Sundmacher, K. Integrated solvent and process design exemplified for a Diels–Alder reaction. *AIChE J.* **2015**, *61*, 147–158. [CrossRef]
18. Harper, P.M.; Gani, R.; Kolar, P.; Ishikawa, T. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilib.* **1999**, *158–160*, 337–347. [CrossRef]
19. Sun, G.; Fan, T.; Sun, X.; Hao, Y.; Cui, X.; Zhao, L.; Ren, T.; Zhou, Y.; Zhong, R.; Peng, Y. In Silico Prediction of O6-Methylguanine-DNA Methyltransferase Inhibitory Potency of Base Analogs with QSAR and Machine Learning Methods. *Molecules* **2018**, *23*, 2892. [CrossRef]
20. Wang, F.; Cheng, H. Computer-aided biocompatible solvent design for an integrated extractive fermentation–separation process. *Chem. Eng. J.* **2010**, *162*, 809–820.
21. Ooi, J.; Ng, D.K.S.; Chemmangattualappil, N.G. Optimal molecular design towards an environmental friendly solvent recovery process. *Comput. Chem. Eng.* **2018**, *117*, 391–409. [CrossRef]
22. Scheffczyk, J.; Fleitmann, L.; Schwarz, A.; Lampe, M.; Bardow, A.; Leonhard, K. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chem. Eng. Sci.* **2017**, *159*, 84–92. [CrossRef]
23. Yee, Q.Y.; Hassim, M.H.; Chemmangattualappil, N.G.; Ten, J.Y.; Raslan, R. Optimization of quality, safety and health aspects in personal care product preservative design. *Process Saf. Environ. Prot.* **2022**, *157*, 246–253. [CrossRef]
24. Satyanarayana, K.C.; Abildskov, J.; Gani, R.A. Computer-aided polymer design using group contribution plus property models. *Comput. Chem. Eng.* **2009**, *33*, 1004–1013. [CrossRef]
25. Guo, W.; Chai, S.; Zhang, L.; Du, J. Computer-Aided Design of Crosslinked Polymer Membrane Using Machine Learning and Molecular Dynamics. *Chem. Ing. Tech.* **2022**, *95*, 447–457. [CrossRef]
26. Zhang, L.; Mao, H.; Liu, L.; Du, J.; Gani, R. A machine learning based computer-aided molecular design/screening methodology for fragrance molecules. *Comput. Chem. Eng.* **2018**, *115*, 295–308. [CrossRef]
27. Ooi, Y.J.; Aung, K.N.G.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Chemmangattualappil, N.G. Design of fragrance molecules using computer-aided molecular design with machine learning. *Comput. Chem. Eng.* **2022**, *157*, 107585. [CrossRef]
28. Radhakrishnapany, K.T.; Wong, C.Y.; Tan, F.K.; Chong, J.W.; Tan, R.R.; Aviso, K.B.; Janairo, J.I.B.; Chemmangattualappil, N.G. Design of fragrant molecules through the incorporation of rough sets into computer-aided molecular design. *Mol. Syst. Des. Eng.* **2020**, *5*, 1391–1416. [CrossRef]
29. Chemmangattualappil, N.G. Development of solvent design methodologies using computer-aided molecular design tools. *Curr. Opin. Chem. Eng.* **2020**, *27*, 51–59. [CrossRef]
30. Zhang, L.; Mao, H.; Liu, Q.; Gani, R. Chemical product design—recent advances and perspectives. *Curr. Opin. Chem. Eng.* **2020**, *27*, 22–34. [CrossRef]

31. Harlacher, T.; Wessling, M. Gas–Gas Separation by Membranes. In *Progress in Filtration and Separation*; Academic Press: Cambridge, MA, USA, 2015; pp. 557–584. [[CrossRef](#)]
32. Liu, Y.; Li, N.; Cui, X.; Yan, W.; Su, J.; Jin, L. A Review on the Morphology and Material Properties of the Gas Separation Membrane: Molecular Simulation. *Membranes* **2022**, *12*, 1274. [[CrossRef](#)]
33. Bicerano, J. *Prediction of Polymer Properties*; CRC Press: Boca Raton, FL, USA, 2002. [[CrossRef](#)]
34. Eichenhofer, M.; Arreguin, S.; Wong, J. *Neurogastroenterology and Motility*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2019; pp. 1–5.
35. Van Krevelen, D.W.; Nijenhuis, K.T. Chapter 1—Polymer Properties. In *Properties of Polymers*; Elsevier: Amsterdam, The Netherlands, 2009; pp. 3–5.
36. Jia, L.; Xu, J. A simple method for prediction of gas permeability of polymers from their molecular structure. *Polym. J.* **1991**, *23*, 417–425. [[CrossRef](#)]
37. Rahman, M.M. Membrane Separation of Gaseous Hydrocarbons by Semicrystalline Multiblock Copolymers: Role of Cohesive Energy Density and Crystallites of the Polyether Block. *Polymers* **2021**, *13*, 4181. [[CrossRef](#)] [[PubMed](#)]
38. Koros, W.J.; Mahajan, R. Pushing the limits on possibilities for large scale gas separation: Which strategies? *J. Memb. Sci.* **2000**, *175*, 181–196. [[CrossRef](#)]
39. Ivanciuc, O. *Electrotopological State Indices*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2007; pp. 85–109. [[CrossRef](#)]
40. Hall, L.H.; Kier, L.B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. *Rev. Comput. Chem.* **2007**, *2*, 367–422. [[CrossRef](#)]
41. Calibration, M.; Kier, L.B. I21 Index of Molecular Flexibility from Kappa Shape Attributes. *Comput. Chem.* **1989**, *8*, 735. [[CrossRef](#)]
42. Martin, T. *User's Guide for T. E. S. T. (Toxicity Estimation Software Tool) Version 5.1 A Java Application to Estimate Toxicities and Physical Properties from Molecular Structure*; US Environmental Protection Agency: Cincinnati, OH, USA, 2020.
43. Prędko, B.; Słowiński, R.; Stefanowski, J.; Susmaga, R.; Wilk, S. ROSE—Software implementation of the rough set theory. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; Springer: Berlin/Heidelberg, Germany, 1998; Volume 1424, pp. 605–608. [[CrossRef](#)]
44. Kagramanov, G.; Gurkin, V.; Farnosova, E. Physical and Mechanical Properties of Hollow Fiber Membranes and Technological Parameters of the Gas Separation Process. *Membranes* **2021**, *11*, 583. [[CrossRef](#)]
45. Eslick, J.C.; Ye, Q.; Park, J.; Topp, E.M.; Spencer, P.; Camarda, K.V. A computational molecular design framework for crosslinked polymer networks. *Comput. Chem. Eng.* **2009**, *33*, 954–963. [[CrossRef](#)]
46. Conte, E.; Martinho, A.; Matos, H.A.; Gani, R. Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity. *Ind. Eng. Chem. Res.* **2008**, *47*, 7940–7954. [[CrossRef](#)]
47. Cao, C.; Lin, Y. Correlation between the glass transition temperatures and repeating unit structure for high molecular weight polymers. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 643–650. [[CrossRef](#)]
48. Fried, J.R. *Polymer Science and Technology*, 3rd ed.; Pearson: London, UK, 2014.
49. Sulchek, T.A.; Friddle, R.W.; Noy, A. Counting and Breaking Single Bonds: Dynamic Force Spectroscopy in Tethered Single Molecule Systems. In *Handbook of Molecular Force Spectroscopy*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 251–272. [[CrossRef](#)]
50. Stevens, M.P. *Polymer Chemistry: An Introduction*, 3rd ed.; Oxford University Press: New York, NY, USA, 1999.
51. Mark, J.E. Polymer Data Polymer Data. *J. Am. Chem. Soc.* **2009**, *131*, 655–657.
52. AlMaadeed, M.A.; Ouederni, M.; Khanam, P.N. Effect of chain structure on the properties of Glass fibre/polyethylene composites. *Mater. Des.* **2013**, *47*, 725–730. [[CrossRef](#)]
53. Mohanty, A.D.; Bae, C. Transition Metal-Catalyzed Functionalization of Polyolefins Containing CC, CC, and CH Bonds. In *Advances in Organometallic Chemistry*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 64, pp. 1–39. [[CrossRef](#)]
54. Hearle, J.W.S. Textile Fibers: A Comparative Overview. In *Encyclopedia of Materials: Science and Technology*; Elsevier: Amsterdam, The Netherlands, 2001; pp. 9100–9116. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.