Membrane and Water Treatment, *Vol. 16, No. 6 (2025) 291-302*
https://doi.org/10.12989/mwt.2025.16.6.291

Prediction of gas permeation through polymers with intrinsic porosity using a hybrid neutral network-particle swarm model

Maroua Henni[1a], Hanaa Hasnaoui[1b], Mohamed Krea[*1] and Denis Roizard[2c]

[1]*Material and Environmental Laboratory, GPE department, Faculty of Technology, University of Medea 26000, Algeria*
[2]*Laboratoire Réactions et Génie des Procédés – CNRS 7274, Université de Lorraine, ENSIC, 1, rue Grandville – BP 20451, 54001 Nancy Cedex, France*

(Received June 25, 2025, Revised October 28, 2025, Accepted October 29, 2025)

Abstract. This study develops a quantitative structure-property relationship (QSPR) model using a hybrid neural network and particle swarm optimization (PSO) to predict the gas separation performance of 120 polymers of intrinsic microporosity (PIMs). Over 5000 descriptors, including topological, constitutional, functional groups, and geometrical properties, were computed using alvaDesc software. Genetic algorithm optimization combined with partial least squares regression was used to select relevant descriptors for predicting PIM permeability to N_2, CH_4, and CO_2. A hybrid neural network model with particle swarm optimization-based backpropagation (PSO-BP) algorithms was used for permeability prediction, and the results were compared to experimental published data. The PSO-BP model showed promising results, with root mean squared error (RMSE) values of 0.0048, 0.000743, and 0.0045 for CO_2, N_2, and CH_4, permeabilities respectively. Key descriptors for predicting PIM permeability are associated with multiple physicochemical properties, including GATS, 3D Morse, TDB, SpMax, MATS, CATS3D, RDF, and ATS descriptors. CO_2 permeability prediction requires more 3D descriptors than N_2 and CH_4.

Keywords: descriptors; gases; particle swarm optimization-based backpropagation; polymers of intrinsic microporosity; quantitative structure-property relationship

1. Introduction

Greenhouse gases are essentially composed of carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2O). The concentrations of these gases are rising swiftly in the atmosphere as a result of industrialization and modernization, contributing to global climate change, as described by Mohammed *et al* (2021). It is well known that CO_2 is mainly responsible for global warming (Metz *et al*. 2005).

A lot of efforts have been made to lower CO_2 emissions into the atmosphere. To further decrease these emissions, we need more effective CO_2-capture methods. Absorption is a well-known technology for reducing CO_2 levels in natural gas, but membrane technology is a great alternative. It is modular, more compact, consumes less energy, has lower operating costs, and offers higher energy efficiency and ease of operation thanks to its flexibility and scalability (Han and Ho 2021).

In the membrane separation process, pressure is often used to force a gas mixture across a membrane, where separation is achieved by the difference in individual gas permeabilities. The membrane's permeability (P_i), described by Fick's law, for a certain gas species i, determines the membrane process performance. For dense membranes, this permeability is very low and the most commonly used unit for permeability is Barrer, which is a non-SI unit named after Barrer. It represents a gas volume of 10^{-10} cm^3 (STP) crossing a membrane having a thickness of 1cm during a time of 1s when the pressure difference between the feed and permeate side is 1cmHg.

Using the solution-diffusion model of gas transport, the permeability of gas separation membranes may be determined by multiplying diffusivity (D) and solubility (S) (Crank, 1975): $Pi = Di \times Si$. When comparing gas i and gas j permeability, another performance measured is the membrane's selectivity between two gases i and j, named α, which is defined as the ratio (Kesting and Fritzsche 1993) $\alpha = Pi/Pj$.

A perfect membrane for separating two gases would ideally have both high permeability and selectivity at the same time. This is a challenge because all membranes have a trade-off, referred to as the Robeson trade-off. This trade-off is between permeability, which is how quickly molecules can pass through the membrane, and selectivity, which defines how well the desired molecules are separated from the rest.

To enhance permeability, polymers of intrinsic micro-porosity (PIMs) have been introduced as materials for films or membranes in different applications, and they are a relatively new type of macromolecule. PIMs generally enhance fractional free volume via inefficient chain packing to increase permeability while simultaneously stiffening the polymer backbone and improving solubility selectivity.

[*]Corresponding author, Professor
E-mail: krea.mohamed@univ-medea.dz
[a] Ph.D. candidate, E-mail: marouahen94@gmail.com
[b] Ph.D., E-mail: ha_hasnaoui@yahoo.fr
[c] Research Director, E-mail: denis.roizard@univ-lorraine.fr

ISSN: 2005-8624 (Print), 2092-7037 (Online)

PIMs can easily be processed into thin films that are highly permeable and selective for gas separations (Solomon *et al*. 2016).

Fang *et al*. have made appeals to the ab initio approach (Monte Carlo and molecular dynamics simulations) to calculate the sorption and diffusion of four gases (H_2, O_2, CH_4, and CO_2) in two membranes made from PIMs (PIM-1 and PIM-7) (Fang *et al*. 2010).

They also combined molecular simulations and ab initio calculations to study CO_2/N_2 permeation and separation in PIMs having different functional groups (cyano, trifluoromethyl, phenylsulfone, and carboxyl). This computational study provides microscopic insights into the role of functional groups in gas permeation and suggests that strong CO_2-philic groups should be chosen to functionalize PIMs films for efficient CO_2/N_2 separation (Fang *et al*. 2011).

Another approach based on supervised machine learning ML is also widely used to understand the relationship between polymer structure and its gas permeability performances.

Inputs of ML methods, based on the chemical structure of polymer repeat units, have been successfully applied to accurately predict numerous polymer properties, including glass transition temperature (Tao *et al*. 2021a, b, Chen *et al*. 2021), thermal conductivity (Wu *et al*. 2019), dielectric constants (Kanakkithodi *et al*. 2016), and organic photovoltaic properties (Sun *et al*. 2019).

The use of the chemical structure of polymer repeating units as an input for ML particularly using artificial neural networks ANN in QSPR method has also been observed to predict several polymer properties, such as the intrinsic viscosity of polymer solutions (Afantitis *et al*. 2006, Gharagheizi 2007), glass transition temperature (Palomba *et al*. 2012), refractive index (García-Domenech and Julián-Ortiz 2002, Xu *et al*. 2004), solubility parameters (Yu *et al*. 2006), and gas permeability through different polymers using the group contribution method (Hasnaoui *et al*. 2017, Wessling *et al*. 1994).

Although ANN is widely considered a predictive technique to solve engineering problems, two significant problems have occurred across the learning process: being trapped in local optima and slow convergence. To overcome these limitations, an effective solution is to design and optimize the weights and biases of the ANN model by using efficient optimization algorithms.

Miscellaneous optimization algorithms (OAs), such as the genetic algorithm (GA), imperialism competitive algorithm (ICA), simulated annealing, and particle swarm optimization (PSO), have been widely utilized to solve nonlinear and complicated engineering problems. These hybrid forecasting models, based on the powerful ability of global search, allow the weights and biases of an ANN network to be determined more effectively, thereby improving ANN performance prediction (Koopialipoor *et al*. 2019). Compared with genetic algorithms and, simulated annealing, PSO algorithm is a global and advanced algorithm with a strong ability to search global optimum. The most important advantages of PSO are the few parameters needed to adjust and its easy implementation (Zhang and Wu 2011).

The goal of this study was to create an efficient and precise prediction method for the permeability of three-gas CO_2, CH_4 and N_2 in PIMs by using only the polymeric structure of these PIMs and a back propagation (BP) ANN trained by the PSO algorithm. This was motivated by the success of ANN models based on PSO and its various variants in predicting gas solubility in polymers (Li *et al*. 2013, Mengshan *et al*. 2017), and, in particular, the BP algorithm, which can approximate continuous functions with acceptable accuracy, and is widely used in nonlinear modeling and approximation functions. It consists to an artificial neural network which represent a mathematical modeling technique that employs regression to obtain relationship between inputs (polymer structure molecular descriptors) and output (permeabilities of the polymer film to studied gas) using a large volume of available experimental data.

In this study, we demonstrate how, from a small set of 120 PIM polymers data and rational choice of the input data feeding ANN-PSO model, by using representative molecular descriptors, calculated using the alvaDesc software (2019), obtained from an optimized molecular chain of three repeating units, we elaborate a performant permeability predicting model.

2. Materials and Methods

The global approach followed to establish our hybrid model is depicted in Fig. 1, which is further detailed in the subsequent steps. To realize this work and achieve the optimization of BP-PSO model parameters, we used MATLAB software, particularly the Neural Network Toolbox Matlab (R2019b).

2.1 Database elaboration

A representative database containing 120 PIMs with their gas permeability, in Barrer at 25 to 35 °C, values is collected from literature for three kinds of gases (N_2, CH_4, CO_2). To better represent all PIMs families in this database, we included different PIMs classes based on their structure, namely ladder PIMs and polyimides of intrinsic microporosity (PIM-Pis).

• Ladder PIMs polymers: Most ladder PIMs are based on benzodioxane-, phenazine- and TB-based linkers (PIM-1, TPIM-1, PIM-TRIP-TB…).

• PIM-Pis: can be further divided into two general groups containing: 1) contorted dianhydride sites and/or 2) contorted diamine sites,

1) PIM-PIs containing dianhydride contortion sites (PIM-PI-1 to PIM-PI-11, KAUST-PI-1 to KAUST-PI-7…) and/or

2) PIM-PIs containing diamine contortion sites (PIM-PI-TB-1, PIM-PI-TB-2, TBDA1-SBI-PI….)

2.2 PIMs molecular descriptors determination

Macroscopic properties of the polymer membranes are closely related to the molecular structures of their constituent backbones. The backbones of PIMs contain aromatic rings connected by spiro-centers, which form a rigid and contorted structure with high free volume and intrinsic microporosity

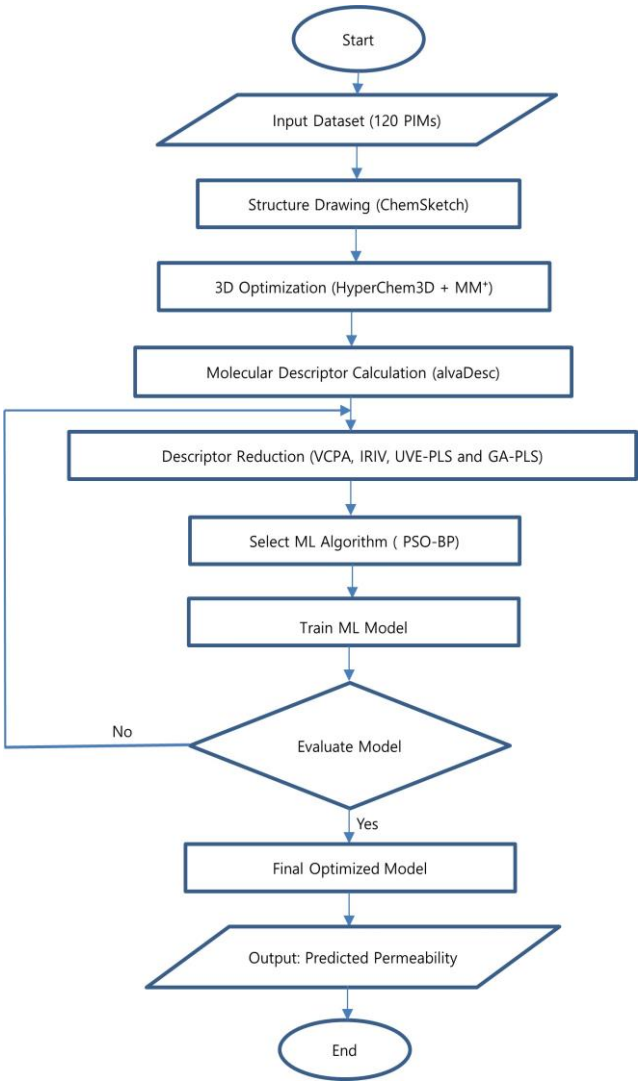

Fig. 1 work process diagram

leading to high permeability as well as good selectivity. To use the chemical information contained in the backbone structure of PIM we have used their molecular descriptors, which are numeric quantities containing chemical information about molecular structures. Depending on the algorithm of calculations or method of experimental determination or concept of the origin, the molecular descriptors can be classified as being one-dimensional (1D), two-dimensional (2D), three-dimensional (3D), or four-dimension (4D) descriptors. This classification is based on the level of molecular representation required for calculating molecular descriptors.

The 1D descriptors represent information calculated from the molecular formula of the molecule, which includes the number and type of atoms and the molecular weight.

The 2D descriptors represent molecular information about constitutional descriptors like size, shape, and electronic distribution in the molecule obtained from the graph of the molecule. These descriptors also represent the molecular fingerprint, encoding property information like atom pairs and fragment descriptors. two-dimensional (2D) materials have been widely used in separation films because of their interlayer structure. (Xuan *et al*. 2023)

The 3D descriptors (Niazi and Mariam 2023) describe mainly properties that are related to the 3D conformation of the molecule in space, such as quantum-chemical calculation-based descriptors, size, steric, surface and volume descriptors. Considering the spatial arrangement of atoms, they comprise autocorrelation descriptors, substituent constants, quantum, chemical descriptors, 3D-MoRSE descriptors, WHIM and GETAWAY descriptors. GETAWAY descriptors block. It is the special eigenvalue of the influence/distance matrix weighted by atomic polarizabilities (Alicja *et al*. 2010).

4D descriptors (Fourches and Ash 2019) encompass properties that change over time or involve spatiotemporal aspects. These dynamic molecular descriptors include electrostatic, hydrophobic or hydrogen bonding potentials, in “all” points of space surrounding the molecule (often on a grid), the GRID and the Comparative Molecular Field Analysis (CoMFA) descriptors.

The calculation of over 5000 of these molecular descriptors is determined by the most recent powerful tool alvaDesc software.

Before determining these molecular descriptors for PIMs, their repeating unit is first drawn using Chemsketch freeware and triplicated to represent an oligomer then converted to .hin extension file. The resulting file is open in HyperChem3D (8.0.7) software and subjected to structural optimization in a two-step: A raw optimization by Molecular Mechanics (MM+) followed by the quantum calculation of molecular electronic structure using Austin Model1 (AM1) HyperChem (8.0.7) at a convergence criterion of 0.01kcal^{-1} mol^{-1}.

Molecular descriptors of each optimized PIM structure were calculated by AlvaDesc software.

2.3 Reduction of descriptors number

Since alvaDesc software gives a complete list of over 5000 molecular descriptors which will be taken as input variables to the ANN model, it is crucial to select only the appropriate descriptors from this complete list of descriptors by removing those that are irrelevant or redundant. This step reduces the number of descriptors to retain only those relevant to the prediction of permeability of PIMs to studied gases. To find the correct size of the relevant sub-space of descriptors (number of retained molecular descriptors for each gas), we have used different variable selection supervised methods, (Mehmood *et al*. 2012) including:

Variable combination population analysis (VCPA)

Iteratively retaining informative variables (IRIV)

Uninformative variable elimination in PLS (UVE-PLS)

Genetic algorithm combined with Partial Least Squares regression (GA-PLS)

2.4 Hybrid neural network

2.4.1 Choice of ANN architecture:

Artificial neural network (ANN) models have been attractive and capture information from data (Abiodun and Jantan 2019), it has been an intelligent system for prediction

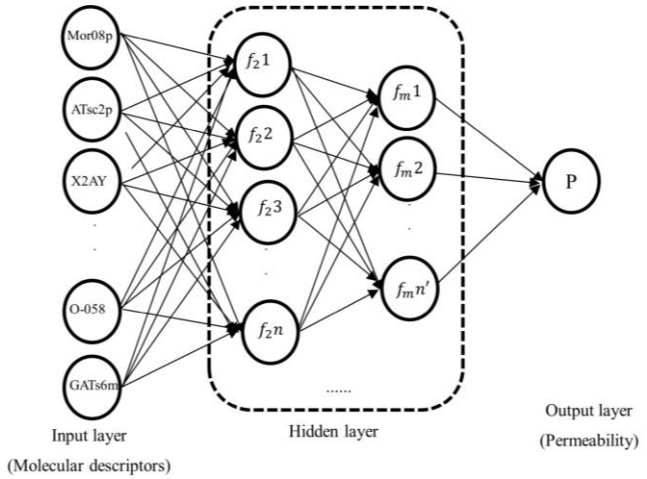

Fig. 2 Scheme of artificial neural networks

(Saikiaa and Baruah 2020) and classification. ANN is composed of many layers, input, output layers and at least one hidden layer in most applications (Saemi *et al*. 2007), A network diagram of a neural network model is presented in Fig. 2 (Wua *et al*. 2021).

$$h = \begin{cases} n + 0.618(n - m), & n \geq m \\ m - 0.618(m - n), & n < m \end{cases} \quad (1)$$

where n represents the input node number (number of retained descriptors for each gas), m is the output node's number (only 1 output layer in our case) and h is the number of nodes in the hidden layer.

2.4.2 Optimization of ANN by PSO

PSO (Kennedy and Eberhart 1995) algorithm was implemented as a training algorithm to accelerate the updating of the weights and biases of multilayer feed-forward network due to its outstanding features, including robustness, simplicity, flexibility, superior convergence characteristics and many others (Ahmadi *et al*. 2015) We presented the flowchart of the combined algorithm ANN-PSO in Fig. 3.

2.4.3 Parameter of particle swarm algorithm

1) Inertia weight w: The role of inertia weight w is to optimize the development and detection ability of PSO algorithm. It controls the influence that the previous speeds of particles have on their current ones. If w is small, it can improve the convergence speed, which is very beneficial for local optimization. When w is large, it is possible to make algorithms optimize the parameters in the global scope, which helps avoid being trapped in local optima. In our case the value of w is adaptive in [0.3, 0.9].

2) Learning factors c1, c2: Learning factors c1, c2 determine the impact of individual extremes and population extremes for the current speeds of the particles in PSO algorithm. In our case, the values of c1 c2 are fixed to 1.5 and 2.0 respectively.

3) The fitness function of PSO: In this study, the PSO algorithm is used to optimize BP neural network initial weights and biases. The mean square error (MSE) function is used as a fitness evaluation function of particles (Cheng

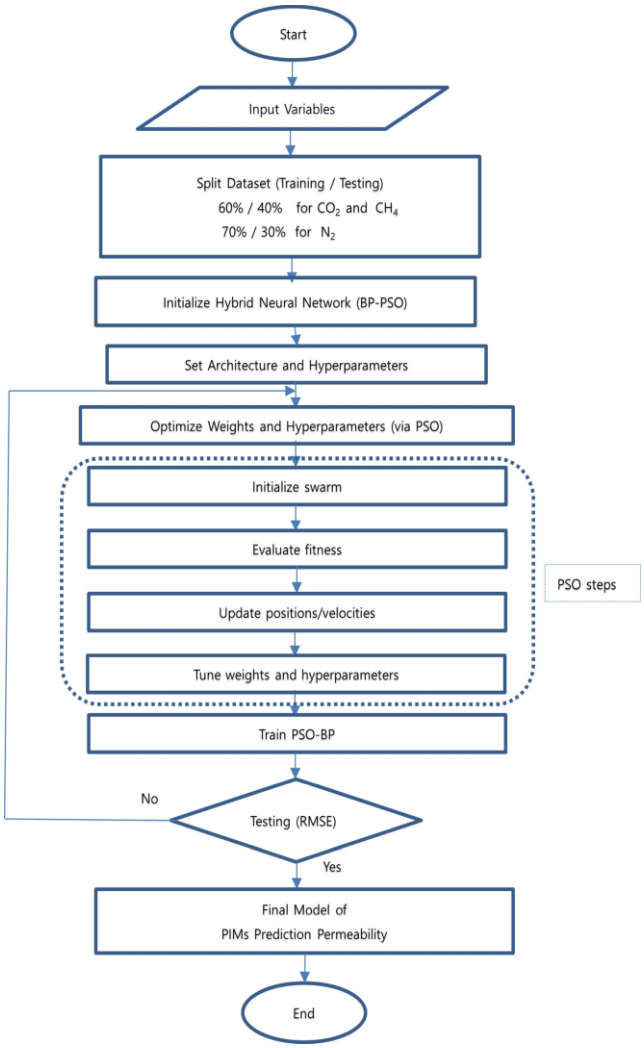

Fig. 3 Flowchart of the BP-PSO methodology

and Feng 2015) Eq. (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(P_{exp_i} - P_{pred_i})^2 \quad (2)$$

2.4.4 Evaluation of AN-PSO model Performances:

Statistical parameters are frequently used to assess an established ANN model's performance. We computed the root mean square error (RMSE), square of the correlation coefficient or the determination factor (R2), and performance factor (C) where their expressions are:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_{exp} - P_{pred})^2} \quad (3)$$

where P_{exp} , P_{pred} , and n are experimental values, predicted values, and the number of samples, respectively.

$$R = \frac{\sum_1^n (P_{exp} - \overline{P_{exp}})(P_{pred} - \overline{P_{pred}})}{\sqrt{\sum_1^n (P_{exp} - \overline{P_{exp}})^2 (P_{pred} - \overline{P_{pred}})^2}} \quad (4)$$

Table 1 Retained relevant descriptor number based on the minimum MSE of different variable selection methods studied in this work

	VCPA	GA-PLS	PLSUVE	IRIV	Ndesc** (by GA-PLS)
N_2	399	266	513	373	19
CH_4	621	593	$1.4*10^3$	950	14
CO_2	$4.26*10^3$	$3.1*10^3$	$4.8*10^3$	$1.1*10^4$	44

**Reduced number of desriptors

Table 2 Optimized structure of the neural networks

Gases	Input layer	Hidden layer		Output layer		Learning algorithm
	Number of neurons	Number of neurons	Activation function	number of neurons	Activation function	
N_2	19	30	Tansig[a]	1	purelin[b]	Train LM[c]
CH_4	14	25	Tansig	1	purelin	Train LM
CO_2	44	66	Tansig	1	purelin	Train LM

[a]hyperbolic tangent functions. [b]linear functions. [c]Levenberg-Marquardt backpropagation

To measure the predictive power of different network topologies, a performance factor C defined by Wessling *et al*. (1994), in the Eq. (5) below:

$$C = \frac{\sum_1^n (P_{pred} - P_{exp})}{\sum_1^n (\overline{P}_{exp} - P_{exp})} \quad (5)$$

where P_{exp} and P_{pred} are the experimental and the calculated value, respectively; $\overline{P_{exp}}$ and $\overline{P_{pred}}$ represent the average of experimental and predicted permeability, respectively; n is the number of compounds in the data set.

The predictive squared correlation coefficient, Q^2, was taken as a measure of the prediction performance of the model, which simply verifies how small the differences are between experimental data and external dataset predictions, can be written as:

$$\boldsymbol{Q^2_{F1}} = \mathbf{1} - \frac{\sum_{i=1}^{n_{EXT}} (P_{exp} - P_{pred})^2}{\sum_{i=1}^{n_{EXT}} (P_{exp} - \overline{P_{exp}}_{TR})^2} \quad (6)$$

where $\boldsymbol{n}$ is the total number of objects in the entire data set, TR= training set, EXT= external prediction set, P_{exp_i}= experimental data values, P_{pred_i} = predicted data values, and $\bar{P}_{exp}$ = average of the experimental data values (Chirico and Gramatica 2012).

3. Results and discussion

3.1 Descriptors determination

After calculating PIMs descriptors using alvaDesc, we obtained over 1398 descriptors, resulting in a matrix of 120 × 1398. This matrix needs cleanup, as it contains insignificant values, through the following steps:

- Filter out non-informative descriptors (i.e., constant descriptors)
- Eliminate all descriptors having more than 75% identical values
- Delete all descriptors with a Relative Standard Deviation (RSD) <0.05.

After these steps we obtain 1143 descriptors, which must be further reduced so that the input variables to the ANN may be as informative as possible for the target variable (permeability) as much as possible to improve model performance and avoid over fitting.

Generally, the thumb rule followed in ANN model training is 10 to 20 samples for each feature, hence, with our 120 samples dataset, we must do further reduction of the descriptor number by variable selection method until we fulfil this rule. Variable reduction may improve model performance, eliminate some useless redundancies from the model, and using a small number of variables necessary for ANN model prediction also means that we are maximizing the influence of each variable in the final model.

In Table 1 We present the best results obtained in terms of mean squared error (MSE), for each gas (N_2, CH_4, CO_2), where the colored values in each row indicate the selected methods that produce minimum MSE with the number of retained descriptors considered as relevant shown in the last column.

We observe that for all gases, GA-PLS gives the best result in terms of minimum MSE among the studied descriptors reduction methods. We also notice that CO_2 requires more descriptors (44 descriptors) followed by N_2 (19 descriptors) and CH_4 with the lowest value (14 descriptors).

3.2 Machine learning (ML) methods to predict permeability of PIMs

The main goal of this study is to evaluate the potential of an optimized neural networks to simulate permeability of PIMs to gases based on existing experimental data. In this study, an architecture of three-layer BP neural network trained by the PSO algorithm (PSO-BP ANN) is optimized to predict gas

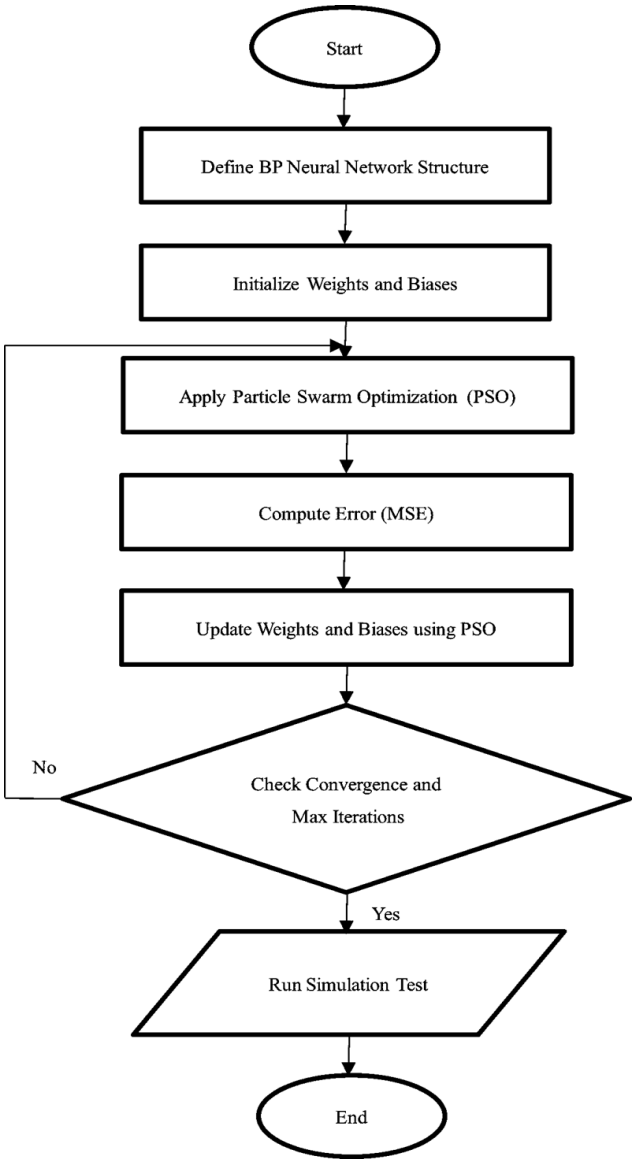

Fig. 4 BP-PSO neural algorithm process

permeability in polymers. The PSO-BP ANN architecture for each gas is Ndesc–h–1, where the corresponding values are given in Table 1. The input layer with Ndesc nodes represents the number of representative descriptors retained for each gas by GA-PLS algorithm from the descriptor reduction step. The number of nodes h in the hidden layer for each gas is determined by Eq. (1) using its appropriate Ndesc value. The output layer containing one node represents gas permeability in PIMs.

The activation function plays an important role in the training of neural networks. It provides the necessary nonlinearity required for the model to learn complex representations. Based on our earlier work (Hasnaoui *et al.* 2017) and several attempts we have chosen the Tansig as activation functions in the hidden layer, "purelin" in output layers and "trainlm" as training algorithms available in the Neural Network Toolbox as mentioned in Table 2.

The flowchart of the PSO-BP Neural Network learning algorithm is shown in Fig. 4.

To optimize the neural network PSO-BP, the PSO algorithm was used to find the optimal initial weights and thresholds for the BP neural network. The particles in the swarm represent different sets of weights, and their positions are updated based on their own and the global best positions (Mulumba *et al.* 2023)

The training stages of the BP neural network are outlined as follows: (Cao *et al.* 2016)

The initial step is to establish the network. The structure of the network, the desired output, and the learning rate are based on the number of descriptors obtained through GA-PLS for each gas.

The PSO algorithm fine-tunes individual solutions for the initial weight and threshold of the network.

The following step is to input the training sample and compute the output of the network layers, then calculate the learning error of the network. We keep adjusting the connection weight values and thresholds of the layers until the RMSE meets the expected standards and we check if the number of iterations has hit the training limit. If either of these conditions is satisfied, the training concludes. If not, we continue with the iterative learning process.

After training, the network was tested on fresh input/output data to evaluate its generalization capabilities. Our BP-PSO model was optimized by several tests and extensive calculation time varying the size of training and testing sets in order to lower the value of RMSE. We have obtained the best split of our dataset into a training dataset n_{tra} containing 60% (n_{tra} = 72 PIMs) and 40% as test n_{test} dataset (n_{test} = 48 PIMs) for CO_2 and CH_4 gas, and 70% (n_{tra} = 84 PIMs) and 30% as test n_{test} dataset (n_{test} = 36 PIMs) for N_2 gas.

Fig. 5 depicts the neural network regression graphs for the experimental and predicted gas permeability for different PIM films by our built PSO-BP model and the correlation between Pexp and Ppre values, where the straight line "Y = X" indicates that the predicted value is equal to the experimental value.

$$Q_{ih} = \frac{|W_{ih}|}{\sum_{i=1}^{ni}|W_{ih}|} \tag{7}$$

$$\mathrm{RI}(\%)_{i=} \frac{\sum_{h=1}^{nh} Q_{ih}}{\sum_{h=1}^{nh}\sum_{i=1}^{ni} Q_{ih}}\ x\ 100 \tag{8}$$

We obtained predicted values very close to experimental data points (practically all are following the straight line of Y = X, for the studied gas, both in the learning and the testing datasets.

We have summarized in Table 3 some statistical parameters representing the fitting goodness of our model. We have achieved a correlation coefficient (R) value for CO_2, N_2 and CH_4 equal to 1, which emphasized the robustness and accuracy of the obtained neural network models.

3.3 Analysis of descriptor type and meaning

The selection of suitable descriptors is critical for constructing reliable and affordable models. It helps to mitigate overfitting and uncovers the key characteristics relevant to a particular target.

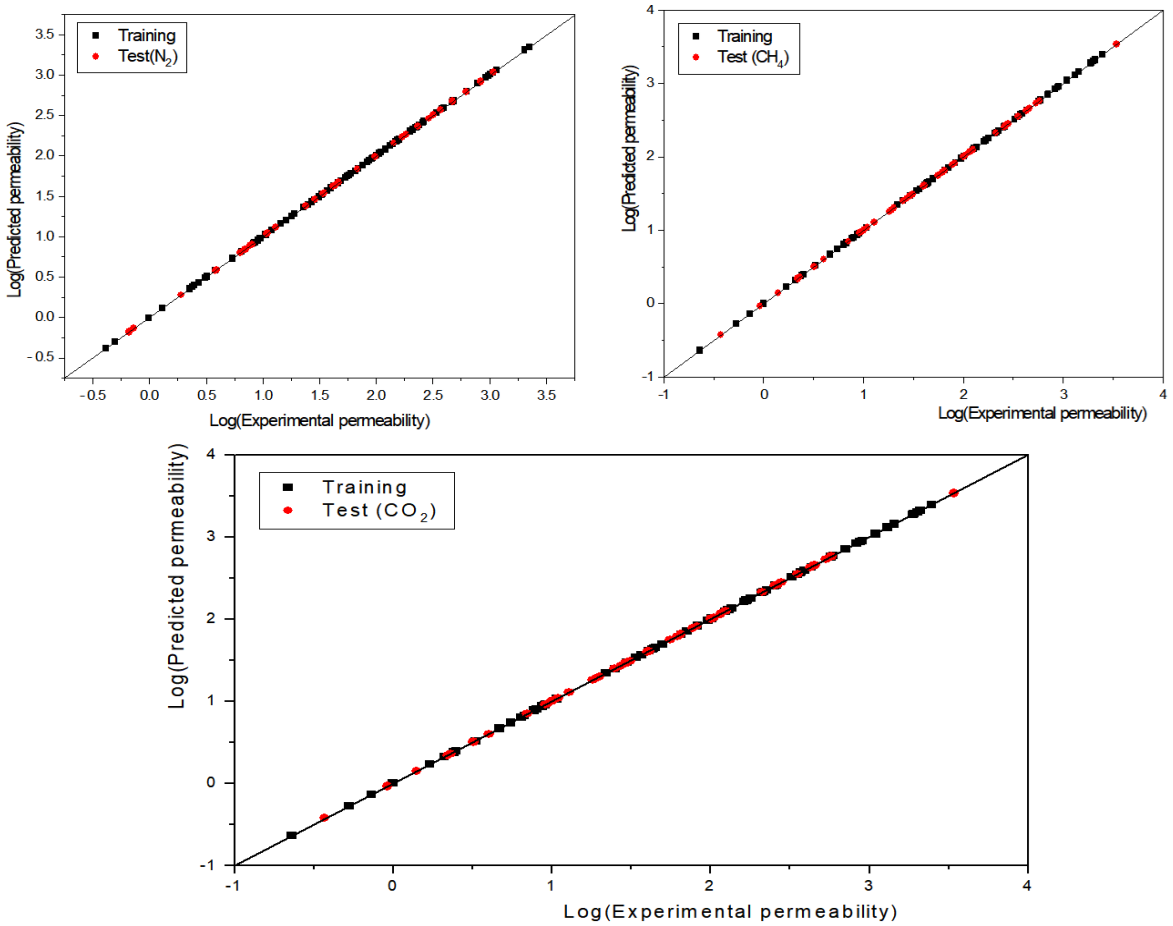

Fig. 5 Correlation of experimental and predicted gas permeability

Table 3 Goodness of fitting indicators of PSO-BP model

Indicator	Gas			Reliability threshold (Chirico and Gramatica 2012)
	CO_2	N_2	CH_4	
R_{Test}	1	1	1	Near or equal to 1
R_{global}	1	1	1	Near or equal to 1
$R^2_{adjusté}$	1	1	1	Near or equal to 1
$RMSE_{Test}$	$7.6*10^{-3}$	$2.70*10^{-4}$	$7.2*10^{-3}$	-
$RMSE_{global}$	$4.8*10^{-3}$	$7.43*10^{-4}$	$4.5*10^{-3}$	-
n_{tra}	72	84	72	-
n_{test}	48	36	48	-

Hence, a sensitivity study was carried out using Matlab software to determine the significance of each descriptor in predicting the permeability of PIMs in the examined gases. This relevance is based on partitioning the connection weights to assess the relative importance of the various inputs, as first proposed by Garson (1991) and repeated by Goh (1995), It involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron, Ge *et al.* (2003) as mentioned by Eqs. (7) and (8).

Where h represent hidden neuron,|W_ih |: the absolute value of the input-hidden layer connection weight, ni numbber: input neuron, nh: hidden neuron number and RI(%): The relative importance of each input variable which represents descriptors importance in our case.

The importance of retained descriptors for predicting PIM permeability, reported by their relative importance for each gas, is shown in descending order in Fig. 6.

There is a notable absence of insignificant descriptors. This means that the GA-based descriptor selection technique has eliminated all unimportant descriptors, and the PSO-BP ANN model considers all descriptors to have nearly comparable relevance. We have collected in Table 4 the most important descriptors used to predict PIM permeability for the studied gases, each associated with multiple physicochemical properties (m: mass, e: electro-

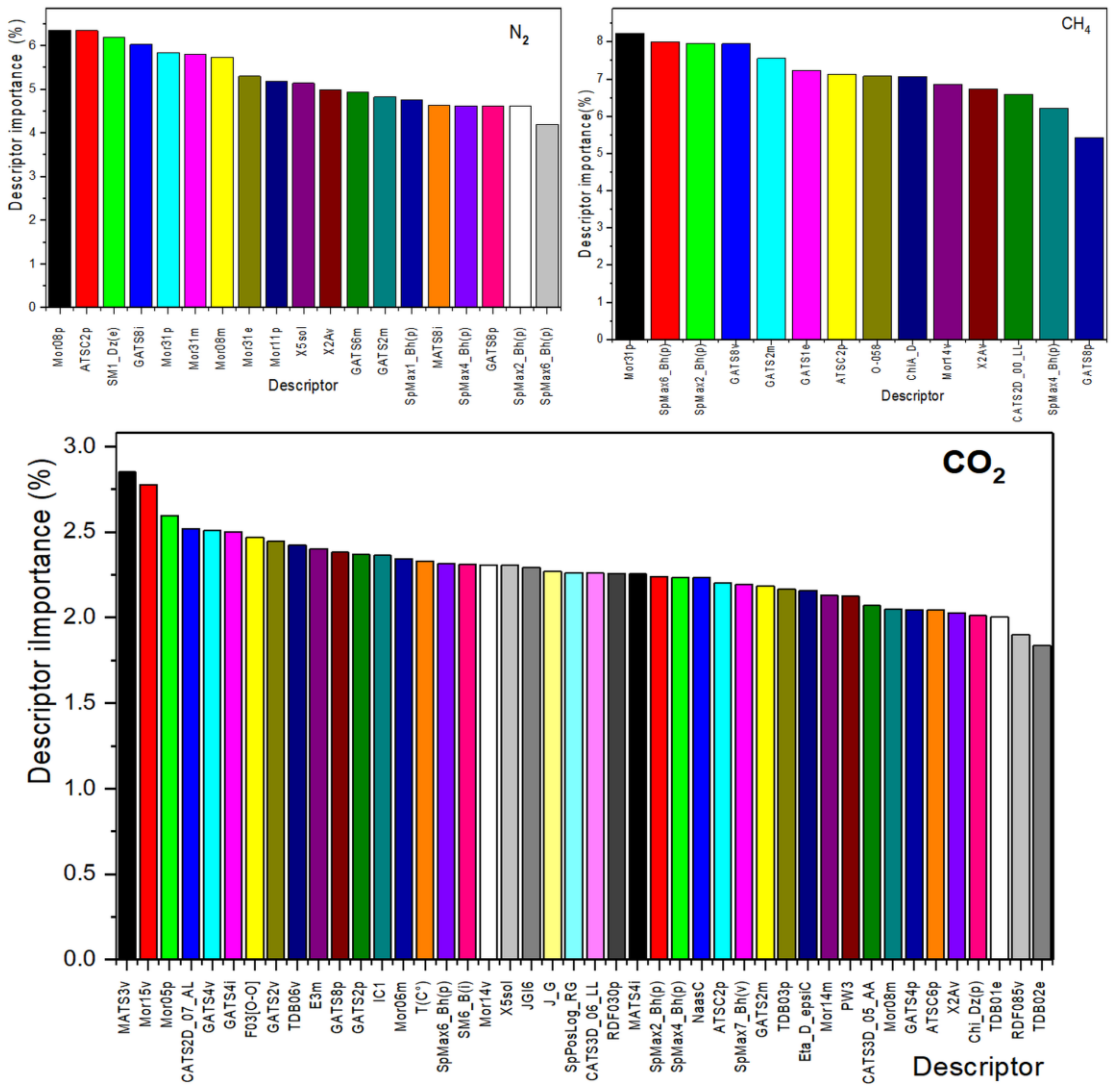

Fig. 6 Importance of each descriptor involved in the prediction of PIMs permeability to different gases (N_2, CH_4 and CO_2)

Table 4 Most important descriptors, used to predict PIM permeability of gasses, related to different chemical properties (m: mass, e: Sanderson electronegativity, p: polarizability, v: Vander Walls volume, AA: acceptor-acceptor and LL: lipophyl-lipophyl)

Descriptors		Gas		
		CO_2	N_2	CH_4
Class (2D)	ATS	2(2p,6p)	1(2p)	1(2p)
	GATS	7(2m, 2p, 2v, 4p, 4i, 4v, 8p)	4(2m, 6m, 8p, 8i)	4(1e, 2m, 8p, 8v)
	MATS	2(4i,3v)	1(8i)	
	SpMax	4(2, 4, 6, Bh(p),7Bh(v))	4(1, 2, 4, 6 Bh(p))	3(2, 4, 6 Bh(p))
Class (3D)	Mor	6(06, 08, 14m, 05p, 14, 15v)	6(31e, 08m, 31m, 08p, 11p, 31p)	2(31p, 14v)
	TDB	4(01e, 02e, 03p, 06v)		
	RDF	2(030p, 085v)		
	CATS3D	2(05_AA, 06_LL)		

negativity, p: polarizability, v: Van der Waals volume, AA: acceptor-acceptor, and LL: lipophil-lipophil). We note the presence of GATS descriptors (7 times) followed by Morse descriptors (6 times), TDB and SpMax (4 times), and

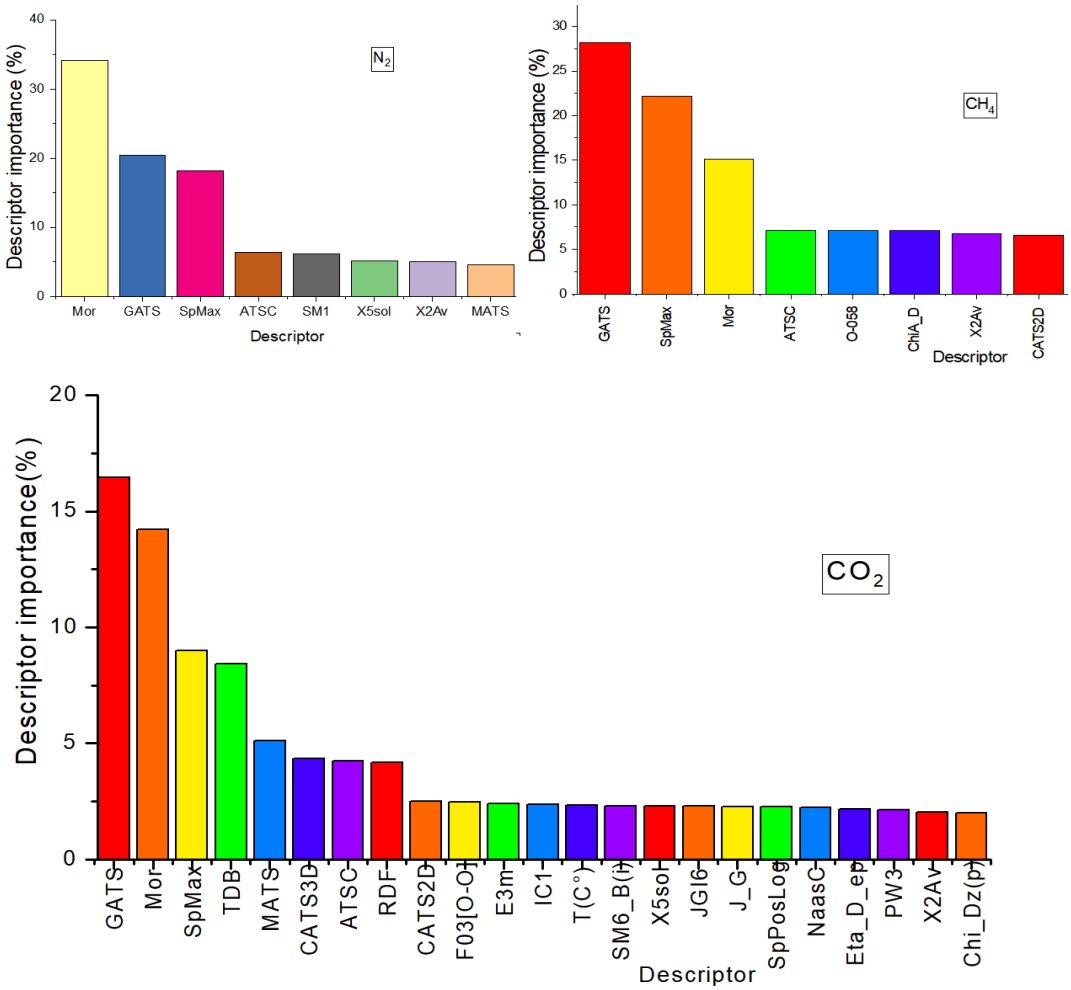

Fig. 7 Importance of different descriptor types in the prediction of PIMs Permeability

MATS, CATS3D, RDF, and ATS (2 times).

We note that CO_2 needs more 3D descriptors (3D Morse, RDF, TDB and CATS3D) followed by N_2 and CH_4 (3D Morse only).

To view the importance of each class of descriptor involved in the prediction of PIMs to studied gases, we have grouped descriptors types and reported them in the Fig 7. In all gas we observe the dominance of three descriptors :3D morse, GATS and SpMax.

The importance of 3D descriptors (especially 3D Morse) reaches 34% in prediction of N_2 gas followed by 31% (distributed on 3D morse, TDB, CATS3D and RDF) in the case of CO_2 and less 15% for CH_4.

It might be challenging to interpret a QSAR model in terms of the precise contribution of substituents and other chemical properties to the predicted activity. (Devinyak *et al*. 2014, Abreu *et al.*2009).

The 2D GATS descriptors, involved in the permeability prediction of the three studied gases, stand for the Geary coefficient descriptor. GATS descriptors are a type of autocorrelation descriptor calculated based on the Geary autocorrelation function which involves measuring the similarity or correlation of a property between different atoms or bonds at varying distances within a molecule. (Łapińska *et al*. 2024) They represent a set of molecular descriptors that describe the spatial distribution of atom or bond properties in a molecule.

They are extracted from representations of compounds, and they encompass important information, such as molecule size, degree of branching, flexibility, the neighborhood of atoms as nitrogen, the effect of electronegative oxygen and halogens atoms (Todeschini and Consonni 2009). They produce information about electrostatic and steric effects, and the overall 2D shape (Gupta *et al*. 2005). GATS descriptors are also involved in compounds with the highest biological activity when studied molecules have halogen atoms (for example F in PIM and Br with large polarizing power in the hydrophobic region (da Silva *et al*. 2021).

GATS and MATS describe how the considered property (permeability in our case) is distributed along the topological structure. They indicate that the presence of polarizable atoms pair (da Silva *et al*. 2021) at topological distance equal to (2, 4 and 8 for CO_2; 2, 6, and 8 for N_2; 1, 2 and 8 for CH_4) that contribute to permeability. Polarizable pair atoms at topological distance bonds cited above is most likely to be achieved between the presence of polarizable atoms in PIMs rings may give a positive contribute to permeability as in the example of some PIM repeat units structure illustrated in Fig. 8.

The SpMax-series descriptors, also called spectral

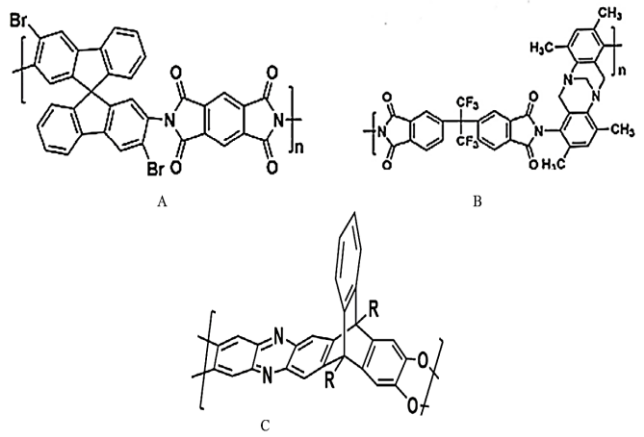

Fig. 8 Example of the presence of polar atom pairs in PIMs PIMS at different lag or topological distances: A) lag 2 or 6 (lag 2: O and N in imide cycle and lag 6: N-N in imide cycle); B) with lag 4 (N in imide and the other N atoms); C) lag 8 (between N and O atoms)

indices of a molecule, are correspondingly considered correlated with the molecular branching and show the great influence of molecular weight and molecular branching on a considered property (Yang *et al*. 2021).

Concerning 3D-MoRSE descriptors which are largely implicated in the prediction of PIMs permeability of PIMs to CO_2, they correspond to the representation of structures based on Electron diffraction. 3D-MoRSE descriptors are based on the concept of obtaining information from the 3D atomic coordinates by the transform used in electron diffraction studies for preparing theoretical scattering curves. They represent a very flexible 3D structure encoding framework for chemoinformatics that depend on interatomic distances subjected to random variations both in nature (molecular vibration) and in computational studies (due to stochastic methods of molecular geometry optimization of molecules in 3 Dimensions) (Devinyak *et al*. 2014). They can reveal the skeleton and substituent information for a molecule space, for example, to describe steric effect and hindrance or structure/activity properties of a molecule.

In addition, the second 3D descriptor involved in the permeability prediction of PIMs to CO_2 gas, TDB is 3D topological distance-based autocorrelation. It is a member of the 3D autocorrelation descriptors which uses both geometric and topological distances to encrypt information about molecular structure. TDB is an index of shape and branching of a given molecule (Todeschini and Consonni 2009).

The third 3D descriptors, RDF (Radial Distribution Function) is a function of distribution of other atoms or molecules from a considered atom or a molecule with the distance as the independent variable. The RDF descriptors of a molecule of n atoms can be interpreted as the probability distribution of finding an atom in a spherical volume of radius R (Ga *et al*. 2008). These 3D descriptors suggest the occurrence of dependence between the permeability of PIMs and the 3D molecular distribution of electronegative atoms (like O and N) and the presence of electron-donating substituents at the benzene rings (Abreu *et al*. 2009). In addition to interatomic distances in the entire molecule, the RDF descriptors include other information, such as bond distances, ring kinds, planar and non-planar systems, and atom types. In CO_2 permeability we have two RDF descriptors at is the Radial Distribution Function weighted by atomic Van der Waals volumes at 3 Å and Radial Distribution Function weighted by atomic polarizability at atomic distance of 8.5Å consecutively.

4. Conclusions

In this investigation, we applied a hybrid artificial neural networks-Particle Swarm Optimization algorithm to forecast the permeability of PIMs films for three gases: N_2, CH_4, and CO_2. The permeability coefficients are predicted exclusively from the microstructure of the repeating unit of PIM, with a judiciously selected set of molecular descriptors acting as inputs to this established feedforward neural network model. The optimized structure of this ANN model is notably simple, comprising a fixed architecture with three layers: input, hidden, and output layers arranged at 19:30:1 for N_2, 14:25:1 for CH_4, and 44:66:1 for CO_2. The application of the Particle Swarm Optimization algorithm in conjunction with this architecture results in outstanding permeability prediction outcomes after the training phase, achieved in a reasonable duration and number of iterations, on a randomly chosen set of PIMs, attaining an RMSE of less than 10^{-3} when tested on a designated set of PIMs for this purpose. This model provided significant insights into which descriptors are most pertinent to the permeability of PIMs to the gases under investigation, particularly 3D descriptors.

References

Abreu, R.M.V., Ferreira, I.C.F.R. and Queiroz, M.J.R.P. (2009), "QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo[b]thiophenes", *Eur. J. Med Chem*., **44**, 1952-1958. https://doi.org/10.1016/j.ejmech.2008.11.011.

Abiodun, O.I. and Jantan, A. (2019), "Comprehensive review of artificial neural network applications to pattern recognition", *IEEE Access*., **7**, 158820-158846. https://doi.org/10.1109/ACCESS.2019.2945545.

Abreu, R., Falcão, S., Calhelha, R.C., Ferreira, I.C.F.R., Queiroz, M.J.R.P. and Vilas-Boas, M. (2009), "Insights in the antioxidant activity of diarylamines from the 2,3-dimethylbenzo[b] thiophene through the redox profile", *J. Electroanal. Chem*., **628**, 43-47. https://doi.org/10.1016/j.jelechem.2009.01.004.

Afantitis, A., Melagraki, G., Sarimveis, H., Koutentis, P.A., Markopoulos, J. and Igglessi-Markopoulou, O. (2006), "Prediction of intrinsic viscosity in polymer-solvent combinations using a QSPR model", *J. Polym*., **47**, 3240-3248. https://doi.org/10.1016/j.polymer.2006.02.060.

Ahmadi, M.A., Soleimani, R., Lee, M., Kashiwao, T. and Bahadori, A. (2015), "Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool", *Petrolum*., **1**, 118-132. https://doi.org/10.1016/j.petlm.2015.06.004.

Alicja, N., Jacek, N. and Stanisaw, K. (2010), "Prediction of retention of uncharged solutes in nanofiltration by means of

molecular descriptors", *Membr. Water Treat.*, **1**(3), 181-192. DOI: http://dx.doi.org/10.12989/mwt.2010.1.3.181.

Alvascience srl (2019), alvaDesc (software for molecular descriptors calculation); Available at: https://www.alvascience.com/

Cao, J., Cui, H., Shi, H. and Jiao, L. (2016), "Big data: A parallel particle swarm optimization-back-propagation neural network algorithm based on mapreduce", *PLoS ONE*., **11**, e0157551. https://doi.org/10.1371/journal.pone.0157551.

Cheng, W. and Feng, P. (2015), "Network traffic prediction algorithm research based on PSO-BP neural network", *ISRME*., 1239- 1243. https://doi.org/10.2991/isrme-15.2015.252.

Chirico, N. and Gramatica, P. (2012), "Real external predictivity of QSAR models, Part 2, new intercomparable thresholds for different validation criteria and the need for scatter plot inspection", *J. Chem. Inf. Model*., **52**, 2044-2058. https://doi.org/10.1021/ci300084j.

Crank, J. (1975), *The Mathematics of Diffusion*, (2nd Edition), Oxford University Press.

Da Silva, A.P., Chiari, L.P.A., Guimaraes, A.R., Honorio, K.M. and da Silva, A.B.F. (2021), "Drug design of new 5-HR antagonists aided by artificial neural networks", *J. Mol. Graph. Model*., **104**, 107844 -107847. https://doi.org/10.1016/j.jmgm.2021.107844.

Devinyak, O., Havrylyuk, D. and Lesyk, R. (2014), "3D-MoRSE descriptors explained", *J. Mol. Graph. Model*., **54**, 194-203. https://doi.org/10.1016/j.jmgm.2014.10.006.

Fang, W., Zhang, L. and Jiang, J. (2010), "Polymers of intrinsic microporosity for gas permeation: A molecular simulation study", *Mol. Simul.*, **36**(12), 992-1003. https://doi.org/10.1080/08927022.2010.498828.

Fang, W., Zhang, L. and Jiang, J. (2011), "Gas permeation and separation in functionalized polymers of intrinsic microporosity: a combination of molecular simulations and Ab initio calculations", *J. Phys. Chem. C.*, **115**, 14123-14130. https://doi.org/10.1021/jp204193g.

Fourches, D. and Ash, J. (2019), "4D- quantitative structure-activity relationship modeling: making a comeback", *Expert. Opin. Drug. Discov*., **14**, 1227-1235. https://doi.org/10.1080/17460441.2019.1664467.

Ga, Z., Fall, Y., Go, G. and Pe, M. (2008), "Radial distribution function descriptors for predicting affinity for vitamin D receptor", *Eur. J. Med. Chem*., **43**, 1360-1365. https://doi.org/10.1016/j.ejmech.2007.10.020.

García-Domenech, R. and Julián-Ortiz, J.V. (2002), "Prediction of indices of refraction and glass transition temperatures of linear polymers by using graph theoretical indices", *J. Phys. Chem. B*., **106**, 1501-1507. https://doi.org/10.1021/jp012360u.

Garson, G.D. (1991), "Interpreting neural network connection weights", *AI Expert*., **6**, 47-51.

Ge, M., Dimopoulos, I. and Lek, S. (2003), "Review and comparison of methods to study the contribution of variables in artificial neural network models", *Ecol. Model*., **160**, 249-264. https://doi.org/10.1016/S0304-3800(02)00257-0.

Gharagheizi, F. (2007), "QSPR analysis for intrinsic viscosity of polymer solutions by means of GA MLR and RBFNN", *Comput. Mater. Sci*., **40**, 159-167. https://doi.org/10.1016/j.commatsci.2006.11.010.

Goh, A.T.C. (1995), "Back-propagation neural networks for modeling complex systems", *Artif. Intell. Eng*., **9**, 143-151. https://doi.org/10.1016/0954-1810(94)00011-S.

Gupta, M.K., Sagar, R., Shaw, A.K. and Prabhakar, Y.S. (2005), "CP-MLR directed QSAR studies on the antimycobacterial activity of functionalized alkenols-topological descriptors in modeling the activity", *Bioorganic. Med. Chem*., **13**, 343-351. https://doi.org/10.1016/j.bmc.2004.10.025.

Han, Y. and Ho, W.S.W. (2021), "Polymeric membranes for CO2 separation and capture", *J. Membr. Sci*., **628**, 119244-119215. https://doi.org/10.1016/j.memsci.2021.119244.

Hasnaoui, H., Krea, M. and Roizard, D. (2017), "Neural networks for the prediction of polymer permeability to gases", *J. Membr. Sci.*, **541**, 541-549. https://doi.org/10.1016/j.memsci.2017.07.031.

HyperChem, R. (1993), for Windows, Molecular Modeling System, Hypercube, Inc. and Autodesk, Ringoes, NJ, U.S.A. https://www.acdlabs.com/resources/freeware/chemsketch/index.php

Kanakkithodi, A.M., Pilania, G., Huan, T.D., Lookman, T. and Ramprasad, R. (2016), "Machine learning strategy for accelerated design of polymer dielectrics", *Sci. Rep*., **6**, 20952. https://doi.org/10.1038/srep20952.

Kennedy, J. and Eberhart, R. (1995), "Particle swarm optimization", *Proceedings of ICNN'95 - International Conference on Neural Networks*, **4**, 1942-1948. https://doi.org/10.1109/ICNN.1995.488968.

Kesting, R.E. and Fritzsche, A.K. (1993), *Polymeric Gas Separation Membranes*, (1st Edition), Chapter 4: Diffusion in a plane sheet, p. 46, Wiley-Interscience, New York, U.S.A.

Koopialipoor, M., Fallah, A., Armaghani, D.J., Azizi, A. and Mohamad, E.T. (2019), "Three hybrid intelligent models in estimating flyrock distance resulting from blasting", *Eng. Comput*., **35**, 243-256. https://doi.org/10.1007/s00366-018-0596-4.

Łapińska, N., Pacławski, A., Szlęk, J. and Mendyk, A. (2024), "Integrated QSAR models for prediction of serotonergic activity: machine learning unveiling activity and selectivity patterns of molecular descriptors", *Pharm.*, **16**, 349-348. https://doi.org/10.3390/pharmaceutics16030349.

Li, M., Huang, X., Liu, H., Liu, B. and Wu, Y. (2013), "Fluid phase equilibria prediction of gas solubility in polymers by back propagation artificial neural network based on self-adaptive particle swarm optimization algorithm and chaos theory", *Fluid. Phase. Equil*., **356**, 11-17. https://doi.org/10.1016/j.fluid.2013.07.017.

MathWorks (1984), MATLAB: The Language of Technical Computing: Version 9. 7.0.1190202(R2019b), MathWorks.

Mehmood, T., Liland, K.H., Snipen, L. and Sæbø, S. (2012), "A review of variable selection methods in partial least squares regression", *Chemometr. Intell. Lab. Syst*., **118**, 62-69. https://doi.org/10.1016/j.chemolab.2012.07.010.

Mengshan, L., Wei, W., Bingsheng, Ch., Yana, W., Xingyuan, H. (2017), "Solubility prediction of gases in polymers based on an artificial neural network: A review", *RSC. Adv*., **7**, 35274-35282. https://doi.org/10.1039/C7RA04200K.

Metz, B., Davidson, O., de Coninck, H., Loos, M. and Meyer, L. (2005), "IPCC special report on carbon dioxide capture and storage, Chapter 5: Underground geological storage", Cambridge University Press, Cambridge, United Kingdom and New York, NY, U.S.A.

Mohammed, S., Gill, A.R., Alsafadi, K., Hijazi, O., Yadav, K.K., Hasan, M.A., Khan, A.H., Islam, S., Pinto, M.M.S.C., Harsanyi, E. (2021), "An overview of greenhouse gases emissions in Hungary", *J. Clean. Prod*., **314**, 127865. https://doi.org/10.1016/j.jclepro.2021.127865.

Mulumba, D.M., Liu, J., Hao, J., Zheng, Y. and Liu, H. (2023), "Application of an optimized PSO-BP neural network to the assessment and prediction of underground coal mine safety risk factors", *Appl. Sci.*, **13**, 5317-5316. https://doi.org/10.3390/app13095317.

Niazi, S.K. and Mariam, Z. (2023), "Recent advances in machine-learning-based hemoinformatics: A comprehensive review", *Int. J. Mol. Sci*., **24**, 11488-11487. https://doi.org/10.3390/ijms241411488.

Palomba, D., Vazquez, G.E. and Díaz, M.F. (2012), "Novel

descriptors from main and side chains of high-molecular-weight polymers applied to prediction of glass transition temperatures", *J. Mol. Graph*., **38**, 137-147. https://doi.org/10.1016/j.jmgm.2012.04.006.

Saemi, M., Ahmadi, M. and Varjani, A.Y. (2007), "Design of neural networks using genetic algorithm for the permeability estimation of the reservoir", *J. Pet. Sci. Eng*., **59**, 97-105. https://doi.org/10.1016/j.petrol.2007.03.007.

Saikiaa, P. and Baruah, R.D. (2020), "Artificial neural networks in the domain of reservoir characterization: A review from shallow to deep models", *Comput. Geosci*., **135**, 104357. https://doi.org/10.1016/j.cageo.2019.104357.

Solomon, M.F.J., Song, Q., Jelfs, K.E., Ibanez, M.M. and Livingston, A.G. (2016), "Polymer nanofilms with enhanced microporosity by interfacial polymerization", *Nat. Mater*., **15**, 760-767. https://doi.org/10.1038/NMAT4638.

Sun, W., Zheng, Y., Yang, K., Zhang, Q., Shah, A.A., Wu, Z., Sun, Y., Feng, L., Chen, D., Xiao, Z., Lu, S., Li, Y. and Sun, K. (2019), "Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials", *Sci. Adv*., **5**, 4275-4624. https://doi.org/10.1126/sciadv.aay4275.

Tao, L., Chen, G. and Li, Y. (2021a), "Machine learning discovery of high-temperature polymers", *Patterns*, **2**, 100225-100228. https://doi.org/10.1016/j.patter.2021.100225.

Tao, L., Varshney, V. and Li, Y. (2021b), "Benchmarking machine learning models for polymer informatics: an example of glass transition temperature", *J. Chem. Inf. Model*., **61**, 5395-5413. https://doi.org/10.1021/acs.jcim.1c01031.

Tattershall, C.E. and Wang, D. (2004), "Solution-processed, organophilic membrane derived from a polymer of intrinsic microporosity", *Adv. Mater*., **16**, 456-459.

Todeschini, R. and Consonni, V. (2009), *Molecular Descriptors for Chemoinformatics: Volume I: Alphabetical Listing/volume II: Appendices, References*, (2nd Edition), John Wiley & Sons.

Wessling, M., Mulder, M.H.V., Bos, A., van der Linden, M., Bos, M. and van der Linden, W.E. (1994), "Modelling the permeability of polymers: A neural network approach", *J. Membr. Sci.*, **86**, 193-198. https://doi.org/10.1016/0376-7388(93)E0168-J.

Wu, S., Kondo, Y, Kakimoto, M.a., Yang, B., Yamada, H., Kuwajima, I., Lambard, G., Hongo, K., Xu, Y., Shiomi, J., Schick, C., Morikawa, J. and Yoshida, R. (2019), "Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm", *npj. Comput. Mater*., **5**, 66. https://doi.org/10.1038/s41524-019-0203-2.

Wua, Z., Xian, Z., Ma, W., Liu, Q., Huang, X., Xiong, B., He, S. and Zhang, W. (2021), "Artificial neural network approach for predicting blood brain barrier permeability based on a group contribution method", *Comput. Methods. Programs. Biomed*., **200**, 105943-105916. https://doi.org/10.1016/j.cmpb.2021.105943.

Xu, J., Chen, B., Zhang, Q. and Guo, B. (2004), "Prediction of refractive indices of linear polymers by a four-descriptor QSPR model", *J. Polym*., **45**, 8651-8659. https://doi.org/10.1016/j.polymer.2004.10.057.

Xuan, L., Zhu, Z., Hengzhang, D., Kuang, M., Yafei, Z. and Bin, L. (2023), "Controlling interlayer spacing of GO membranes via the insertion of GN for high separation performance", *Membr. Water Treat.*, **14**(3), 107-114. https://doi.org/10.12989/mwt.2023.14.3.107.

Yang, L., Sang, C., Wang, Y., Liu, W., Hao, W., Chang, J. and Li, J. (2021), "Development of QSAR models for evaluating pesticide toxicity against skeletonema costatum", *Chemosphere*, **285**, 131456-131463. https://doi.org/10.1016/j.chemosphere.2021.131456.

Yu, X., Wang, X., Wang, H., Li, X. and Gao, J. (2006), "Prediction of solubility parameters for polymers by a QSPR model", *QSAR. Combinat. Sci*., **25**(2), 156-161. https://doi.org/10.1002/qsar.200530138.

Zhang, Y. and Wu, L. (2011), "Crop classification by forward neural network with adaptive chaotic particle swarm optimization", *J. Sens*., **11**, 4721-4743. https://doi.org/10.3390/s110504721.

Zhao, M., Zhang, C. and Weng, Y. (2023), "Improved artificial neural networks (ANNs) for predicting the gas separation performance of polyimides", *J. Membr. Sci*., **681**, 121765-121780. https://doi.org/10.1016/j.memsci.2023.121765.

JL