

## Article

# Artificial Neural Network Modeling of Glass Transition Temperatures for Some Homopolymers with Saturated Carbon Chain Backbone

Elena-Luiza Epure \* , Sîziana Diana Oniciuc , Nicolae Hurduc and Elena Niculina Drăgoi \* 

“Cristofor Simionescu” Faculty of Chemical Engineering and Environmental Protection, Gheorghe Asachi Technical University, 73 Prof. Dr. Doc. D. Mangeron Street, 700050 Iasi, Romania; oniciuc.sinziana30@gmail.com (S.D.O.); Nicolae.Hurduc@tuiasi.ro (N.H.)

\* Correspondence: lepure@tuiasi.ro (E.-L.E.); elenan.dragoi@gmail.com (E.N.D.)

**Abstract:** The glass transition temperature ( $T_g$ ) is an important decision parameter when synthesizing polymeric compounds or when selecting their applicability domain. In this work, the glass transition temperature of more than 100 homopolymers with saturated backbones was predicted using a neuro-evolutive technique combining Artificial Neural Networks with a modified Bacterial Foraging Optimization Algorithm. In most cases, the selected polymers have a vinyl-type backbone substituted with various groups. A few samples with an oxygen atom in a linear non-vinyl hydrocarbon main chain were also considered. Eight structural, thermophysical, and entanglement properties estimated by the quantitative structure–property relationship (QSPR) method, along with other molecular descriptors reflecting polymer composition, were considered as input data for Artificial Neural Networks. The  $T_g$ 's neural model has a 7.30% average absolute error for the training data and 12.89% for the testing one. From the sensitivity analysis, it was found that cohesive energy, from all independent parameters, has the highest influence on the modeled output.

**Keywords:** glass transition temperature; artificial neural networks; QSPR; homopolymers; Bacterial Foraging Optimization



**Citation:** Epure, E.-L.; Oniciuc, S.D.; Hurduc, N.; Drăgoi, E.N. Artificial Neural Network Modeling of Glass Transition Temperatures for Some Homopolymers with Saturated Carbon Chain Backbone. *Polymers* **2021**, *13*, 4151. <https://doi.org/10.3390/polym13234151>

Academic Editors: Riccardo Concu and Michael González-Durruthy

Received: 2 November 2021

Accepted: 25 November 2021

Published: 27 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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

Due to their high specific strength, lightweight, high performance, or easy processability, polymers are often chosen as the best option in different applications. Even though the high number of synthesized polymers is impressive, nowadays there is still continuous competition to create new polymers with enhanced or specific end-use properties [1–4]. However, the maximum utility potential of these materials was not reached. The determination of the relationship between structure and properties can lead to advances in the field and can help in the endeavor to find materials with specific properties. Nevertheless, in the case of polymers, due to their complexity, this attempt is difficult and time-consuming.

A key parameter of polymers, the glass transition temperature ( $T_g$ ), could be perceived either as an indicator of amorphous content, different degrees of polymerization, the presence of moisture or as additives in these materials. Still, the  $T_g$  is not that easy to be determined neither in an experimental or a theoretical manner, because of the multiple factors that control this temperature. In view of the fact that the transition in a rubbery state occurs alongside a temperature range, assigning a unique value is problematic, especially for the wider domains. The most popular instrumental technique for the investigation of  $T_g$  is Differential Scanning Calorimetry (DSC), yet it was observed that this method has its limitations. Computational methods and, in particular, molecular dynamic simulations are well suited to the characterization of the structure and dynamics of the polymers at an atomic level. The solvent, the confinement, the difunctional or trifunctional cross-linkers, and other effects on the glass transition were spotlighted by molecular dynamics

simulations (MD) [5–8]. However, the MD method is time-consuming, needs powerful computational resources, and the results have to be carefully selected, mainly when special cases are encountered, like wide transition region [9]. Under these conditions, another easy-to-implement instrument, suitable for Tg prediction was sought. At this point, an Artificial Neural Network approach (ANN), another kind of computational modeling method, seems to be the best way to achieve the desired tasks, knowing the machine learning flexibility to address and solve problems from the most distinct fields of the real world (medicine, forecasting, defense, economics, entertainment, and so on). The main advantage of ANNs is their ability to determine a non-linear complex model that correlates the output with different input data for complicated systems or processes, without requiring a prior model. Capable to extract subtle information, this modeling method could provide from sensitivity analysis, the variables' contribution to the output data. A sustainable materials development, an accelerated evaluation, or optimization of products could be thus addressed with machine learning, easily overcoming conventional synthesis and characterizing methods [10–13]. Examples of compounds for which ANNs were successfully used to predict Tg include multicomponent oxide glasses [14], ionic liquids [15], polystyrenes [16–18], organic electroluminescent devices materials [19], polyhydroxyalkanoate homo- and copolymers [20], or aromatic polyamides and polybenzimidazoles [21].

Despite the fact that a large amount of data is found in different sources, a unique model that connects the structure and the Tg of polymers was not developed yet. The aim of this study is to take a step further by building a mathematical model, universally available, for the prediction of the glass transition temperature of different kinds of homopolymers with saturated backbones. In this regard, a neuro-evolutive strategy that combines artificial neural networks with a modified bio-inspired metaheuristic represented by the Bacterial Foraging Optimization was applied. The classical Bacterial Foraging Optimization algorithm will be referred to as BFO, while its modified version proposed in this work will be referred to as mBFO. The motivation for choosing ANN-mBFO in detriment of the established classical strategies relies on the following: (i) its easiness of use compared with the phenomenological software modeling approaches that require extensive knowledge; (ii) the generalization capabilities of ANNs strategies that can extract useful information from data; (iii) the efficiency of BFO to solve complex systems. ANNs are simplified mathematical models that imitate the manner in which the brain functions. Although there are significant differences between ANNs and their biological counterpart (in terms of complexity, learning and adaptive mechanisms and capabilities), ANNs proved to be efficient alternative tools that can be successfully used in various applications [22].

In this work, mBFO acts as an optimizer for the ANN architecture and parameters. This strategy for ANN determination belongs to the neuro-evolutive class. The motivation for this selection is based on the following aspects: (i) a simultaneous architecture and internal parameters optimization can be performed; (ii) the shortcomings of gradient descent training algorithms are overcome; (iii) evolution is another form of adaptation. BFO [23] is a population-based bio-inspired metaheuristic that simulates the foraging behavior of *E. coli* bacterium. In combination with machine learning approaches (such as ANNs and Support Vector Machines), it was successfully applied to solve various problems, for instance, rock mass classification [24], Charcot-Matie-tooth disease [25], identification and classification of plant disease [26,27], compact fractal antenna design [28].

The novelty of this work relies on the identification of a simpler mathematical model for Tg that can efficiently replace the complex phenomenological models and the experimental trials, based on a new version of BFO with improved performance.

## 2. Materials and Methods

### 2.1. Database Construction

The current work focuses on developing a neural model that can predict Tg for homopolymers with a saturated backbone. From scientific literature or platforms, only the homopolymers with reported experimental Tg were considered. In most cases, the

selected polymers have a  $-\text{CH}_2-\text{CR}_1\text{R}_2-$  carbon chain (where  $\text{R}_1 = -\text{H}, -\text{F}, -\text{Cl}, -\text{CH}_3$  or  $-\text{CN}$ ;  $\text{R}_2 = -\text{F}, -\text{Cl}, \text{alkyl}, \text{cycloalkyl}, -\text{CH}=\text{CH}_2, -\text{COOH}, -\text{COOR}, -\text{C}_6\text{H}_6, \text{substituted benzene}, -\text{C}(=\text{O})\text{NRR}', -\text{OR}, -\text{CN}$  or pyridine). Therefore, the database consists of polymers from different families, such as polyolefins, halogenated polyolefins, polyacrylates, polymethacrylates, polycyanoacrylates, polystyrenes/halogenated polystyrenes, poly(vinyl ethers) and polyacrylamides. The poly(vinyl alcohol), one polyglycol and a few polyoxyalkylenes were also considered in the database. The last homopolymers have an oxygen atom in the backbone. The homopolymers considered for ANN are listed in Table S1 (Supplementary Materials).

The structural information used as input data for the ANNs are represented by the number of atoms of the structural units, the weight of the main chain from the structural units (M CB), the weight of non-hydrogen atoms in the main chain of repeat units (M CB-H), and a series of indices: the number of grafts (linear), the presence of a cyclic structure in the side chain (cyc CL), the aromaticity of the side chain (ar CL), the heteroatom presence in the main or side chain (h CB, h CL). The values of the linear indicator could be 0, 1, and 2, depending on the number of grafts. The other indices, cyc CL, ar CL, h CB, h CL, are Boolean types, having an assigned value of 1 if the condition is met.

Other parameters for ANN were those obtained in an empirical manner with the help of the QSPR method through the Synthia module of the Materials Studio software (v 4.0, Accelrys, San Diego, CA, USA) connectivity index  $0\chi$ , connectivity index  $1\chi$ , van der Waals volume, molar volume at 1 and 298K, density, cohesive energy, and entanglement molecular weight. This empirical method is a rapid property-prediction technique, avoiding molecular dynamics, a complex and time-consuming method.

The zeroth-order (atomic) connectivity indices  $0\chi$  and the first-order (bond) connectivity indices  $1\chi$  are defined as polymer chain connectivity indices, summing the information over the vertices, respectively the edges of the hydrogen-suppressed graph [29].

Van der Waals volume, molar volume at 298 K or even at 1 K expresses volumetric properties. Being a function of the temperature, the molar volume was calculated for 1K to take into account the effects of the frozen-in packing volume at near absolute zero temperature. The van der Waals volume reports the real space occupied by the atoms, without considering the empty space created by the packing of the polymer chain.

A significant parameter that characterizes the polymers is density that indirectly covers the chain packing imposed by tacticity, the degree of grafting or the grafted chain length. The influence of temperature on the density was switched to the two molar volumes previously mentioned.

As it quantifies the intermolecular forces, cohesive energy was considered as an input for the current ANN. This term comprises some factors like crystallinity, dipole moment, and the presence of hydrogen bonds.

Entanglement molecular weight measures the average molecular weight between the entanglements, which relates to the mobility of the chains. This parameter is especially important for the polymers in the rubbery state. The properties were computed for polymers having  $M_w = 10,000$  a.m.u.

## 2.2. Bacterial Foraging Optimization

The modeling procedure is based on ANN architecture and internal parameter optimization using a modified BFO (mBFO) algorithm.

In its standard form, as proposed in [23], BFO imitates the concept of the social foraging behavior of bacteria colonies. In the search for food, the bacterium performs small moves in order to maximize the difference between the energy gained from the food and the one burnt. This searching process is referred to as chemotaxis and in the BFO algorithms is simulated through swimming and tumbling using the flagella [30]. Besides chemotaxis, the behavior of *E. coli* includes a communication aspect. When stimulated, the bacteria cells release an attractant that allows their aggregation in groups with high density. This process is known as swarming. Moreover, the healthier bacteria undergo asexual

reproduction, the cell splitting into two bacteria. This aspect represents the reproduction step. In the final step of the algorithm (elimination and dispersal), some of the cells have a small probability of being randomly eliminated and new replacements are generated. The simplified schema of the BFO algorithm is presented in Figure 1.

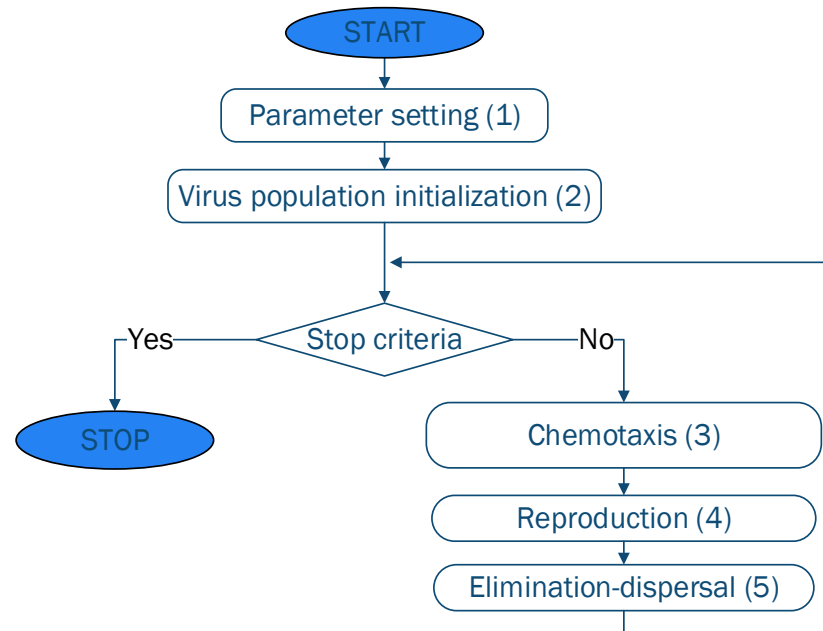


Figure 1. Simplified schema of the BFO algorithm.

The steps indicated in Figure 1 are detailed below:

(1) Parameter setting. Before starting the algorithm, the control parameters of the algorithm are set. BFO has the following parameters that direct the search: population size ( $S$ ), the life-span of each bacterium controlled in the chemotactic loop ( $N_c$ ), the swimming length ( $N_s$ ), the size of the step for the tumble movement ( $C_i$ ), the number of reproduction loops ( $N_{re}$ ), number of elimination-dispersal events ( $N_{ed}$ ), the probabilities with which the population is subjected to elimination-dispersal events ( $p_{ed}$ ) [30].

(2) Population initialization. Based on the limits of the parameters of the problem being solved, the initial population is generated through the use of random number generators. This strategy is similar to the initialization procedure usually employed by the majority of bio-inspired metaheuristics (Equation (1))

$$x_i = \min_i + \text{rand} \times (\max_i - \min_i) \quad (1)$$

where  $x_i$  represents the  $i^{\text{th}}$  parameter of the individual  $x$  and  $\min$  and  $\max$  represent the minimum and the maximum allowed limits for the  $i^{\text{th}}$  parameter.

(3) The chemotaxis step focuses on the movement of bacteria based on the perceived concentration of chemical gradients. The movement is composed of two types of rotation: (i) clockwise, which generates a forward movement; and (ii) counter-clockwise, which generates a tumbling movement. The forward movement occurs when a rich gradient is found and the tumbling is performed when a richer gradient is found.

(4) Reproduction focuses on creating copies for 50% of the healthiest individuals. The health of an individual measures its suitability to the environment (how many nutrients have been found and how it was able to avoid noxious substances). This is calculated using a function that in the standard Evolutionary Algorithms terminology is referred to as fitness function.

(5) In the elimination-dispersal step, some of the individuals are replaced with new individuals. This replacement occurs with a  $p_{ed}$  probability. Its aim is to refresh the population and to avoid the cases when all individuals are concentrated in a particular area

(local minima). In the initial BFO algorithm,  $p_{ed}$  is fixed and thus, all individuals have the same replacement probability. Thus, there is the possibility of replacing very good solutions and introducing new worse ones that are far from the optimum. Therefore, in this work, the  $p_{ed}$  is modified adaptively, this being one of the specific characteristics of modified BFO (mBFO) (Equation (2)). The aim of this change is to eliminate the issues introduced by the fixed  $p_{ed}$  value and to ensure that the good solutions have a lower probability of being replaced.

$$P_{ed} = \begin{cases} \frac{fit_{min} + fit_{avg}}{fit_{max} + fit_{min}}, & \text{if the objective is fitness minimization} \\ \frac{fit_{max} - fit_{avg}}{fit_{max} - fit_{min}}, & \text{if the objective is fitness maximization} \end{cases} \quad (2)$$

where  $fit_{min}$  represents the minimum of the fitness of all individuals,  $fit_{max}$  the maximum and  $fit_{avg}$  the average.

Another modification performed in this work consists of the introduction of an additional step that performs a local search. The objective is to locally improve the so far best solution using a combination of Random Search Algorithm and Opposition Based Learning (OBL) [31]. This idea of locally improving the solutions allows the algorithm to additionally exploit an identified region and improves the probability of finding solutions with high performance [32].

### 2.3. Artificial Neural Networks

The resulting mBFO algorithm is applied to optimize the ANN that models the considered process. The type of ANN used in this work is represented by a Feed Forward Multilayer Perceptron (MLP). MLP is one of the most used types of ANNs [33]. A MLP is formed of perceptrons organized into layers and each layer is fully connected with the next layer. As this type of organization is not compatible with the structure of the mBFO individual (which is a vector of real numbers), a direct encoding is applied to transform the ANN (also known in the neuro-evolutionary terminology as the phenotype) into the genotype.

When combining meta-heuristics with ANNs, evolution can be performed at three levels: (i) connection weights (which corresponds to the training phase and is formulated as the minimization of the Mean Squared Error (MSE)—described by Equation (3)); (ii) architecture (which corresponds to the process of identifying the optimal topology); and (iii) learning rules (which aims to adapt the learning rules using the evolutionary process) [34]. In the current work, mBFO is applied for a simultaneous optimization of connection weights and architecture. Thus, the fitness function in mBFO is computed based on the MSE determined in the training phase. Moreover, distinctively from other neuro-evolutionary strategies, where the activation function is set per layer (which means that all the neurons in the same layer have the same activation function), in the proposed variant, each neuron can have its specific activation function.

$$MSE = \frac{1}{n} \sum (Y - \hat{Y})^2 \quad (3)$$

where  $n$  represents the total number of exemplars (data) for which the MSE is computed,  $Y$  represents the known output, and  $\hat{Y}$  is the predicted output.

The complete list of functions a neuron can have is as follows: sigmoid, bipolar sigmoid, logarithmic sigmoid, tangent sigmoid, sinus, radial basis, logistic, tangent hyperbolic, bent identity, soft exponential, linear, leaky relu, and relu. For the functions that need parameters (i.e., sigmoid, bipolar sigmoid, soft exponential), additional memory is allocated. In the end, the genotype (vector of real values), which is the direct encoding of the ANN (phenotype), has the following structure:

In Figure 2,  $N_{hl}$  indicates the number of hidden layers,  $N_{n1}$  and  $N_{n2}$  the number of neurons in the first and, respectively, the second hidden layer,  $W$  are the weights,  $B$  are the biases,  $A_f$  represents the activation functions, and  $PA_f$ —the parameters of the activation functions. The limitation to only two hidden layers is imposed based on practical



aspects: the higher the number of layers, the higher the number of weights, biases, and activation functions.

Nhl	Nn1	Nn2	W1	...	Wn	B1	...	Bm	Af1	...	Afk	PAf1	...	PAfk
-----	-----	-----	----	-----	----	----	-----	----	-----	-----	-----	------	-----	------

**Figure 2.** Structure of the genotype representing the evolved ANN parameters.

#### 2.4. Sensitivity Analysis

Sensitivity analysis (SA) for ANNs refers to the determination of the influence of inputs and/or weight perturbations on the model output [35]. SA can be used to determine the importance of model inputs, testing the model conceptualization and improving its structure [36]. There are multiple strategies that can be used to perform SA, among which “clamping” and “Pad” are the most known. Clamping consists of comparing the errors obtained when the original data is applied, with the errors obtained when the inputs of interest are set to a fixed value [37]. In the case of Pad, the sensitivity of each output in relation to a specific input is defined as a differential coefficient. Pad uses partial derivatives and allows to identify the output for small input variations and to perform a classification of the relative contribution of each variant [38]. In this work, the Pad method is used. The exact method of computing the sensitivity is presented in a previous article [39].

### 3. Results

#### 3.1. Database

After selecting the published information, the database was constructed with 22 polyolefins, 7 halogenated polyolefins, 13 polystyrene-type compounds, 19 polyacrylates, 29 polymethacrylates, 4 polycyanoacrylates, and 12 homopolymers with hydroxyl or ether groups (Tables S1 and S2). Moreover, six compounds that cannot be included in the previously mentioned classes, the majority of them being polyacrylamides with vinyl backbone are included in the subclass “others”. Where a temperature range was reported, the lowest value was chosen. Moreover, if multiple single values were found, the value with the most bibliographic references was selected.

The dataset regarding QSPR properties selected as input for ANN can be accessed at [https://elenadragoi.ro/CV/Documents/BD\\_ANN\\_polymers.xlsx](https://elenadragoi.ro/CV/Documents/BD_ANN_polymers.xlsx), accessed on 23 November 2021. Each one of the inputs has proven its utility, mostly in the case where almost all the parameters are identical. For instance, poly (2-chlorostyrene) and poly (3-chlorostyrene) are different only by the entanglement molecular weight parameter; or poly (3-chlorostyrene) and poly (4-chlorostyrene) are differentiated by their cohesive energy value.

The structural parameters that complete the QSPR ones are important for a beneficial “learning” of the polymers by ANN. The argument that sustains this information is the following: the presence of an aromatic structure in the side chain is indicated by the arCL parameter, and this aromaticity cannot be discovered through zeroth-order (atomic) connectivity indices  $^0\chi$  or the first-order (bond) connectivity indices  $1\chi$ .

For the properties calculated by the QSPR method, the following standard deviations were reported (BIOVIA, Materials Studio): van der Waals volume 1.8%, molar volume 2.2–2.5%, density 2.5%, cohesive energy (Fedors) at 298 K 3.9%, entanglement molecular weight 30%.

#### 3.2. Artificial Neural Network Modeling

In this work, three modeling situations were considered: (i) Case 1, corresponding to the groups of polyolefins and halogenated polyolefins (with a total of 27 samples); (ii) Case 2, corresponding to the groups of polymethacrylates and polycyanoacrylates (with a total of 32 samples); and (iii) Case 3, that corresponds to the entire dataset and includes all the considered homopolymers (with a total of 104 samples). For all three cases, the same type of ANN model is established (feed-forward multilayer perceptron) using the same settings for the mBFO (number of function evaluations =  $5 \cdot 10^5$ , population size = 20).

Prior to model determination, the available data underwent a processing stage, where it was normalized, randomized, and split. In the normalization phase, a Min-Max approach was used and its role was to transform all data in the  $(-1, 1)$  interval to level the influence of the inputs with a higher order of magnitude. The role of randomization and splitting is to ensure that in the training and testing sets, randomly distributed points are assigned. In total, 75% of the data was assigned to the training set and 25% to the testing set. The training set is used in the determination of the ANNs, while the testing set is used to measure the generalization capabilities of the identified models.

Since mBFO has a stochastic nature, for each case, a set of 50 simulations were performed, the statistics for all the models obtained being presented in Table 1, where Architecture indicates the arrangement and the number of neurons in layers in the format inputs: neurons in the first hidden layer: neurons in the second hidden layer: outputs. The difference in the number of input neurons is given by the fact that for Cases 1 and 2, some of the inputs were fixed.

**Table 1.** Statistics for the models determined.

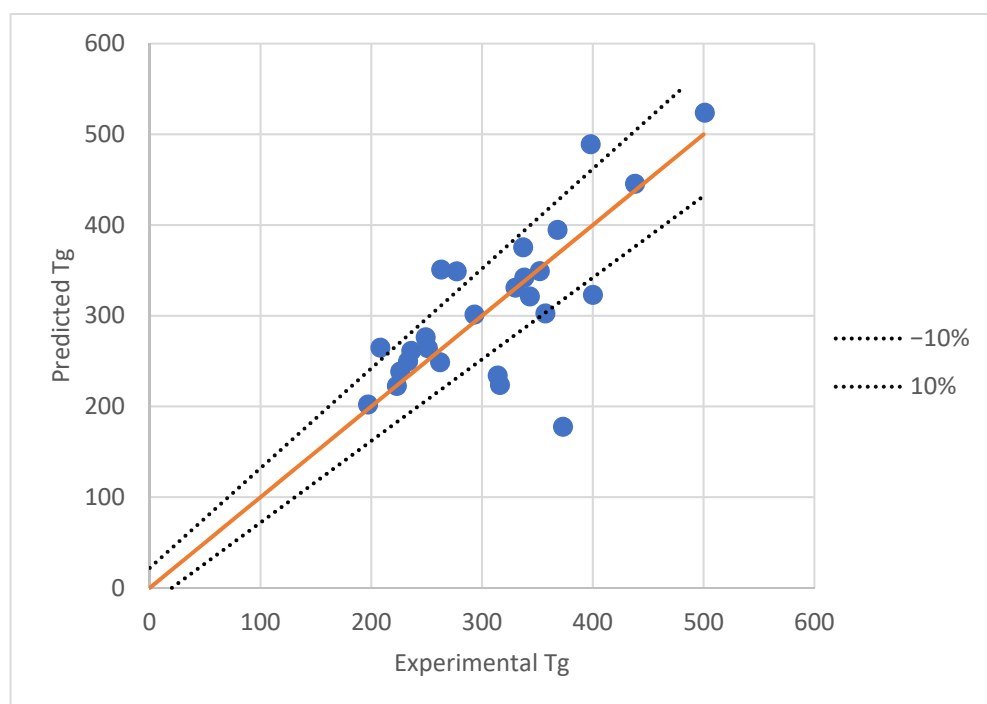
Case	Statistic Indicator	Fitness	MSE Training	MSE Testing	Correlation (R) Training	Correlation (R) Testing	Architecture
1	Best	349.9282	0.002858	0.041977	0.976802	0.530663	18:13:01
	Worst	150.3113	0.006653	0.087952	0.944351	-0.15623	18:07:01
	Average	187.8869	0.00552	0.086771	0.954337	-0.05816	
2	Best	1125.606	0.000888	0.039731	0.988932	0.2676	15:08:01
	Worst	213.079	0.00469	0.02958	0.93499	0.12272	15:04:01
	Average	813.29	0.001905	0.033154	0.974491	0.193581	
3	Best	164.8208	0.006067	0.028287	0.920757	0.726176	21:07:01
	Worst	116.8844	0.008555	0.058398	0.88678	0.539156	21:04:01
	Average	132.7765	0.007608	0.049029	0.899501	0.445961	

In Table 1, the suitability of the models was determined on the fitness function and the correlation (R) is computed using Equation (4). The higher the value for the fitness function, the better the model performed in the training phase. As it can be observed, for the best models, in the training phase, acceptable MSE values were obtained. However, for the testing phase, for Cases 1 and 2, the correlation is very low, indicating that the identified model had a low generalization capability and was not able in all the cases to model the variation of  $T_g$ . These findings were also supported by the average absolute error (AAE) computed in the testing phase). AAE was 15.79% in Case 1, 21.39% in Case 2, and 12.89% in Case 3. Thus, although the objective of Cases 1 and 2 was to identify improved models for compounds with specific groups, the results showed that the general model (Case 3) is better in terms of generalization and has a better testing performance.

$$R = \sqrt{R^2} = \frac{\sum \text{dif}Y_i \sum \text{dif}\hat{Y}_i}{\sqrt{\sum \text{dif}Y_i^2} \sqrt{\sum \text{dif}\hat{Y}_i^2}} \quad (4)$$

where  $R^2$  indicates the goodness of fit and  $\text{dif}Y$  represent the difference between  $Y$  and its mean,  $\text{dif}\hat{Y}$  represent the difference between  $\hat{Y}$  and its mean and  $Y, \hat{Y}$  have the same meaning as for Equation (3).

The experimental data versus the predicted values obtained by the best ANN in Case 3 is presented in Figure 3.



**Figure 3.** Comparison between experimental and ANN predictions for the testing data in the case of the best model obtained for all the data.

As it can be observed, there are a few points where there are significant differences (error > 10%).

The elimination of the three points with the highest error in the AAE determination leads to a value of 8.47%, which is considered acceptable taking into account the high difficulty to reduce a range of temperature to a unique value for T<sub>g</sub>.

To validate the model obtained in Case 3, a series of additional samples were collected and used to demonstrate its generalization capability. The AAE obtained for this additional validation set was 15.096% and the correlation was 0.901. These results indicate that the model is able to perform well on unseen data that correspond to the considered criteria and belong to the modeled classes.

In order to determine the influence of each independent parameter on the modeled output, the best model resulting for Case 3 was examined using a sensitivity analysis procedure. The sensitivity values associated with each parameter are presented in Table 2. The higher values of sensitivity indicate a higher influence of the specific parameter in the model.

**Table 2.** Sensitivity analysis for the best model in Case 3.

Order	Input	Sensitivity	Order	Input	Sensitivity
1	Cohesive energy	10.238	12	Molar volume, 1 K	3.668
2	H	9.974	13	Connectivity index $^0\chi$	3.479
3	Density	9.314	14	ar CL	3.102
4	N	9.183	15	h CL	3.074
5	M CB	8.023	16	O	2.920
6	Linear	5.185	17	vdW volume	2.386
7	F	5.060	18	Molar volume, 298 K	2.080
8	cyc CL	5.043	19	M CB-H	1.398
9	C	4.836	20	h CB	0.993
10	Cl	4.189	21	Connectivity index $^1\chi$	0.951
11	Entanglement molecular weight	3.9295			



The mathematical relations describing the best ANN model that can well identify Tg for all the types of compounds considered are presented in the Table S3 (in a Python format) and can be accessed using the following link: [https://elenadragoi.ro/CV/Documents/model\\_from\\_ANN\\_polymers.py](https://elenadragoi.ro/CV/Documents/model_from_ANN_polymers.py), accessed on 23 November 2021.

#### 4. Discussion

An analysis of the points that do not fit in the interval  $-10\%$ ,  $+10\%$  from Figure 3 indicate some surprising results, especially since the highest reported value of Tg was chosen for polyacrylonitrile (from class “other”). Poly(chlorotrifluoroethylene), which has the biggest error, has also reported 296 K as Tg data (Aldrich). Still, the deviation remains significant, at 40%. A plausible explication for the high errors of poly(tert-butyl acrylate) and poly(tert-butyl methacrylate), would be that both structures are found in the testing set, and ANN had only one structure with this kind of alkyl group in the training phase, namely the poly (N-tert-butylacrylamide).

In the case of the sensitivity analysis performed, the results are in correlation with the theoretical data. It was stated that cohesive energy is a fundamental property of polymers [29], so the fact that this is the parameter with the highest sensitivity for the neural model is no surprise at all.

The packing regularity can be indirectly found both in the density value and the difference between the molar volume at 298 K and 1 K. However, the density provides information about the degree of grafting or the grafted chain length, so as to understand its favorable position in this ranking. The molar volumes at 298 K and 1 K were thought of as parameters that deliver temperature influence on the polymer behavior. It seems that their dependence and also the van der Waals volume disqualify them regarding their importance for ANN.

Since the entanglement molecular weight is an important parameter only for the polymers in a rubbery state at room temperature, the low sensitivity for the whole series of polymers is understood. The parameters with the lowest influence on the current model are represented by h CB and connectivity index  $1\chi$ . These results can then be used further to prune the ANN model.

#### 5. Conclusions

The capacity to quantitatively predict the Tg remains a significant obstacle in polymer design. A machine-learning model is a great tool for solving this problem. In the current work, the Tg prediction was performed using a neuro-evolutionary strategy combining ANNs and mBFO, a modified version of BFO. The performed modifications in mBFO include the transformation of the  $p_{ed}$  fixed variable into an adaptive one based on the fitness function of all individuals in the population and a local search procedure that combines the Local Search algorithm and Opposition Based Learning. The goal of these modifications is to improve the general performance of the algorithm and its capability to efficiently determine good solutions. The role of mBFO was to optimize all the characteristics of the ANNs. A series of structural parameters like the number of atoms, connectivity indices, van der Waals and molar volume, along with cohesive energy and entanglement molecular weight were considered as input for the ANN. The sensitivity analysis revealed that the cohesive energy followed by density and weight of the main chain from structural units significantly affect the Tg values.

Simulations in various cases, focusing on different types of compounds, showed that the mBFO-ANN combination is able to find good models for Tg. However, in terms of testing performance, the best model obtained was the one determined in the case in which all the polymers from the database were taken into consideration, and not for individual groups of polymers. The MSE obtained for the training data was 0.006 and 0.028 for the testing data. The testing AAE was 12.89%, with a few exemplars that had a higher error. When eliminating only three cases, the AAE was reduced to 8.47%. Moreover, an additional set of data was gathered and used for validation. The results obtained indicate

that the model was able to provide acceptable predictions even in those cases. Thus, the generalization capability of the identified ANN was demonstrated and the model can be successfully utilized for other compounds from the same classes. Thus, for situations where the T<sub>g</sub> is not known and difficult to determine, the model can provide fast predictions based on the known dependent parameters.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/polym13234151/s1>, Table S1: The experimental T<sub>g</sub> values of the polymers used as learning and testing data, Table S2: The dataset used for validating the ANN model. Table S3: The ANN model obtained for all homopolymers considered.

**Author Contributions:** Conceptualization, E.-L.E.; methodology, E.-L.E. and E.N.D.; software, E.-L.E., N.H. and E.N.D.; validation, E.-L.E., N.H. and E.N.D.; formal analysis, E.-L.E.; investigation, E.-L.E., E.N.D. and S.D.O.; resources, E.-L.E. and N.H.; data curation, E.-L.E., E.N.D. and S.D.O.; writing—original draft preparation, E.-L.E. and E.N.D.; writing—review and editing, E.-L.E.; visualization, E.-L.E. and E.N.D.; supervision, E.-L.E.; project administration, E.-L.E.; funding acquisition, E.-L.E. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by a publication grant of the TUIASI, project number GI /P18/2021. (TUIASI—“Gheorghe Asachi” Technical University of Iasi).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in the Supplementary Materials.

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

## References

1. Wickramaarachchi, K.; Sundaram, M.M.; Henry, D.J.; Gao, X. Alginate biopolymer effect on the electrodeposition of manganese dioxide on electrodes for supercapacitors. *ACS Appl. Energy Mater.* **2021**, *4*, 7040–7051. [[CrossRef](#)]
2. Sava, I.; Sacarescu, L.; Stoica, I.; Apostol, I.; Damian, V.; Hurduc, N. Photochromic properties of polyimide and polysiloxane azopolymers. *Polym. Int.* **2009**, *58*, 163–170. [[CrossRef](#)]
3. Chiriac, A.P.; Rusu, A.G.; Nita, L.E.; Macsim, A.-M.; Tudorachi, N.; Rosca, I.; Stoica, I.; Tampu, D.; Aflori, M.; Doroftei, F. Synthesis of Poly (Ethylene Brassylate-Co-squaric Acid) as Potential Essential Oil Carrier. *Pharmaceutics* **2021**, *13*, 477. [[CrossRef](#)]
4. Epure, E.-L.; Stoica, I.; Albu, R.M.; Hulubei, C.; Barzic, A.I. New Strategy for Inducing Surface Anisotropy in Polyimide Films for Nematics Orientation in Display Applications. *Nanomaterials* **2021**, *11*, 3107. [[CrossRef](#)]
5. Yoshioka, A.; Tashiro, K. Solvent effect on the glass transition temperature of syndiotactic polystyrene viewed from time-resolved measurements of infrared spectra at the various temperatures and its simulation by molecular dynamics calculation. *Macromolecules* **2004**, *37*, 467–472. [[CrossRef](#)]
6. Wu, C. Glass transition in single poly(ethylene oxide) chain: A molecular dynamics simulation study. *J. Polym. Sci. Part B Polym. Phys.* **2017**, *55*, 178–188. [[CrossRef](#)]
7. Li, C.; Coons, E.; Strachan, A. Material property prediction of thermoset polymers by molecular dynamics simulations. *Acta Mech.* **2014**, *225*, 1187–1196. [[CrossRef](#)]
8. Ma, X.; Zheng, F.; van Sittert, C.G.; Lu, Q. Role of intrinsic factors of polyimides in glass transition temperature: An atomistic investigation. *J. Phys. Chem. B* **2019**, *123*, 8569–8579. [[CrossRef](#)]
9. Wu, C. Re-examining the procedure for simulating polymer T<sub>g</sub> using molecular dynamics. *J. Mol. Modeling* **2017**, *23*, 270. [[CrossRef](#)]
10. Hardian, R.; Liang, Z.; Zhang, X.; Szekely, G. Artificial intelligence: The silver bullet for sustainable materials development. *Green Chem.* **2020**, *22*, 7521–7528. [[CrossRef](#)]
11. Ding, W.; Lu, Y.; Peng, X.-L.; Dong, H.; Chi, W.; Yuan, X.; Sun, Z.-Z.; He, H. Accelerating evaluation of the mobility of ionic liquids modulated PEDOT flexible electronics using machine learning. *J. Mater. Chem. A* **2021**, *9*, 25547–25557. [[CrossRef](#)]
12. Yan, Y.; Mattisson, T.; Moldenhauer, P.; Anthony, E.J.; Clough, P.T. Applying machine learning algorithms in estimating the performance of heterogeneous, multi-component materials as oxygen carriers for chemical-looping processes. *Chem. Eng. J.* **2020**, *387*, 124072. [[CrossRef](#)]
13. Kumar, S.; Ignacz, G.; Szekely, G. Synthesis of covalent organic frameworks using sustainable solvents and machine learning. *Green Chem.* **2021**, *23*, 8932–8939. [[CrossRef](#)]
14. Cassar, D.R.; De Carvalho, A.C.; Zanotto, E.D. Predicting glass transition temperatures using neural networks. *Acta Mater.* **2018**, *159*, 249–256. [[CrossRef](#)]

15. Valderrama, J.O.; Campusano, R.A.; Rojas, R.E. Glass transition temperature of ionic liquids using molecular descriptors and artificial neural networks. *C. R. Chim.* **2017**, *20*, 573–584. [[CrossRef](#)]
16. Yu, X.; Huang, X. A quantitative relationship between T<sub>g</sub>s and chain segment structures of polystyrenes. *Polímeros* **2017**, *27*, 68–74. [[CrossRef](#)]
17. Paolo, B.; Pingue, P. Glass Transition Temperature in Polystyrene Supported Thin Films: A SPM-based Investigation of the Role of Molecular Entanglement. *J. Polym. Sci. Part B Polym. Phys.* **2013**, *51*, 1149–1156.
18. Lee, H.; Ahn, H.; Naidu, S.; Seong, B.S.; Ryu, D.Y.; Trombly, D.M.; Ganesan, V. Glass transition behavior of PS films on grafted PS substrates. *Macromolecules* **2010**, *43*, 9892–9898. [[CrossRef](#)]
19. Kim, Y.S.; Kim, J.H.; Kim, J.S.; No, K.T. Prediction of glass transition temperature (T<sub>g</sub>) of some compounds in organic electroluminescent devices with their molecular properties. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 75–81. [[CrossRef](#)] [[PubMed](#)]
20. Jiang, Z.; Hu, J.; Marrone, B.L.; Pilania, G.; Yu, X.B. A deep neural network for accurate and robust prediction of the glass transition temperature of polyhydroxyalkanoate homo- and copolymers. *Materials* **2020**, *13*, 5701. [[CrossRef](#)] [[PubMed](#)]
21. Yu, X. Support vector machine-based QSPR for the prediction of glass transition temperatures of polymers. *Fibers Polym.* **2010**, *11*, 757–766. [[CrossRef](#)]
22. Maul, T.H. Improving Neuroevolution with Complementarity-Based Selection Operators. *Neural Process. Lett.* **2016**, *44*, 887–911. [[CrossRef](#)]
23. Passino, K.M. Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Syst. Mag.* **2002**, *22*, 52–67.
24. Zheng, S.; Jiang, A.; Yang, X.; Luo, G.; Pham, B.T. A New Reliability Rock Mass Classification Method Based on Least Squares Support Vector Machine Optimized by Bacterial Foraging Optimization Algorithm. *Adv. Civ. Eng.* **2020**, *2020*, 3897215. [[CrossRef](#)]
25. Al-Kheraif, A.A.; Hashem, M.; Al Esawy, M.S.S. Developing Charcot–Marie–Tooth disease recognition system using bacterial foraging optimization algorithm based spiking neural network. *J. Med. Syst.* **2018**, *42*, 192. [[CrossRef](#)] [[PubMed](#)]
26. Chouhan, S.S.; Kaul, A.; Singh, U.P.; Jain, S. Bacterial foraging optimization based radial basis function neural network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards plant pathology. *IEEE Access* **2018**, *6*, 8852–8863. [[CrossRef](#)]
27. Teja, V.S.; Srinivas, C.; Radhika, P. Plant Disease Detection and Classification Using Bacteria Foraging Optimization Algorithm Through Convolution Neural Network. *J. Comput. Theor. Nanosci.* **2020**, *17*, 3567–3576. [[CrossRef](#)]
28. Dhaliwal, B.S.; Pattnaik, S.S. BFO–ANN ensemble hybrid algorithm to design compact fractal antenna for rectenna system. *Neural Comput. Appl.* **2016**, *28*, 917–928. [[CrossRef](#)]
29. Bicerano, J. *Prediction of Polymer Properties*; CRC Press: Boca Raton, FL, USA, 2002.
30. Zhu, G.-Y.; Zhang, W.-B. Optimal foraging algorithm for global optimization. *Appl. Soft Comput.* **2017**, *51*, 294–313. [[CrossRef](#)]
31. Tizhoosh, H.R. Opposition-Based Learning: A New Scheme for Machine Intelligence. In Proceedings of the International Conference on Computational Intelligence for Modeling, Control and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, Vienna, Austria, 28 November 2005; pp. 695–701.
32. Sharma, A.; Sharma, H.; Bhargava, A.; Sharma, N.; Bansal, J.C. Optimal placement and sizing of capacitor using Limaçon inspired spider monkey optimization algorithm. *Memetic Comput.* **2016**, *9*, 311–331. [[CrossRef](#)]
33. Salcedo-Sanz, S. Modern meta-heuristics based on nonlinear physics processes: A review of models and design procedures. *Phys. Rep.* **2016**, *655*, 1–70. [[CrossRef](#)]
34. Islam, M.; Yao, X. Evolving Artificial Neural Network Ensembles. In *Computational Intelligence: A Compendium*; Fulcher, J., Jain, L., Eds.; Studies in Computational Intelligence; Springer: Berlin/Heidelberg, Germany, 2008; Volume 115, pp. 851–880.
35. Yeung, D.S.; Cloete, I.; Shi, D.; Ng, W.W.Y. *Sensitivity Analysis for Neural Networks*; Natural Computing Series; Springer: Berlin/Heidelberg, Germany, 2010.
36. Chen, G.Z.; Wang, J.Q.; Li, R.Z. Parameter Identification of the 2-Chlorophenol Oxidation Model Using Improved Differential Search Algorithm. *J. Chem.* **2015**, *2015*, 313105. [[CrossRef](#)]
37. Montano, J.J.; Palmer, A. Numeric sensitivity analysis applied to feedforward neural networks. *Neural Comput. Appl.* **2003**, *12*, 119–125. [[CrossRef](#)]
38. Gevrey, M.; Dimopoulos, I.; Lek, S. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Model.* **2003**, *160*, 249–264. [[CrossRef](#)]
39. Dragoi, E.-N.; Curteanu, S.; Galaction, A.-I.; Cascaval, D. Optimization methodology based on neural networks and self-adaptive differential evolution algorithm applied to an aerobic fermentation process. *Appl. Soft Comput.* **2013**, *13*, 222–238. [[CrossRef](#)]