J.Natn.Sci.Foundation Sri Lanka 2025 53 (2): 165 - 172
DOI: http://dx.doi.org/10.4038/jnsfsr.v53i2.12381

RESEARCH ARTICLE

Deep Learning

A comparison of optimizers in a PyTorch based artificial neural network to predict normal boiling points of alkanes

MZ Afzal* **and SS Siddiqi**
Department of Mathematics, University of Central Punjab, Lahore, Pakistan.

Submitted: 01 October 2024; Revised: 12 March 2025; Accepted: 30 April 2025

Abstract: In this study, we explore the effectiveness of various optimization algorithms for predicting the normal boiling points of alkanes using PyTorch-based neural network models. Alkane molecular structures were uniquely encoded into a machine-readable format and used to train both an artificial neural network (ANN) and a graph neural network (GNN) model. We compared the performance of three optimizers; Stochastic Gradient Descent (SGD), Adam, and Rprop, across two loss functions (L1Loss and MSELoss) and multiple learning rates (0.01, 0.001, and 0.0001). Our results indicate that Adam consistently yields the lowest mean squared error (MSE) values across both models, highlighting its robust performance. Notably, in the ANN framework, Rprop demonstrated rapid convergence and outperformed SGD, whereas in the GNN model, SGD showed superior performance compared to Rprop. The analysis of MSE values further corroborated these findings, emphasizing the critical role of optimizer selection in achieving both high accuracy and model stability. These insights provide valuable guidance for future applications of neural networks in chemical property prediction and underscore the importance of fine-tuning optimization parameters in machine learning workflows.

Keywords: Alkanes, boiling points of alkanes, neural network, optimizers, PyTorch.

INTRODUCTION

The (normal) boiling point of a substance is defined as the temperature at which it boils under standard atmospheric pressure. The common methods to determine the boiling points of substances are simple distillation, fractional distillation, reflux distillation, steam distillation, microscale boiling point determination, ebulliometry, and thermometric titration.

The boiling point of a substance holds significant importance in various fields of science and industry due to several reasons. It can be used to determine the purity of a substance. It is used to narrow down the possible identities of a compound. Knowledge of the boiling points of substances is crucial for controlling the separation and purification of chemical mixtures. The boiling point of a substance is indicative of its volatility and vapor pressure at a given temperature. Understanding the boiling points of substances helps in assessing the potential impacts of evaporation, condensation, and atmospheric dispersion on the environment in case of spills or releases. Knowledge of boiling points is essential for the handling and storage of chemicals safely.

Beside laboratory methods, quantitative structure-property relationship (QSPR) models became popular to predict normal boiling points after the seminal paper of Wiener (1947). In order to avoid an unusually long list of work in this direction, we are going to give references of papers that are published in the last twenty years. The normal boiling points of aldehydes, esters, and ketones were computed by Duchowicz *et al*. (2002). A wonderful review about predicting boiling point, melting point, and vapour pressure with QSPR models was published

* Corresponding author (l1f22phma0001@ucp.edu.pk; https://orcid.org/0009-0005-4472-2069)

by Dearden (2003). Using a QSPR technique, the boiling point of polycyclic aromatic hydrocarbons were studied by de Lima Ribeiro and Ferreira (2003). The normal boiling points of some organic molecules were computed by Gonz´alez *et al.* (2004). Using topological indices, the normal boiling points along with molar refractivities of a large set of organic compounds are found in the paper by Ghavami *et al.* (2009). SMILES-based optimal descriptors were used to find normal boiling points of acyclic and cyclic hydrocarbons by Toropov *et al.* (2010). Utilizing the electronegativity topological descriptor, Yi-min forecasted normal boiling points of alkanes, unsaturated hydrocarbons, and alcohols (Yi-min *et al.*, 2013). With the help of semiempirical quantum chemistry, Saadi predicted the boiling points of amines (Saadi *et al.*, 2015). The boiling points of n-alkanes were related to critical constants by Messerly *et al.* (2017). Using various topological indices, the normal boiling points of alcohols and phenols were predicted by Arjmand and Shafiei (2018). A new graph-based parameter, conduction, is introduced to develop a single-parameter model for accurately predicting alkane boiling points, offering a reliable alternative to experimental data (Mukwembi & Nyabadza, 2021). QSPRpred is an open-source Python toolkit for building, analyzing, and reproducing robust QSPR models, enabling seamless prediction of molecular properties from SMILES strings (van den Maagdenberg *et al.*, 2024). A linear regression model accurately predicts thermochemical properties of alkanes (> C10) for studying hydro-isomerization equilibria in zeolites, outperforming traditional group contribution methods (Sharma *et al.*, 2024).

Predicting boiling points using neural networks has evolved significantly over time, with advancements in both model architecture and data availability. It was Cherqaoui who first applied neural networks to find boiling points (Cherqaoui & Villemin, 1994; Cherqaoui *et al*., 1994). Using the molecular structure and a neural network, Goll predicted normal boiling points of some organic compounds (Goll & Jurs, 1999). A novel group contribution-based neural network model accurately predicts molecular boiling points across a wide temperature range, outperforming Joback's equation (Sumie & Umpei, 2010). Gharagheizi designed a three-layer feed-forward neural network along with the Levenberg–Marquardt optimization technique to predict normal boiling points of 17,768 chemical compounds (Gharagheizi *et al.*, 2013). High-dimensional neural network (HDNN) potentials and systematic molecular fragmentation methods are compared for predicting alkane energies, with HDNN achieving higher accuracy against coupled cluster references (Gastegger *et al.*, 2016). By training an artificial neural network (ANN), Jin predicted normal boiling points of oxygen-containing and hydroxyl compounds (Jin & Bai, 2016a, 2016b, 2016c, 2016d). Neural networks trained on fragmented data predict physical properties of alkanes by leveraging property correlations, achieving high accuracy across various thermodynamic parameters (Santak & Conduit, 2019). By designing a multiple linear regression and a multi-layer perceptron neural network, Fissa predicted normal boiling points of pure hydrocarbons (Fissa *et al.*, 2019). Using the Wang-Landau simulations and machine learning techniques, Groven predicted boiling points and critical points of polycyclic aromatic hydrocarbons (Groven *et al*., 2019). Using a graph convolutional neural network, Qu predicted normal boiling points of several organic compounds (Qu *et al*., 2022). A multi-layer perceptron model using molecular group descriptors accurately predicts thermophysical properties of fluorine/chlorine-containing refrigerants, with SHAP analysis confirming feature contributions (Li *et al.*, 2022). Using an AI approach, Liu and Maryamto predicted normal boiling points of refrigerants (Liu *et al.*, 2023). Lastly, A machine learning model was developed to accurately predict the boiling temperature of environmentally friendly insulation materials, aiding their screening and optimization (Tang *et al.*, 2024).

We can efficiently train neural networks to discern the intricate relationships between molecular characteristics and boiling points, which ultimately enables us to predict the normal boiling points of alkanes with remarkable accuracy.

By leveraging PyTorch's robust framework and by using topological structures, this work predicts normal boiling points of 150 isomeric alkanes, which were extracted from NIST Chemistry Webbook https://webbook.nist.gov/ and analyzes the performance of several optimizers with different loss functions and learning rates.

MATERIALS AND METHODS

Data encoding

Consider an alkane isomer with a chain of at most 10 carbon atoms. We will represent this alkane isomer using a 10-digit number. First of all, carbon atoms in the straight chain are labelled, followed by branched ones. Each digit now corresponds to a number of carbon atoms connected to this carbon. The following example illustrates the encoding procedure for 2, 4-dimethyl-3-ethylpentane (Figure 1).

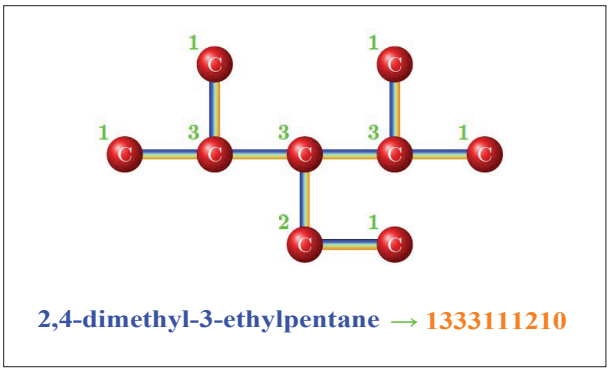

Figure 1: Conversion of an alkane to a machine-readable code

Model architecture of ANN

Neural networks are computational models inspired by the human brain's structure, designed to recognize patterns and solve complex problems. They consist of interconnected layers of neurons that learn from data through training, enabling them to make predictions and decisions. This adaptive learning capability has led to significant advancements in fields such as image recognition, natural language processing, and predictive analytics. An example schematic of an ANN is depicted in Figure 2.

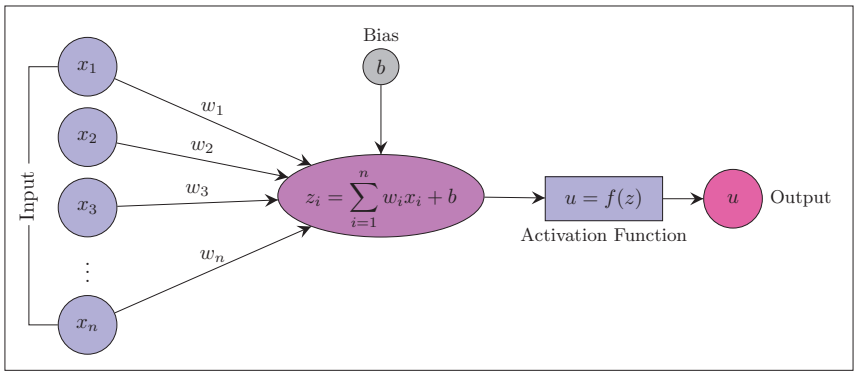

Figure 2: A schematic of a single artificial neuron, illustrating how the weighted sum of inputs plus bias is passed through an activation function to produce the output.

PyTorch to model a 3-layer ANN:

Input Layer: The input layer consists of 10 neurons, each representing one input feature. These neurons receive the input data and pass it to the next layer.

Hidden Layer: The hidden layer contains 7 neurons. Each neuron in the hidden layer receives inputs from all the neurons in the input layer. We used the sigmoid function as the activation function for each neuron in the hidden layer. This layer captures patterns and features in the input data.

Output Layer: The last layer of the network consists of a single neuron. It receives a weighted sum from all the neurons in the hidden layer. The output neuron predicts the normal boiling point of the corresponding alkane. The model is depicted in Figure 3.

Model architecture of graph neural network (GNN)

The GNN model used to predict the boiling point is depicted in Figure 4. This model is a GNN designed to predict the boiling points of chemical compounds using molecular structures represented as graphs. Each molecule is transformed into a graph where atoms serve as nodes with atomic number and degree as features, while bonds act as edges with bond type attributes. Additionally, molecular descriptors, as described in data encoding above, are included to enhance predictive accuracy. The GNN architecture comprises two graph convolutional layers (GCNConv) that extract relational features, followed by a fully connected layer that combines the graph embeddings with the molecular descriptors for final regression. The model is trained using both L1Loss and MSELoss loss functions and optimized using three optimizers: SGD, Adam, and Rprop over 3000 epochs. The best model is saved and used for inference, where predictions are exported to an Excel file.

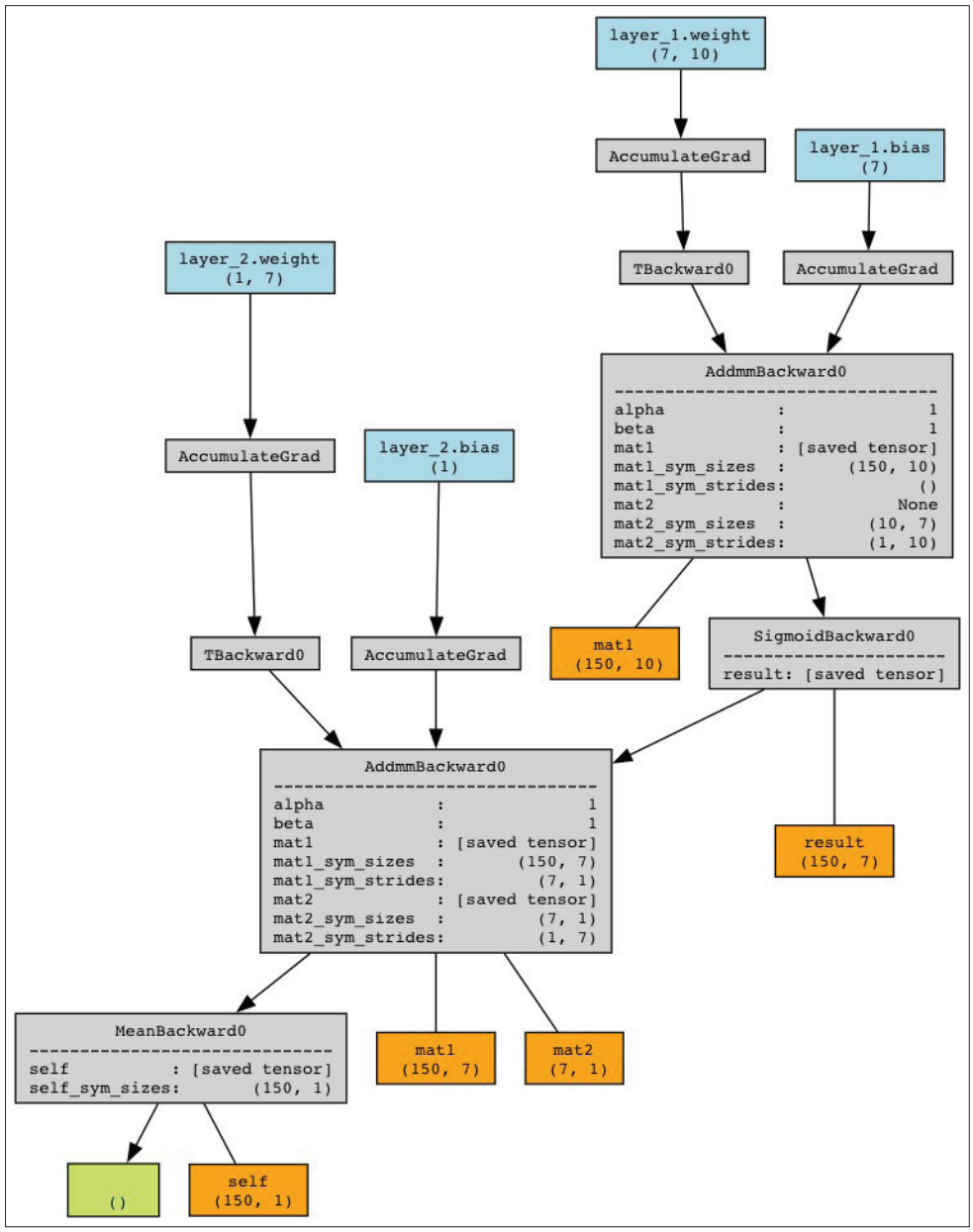

Figure 3: mat1 (input matrix) 150×10: 150-alkanes, each having 10-digit code; weight matrix (7×10) applied at the first layer. This results in a (150×7) matrix on which the sigmoid activation is applied. On the second layer weight matrix (1×7) is applied. This results in the final output matrix (150×1). The small grey rectangles store information about the operation applied at each layer, this helps the optimizer to keep track of the gradient descent operations. The α and β parameters are the values used by the optimizer during the backward pass. The stride parameter indicates that PyTorch inherently implements a convolution mechanism to perform matrix multiplications.

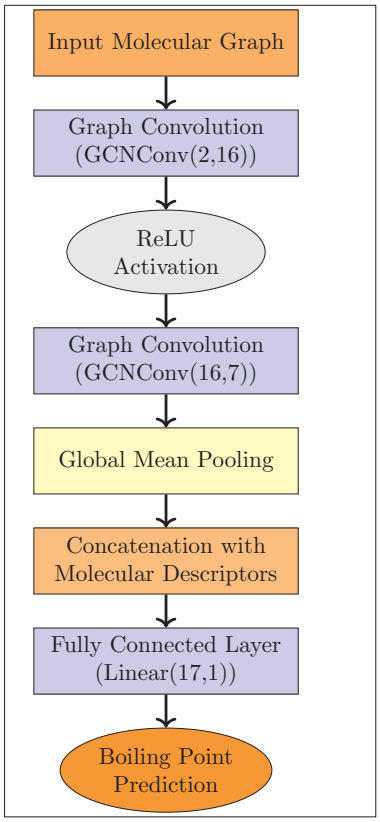

Figure 4: GNN Boiling Point Flowchart

RESULTS AND DISCUSSION

In this section, we analyze the role of three optimizers, SGD, Adam, and Rprop, in the performance of our proposed models, ANN and GNN, in predicting the boiling points. The models were trained using each of the three optimizers and evaluated with two loss functions, L1Loss and MSELoss, across three learning rates (0.01, 0.001, and 0.0001). By systematically comparing these variations, we aim to assess the impact of different training configurations on prediction accuracy. The results provide insights into the effectiveness of ANN and GNN in capturing the underlying structure-property relationships in alkanes and highlight the most suitable model configurations for accurate boiling point prediction.

It can be seen in Figure 5 that Adam outperforms in both models. In the ANN model, Rprop proves more effective than SGD, while in the GNN model, SGD shows superior performance to Rprop. Note that Figure 5 compares only the validation dataset. Although the training loss continuously decreases, it is the validation performance that more accurately reflects potential overfitting, as improvements on the training data do not necessarily translate to better generalization.

Table 1: MSE values for ANN and GNN under three optimizers (SGD, Adam, Rprop), two loss functions (L1Loss, MSELoss), and three learning rates (0.01, 0.001, 0.0001).

Loss Function	Learning Rate	ANN			GNN		
		SGD	Adam	Rprop	SGD	Adam	Rprop
L1	0.01	82.37	24.26	1322.45	109.69	12.91	1756.40
	0.001	2268.88	40.49	1180.48	67.68	19.59	1944.94
	0.0001	2541.31	235.00	1277.06	95.90	30.56	1825.25
MSE	0.01	94.97	7.98	520.13	32.96	14.13	1372.73
	0.001	2435.04	80.14	465.91	219.83	23.80	1382.61
	0.0001	2193.27	26.94	535.52	54.35	21.19	1579.77

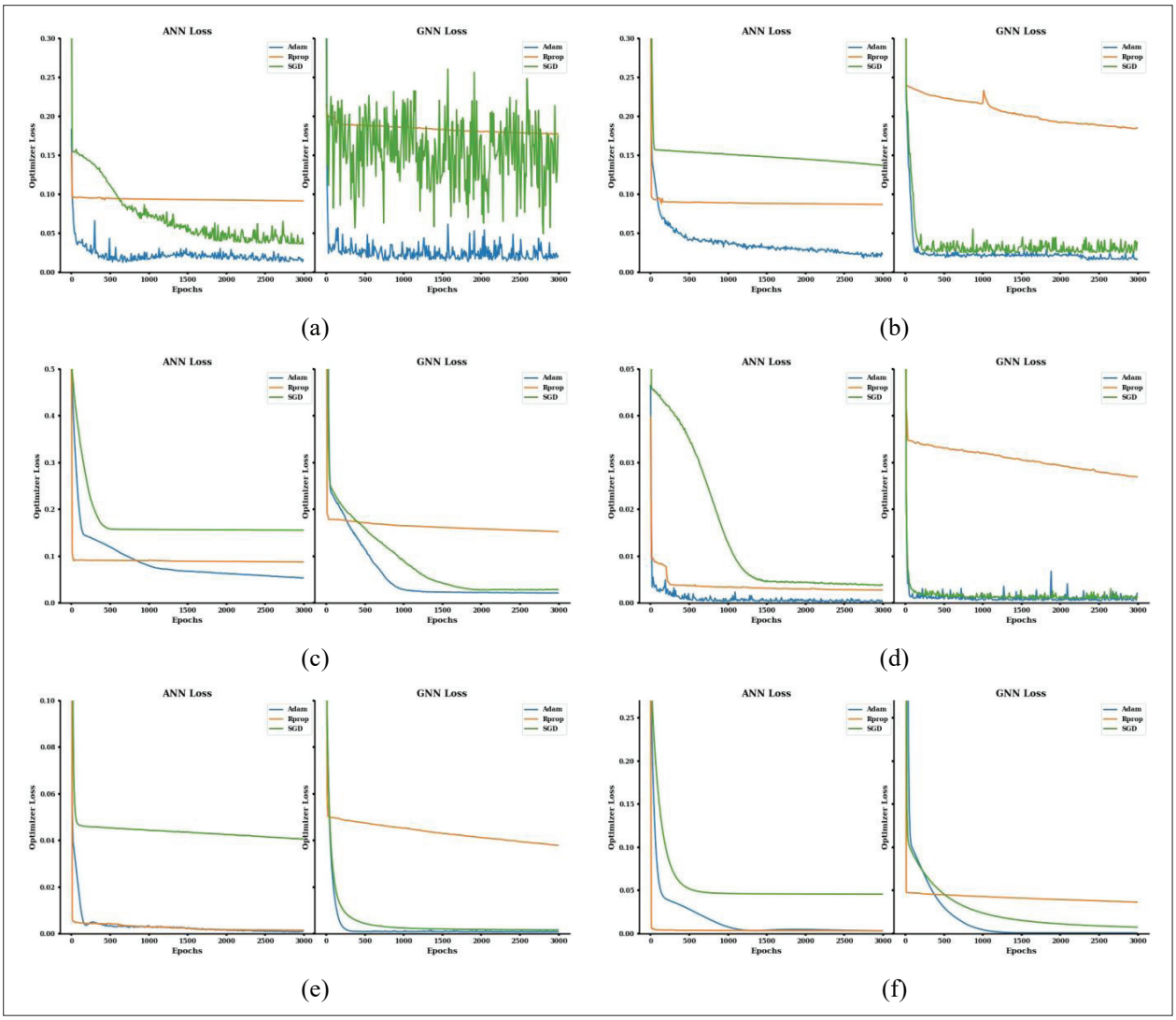

Figure 5: (a)-(c) use L1Loss, while (d)-(f) use MSELoss, both with learning rates 0.01, 0.001, and 0.0001.

Statistical analysis

Table 1 presents the mean squared error (MSE) results obtained from the ANN and GNN models when trained using three optimizers (SGD, Adam, and Rprop) in combination with two loss functions (L1Loss and MSELoss) across three learning rates (0.01, 0.001, and 0.0001). Notably, these MSE values are calculated on the entire dataset—including the training, validation, and test sets—offering a comprehensive evaluation of model performance. Overall, Adam tends to yield lower MSE values compared to SGD and Rprop for both models, though the extent varies with the choice of loss function and learning rate. Rprop exhibits larger variability in some cases, while SGD occasionally achieves competitive performance, particularly in the GNN model. Furthermore, the trends observed in Table 1 align with those in the Figure 5, which focused solely on validation data, reinforcing the consistency of the observed optimizer performances. These findings underscore the importance of carefully selecting the optimizer, loss function, and learning rate to achieve optimal boiling point prediction accuracy.

CONCLUSION

This study demonstrates that neural networks, when effectively optimized using PyTorch, can accurately predict the normal boiling points of alkanes based on their molecular descriptors. Our experiments with both ANN and GNN architectures, trained with three optimizers (SGD, Adam, and Rprop), two loss functions (L1Loss and MSELoss), and various learning rates, reveal that overall, Adam consistently yields the lowest mean squared error. Notably, while Rprop outperforms SGD in the ANN model, the opposite trend is observed in the GNN model, where SGD shows superior performance over Rprop. For example, under MSELoss with a learning rate of 0.01, the ANN model achieved an MSE of 7.98 with Adam, compared to 94.97 with SGD and 520.13 with Rprop; similarly, the GNN model recorded an MSE of 14.13 with Adam, versus 32.96 with SGD and 1372.73 with Rprop. These findings underscore the potential of advanced neural network techniques in chemical property prediction, paving the way for their broader application in fields such as materials science and chemical engineering.

REFERENCES

Arjmand, F. & Shafiei, F. (2018). Prediction of the normal boiling points and enthalpy of vaporizations of alcohols and phenols using topological indices. *Journal of Structural Chemistry*, *59*, 748–754. Publisher: Springer. https://doi.org/10.1134/S0022476618030393

Cherqaoui, D. & Villemin, D. (1994). Use of a neural network to determine the boiling point of alkanes. *Journal of the Chemical Society, Faraday Transactions*, *90*(1), 97–102. Publisher: The Royal Society of Chemistry. https://doi.org/10.1039/ft9949000097

Cherqaoui, D., Villemin, D., Mesbah, A., Cense, J.-M., & Kvasnicka, V. (1994). Use of a neural network to determine the normal boiling points of acyclic ethers, peroxides, acetals and their sulfur analogues. *Journal of the Chemical Society, Faraday Transactions*, *90*(14), 2015–2019. https://doi.org/10.1039/ft9949002015

Dai, Y.-m., Zhu, Z.-p., Cao, Z., Zhang, Y.-f., Zeng, J.-l., & Li, X. (2013). Prediction of boiling points of organic compounds by QSPR tools. *Journal of Molecular Graphics and Modelling*, *44*, 113–119. Publisher: Elsevier. https://doi.org/10.1016/j.jmgm.2013.04.007

De Lima Ribeiro, F. A. & Ferreira, M. M. C. (2003). QSPR models of boiling point, octanol–water partition coefficient and retention time index of polycyclic aromatic hydrocarbons. *Journal of Molecular Structure: THEOCHEM*, *663*(1-3), 109–126. Publisher: Elsevier. https://doi.org/10.1016/j.theochem.2003.08.107

Dearden, J. C. (2003). Quantitative structure-property relationships for prediction of boiling point, vapor pressure, and melting point. *Environmental Toxicology and Chemistry: An International Journal*, *22*(8), 1696–1709. Publisher: Wiley Online Library. https://doi.org/10.1897/01-363

Duchowicz, P., Castro, E. A., & Toropov, A. (2002). Qspr Modeling of Normal Boiling Point of Aldehydes, Ketones and Esters by Means of Nearest Neighboring Codes Correlation Weighting. *Science Direct Working Paper*, (S1574-0331):04. https://ssrn.com/abstract=2969228

Fissa, M. R., Lahiouel, Y., Khaouane, L., & Hanini, S. (2019). QSPR estimation models of normal boiling point and relative liquid density of pure hydrocarbons using MLR and MLP-ANN methods. *Journal of Molecular Graphics and Modelling*, *87*, 109–120. Publisher: Elsevier. https://doi.org/10.1016/j.jmgm.2018.11.013

Gastegger, M., Kauffmann, C., Behler, J., & Marquetand, P. (2016). Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes. *The Journal of chemical physics*, *144*(19). https://doi.org/10.1063/1.4950815

Gharagheizi, F., Mirkhani, S. A., Ilani-Kashkouli, P., Mohammadi, A. H., Ramjugernath, D., & Richon, D. (2013). Determination of the normal boiling point of chemical compounds using a quantitative structure–property relationship strategy: Application to a very large dataset. *Fluid Phase Equilibria, 354*, 250–258. Publisher: Elsevier. https://doi.org/10.1016/j.fluid.2013.06.034

Ghavami, R., Najafi, A., & Hemmateenejad, B. (2009). QSPR studies on normal boiling points and molar refractivities of organic compounds by correlation-ranking-based PCR and PC–ANN analyses of new topological indices. *Canadian Journal of Chemistry*, *87*(11), 1593–1604. https://doi.org/10.1139/V09-109

Goll, E. S. & Jurs, P. C. (1999). Prediction of the normal boiling points of organic compounds from molecular structures with a computational neural network model. *Journal of chemical information and computer sciences*, *39*(6), 974–983. Publisher: ACS Publications. https://doi.org/10.1021/ci9900711

Gonz´alez, M. P., Toropov, A. A., Duchowicz, P. R., & Castro, E. A. (2004). QSPR calculation of normal boiling points of organic molecules based on the use of correlation weighting of atomic orbitals with extended connectivity of zero- and first-order graphs of atomic orbitals. *Molecules*, *9*(12), 1019–1033. Publisher: MDPI. https://doi.org/10.3390/91201019

Groven, S. D., Desgranges, C., & Delhommelle, J. (2019). Prediction of the boiling and critical points of polycyclic aromatic hydrocarbons via Wang-Landau simulations and machine learning. *Fluid Phase Equilibria*, *484*, 225–231. https://doi.org/10.1016/j.fluid.2018.11.030

Jin, L. & Bai, P. (2016a). Modelling of normal boiling points of hydroxyl compounds by radial basis networks. *Mod. Chem.*, *4*(2), 24–29. https://doi.org/10.11648/j.mc.20160402.12

Jin, L. & Bai, P. (2016b). Prediction of the normal boiling point of oxygen containing organic compounds using quantitative structure–property relationship strategy. *Fluid Phase Equilibria*, *427*, 194–201. Publisher: Elsevier. https://doi.

org/10.1016/j.fluid.2016.07.015

Jin, L. & Bai, P. (2016c). QSPR study on normal boiling point of acyclic oxygen containing organic compounds by radial basis function artificial neural network. *Chemometrics and Intelligent Laboratory Systems*, *157*, 127–132. Publisher: Elsevier. https://doi.org/10.1016/j.chemolab.2016.07.007

Jin, L. & Bai, P. (2016d). QSPR study on normal boiling point of acyclic oxygen containing organic compounds by radial basis function artificial neural network. *Chemometrics and Intelligent Laboratory Systems*, *157*, 127–132. Publisher: Elsevier. https://doi.org/10.1016/j.chemolab.2016.07.007

Li, Q., Ren, J., Liu, Y., & Zhou, Y. (2022). Prediction of critical properties and boiling point of fluorine/chlorine-containing refrigerants. *International Journal of Refrigeration*, *143*, 28–36. https://doi.org/10.1016/j.ijrefrig.2022.06.024

Liu, B. & Nouroddin, M. K. (2023). Application of artificial intelligent approach to predict the normal boiling point of refrigerants. *International Journal of Chemical Engineering*, 2023. https://doi.org/10.1155/2023/6809569

Messerly, R. A., Knotts IV, T. A., Giles, N. F., & Wilding, W. V. (2017). Developing an internally consistent set of theoretically based prediction models for the critical constants and normal boiling point of large n-alkanes. *Fluid Phase Equilibria*, *449*, 104–116. https://doi.org/10.1016/j.fluid.2017.06.014

Mukwembi, S. & Nyabadza, F. (2021). A new model for predicting boiling points of alkanes. *Scientific reports*, *11*(1), 24261. https://doi.org/10.1038/s41598-021-03541-z

Qu, C., Kearsley, A. J., Schneider, B. I., Keyrouz, W., & Allison, T. C. (2022). Graph convolutional neural network applied to the prediction of normal boiling point. *Journal of Molecular Graphics and Modelling*, *112*, 108149. Publisher: Elsevier. https://doi.org/10.1016/j.jmgm.2022.108149

Saadi, S., Asrin, B., & Amin, R. (2015). Prediction the normal boiling points of primary, secondary and tertiary liquid amines from their molecular structure descriptors. *CMST*, *21*(4), 201–210. Publisher: PSNC, Poznan Supercomputing and Networking Center. https://doi.org/10.12921/cmst.2015.21.04.004

Santak, P. & Conduit, G. (2019). Predicting physical properties of alkanes with neural networks. *Fluid Phase Equilibria*, *501*, 112259. https://doi.org/10.1016/j.fluid.2019.112259

Sharma, S., Sleijfer, J. J., Op de Beek, J., van der Zeeuw, S., Zorzos, D., Lasala, S., Rigutto, M. S., Zuidema, E., Agarwal, U., Baur, R., *et al.* (2024). Prediction of thermochemical properties of long-chain alkanes using linear regression: Application to hydroisomerization. *The Journal of Physical Chemistry B*, *128*(39), 9619–9629. https://doi.org/10.1021/acs.jpcb.4c05355

Sumie, T. & Umpei, N. (2010). Generation of a novel equation for molecular boiling point estimation using a neural network. *Journal of Computer Chemistry, Japan*, *9*(2), 73–78. https://doi.org/10.2477/jccj.H2128

Tang, N., Tan, J., Sun, D., Li, L., & Li, X. (2024). Boiling temperature prediction model of environmental friendly insulation molecules based on machine learning. In 2024 7th International Conference on Electric Power Equipment-Switching Technology (ICEPE-ST), pages 795–799. IEEE. https://doi.org/10.1109/ICEPE-ST61894.2024.10792464

Toropov, A., Toropova, A., & Benfenati, E. (2010). QSPR modelling of normal boiling points and octanol/water partition coefficient for acyclic and cyclic hydrocarbons using SMILES-based optimal descriptors. *Central European Journal of Chemistry*, *8*, 1047–1052. Publisher: Springer. https://doi.org/10.2478/s11532-010-0072-5

Van den Maagdenberg, H. W., Šícho, M., Araripe, D. A., Luukkonen, S., Schoenmaker, L., Jespers, M., B´equignon, O. J., Gonz´alez, M. G., van den Broek, R. L., Bernatavicius, A., *et al.* (2024). Qsprpred: a flexible open-source quantitative structure-property relationship modelling tool. *Journal of Cheminformatics, 16*(1), 128. https://doi.org/10.1186/s13321-024-00908-y

Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, *69*(1), 17–20. Publisher: ACS Publications. https://doi.org/10.1021/ja01193a005