Physics-informed machine learning for predicting temperature-dependent chemical properties

Mahyar Rajabi Kochi,[1,2] Hanie Rezaei,[3] Sartaaj Takrim Khan,[1,2] Bhanu Mamillapalli,[1] Maryam Ebrahimiazar,[3] Haoming Ye,[3] Rose Moosavian,[3] Mohammad Zargartalebi,[3] David Sinton,[3] Seyed Mohamad Moosavi[1,2,*]

[1]Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, ON, M5S 3E5, Canada
[2] Vector Institute for Artificial Intelligence, Toronto, ON, M5G 0C6, Canada
[3] Mechanical & Industrial Engineering, University of Toronto, Toronto, ON, M5S 3G8, Canada
*mohamad.moosavi@utoronto.ca

1. Learning Curve

Fig. S1 is depicting learning curve as a collection of data points that in this case describe how the performance of physics-informed model's performance is related to training sample sizes. These learning curve can typically be divided into three sections: In the first section, the model's performance improves rapidly with an increase in the size of the training set; the second section is characterized by a turning point where the increase in performance is less rapid and a final section where the model has reached its efficiency threshold, i.e. no (or only marginal) improvement in performance is observed with increasing training set size, both regarding training and validation data. The differences between the predicted and actual dynamic viscosity in train and validation datasets are found to be in the range of 0.2 ± 0.01 and 0.34 ± 0.01 when train set size is greater than 342. Any increase beyond this value does not enhance accuracy significantly.

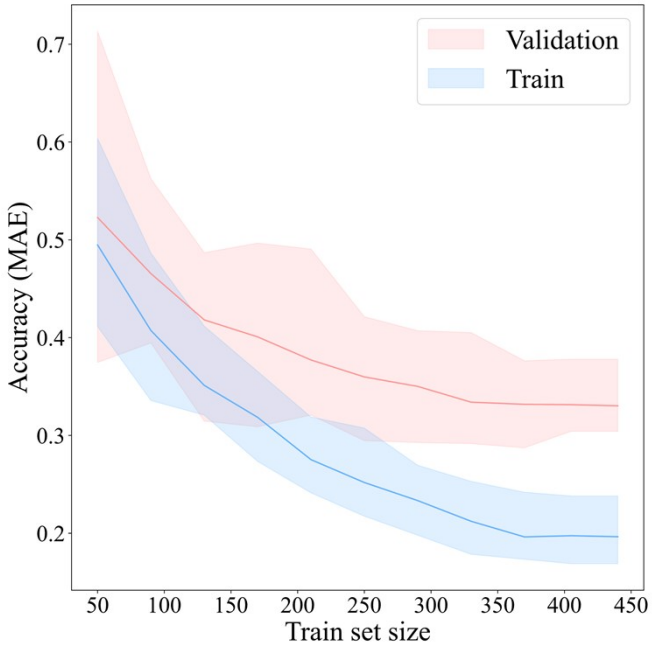


Fig. S1. Learning curve of physics-informed model in case of dynamic viscosity prediction. The method is tested on several relatively small data subsets (n = 50 to 450). The differences between the predicted and actual dynamic viscosity in train and validation datasets are found to be in the range of 0.2 ± 0.01 and 0.34 ± 0.01 when train set size is greater than 342.

2. Benchmarking physics-informed model across temperature ranges

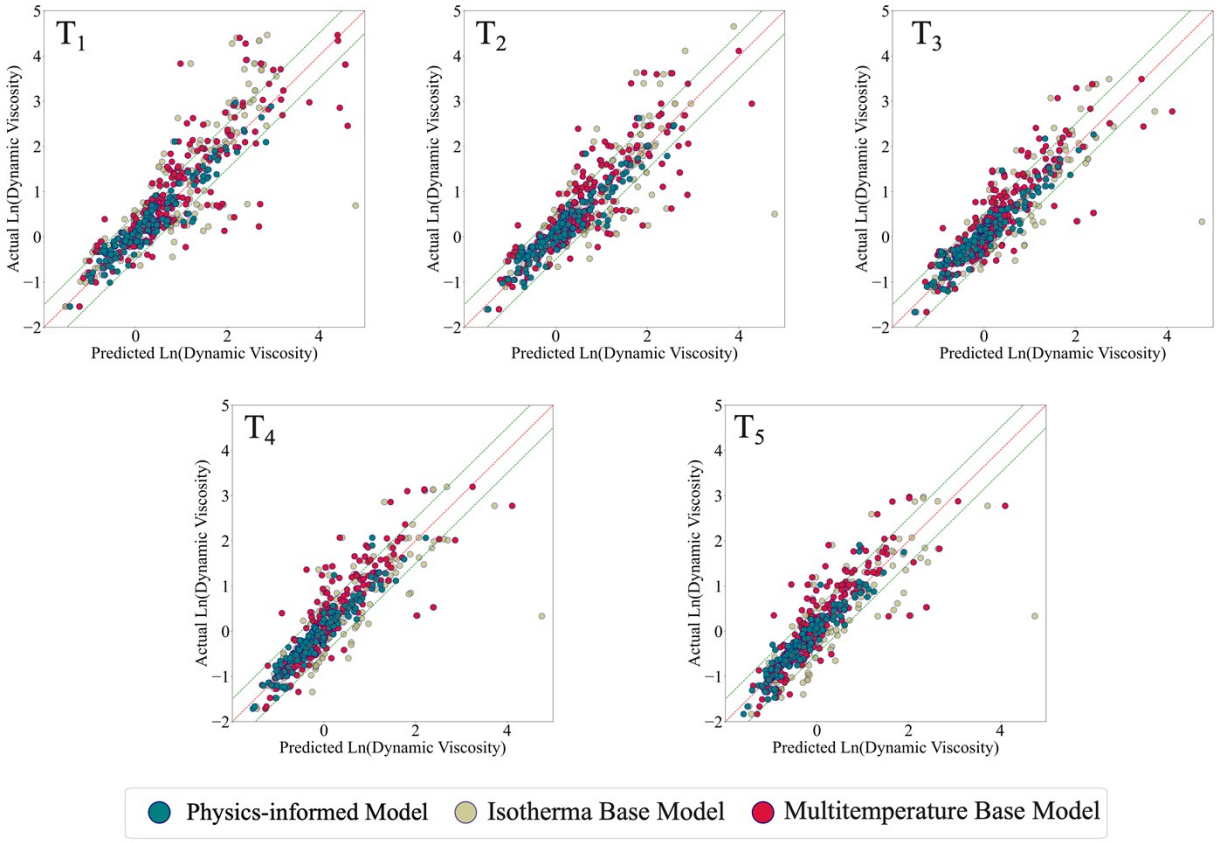


Fig. S2. Prediction accuracy comparison across temperature. Comparing prediction accuracy of the physics-informed model with 2 different base models at five temperature levels using visual parity plots. Temperature distribution at each level can be seen at Fig. 2C.

3. Prediction accuracy comparison at an unseen temperature

We use 70% of the dataset for training and reserve the remaining 30% as an unseen test set. While this split ensures chemical diversity in the test data, all temperature levels are still included in the training set, meaning the model is evaluated on the similar temperature range it has seen during training. To assess the model's ability to generalize to unseen temperature conditions, we introduce an additional split: the model is trained using data from four temperature levels (T_2, T_3, T_4, and T_5) and evaluated on the withheld temperature level T_1. Comparison with the base model shows that the physics-informed model performs well even when tested on temperatures outside the training distribution (Fig. S3).

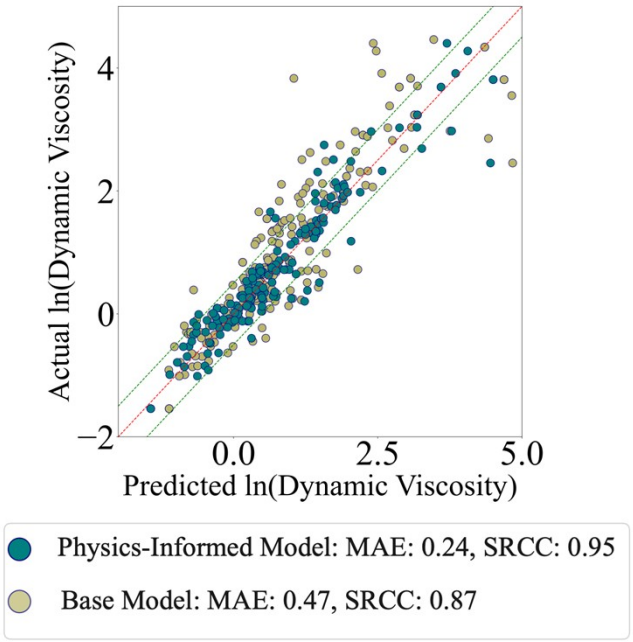


Fig. S3. Prediction accuracy comparison at an extrapolated temperature. Parity plots comparing the prediction accuracy of the physics-informed model and the base model, both trained on datasets containing four temperature levels (T_2, T_3, T_4, and T_5) and evaluated at the unseen temperature level T_1.

4. Model Evaluation Across Chemical Classes

To evaluate the generalizability of the physics-informed model across different chemical structures, we select three key criteria for classification and analysis. First, the chemicals are categorized based on the nature of their hydrocarbon portions, distinguishing between alicyclic, aliphatic, and aromatic compounds. For alicyclic and aliphatic hydrocarbons, the model demonstrates high confidence with over 65% of predictions flagged as certain, strong correlations between predicted and actual values with SRCC of 0.95, and low prediction errors with MAE values of 0.25 and 0.21, respectively (Fig. S4A and Table 1). The model shows an SRCC of 0.93 and the MAE of 0.27 for aromatic species.

Second, we classify the molecules by the type of heteroatoms, including oxygenated, nitrogenous, sulfuric, hydrocarbon, halogenated species. The materials with no or more than one heteroatom are categorized as pure hydrocarbon and multi-heteroatoms, respectively. The physics-informed model demonstrates strength in predicting the dynamic viscosity of both nitrogen- and oxygen-containing compounds. For nitrogenous fluids, primarily amines and nitro types, the model exhibits a 71% predictive certainty, underscored by an SRCC of 0.94. For oxygen-containing compounds, which include a variety of functional groups such as esters, alcohols, carboxylic acids, ketones, and ethers, the model achieves the SRCC of 0.9. For halogenated hydrocarbons, however, the SRCC drops to 0.74. This decline arises from the limited number of halogenated samples (~3% of the dataset). Their distribution is also skewed, as these compounds exhibit the lowest average viscosities within a narrow range (Fig. S4B)[1,2].

As a third criteria, we consider the molecular weight (MW), dividing the chemicals into three groups: MW < 150 $g.mol^{-1}$, 150 $g.mol^{-1}$ < MW < 250 $g.mol^{-1}$, and MW > 250 $g.mol^{-1}$. The physics-informed model demonstrates high predictive accuracy across all classes of molecular weight in the dataset, with SRCC ranging from 0.91 to 0.93 and MAE from 0.19 to 0.23. Notably, predicting dynamic viscosity in higher molecular weight species. (molecules with MW > 250 $g.mol^{-1}$) is inherently more challenging compared to low MW molecules due to their complex molecular interactions and structural characteristics. Higher molecular weight species often exhibit extensive chain entanglements, increased van der Waals forces, and more pronounced steric effects, all

of which contribute to a more intricate and less predictable viscosity behavior. Despite these challenges, the physics-informed model performs well for high MW chemicals, with SRCC 0.93, even though they only constitute 8% of the training dataset (Fig. S4C).

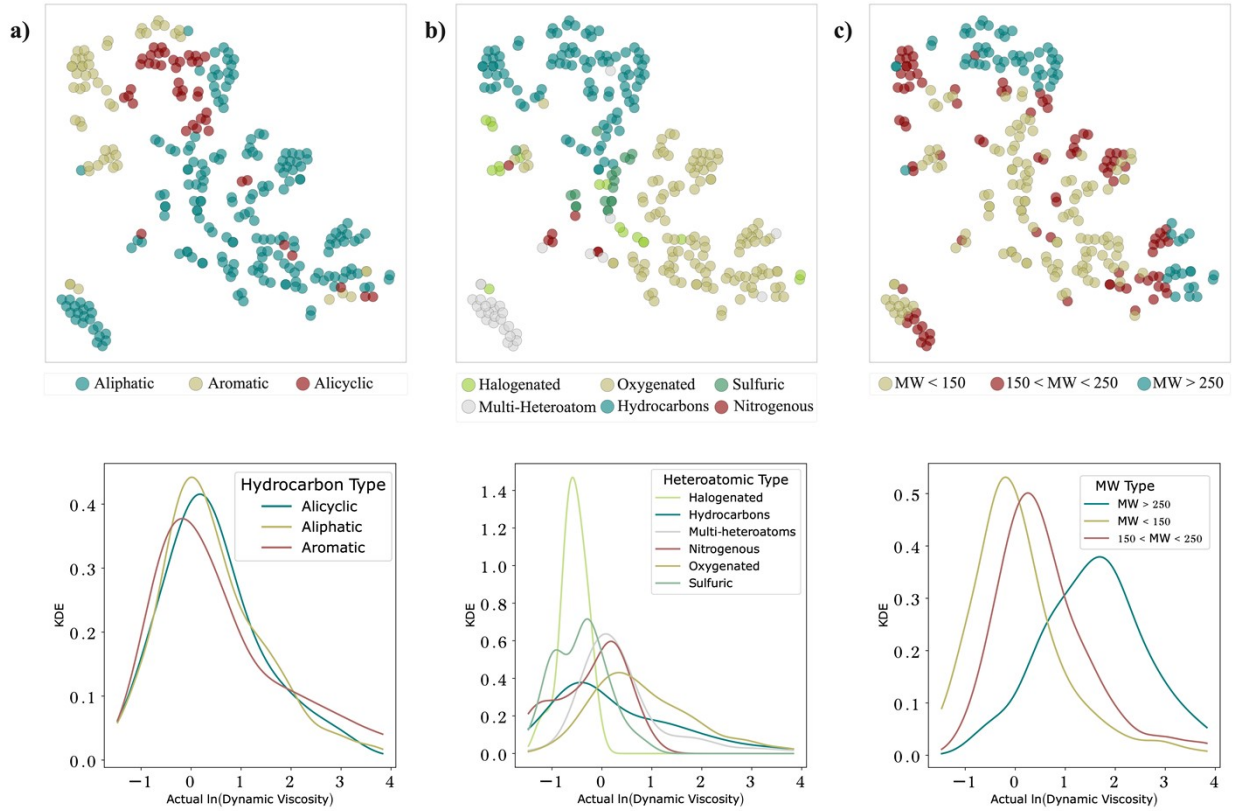


Fig. S4 t-SNE projection of the test dataset with respect to chemical diversity of the dataset. Prediction accuracy has been reported based on three chemistry-based criteria including (**A**) hydrocarbon type (alicyclic, aromatic, and aliphatic), (**B**) heteroatom type (halogen, oxygen, nitrogen, sulfur, pure hydrocarbons and materials with multiple heteroatoms), and (**C**) molecular weight (MW < 150 g.mol^{-1}, 150 g.mol^{-1} < MW < 250 g.mol^{-1}, and MW > 250 g.mol^{-1}) accompanied by kernel density estimates (KDE) for each class of materials.

5. Uncertainty Assessment

Ensemble approaches have been widely used across various fields to enhance prediction accuracy and quantify uncertainty. The core idea is that individual models within an ensemble can complement one another by leveraging their strengths and mitigating weaknesses. As a result, aggregating their outcomes often boosts overall model performance[3].

Fig. S5 illustrates a typical ensemble model framework for dynamic viscosity prediction. A training dataset consisting of chemical descriptors, temperature data points, and their associated dynamic viscosity values is used to build the individual models in the ensemble. Once constructed, the ensemble predicts dynamic viscosity for unseen test data, using the numerical descriptors. The effectiveness of an ensemble approach relies on two critical components: creating diversity among individual models and effectively aggregating their predictions [4].

To achieve diversity within an ensemble, various strategies can be employed, such as using different modeling techniques, applying different parameter settings within the same model type, or training models on varied datasets. Techniques like bootstrapping and boosting are commonly utilized for this purpose [5]. In this work, we adopt bagging (Bootstrap Aggregating) to construct 20 individual models. bagging works by training each model on different versions of the training dataset, generated through random sampling with replacement. This ensures that each model is trained on a slightly different subset of the data, enhancing diversity within the ensemble.

The aggregation of individual model outcomes is the second critical component of an ensemble. Effective aggregation strategies include statistics-based methods (e.g., simple average (SA)) and performance-based approaches, such as globally weighted average (GWA), locally weighted average (LWA), or local fusion (LF) [6]. In practice, this involves assigning a weight to each individual model's prediction and computing a weighted average for the final ensemble prediction. For example, if n represents the individual models (where n=1, 2…,20), the dynamic viscosity prediction is aggregated as a weighted average of their outputs.

$$\bar{\mu} = \frac{\sum_{i=1}^{N} w_i \mu_i}{\sum_{i=1}^{N} w_i}$$ Equation S1

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mu_i - \bar{\mu})^2}$$ Equation S2

Where $\bar{\mu}$ is aggregated viscosity of ensemble of the models, μ_i viscosity of an individual model, w_i is weight associated with each model, σ is standard deviation and N is total number of ensembles of the models (N = 20). In this study, we aggregate all the outcomes of the individual models with equal weights, which makes them contribute equally to the final prediction. Furthermore, each model predicts dynamic viscosity indirectly by estimating the coefficients of the Arrhenius equation, namely activation energy and frequency factor in the pre-final layer (Fig. 1). A material is flagged as certain if the standard deviation of the predicted Arrhenius frequency factors does not exceed a predefined threshold; otherwise, it is filtered out and flagged as uncertain (Fig. S5B). In this work, a threshold of 0.6 was applied to the variance of ln(A). We note that using bagging variance directly on viscosity levels is less suitable, since viscosity values scale strongly with temperature, making the threshold arbitrary and dependent on the chosen temperature range. Accordingly, this threshold is dataset- and problem-specific and should not be interpreted as a universal or generally transferable criterion.

For the case that we have used other non-Arrhenius equations, we follow the same workflow discussed above. Standard deviation of the predicted pre-exponential factor across the ensemble is used as a measure of predictive uncertainty. This pre-factor is shared across all equations and therefore provides a common basis for comparison.

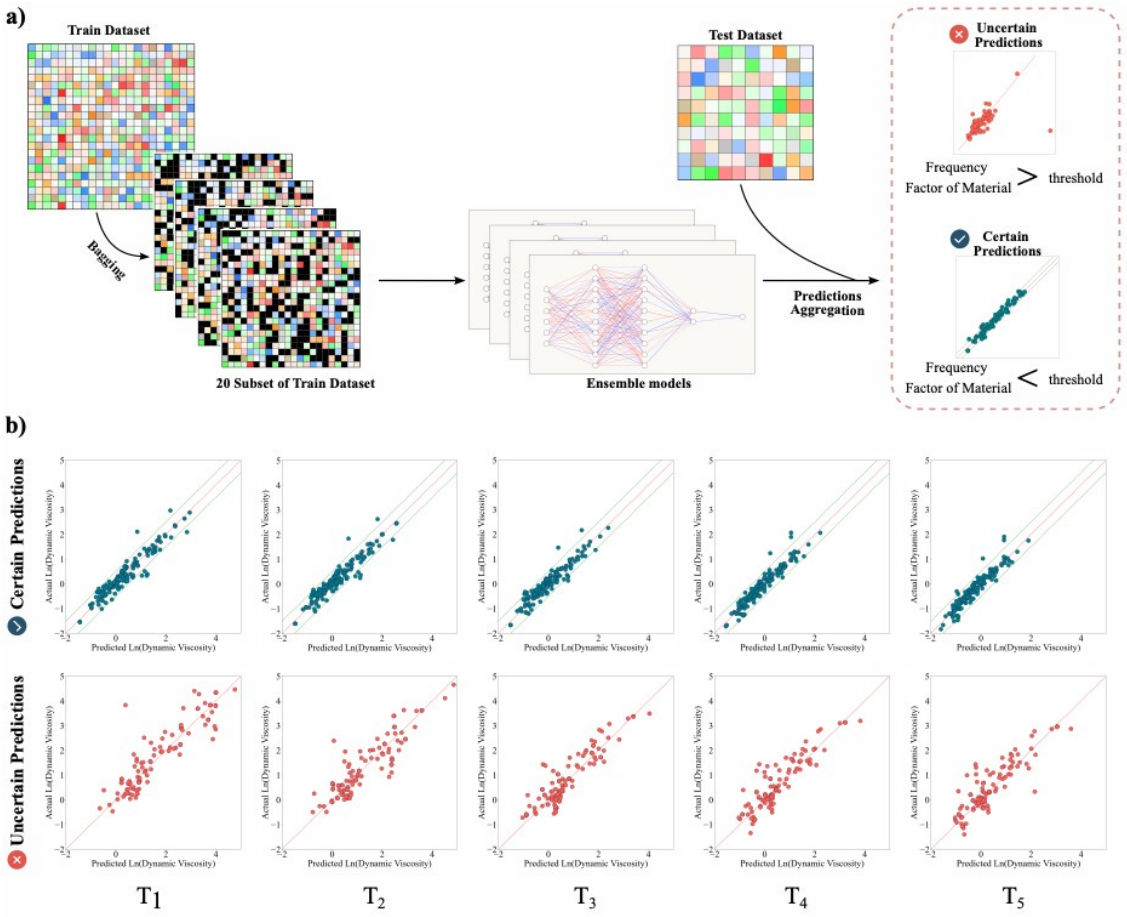


Fig. S5. Schematic of uncertainty assessment through training ensemble of the ANN models. (**A**) An ensemble of 20 artificial neural network models is trained using the bagging technique, where each model is trained on a different bootstrap sample of the original dataset. The training datasets are generated by randomly sampling with replacement, ensuring diversity among the models. Each ANN predicts dynamic viscosity indirectly by estimating the coefficients of the Arrhenius equation, namely activation energy and frequency factor. A material is flagged as certain if the standard deviation of the frequency factor, predicted using the ensemble of ANN models, does not exceed a predefined threshold. The final dynamic viscosity of certain materials is then aggregated using weighted averaging strategy and reported. (**B**) certain and uncertain predictions of physics-informed model for test dataset.

6. Cross-reference benchmarking of model at 288.15 and 298.15 K

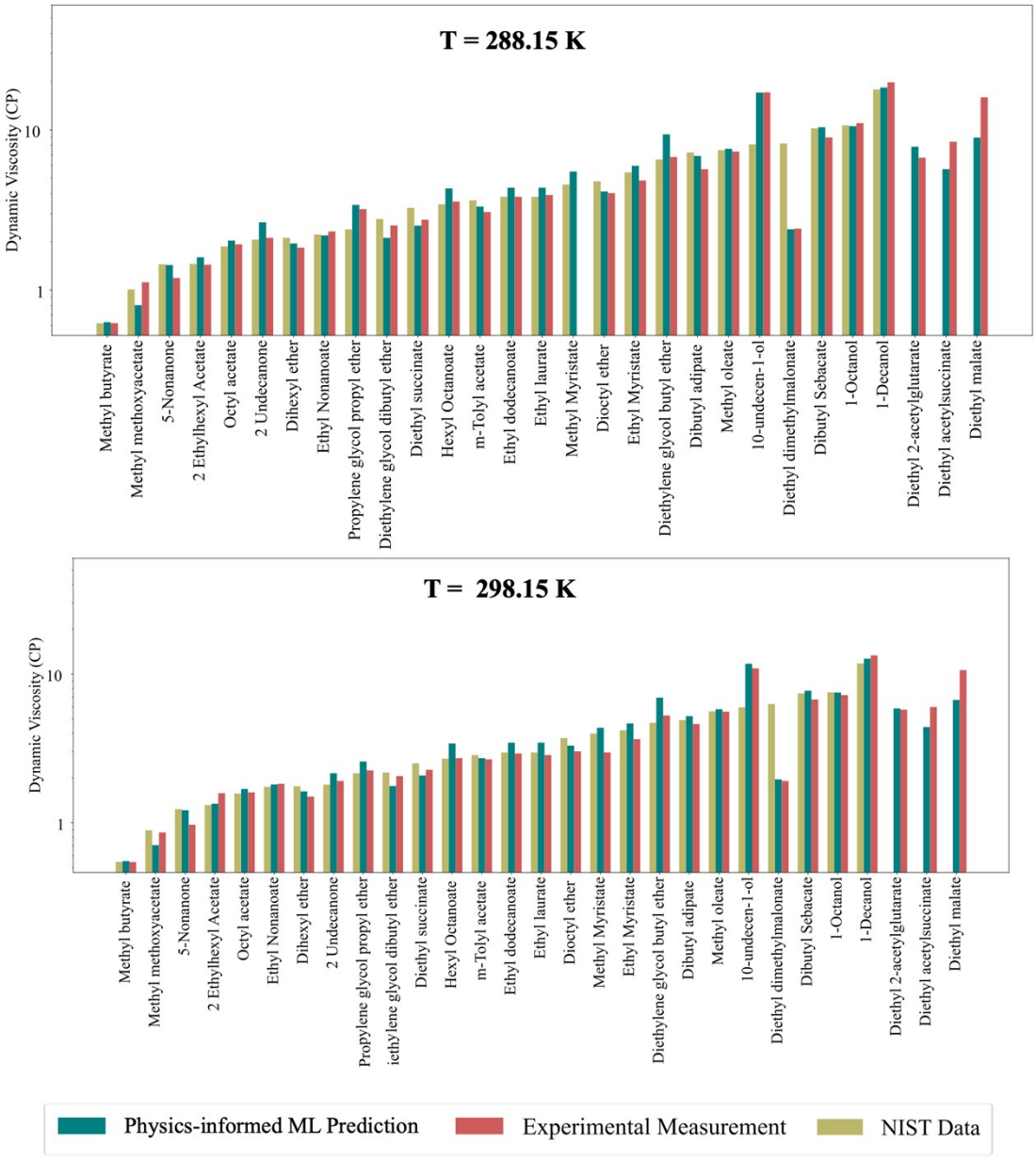


Fig. S6. Physics-informed model verification on thermal fluids. Verifying predictions of the physics-informed model against results obtained from experimental measurement and NIST database for a subset of test data comprising potential thermal fluids at 288.15 and 298.15 K.

7. List of the fluids simulated in the cooling system

Table S1. List of SMILES, chemical class, and hydrogen-bond donor capability of 50 fluids simulated in the cooling system. We obtain data from PubChem[7] and cross-verified with the NIST Standard Reference Database[8].

No	SMILES	Class	Hydrogen bonding donor capability
1	CCCCCCCCCCCCCC(=O)OC	Esters	No
2	CCCCCCCCCCCCCC(=O)OCC	Esters	No
3	CCCCCCCCC(=O)OCC	Esters	No
4	COC(=O)CCCCCCC(=O)OC	Esters	No
5	CCCCOC(=O)CCCCCCCCC(=O)OCCCC	Esters	No
6	CCCCCCCCCCCC(=O)OCC	Esters	No
7	CCCCOCCOCCOCCCC	Esters	No
8	CCCCOCCOCCO	Alcohols	Yes
9	CCOC(=O)CCC(=O)OCC	Esters	No
10	OCCCCCCCCCCC=C	Alcohols	Yes
11	CCCCC(=O)CCCC	Ketones	No
12	CC(=O)Oc1cccc(C)c1	Esters	No
13	CCCCCCCC(=O)OCCCCCC	Esters	No
14	CCCCCCCCOCCCCCCCC	Ethers	No
15	CC(C)COC(=O)CCCCC(=O)OCC(C)C	Esters	No
16	CCCCCCCC\C=C/CCCCCCCC(=O)OC	Esters	No
17	CCCC(=O)OC	Esters	No
18	CCCCCCCCOC(C)=O	Esters	No
19	COCC(=O)OC	Esters	No

20	CCCCCCCC(C)=O	Ketones	No
21	CCCOCC(C)O	Ethers	Yes
22	CCOC(=O)C(C)(C)C(=O)OCC	Esters	No
23	CCCCOC(=O)CCCCC(=O)OCCCC	Esters	No
24	CCCCCCCCO	Alcohols	Yes
25	CCCCCCCCCCO	Alcohols	Yes
26	CCCCCCCCCCCC(=O)OCC	Esters	No
27	CCCCCCCCCC(C)=O	Ketones	No
28	CCCCC(CC)COC(C)=O	Esters	No
29	CCCCC(CC)COC(=O)CCCCC(=O)OCC(CC)CCCC	Esters	No
30	CCCCCCOCCCCCC	Ethers	No
31	CCCCCCCCOC(=O)CCCCC(=O)OCCCCCCCC	Esters	No
32	C1CCc2ccccc2C1	Hydrocarbons	No
33	CCCCCCCCc1cc2CC(C)CCc2cc1C	Hydrocarbons	No
34	OCCO	Alcohols	Yes
35	C1COCO1	Ethers	No
36	CCCCCCC=C	Hydrocarbons	No
37	CCCS	Organosulfur	No
38	CC(C)CC(=O)CC(C)C	Ketones	No
39	CC(C)CCCCC(C)C	Hydrocarbons	No
40	CCCCOCCO	Alcohols	Yes
41	CCCCC(C)O	Alcohols	Yes
42	COCCO	Ethers	Yes
43	CC1CCCCC1O	Alcohols	Yes

44	CC(C)(C)S	Organosulfur	No
45	CCCCC1CCCC1	Hydrocarbons	No
46	C1CCC2CCCCC2C1	Hydrocarbons	No
47	CCCCCCCCCC(O)=O	Alcohols	Yes
48	CCCCCCCCCCCCCCCCCO	Alcohols	Yes
49	CCCCCCCCCCCO	Alcohols	Yes
50	OCCOCCOCCO	Alcohols	Yes

Table S2. Thermophysical properties of simulated fluids obtained from NIST database[8,9] at room temperature.

No	SMILES	Thermal Conductivity (W/m.K)	Density (kg/m^3)	Heat Capacity (J/Kg.K)
1	CCCCCCCCCCCCCC(=O)OC	0.16	861.0	1507.0
2	CCCCCCCCCCCCCC(=O)OCC	0.16	856.9	1515.8
3	CCCCCCCCC(=O)OCC	0.15	860.6	1459.0
4	COC(=O)CCCCCCC(=O)OC	0.16	1013.7	1282.5
5	CCCCOC(=O)CCCCCCCCC(=O)OCCCC	0.18	934.0	1419.5
6	CCCCCCCCCCCC(=O)OCC	0.16	855.8	1497.2
7	CCCCOCCOCCOCCCC	0.15	878.3	1525.4
8	CCCCOCCOCCO	0.16	948.4	1482.1
9	CCOC(=O)CCC(=O)OCC	0.15	1035.3	1220.7
10	OCCCCCCCCCC=C	0.15	911.4	1576.4
11	CCCCC(=O)CCCC	0.16	817.1	1550.5
12	CC(=O)Oc1cccc(C)c1	0.14	1090.4	1127.6
13	CCCCCCCC(=O)OCCCCCC	0.16	857.2	1497.2
14	CCCCCCCCOCCCCCCCC	0.16	802.9	1625.2
15	CC(C)COC(=O)CCCCC(=O)OCC(C)C	0.16	946.1	1360.7
16	CCCCCCCC\C=C/CCCCCCCC(=O)OC	0.14	870.4	2247.1

17	CCCC(=O)OC	0.13	892.4	1287.9
18	CCCCCCCCOC(C)=O	0.15	863.8	1442.1
19	COCC(=O)OC	0.15	1045.9	1195.5
20	CCCCCCCC(C)=O	0.16	817.2	1531.1
21	CCCOCC(C)O	0.14	881.3	1495.5
22	CCOC(=O)C(C)(C)C(=O)OCC	0.14	1156.6	1253.0
23	CCCCOC(=O)CCCCC(=O)OCCCC	0.17	956.0	1365.9
24	CCCCCCCCO	0.16	822.3	1504.3
25	CCCCCCCCCCO	0.16	826.9	1520.1
26	CCCCCCCCCCCC(=O)OCC	0.16	858.2	1497.2
27	CCCCCCCCCC(C)=O	0.16	822.2	1546.8
28	CCCCC(CC)COC(C)=O	0.13	870.6	1438.2
29	CCCCC(CC)COC(=O)CCCCC(=O)OCC(CC)CCCC	0.17	925.7	1453.4
30	CCCCCCOCCCCCC	0.14	792.3	1612.7
31	CCCCCCCCOC(=O)CCCCC(=O)OCCCCCCCC	0.18	1100.5	1457.0
32	C1CCc2ccccc2C1	0.13	965.1	1157.3
33	CCCCCCCCc1cc2CC(C)CCc2cc1C	0.11	898.4	1432.4
34	OCCO	0.26	1109.9	1334.0

35	C1COCO1	0.14	1058.6	1004.3
36	CCCCCCC=C	0.13	711.1	1584.7
37	CCCS	0.07	835.8	1253.2
38	CC(C)CC(=O)CC(C)C	0.14	802.2	1541.1
39	CC(C)CCCCC(C)C	0.12	719.7	1639.7
40	CCCCOCCO	0.15	896.0	1501.2
41	CCCCC(C)O	0.13	810.1	1570.3
42	COCCO	0.19	960.4	1409.7
43	CC1CCCCC1O	0.13	926.2	1335.8
44	CC(C)(C)S	0.05	794.7	1351.8
45	CCCCC1CCCC1	0.12	779.9	1459.9
46	C1CCC2CCCCC2C1	0.11	882.6	1285.6
47	CCCCCCCCCC(O)=O	0.15	1017.6	1461.2
48	CCCCCCCCCCCCCCCCO	0.19	967.0	1570.0
49	CCCCCCCCCCCO	0.17	829.2	1560.8
50	OCCOCCOCCO	0.19	1119.5	1404.1

8. Benchmarking model on pure hydrocarbon dataset

Panwar et al. compiled a dataset of 305 pure hydrocarbons ranging from C_8 to C_{50} [10]. These data were sourced from the American Petroleum Institute (API) Research Project 42, which provided a spectrum of molecular structures, including n-paraffins, branched-paraffins, 1-olefins, branched-olefins, nonfused ring naphthenes, fused ring naphthenes, nonfused ring aromatics, and fused ring aromatics [10]. The molecular weights of these compounds span from 110.20 to 703.30 $g.mol^{-1}$, while viscosity values range from 0.29 cP to 2.00×10^4 cP. Temperature coverage extends from 0 to 135 °C, yielding a total of 1292 viscosity data points. The aim of Panwar et al.'s[10] study was to predict the dynamic viscosity of these hydrocarbons at distinct temperature intervals. In each of their six developed models, temperature was treated as a direct input feature alongside additional static or dynamic descriptors acquired from molecular dynamics simulation. After identifying the most significant predictors, the authors partitioned their dataset into training, validation, and test sets using a random split. They then employed Gaussian Process Regression (GPR) to build the predictive models. Using the same training dataset, we develop a predictive physics-informed model. In contrast to their approach—where temperature is a direct input—we incorporate temperature through the Arrhenius equation. Testing this model on their test dataset demonstrates that employing a physics-informed pipeline not only improves accuracy metrics (Fig. S7) but also yields an explicit Arrhenius-type equation for each material. This explicit form facilitates integration into simulation workflows and transforms the otherwise black-box prediction process into a more physics-informed approach.

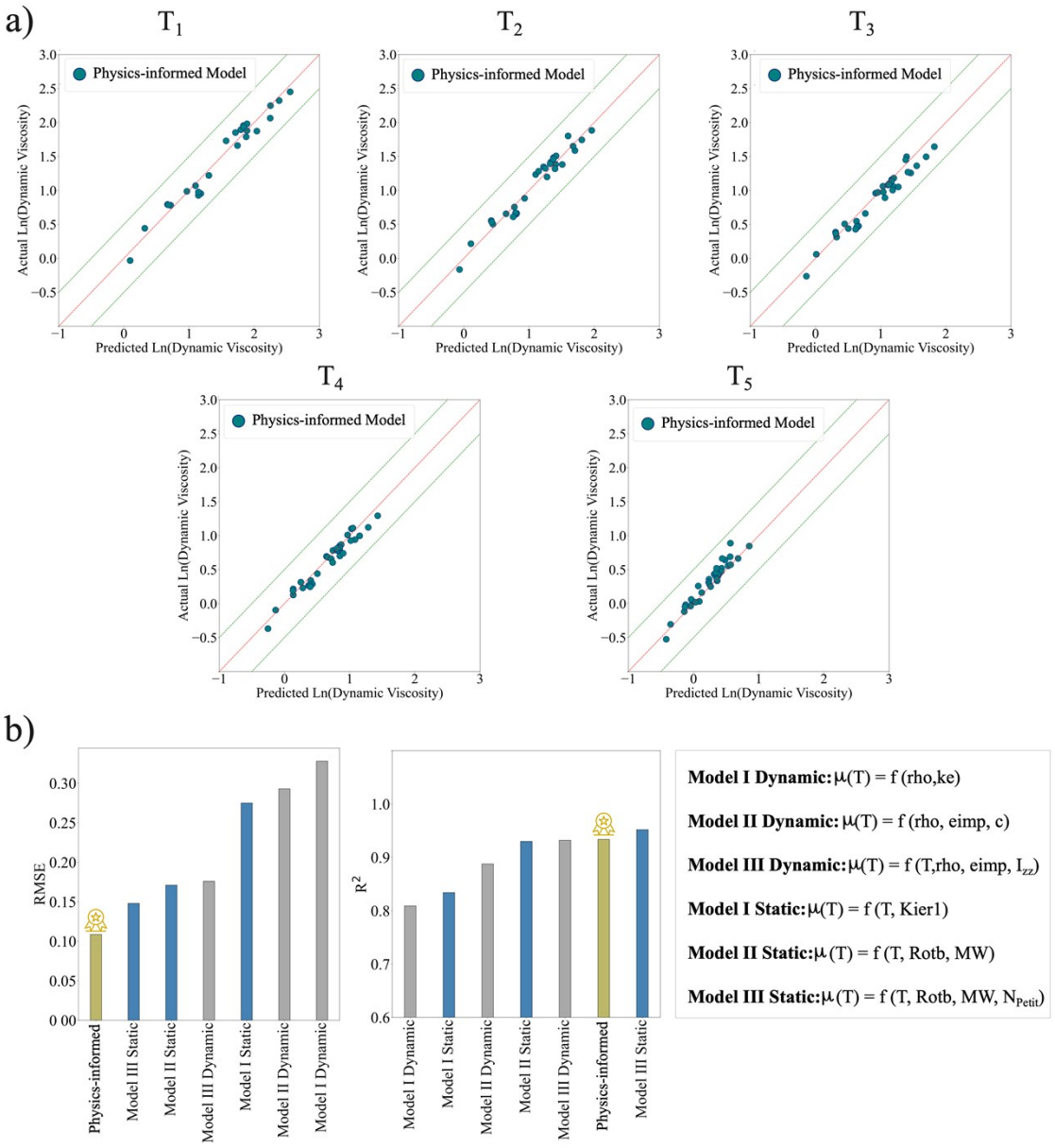


Fig. S7 Benchmarking the physics-informed model against the dynamic viscosity predictor by Panwar et al.[10] (**A**) Parity plots showing the predictions of the physics-informed model compared to true values. For this benchmarking, the same training and testing data reported by Panwar et al. at five temperature levels is used. (**B**) Comparison of the accuracy of the physics-informed model with six models developed by Panwar et al., each utilizing different feature sets.

9. Physics-informed model application in infinity diffusion coefficient prediction of materials in water

Infinity diffusion coefficient of materials in water is a temperature-dependent property, fundamentally related to molecular motion and the overcoming of energy barriers within a fluid. To demonstrate how the pipeline is extendable to other properties, a dataset containing 348 data points at three different temperatures has been collected from DIPPR dataset. We show that Arrhenius-type relationship is valid for more than 98% of the fluids within the dataset (Fig. S8). Training a feedforward neural network model using MORDRED-based numerical descriptors results in a model with robust performance across a wide temperature range from 304 K to 364 K. The high SRCC (0.98-0.99) and low RMSE values (0.05-0.08) across all temperature ranges highlight the reliability of the physics-informed model and underscore its transferability to different temperature conditions (Fig. 6A). Moreover, these results suggest that this predictive procedure is applicable to any Arrhenius-type properties, demonstrating its broad potential beyond dynamic viscosity prediction.

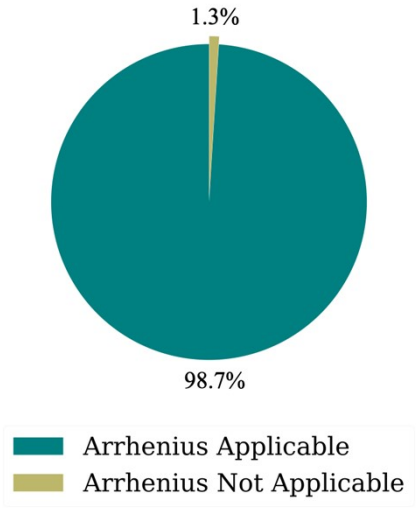


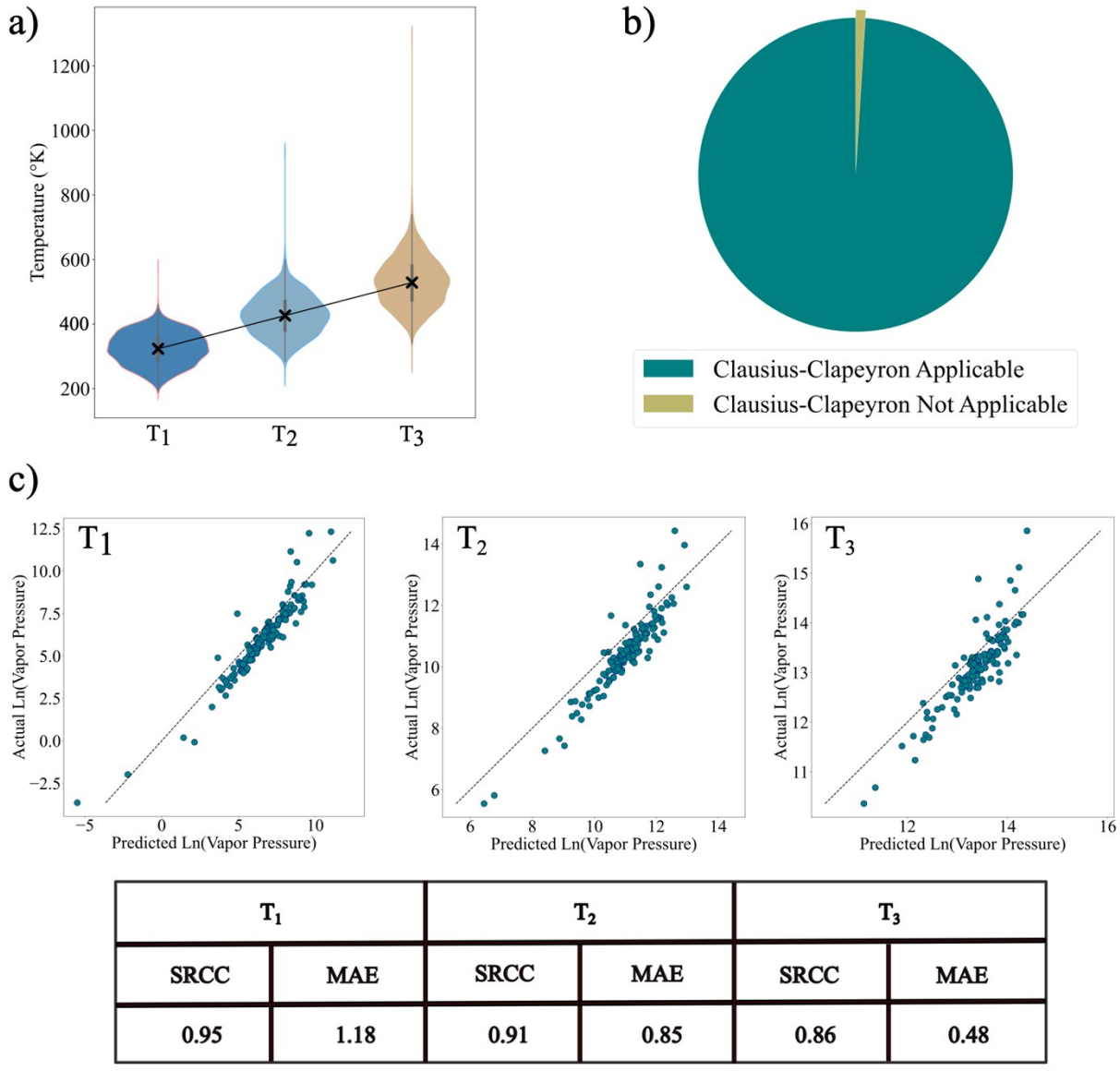
Fig. S8. Validation of the Arrhenius equation for 348 materials in the collected dataset for infinity diffusion coefficient of materials in water as the solvent.

10. Physics-informed, graph-based model for predicting vapor pressure

Vapor pressure of a fluid is one of the important thermophysical properties of fluids correlated with temperature through the Clausius-Clapeyron equation, where the coefficients represent the enthalpy of vaporization and a reference vapor pressure, capturing the thermodynamic equilibrium between liquid and gas phases. We collect a dataset of vapor pressure for approximately 1,400 materials from DIPPR dataset[11], spanning temperatures from 190.5 K to 1288 K, is analyzed, confirming the validity of the equation for over 99% of materials (See Fig. S9). Using numerical descriptors for material featurization, a model is trained to predict the coefficients of the Clausius-Clapeyron equation, achieving an SRCC above 0.91 across three temperature ranges (Fig. 6B and C). This demonstrates that the physics-informed property predictor pipeline is independent of the specific correlative equation used.

To better capture structural information about molecules and reduce inductive bias, a GNN based architecture is developed to predict vapor pressure of organic fluids. We use a directed message passing neural network, adapted directly from the published Chemprop library [12]. However, to allow our model to predict vapor pressure in a range of temperatures, the original framework is modified to use the linearized Clausius-Clapeyron equation to make predictions. The network now has two output nodes, analogous to the two coefficients in the equation. To fine tune the model, ray tune is used to iterate through set ranges of specified hyperparameters. The list of hyperparameters to optimize, as well as their search spaces is taken directly from the defaults provided in the Chemprop library (for a detailed list see Table S3). 1200 trials are run for 50 epochs each, with the first 600 being randomly sampled from the default hyperparameter search spaces. The next 600 are chosen by the FIFO scheduler from the ray library. The best trained model is then evaluated on the test set. Using graph-based representations, the model achieves acceptable accuracy, with an SRCC of approximately 0.95 at low temperatures and 0.86 at higher temperatures, showing that model performs better at lower temperatures (Fig. S9C). This highlights the flexibility of the pipeline to adapt to different representations,

making it a versatile tool for predicting a variety of temperature-dependent properties across diverse datasets and representations.



T_1		T_2		T_3	
SRCC	MAE	SRCC	MAE	SRCC	MAE
0.95	1.18	0.91	0.85	0.86	0.48

Fig. S9 Physics-informed model for vapor pressure prediction. (**A**) Temperature distribution of test materials across three levels. The temperature levels vary for each material due to differences in available temperature data. (**B**) Validation of the Clausius-Clapeyron equation for approximately 1,400 materials in the collected dataset. (**C**) Parity plots and accuracy metrics for

predicting vapor pressure across three temperature ranges using a graph-based representation derived from Chemprop for material representation.

Table S3. List of hyperparameters optimized in Chemprop using ray tune

Hyperparameter	Search Space	Interval Type
Activation Function (In MPNN)	ReLU, LeakyReLU, SELU, PReLU, ELU, TANH	Discrete
Graph Level Aggregation Function	sum, mean, norm	Discrete
Aggregation norm (only used for norm aggregation function)	[1,200]	Integers
Batch size	[16,256]	Powers of 2 (16, 32, 64…)
Message passing depth	[2,6]	Integers
dropout	[0, 0.4]	Multiples of 0.05
FFN hidden dim	[300, 2400]	Multiples of 100
MPNN hidden dim	[300, 2400]	Multiples of 100
FFN num layers	[1,3]	Integers
Max learning rate	[1e-4, 1e-2]	Continuous, log distributed
Initial learn rate (ratio of max)	[1e-2, 1]	Continuous, log distributed
Final learning rate (ratio of max)	[1e-2, 1]	Continuous, log distributed
Warmup Epochs (epochs where learn rate increments from initial to max)	[1, train epochs // 2]	Integer

11. Physics-informed model benchmarking at high unseen temperatures

Benchmarking against unseen temperature ranges is a vital test of extrapolation, as a robust model must predict dynamic viscosity outside the conditions used for training. We therefore withhold the highest temperature range (T_5) entirely from training, training the models only on intermediate-to-high temperature ranges (T_1–T_4), and then predict dynamic viscosities in T_5. A comparison of the two extrapolation experiments—where either the lowest (T_1) or the highest (T_5) temperature range is withheld—highlights systematic differences in model performance (Figs. 4 & S10). When extrapolating to T_1, Arrhenius and VFT clearly outperform MYEGA and AM, reflecting their ability to capture the steep super-Arrhenius rise in viscosity at low temperatures. In contrast, when extrapolating to T_5, Arrhenius and MYEGA perform comparably well, while VFT shows reduced accuracy and AM again performs weakest. This reduced performance of VFT at high temperatures arises from its divergence-based form, which is well suited to describing cooperative slowdown near the glass transition but tends to overestimate curvature in the Arrhenius-like regime at elevated temperatures. Taken together, these results show that the physical basis of each equation strongly influences extrapolation: VFT excels near glass transition temperature but underperforms at high temperature, Arrhenius provides stable predictions across regimes, MYEGA interpolates smoothly between the two, and AM lacks flexibility in both extremes.

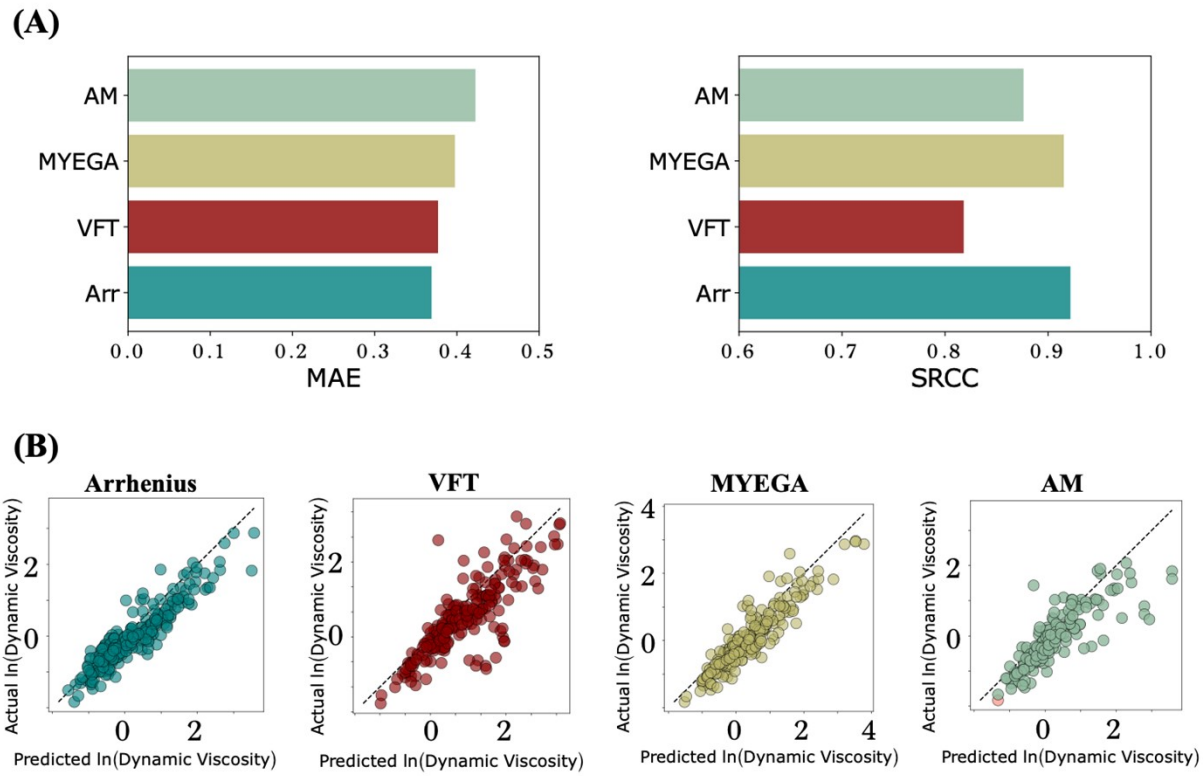


Fig S10. Benchmarking physics-informed viscosity models. (A) Accuracy metrics (MAE and SRCC) evaluated on the highest temperature range (T_5), which is excluded from training and remained completely unseen to the models. Arrhenius and VFT based models perform much better. **(B)** Parity plots of predicted vs. experimental dynamic viscosity highlight closer agreement for Arrhenius and VFT based models compared to MYEGA and AM.

12. Benchmarking Against Established Machine Learning Models

Benchmarking our approach against a set of models is important to ensure that the predictive performance we observe is not model-specific. To this end, we compare our method with several established machine learning approaches in which chemical descriptors are entered directly as numerical inputs and temperature is included explicitly as an additional feature, without any physics-informed architecture. These baselines include Random Forests (RF), an ensemble of decision trees; Support Vector Regression (SVR), which leverages kernel-based flexibility (here with an RBF kernel) to approximate nonlinear functional forms; Gaussian Process regression (GP), which places a prior distribution over functions to provide probabilistic predictions and quantify uncertainty; and XGBoost, a gradient-boosted tree method widely used for structured tabular data.

We also use a physics-informed kernel ridge regressor (PI-KRR) as one of the benchmarking models, which goes beyond treating temperature as a generic feature. This model explicitly embeds the Arrhenius functional form into the kernel basis, enforcing a linear dependence on inverse temperature while allowing descriptor-dependent flexibility through kernel similarity. In this sense, PI-KRR is ‘physics-informed’ because it constrains the hypothesis space to functions consistent with the known temperature–viscosity relationship, while still adapting to data through a residual kernel term. This semiparametric construction, embedding known physics into the model and allowing a kernel method or Gaussian process to capture the discrepancy, has been applied in other fields such as fluid dynamics and materials mechanics. However, to the best of our knowledge, this is the first application where the Arrhenius equation is explicitly embedded to ensure transferability of viscosity predictions across temperature ranges.

The results highlight the importance of explicitly incorporating physical equations when modeling temperature-dependent properties (Figs. S11 and S12). While XGBoost achieves comparable performance to the physics-informed neural network (ARR-NN) at high unseen temperatures, the advantage of embedding physics becomes evident at low unseen temperatures, where the viscosity range broadens, and the physics-based model significantly outperforms purely data-driven approaches. Interestingly, the physics-informed kernel ridge regression (PI-KRR) performs worse than ARR-NN at both

temperature extremes, suggesting that directly embedding the Arrhenius equation at the readout layer of a neural network provides better flexibility and expressiveness than constraining the hypothesis space through kernel similarity alone. This finding indicates that neural architectures may be better suited than kernel methods for capturing residual deviations from physically motivated functional forms.

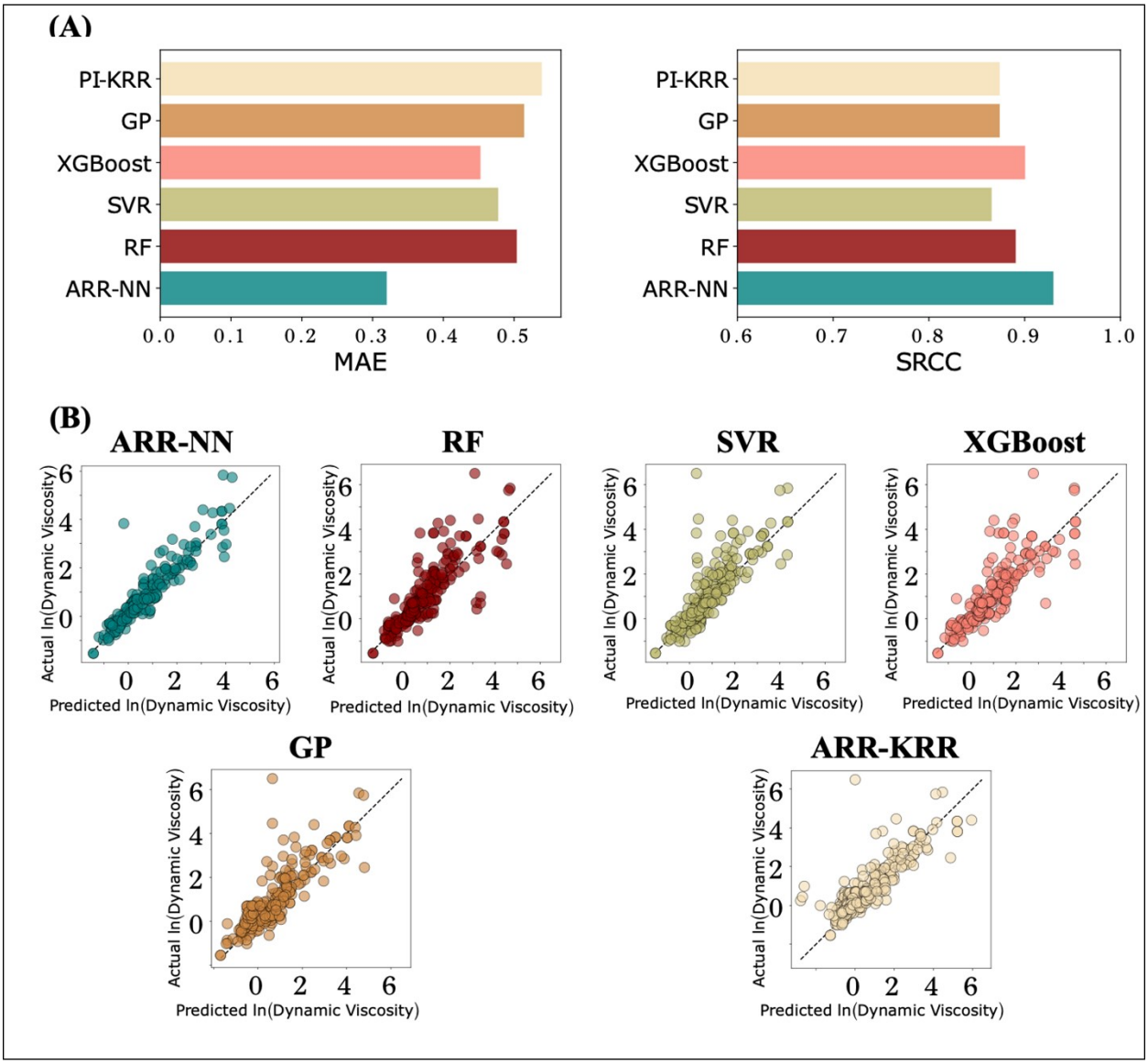


Fig S11. Benchmarking established ML models. (A) Accuracy metrics (MAE and SRCC) evaluated on the lowest temperature range (T_1), which is excluded from training and remained completely unseen to the models. Arrhenius based neural network model performs better. **(B)** Parity plots of predicted vs. experimental dynamic viscosity.

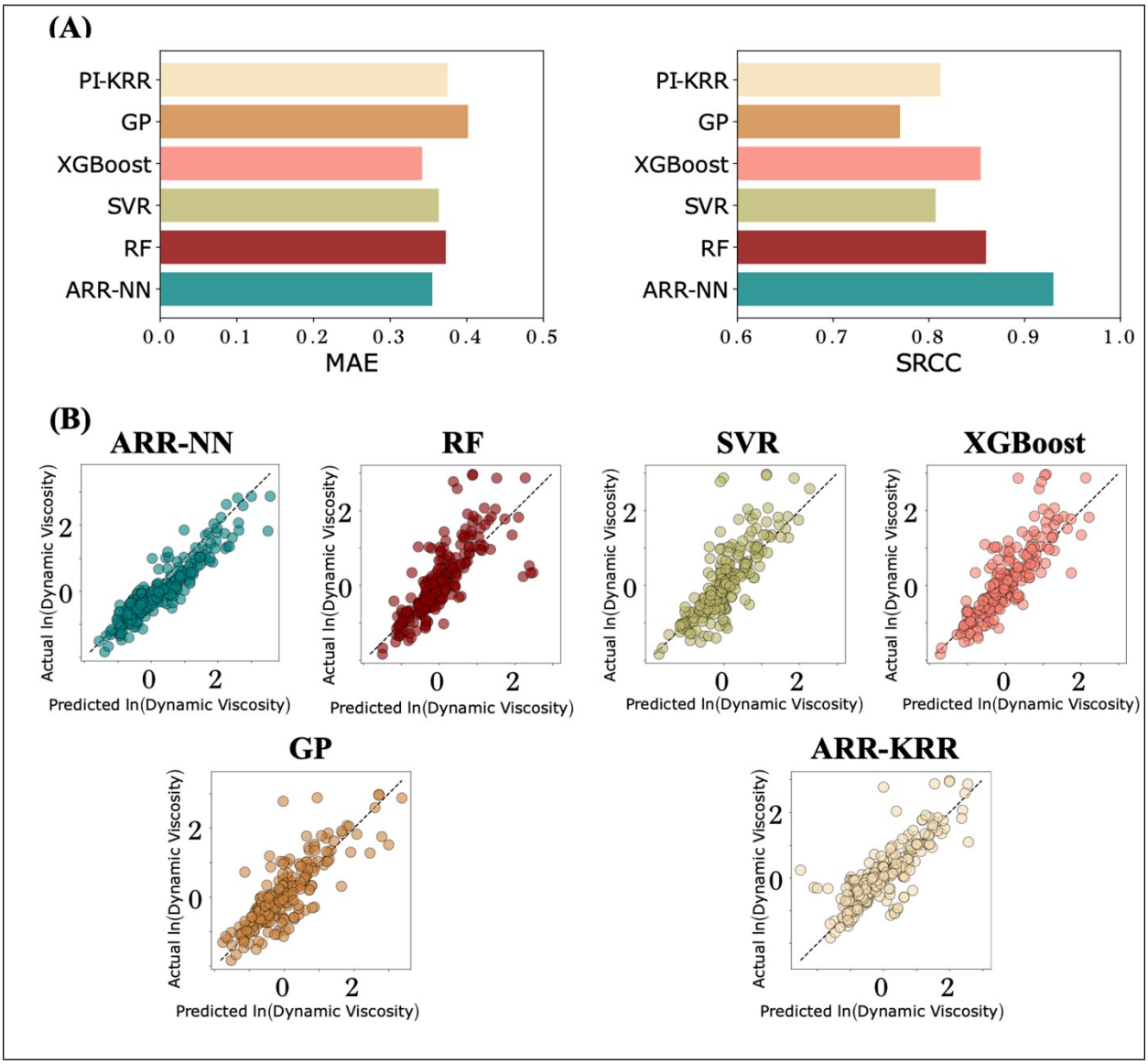


Fig S12. Benchmarking established ML models. (A) Accuracy metrics (MAE and SRCC) evaluated on the highest temperature range (T_5), which is excluded from training and remained completely unseen to the models. Arrhenius based neural network and XGBoost models show similar performances. **(B)** Parity plots of predicted vs. experimental dynamic viscosity.

13. Feature Importance Analysis

From a pool of ~1,600 descriptors, a final set of 30 minimizes validation loss; this set is obtained via XGBoost-based feature selection. We split descriptors by the kind of physical signal they primarily encode. Structural–topological descriptors are those that scale with molecular size or capture the scaffold's internal "geometry" (how big, long, or branched the graph is) often via property-free topology (information content, complexity, path/edge counts) or size-type weights such as van der Waals volume. These features show "how much molecule" and "how it's arranged," so they reflect size and framework texture rather than how the molecule interacts with adjacent species. Polarity–interaction descriptors, in contrast, encode where electrons and charges reside (partial charges, electronegativity, ionization potential), how easily the cloud distorts (polarizability), and the presence of interaction motifs or surfaces (H-bond donors/acceptors, charge- or logP-partitioned surface areas). These features show "how the surface behaves," i.e., the propensity for electrostatic, hydrogen-bonding, hydrophobic, dispersion, or induction interactions. In short, if a descriptor mainly reports size/extent or near-neighbor structural texture, we place it in structural–topological; if it maps electronic character, functional groups, or interaction-partitioned surface, we place it in polarity–interaction. Prior to ML-based feature selection, we had removed highly correlated features (pairwise correlation $|r| < 0.9$), so membership in the same class does not imply collinearity; rather, it reflects conceptual affinity (descriptors that speak to similar aspects of a molecule from different perspectives). The best set of features and their associated classes are summarized in Table S4.

Table S4. List of Mordred descriptors retained after XGBoost feature selection, grouped into Structural–topological and Polarity–interaction classes.

Structural–topological Features	
Feature	**Description**
ATSC0c	Centered autocorrelation (lag 0) with charge weight; size/composition summary.
ATSC0d	Centered autocorrelation (lag 0) with dipole-like weight; scaffold composition snapshot.
ATSC0s	Centered autocorrelation (lag 0) with E-state weight; substituent imprint of the scaffold.
ATSC1v	Autocorrelation (lag 1) with vdW volume weight; near-neighbor size/extent pattern.
ETA_shape_p	Global shape/packing descriptor; summarizes occupied molecular volume.
MATS1dv	Moran autocorrelation (lag 1) with valence-related weight; local size/valence texture.
MATS1i	Moran autocorrelation (lag 1) with ionization-potential weight; pattern regularity along bonds.
MATS1se	Moran autocorrelation (lag 1) with Sanderson electronegativity; short-range polarity texture of scaffold.
MATS1v	Moran autocorrelation (lag 1) with vdW volume; local size/packing correlation across bonds.
RotRatio	Flexibility ratio (rotors vs size); indirect proxy for chain length/branching.
AATS2d	Averaged autocorrelation (lag 2) with dipole-related weight; polarized patterns driving electrostatics.
AATS4se	Averaged autocorrelation (lag 4) with Sanderson electronegativity; longer-range polarity layout.
Polarity–interaction Features	
AMID_O	Amide carbonyl oxygen flag; strong H-bond acceptor hotspot.
Axp-4d	GETAWAY-type, polarizability-weighted (distance 4); inducible dipoles and dispersion.

BertzCT	Graph complexity; diversifies contact surface and multipoint complementarity.
CIC0	Complementary information content; topological diversity enhancing interaction opportunities.
GATS2i	Geary autocorrelation (lag 2) with ionization potential; donor/acceptor behavior patterning.
GGI4	Galvez topological charge index; charge transfer/distribution capacity.
IC2	Information content (order 2); branching/asymmetry increasing potential surface complementarity.
JGI7	Mean topological charge index (order 7); long-range charge distribution effects.
JGI8	Mean topological charge index (order 8); long-range charge distribution effects.
MDEC-22	Distance/edge-count index; increases encounter opportunities and dispersion contact.
MID_O	Oxygen environment flag; polar sites enabling H-bonding.
PEOE_VSA_1	Charge-binned van der Waals surface area; maps charged surface patches.
PEOE_VSA_10	Charge-binned van der Waals surface area; maps charged surface patches.
SIC2	Structural information content (order 2); micro-surface diversity for complementary interactions.
SlogP_VSA3	Surface area in logP bin; hydrophobic contact propensity.
nHBDon	Count of hydrogen-bond donors; explicit interaction capacity.
piPC8	Path-count index (length 8); extends shape reach and nonspecific dispersion contacts.

Fig. S13 presents SHAP (SHapley Additive exPlanations) plots, which quantify the contribution of individual molecular descriptors to the prediction of target properties. By ranking descriptors according to their SHAP values, the analysis provides a clear, quantitative visualization of feature importance. As a supplement to Figure 5 in the main text, this figure adds numerical detail to the comparative analysis of model behavior and descriptor influence.

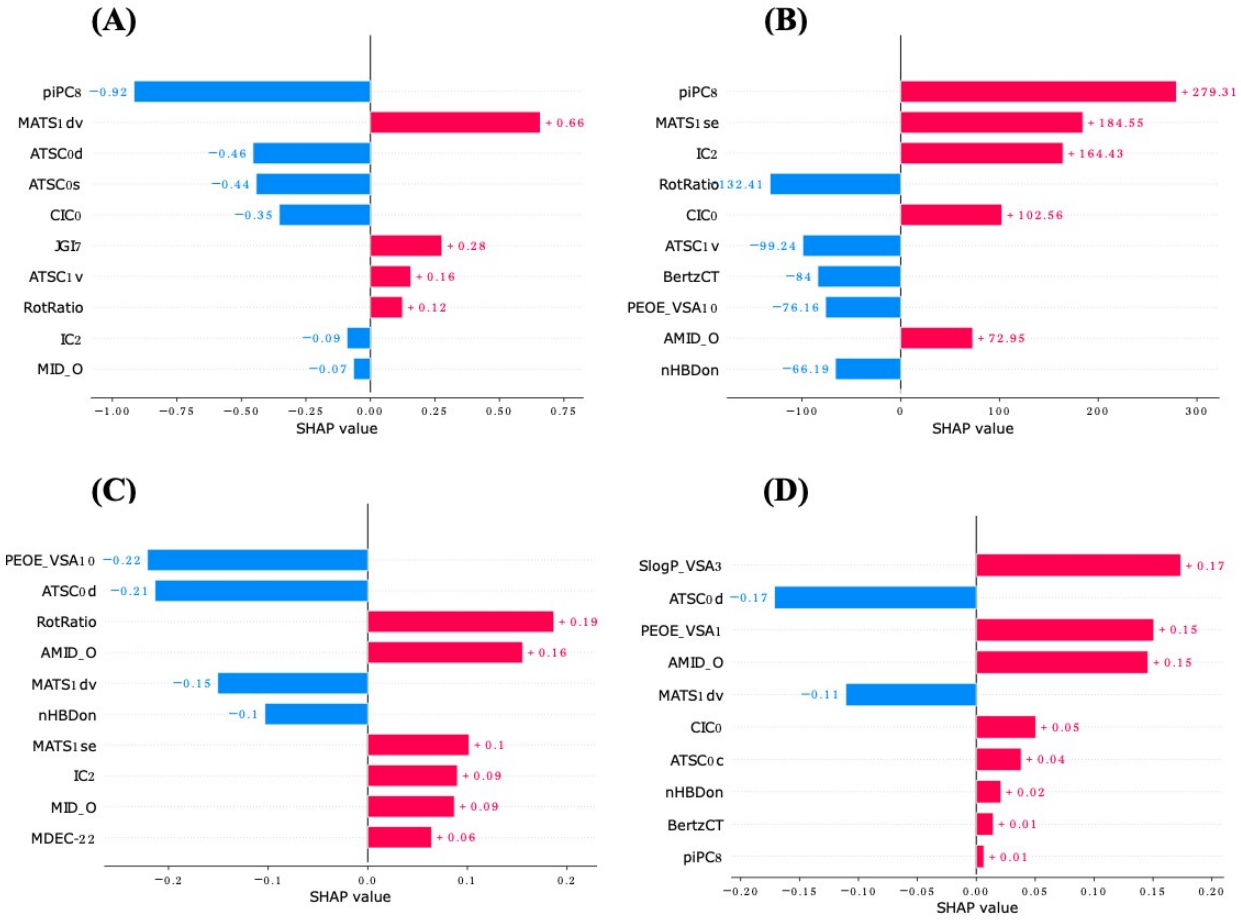


Fig. S13. Understanding the perceived feature importance in property prediction. SHAP values of top 10 descriptors with highest importance in predicting coefficients of the Arrhenius equation: **(A)** the frequency factor and **(B)** activation energy, and dynamic viscosity using: **(C)** the physics-informed model **(D)** the baseline model.

14. Connection Between Device Performance and Temperature Dependence of Dynamic Viscosity

We leverage the interpretability of this machine learning model to understand how chemical structure affects fluid performance. According to the Arrhenius equation, fluids with higher activation energy are more sensitive to temperature changes. In Fig. S14, we show a strong correlation between activation energy and system performance, specifically in terms of pressure drop. Assuming constant viscosity leads to systematic overestimation of pump work, with pressure-drop errors reaching up to 160 Pa (Fig. S14). This effect is most pronounced in fluids with high activation energy (Ea > 2,000 J.mol^{-1}) that typically exhibit strong intermolecular interactions, such as alcohols and carboxylic acids.

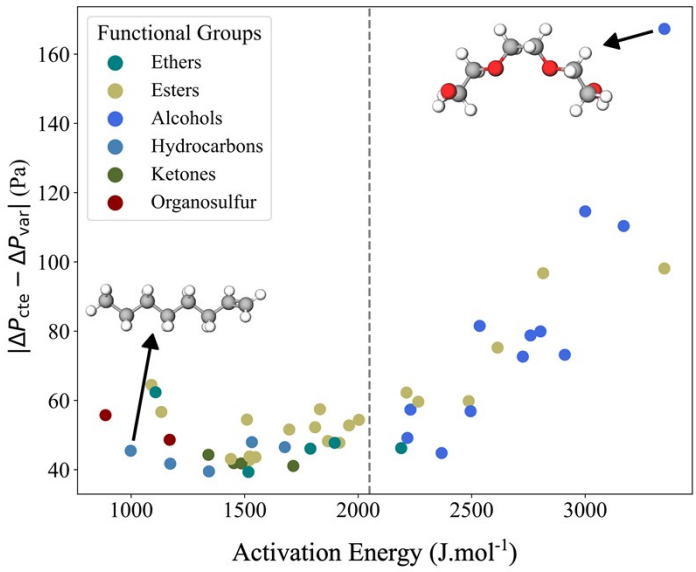


Fig. S14. Relationship between activation coefficient and fluid pressure drop. The difference in fluid pressure drop across the bank of tubes when switching from simplified constant dynamic viscosity to real varying dynamic viscosity, plotted against activation energy, which indicates the sensitivity of each fluid's dynamic viscosity to temperature.

15. Evaluation of Predictive Uncertainty with Calibrated Metrics

To move beyond accuracy measures and assess model reliability, we evaluate the predictive distributions using a set of calibrated uncertainty metrics. Specifically, we consider coverage probability, sharpness, the Gaussian negative log-likelihood (NLL), and the continuous ranked probability score (CRPS). These complementary metrics quantify not only the accuracy of the predictive mean but also the calibration and informativeness of the associated uncertainty estimates, providing a comprehensive view of probabilistic performance that is not captured by point-error measures alone.

Across all five temperature conditions, the physics-informed model demonstrates superior performance according to all four criteria. Its empirical coverage aligns more closely with nominal levels, its predictive intervals are sharper while remaining calibrated, and both NLL and CRPS are uniformly lower relative to the baselines. The multitemperature base model achieves reasonable calibration but at the cost of wider intervals, whereas the isothermal base model consistently undercovers and performs worse on likelihood- and distribution-based scoring rules. Collectively, these results show that physics-informed constraints improve the balance between calibration and sharpness, yielding more reliable and efficient uncertainty quantification than purely data-driven baselines.

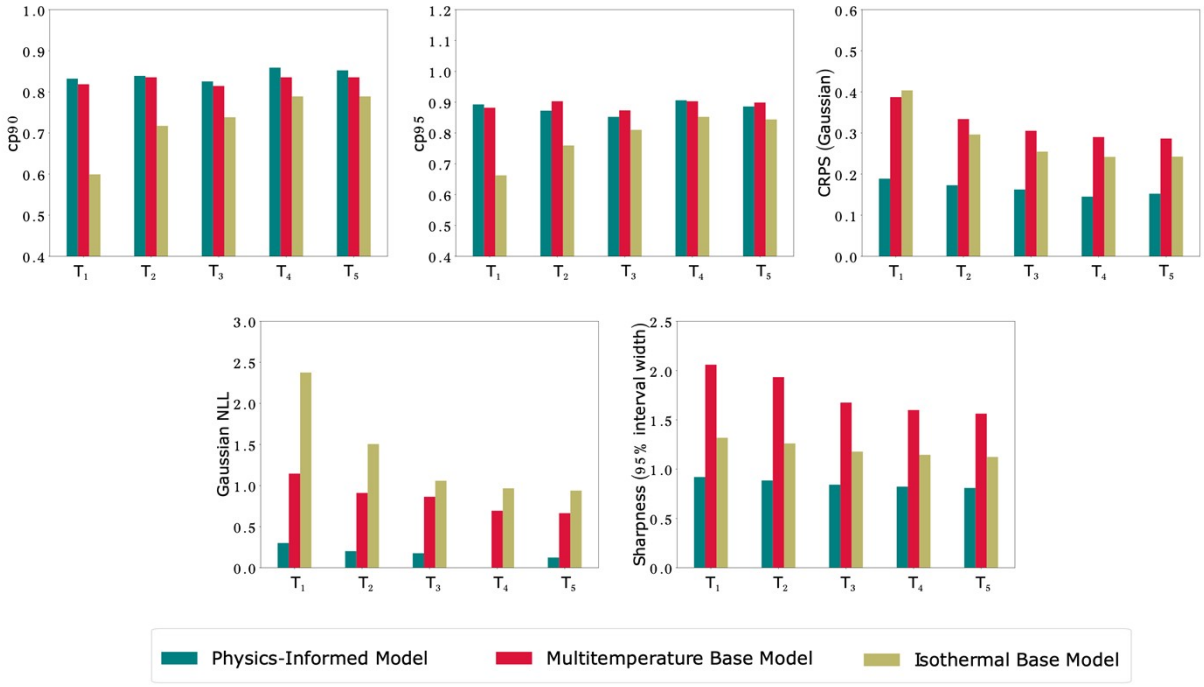


Fig. S15. Calibrated metrics evaluation across five temperature conditions (T_1–T_5). The physics-informed model is compared against the multitemperature base model and the isothermal base model using coverage probability, sharpness, Gaussian negative log-likelihood (NLL), and continuous ranked probability score (CRPS). The physics-informed model achieves sharper and better-calibrated predictive intervals with consistently lower probabilistic scores, while also assigning high uncertainty to erroneous predictions.

16. Physics-Informed Machine Learning for Binary Mixtures

Mixtures are widely used in thermal management applications, where their thermophysical properties, like pure fluids, play a critical role in determining overall performance. Unlike pure compounds, the number of possible mixtures increases exponentially with the number of components, making systematic characterization both experimentally demanding and computationally nuanced. In this context, solving the temperature dependence of mixture properties becomes especially important, as accurate predictive models must generalize across diverse compositions and operating conditions. This motivates the extension of physics-informed machine learning models, which have shown promise for pure compounds, into the domain of mixtures.

We use dynamic viscosity data of binary mixtures published by Bilodeau et. al.[13], which cover 1,628 pure compounds and 34,374 binary mixtures. We generate physicochemical descriptors from SMILES strings of the pure compounds using RDKit python package and standardize the features after removing those with missing values or zero variance. To represent mixtures, we concatenate the embeddings of the two constituent compounds and incorporate their mole fractions as additional input features.

The dataset contains 7,313 unique binary mixtures. We first assess whether the Arrhenius equation can describe the viscosity–temperature relationship across this dataset. We find that the equation holds for nearly all mixtures, with 99.8% of the cases well represented by the Arrhenius form. Building on this, temperature enters the models under two scenarios: in the physics-informed approach, we exclude temperature from the input and instead capture its effect through an Arrhenius-inspired neural network readout layer; in the baseline approach, we include temperature directly as a feature, allowing the model to learn the temperature dependence from data alone.

The results in Table S5 demonstrate that the physics-informed approach works better than the baseline, and, importantly, that it is not limited to pure components. As long as the Arrhenius equation remains valid, the framework naturally extends to more complex cases such as mixtures. In this setting, the neural network learns the chemical contributions of each component from molecular descriptors, while the Arrhenius form couples these learned embeddings to the thermodynamic temperature dependence. This

separation of chemistry and thermodynamics provides a principled way to generalize predictive models beyond pure compounds and into the exponentially larger space of mixtures.

Table S5. Comparison of the physics-informed ML and the baseline model for predicting binary mixture dynamic viscosities. The physics-informed model achieves higher SRCC and R^2 while reducing prediction error, highlighting the benefit of embedding the Arrhenius equation into the prediction framework of mixtures.

Model	SRCC	MAE	MSE	R^2
Physics-Informed ML Model	0.98	0.9	0.07	0.88
Baseline	0.97	0.1	0.13	0.76

17. Performance Evaluation of Models Across 5 Temperature Levels

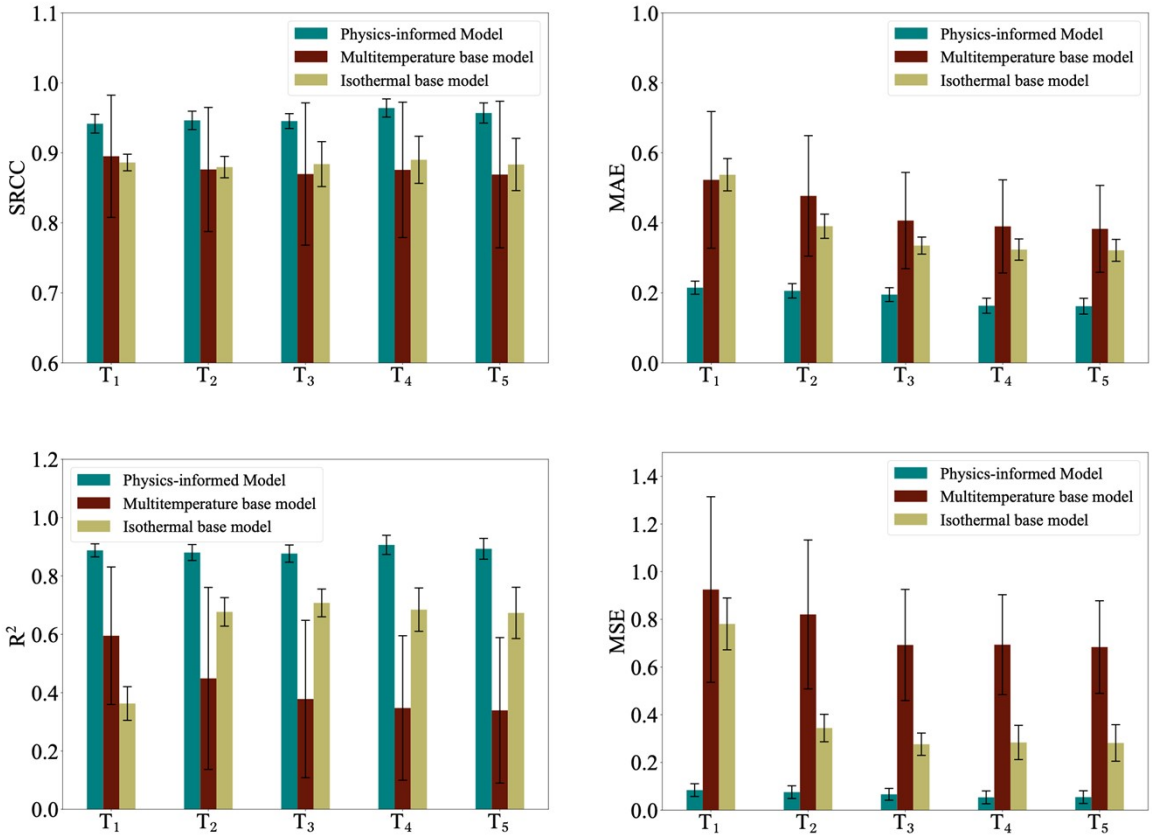


Fig. S16. Performance comparison of physics-informed, multitemperature, and isothermal base models across five temperatures (T_1–T_5). Error bars indicate the standard deviation of predictions computed from an ensemble of 20 models trained on different subset of training datasets.

18. Correlation of Arrhenius Parameters with Molecular Polarity

The results presented in Fig. 5 demonstrate that the physics-informed ML model does not merely memorize training data but instead learns chemically meaningful relationships that govern fluid behavior. By integrating structural–topological and polarity–interaction descriptors, the model captures both the baseline viscosity and its temperature dependence through the learning of activation energy and pre-exponential factors which are the two fundamental components of the Arrhenius equation. These components are distinct in nature and depend on different molecular characteristics: polarity- interaction descriptors play a more significant role in predicting the activation energy, whereas structural-topological features have greater influence on the pre-exponential factor. This distinction reflects the different physical origins of these two parameters. As a complementary analysis, the connection between the Arrhenius coefficients and a quantitative indicator of molecular polarity is further examined to validate this interpretation.

To test this correlation, a subset of compounds common to both the viscosity dataset collected in this study and the QM9 quantum chemistry database[14,15] is identified. A total of 138 compounds are common to both datasets, identified through matching SMILES strings. Polarizability is selected as the quantitative measure of molecular polarity, as it reflects the ease with which a molecule's electron cloud can be distorted by external electric fields or neighboring dipoles. Higher polarizability corresponds to greater molecular deformability and stronger induced dipole interactions, both of which are closely related to polarity and its influence on viscous behavior.

For each compound, the pre-exponential factor (A) and activations energy (E_a) values are correlated with molecular polarizability using the Spearman rank correlation coefficient (ρ). This nonparametric approach captures monotonic relationships without assuming linearity and statistical significance is assessed at a 95% confidence level ($\alpha = 0.05$). The results at Fig S17 show a moderate and statistically significant positive correlation between the activation energy and molecular polarizability ($\rho = 0.45$, p-value = 8.1×10^{-8}). In contrast, the pre-exponential factor exhibits a weak and statistically insignificant correlation ($\rho = -0.19$, p-value = 0.03). These findings indicate that molecules

with higher polarity, as reflected by greater polarizability, tend to exhibit higher activation energies. This trend supports the physical interpretation that stronger dipolar and induced-dipole interactions increase the energy barrier for molecular motion. In contrast, the pre-exponential factor appears to be primarily governed by entropic and geometric–topological characteristics of the molecules and does not show a clear correlation with molecular polarity. The present analysis provides an initial insight into this relationship; however, a more comprehensive investigation would require a larger dataset containing experimentally measured polarity and topological descriptors to confirm these correlations quantitatively.

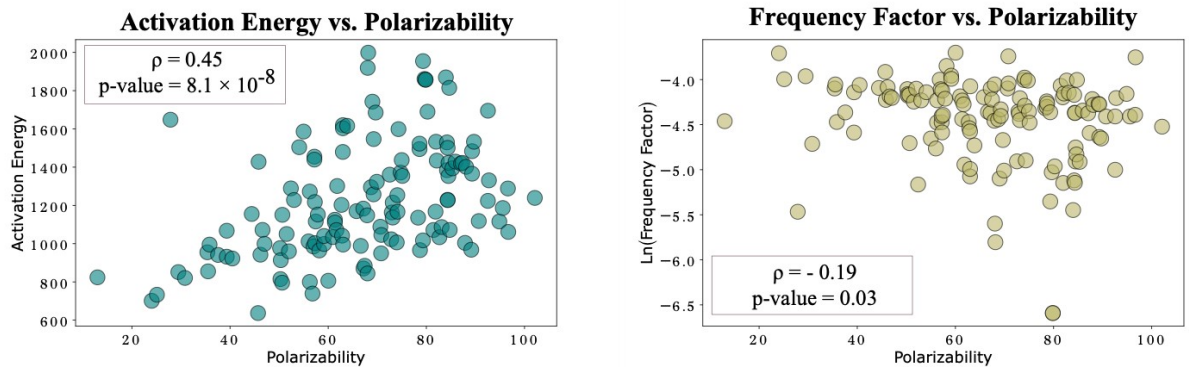


Fig. S17. **Correlation between the predicted Arrhenius parameters and molecular polarizability for 138 compounds common to the viscosity and QM9 datasets. (Left)** Activation energy shows a moderate positive correlation with polarizability (ρ=0.45, p-value = 8.1 × 10^{-8}), indicating that molecules with higher polarity tend to exhibit larger energy barriers for viscous flow. **(Right)** The pre-exponential factor displays a weak and negative correlation with polarizability (ρ=−0.19, p-value = 0.03), suggesting that it is primarily influenced by entropic and structural-topological effects rather than molecular polarity.

19. Cross-checking ML predictions with NIST and experimental data

For thermal fluids at T1 (288 K), we cross-validated the physics-informed ML model predictions against two independent references (experimental measurements and NIST data) using parity plots and standard accuracy metrics. As shown in the Figure S18, ML predictions exhibit excellent agreement with experimental measurements, achieving a high rank correlation (SRCC = 0.98) and low error (MAE = 0.117), indicating both strong accuracy and correct trend capture. The agreement between NIST and experimental data is also good (SRCC = 0.92, MAE = 0.166), though slightly weaker than the ML–experimental comparison. In contrast, the ML–NIST comparison shows increased scatter (SRCC = 0.90, MAE = 0.198), suggesting that discrepancies between ML and NIST values are primarily driven by limitations or inconsistencies in reference data rather than ML model error. Overall, these results demonstrate that the physics-informed ML model reliably reproduces experimental viscosity values and, in cases of disagreement, experimental data support the ML predictions over tabulated NIST values.

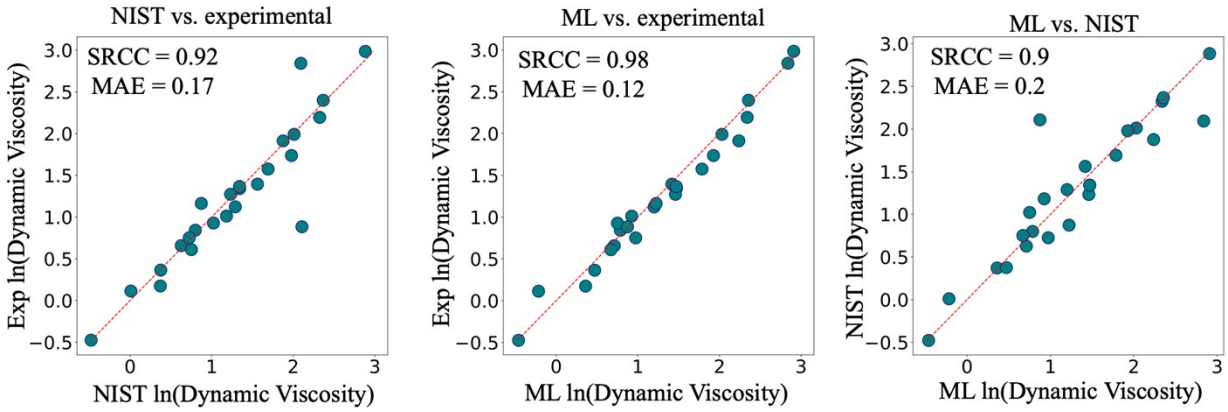


Figure S18. ML prediction cross checking with 2 different references. Parity plots comparing physics-informed ML predictions of ln(dynamic viscosity) with experimental measurements and NIST reference data for thermal fluids at T1 (288 K).

20. Leave-one-class-out chemical extrapolation

To assess the model's ability to extrapolate to unseen chemical classes, we performed explicit leave-one-class-out (LOCO) evaluations. We previously evaluated model performance by stratifying predictions after a random 70/30 split (Table 1). In the LOCO setting, however, all compounds belonging to a given chemical class are completely excluded from training and used solely for testing. The corresponding results are shown in Table S6. As expected, model performance decreases relative to random splits, reflecting the difficulty of predicting viscosity for chemistries absent from the training data. The magnitude of this performance drop is class-dependent. Hydrocarbon subclasses (alicyclic, aliphatic, and aromatic) and high–molecular-weight compounds (MW > 250 $g \cdot mol^{-1}$) retain relatively strong rank correlation (SRCC ≈ 0.9), indicating that the model captures transferable structure–property relationships within these regimes. In contrast, heteroatom-rich classes—particularly nitrogenous and halogenated compounds—exhibit a more pronounced reduction in performance. These results confirm that while the model generalizes well within chemically related domains, extrapolation to entirely unseen functional chemistries remains challenging. Overall, the LOCO analysis provides a realistic lower bound on model performance and clearly delineates the current limits of chemical generalizability.

Table S6. Leave-one-class-out evaluation of model performance across chemistry-based classes. For each case, all compounds belonging to the specified class were excluded from training and used solely for testing. Performance is reported using SRCC and MAE, providing a stringent assessment of chemical extrapolation.

Chemical Criteria		**SRCC**	**MAE**
Hydrocarbon	Alicyclic	0.9	0.5
	Aliphatic	0.88	0.42
	Aromatic	0.93	0.65
Heteroatom	Oxygenated	0.64	0.65
	Nitrogenous	0.53	0.71
	Sulfuric	0.6	0.32
	Halogenated	0.47	0.59
	Hydrocarbon	0.89	0.87
	Multi-heteroatom	0.69	0.58
Molecular weight ($g.mol^{-1}$)	MW < 150	0.62	0.54
	150 < MW < 250	0.78	0.41
	MW > 250	0.91	0.29

21. GNN-based representation for viscosity prediction

We evaluated graph based molecular representations for dynamic viscosity prediction across all five temperature ranges (T1–T5). As summarized in Table S7, the GNN achieves consistently good performance, with high accuracy across temperatures (SRCC = 0.94–0.96) and a monotonic reduction in error as temperature increases (MAE decreasing from 0.28 at T1 to 0.16 at T5). Overall performance (SRCC = 0.95, MAE = 0.22) is comparable to that obtained using physics-informed models built on Mordred descriptors (Figure 4), with no statistically significant improvement observed across temperature regime. This similarity indicates that, for viscosity prediction within the chemical space studied here, both representations capture largely overlapping structure–property information.

Table S7. Performance of the GNN-based and Mordred-based models for dynamic viscosity prediction across temperature ranges T1–T5. Reported SRCC and MAE indicate consistently strong performance across all temperatures, with overall accuracy comparable to models using Mordred descriptors (see Figure 4).

Graph based representation						
Metric	**T1**	**T2**	**T3**	**T4**	**T5**	**Overall**
SRCC	0.94	0.95	0.95	0.96	0.96	0.95
MAE	0.28	0.24	0.22	0.19	0.16	0.22
Mordred numerical representation						
SRCC	0.93	0.93	0.94	0.95	0.96	0.94
MAE	0.26	0.24	0.22	0.19	0.18	0.22

22. Arrhenius baseline for unseen-temperature prediction

We fit Arrhenius equation as baseline independently for each material using viscosity measurements at four temperatures (T_2–T_5). Specifically, the logarithm of viscosity is regressed against inverse temperature, following the standard Arrhenius equation. The fitted model is then used to extrapolate viscosity at the held-out temperature T_1, which is not included during fitting. Predicted viscosities are compared against experimental values at T_1 across all materials, and performance is quantified using MAE and SRCC.

The Arrhenius extrapolation demonstrates excellent agreement with experimental viscosities at the held-out temperature T_1, yielding a SRCC of 0.99 and a MAE of 0.09 (Fig. S19). The parity plot shows that predictions closely follow the ideal parity line across several orders of magnitude in viscosity, indicating that the Arrhenius equation provides a highly effective description of temperature dependence when multiple measurements for the same material are available.

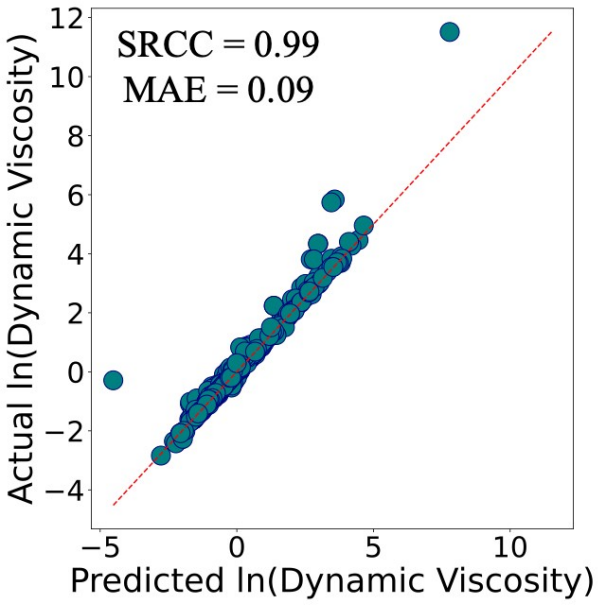


Figure S19. Arrhenius equation fitting as a baseline for unseen-temperature viscosity prediction.

23. Calibration performance under leave-one-class-out evaluation

Figure S20 shows the CP90 calibration metric across temperature ranges (T1–T5) for train/test splitting based on leave-one-class-out strategy, in which all compounds belonging to a given chemistry class are entirely excluded from training and used only for testing. Despite this stringent out-of-distribution setting, the model exhibits consistently strong calibration across all classes, with CP90 values generally improving at higher temperatures. Among hydrocarbon subclasses (alicyclic, aliphatic, aromatic), calibration performance is comparable, indicating that uncertainty estimates generalize well across distinct carbon frameworks. Leaving molecular weight classes reveals systematically higher CP90 values for heavier compounds (MW > 250 $g.mol^{-1}$), while lighter molecules show modestly reduced calibration, particularly at lower temperatures. When grouped by heteroatom chemistry, hydrocarbons and multi-heteroatom compounds show the most reliable calibration, whereas oxygenated, nitrogenous, and sulfur-containing species exhibit slightly lower CP90, most prominently at T1, reflecting increased predictive uncertainty for chemically and thermodynamically challenging regimes. Overall, we can make this claim that the proposed uncertainty model remains well-calibrated even when extrapolating to entirely unseen chemical classes, supporting its robustness for chemically out-of-distribution predictions.

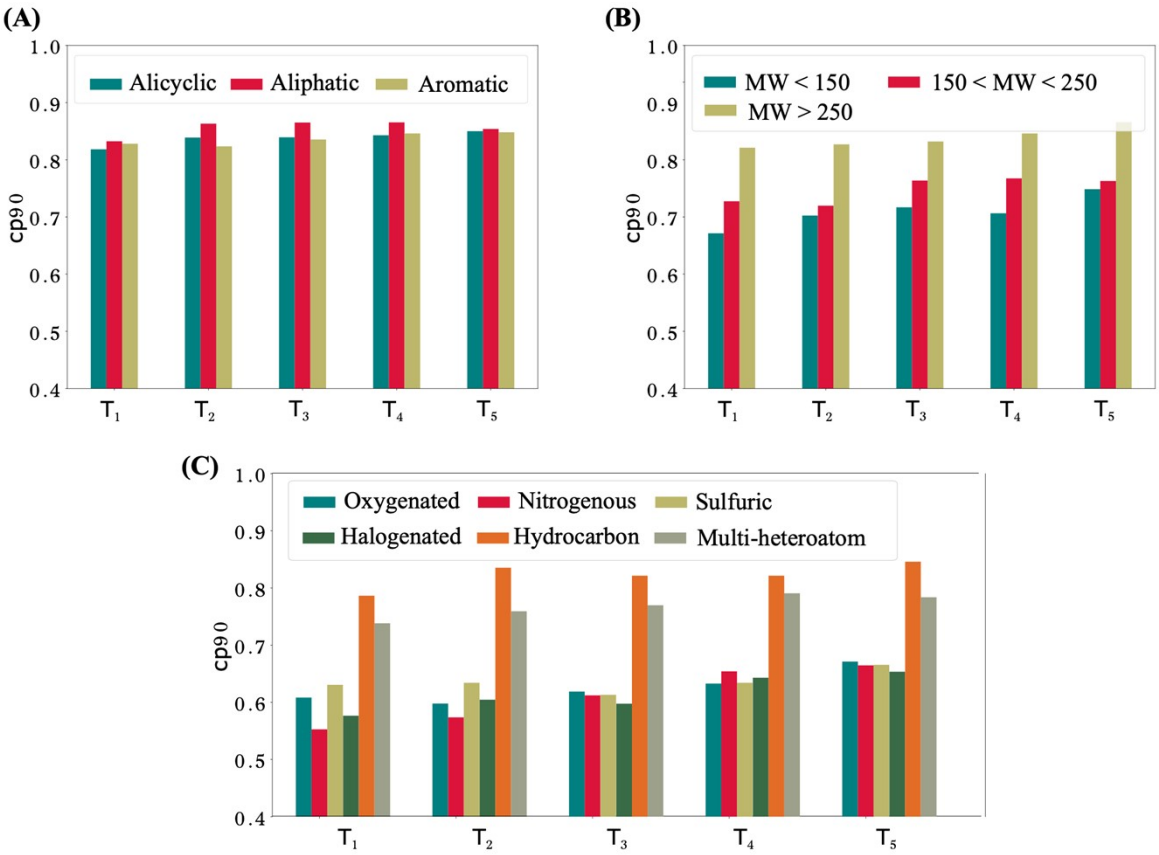


Figure S20. Calibration under leave-one-class-out evaluation. CP90 calibration across temperature ranges (T1–T5) for chemically stratified test sets generated using leave-one-class-out splitting for **(A)** hydrocarbon subclass, **(B)** molecular-weight and **(C)** heteroatom classes. Higher CP90 indicates improved uncertainty calibration.

24. Sensitivity analysis of the uncertainty threshold

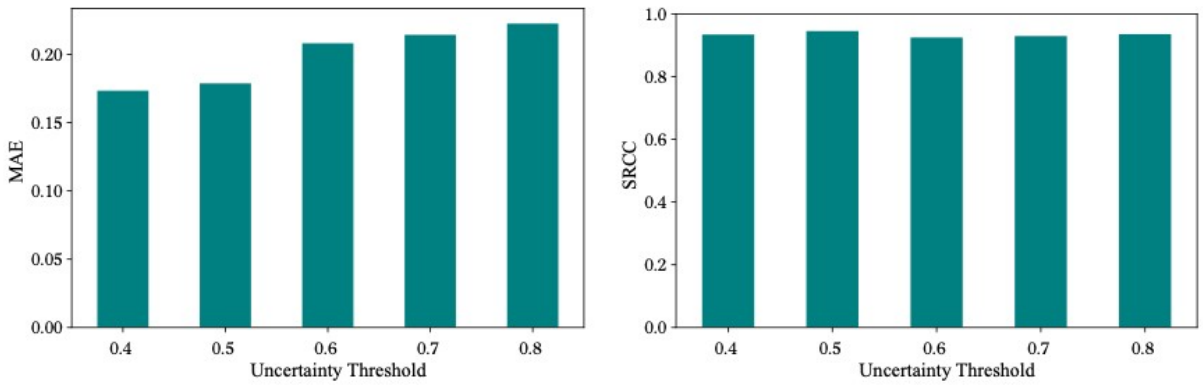


Figure S21. Sensitivity analysis of the uncertainty threshold. Prediction accuracy evaluated at five uncertainty levels ranging from 0.4 to 0.8 with increments of 0.1. The bars represent the average performance metrics across five temperature levels on the same test dataset.

References

1 L. C. Gilday, S. W. Robinson, T. A. Barendt, M. J. Langton, B. R. Mullaney and P. D. Beer, *Chem Rev*, 2015, **115**, 7118–7195.

2 G. Wang, Z. Chen, Z. Xu, J. Wang, Y. Yang, T. Cai, J. Shi and W. Zhu, *J Phys Chem B*, 2016, **120**, 610–620.

3 M. K. Tiwari and C. Chatterjee, *J Hydrol (Amst)*, 2010, **382**, 20–33.

4 S. Kumar, M. K. Tiwari, C. Chatterjee and A. Mishra, *Water resources management*, 2015, **29**, 4863–4883.

5 S. Al-Dahidi, P. Baraldi, E. Zio and L. Montelatici, *Sustainability*, 2021, **13**, 6417.

6 P. P. Bonissone, F. Xue and R. Subbu, *Appl Soft Comput*, 2011, **11**, 1529–1539.

7 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang and E. E. Bolton, *Nucleic Acids Res*, 2025, **53**, D1516–D1525.

8 C. D. M. Vladimir Diky Alexander Y. Smolyanitsky Ala Bazyleva Robert D. Chirico Joe W. Magee Yauheni Paulechka Andrei F. Kazakov Scott A. Townsend Eric W. Lemmon Michael D. Frenkel Kenneth G. Kroenlein, *NIST/Thermodynamics Research Center (TRC)*, preprint, https://app.knovel.com/hotlink/toc/id:kpLTI00007/nist-standard-reference/nist-standard-reference.

9 P. J. Linstrom and W. G. Mallard, *J Chem Eng Data*, 2001, **46**, 1059–1063.

10 P. Panwar, Q. Yang and A. Martini, *J Chem Inf Model*.

11 S. by Aic. Design Institute for Physical Properties, *Design Institute for Physical Property Research/AIChE*, preprint, https://app.knovel.com/hotlink/toc/id:kpDIPPRPF7/dippr-project-801-full/dippr-project-801-full.

12 K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley and M. Mathea, *J Chem Inf Model*, 2019, **59**, 3370–3388.

13 C. Bilodeau, A. Kazakov, S. Mukhopadhyay, J. Emerson, T. Kalantar, C. Muzny and K. Jensen, *Chemical Engineering Journal*, 2023, **464**, 142454.

14 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *Sci Data*, 2014, **1**, 1–7.

15 L. Ruddigkeit, R. Van Deursen, L. C. Blum and J.-L. Reymond, *J Chem Inf Model*, 2012, **52**, 2864–2875.