Energy Proceedings
Vol 64, 2026
ISSN 2004-2965

Machine Learning-driven Melting Point Prediction for Heat Recovery Materials[#]

Lizhong Yang[1†], Xinyu You [2†], Fei Xiao [1], Long Gao [3], Alessandro Romagnoli [4*]

1 College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

2 School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore, Singapore

3 School of Energy and Power Engineering, Northeast Electric Power University, Jilin 132012, China

4 School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore

(*Corresponding Author: a.romagnoli@ntu.edu.sg)

ABSTRACT

Facing the challenge of growing heating and cooling demand, more heat recovery materials used to capture and transfer heat should be developed. Melting point prediction is one of the most essential steps of machine-learning-driven material discovery. Existing methods rely on large datasets, often with the help of density function theory and molecular dynamics simulations, or complex deep learning methods. Here, we developed a multilayer perceptron model for predicting the melting point of the materials using a small but reliable dataset. By comparing with classical machine learning models, including ridge regression, random forest, XGBoost, and support vector machines, the multilayer perceptron model with transfer learning from a much larger dataset not dedicated to heat recovery materials demonstrated a flexible compromise balancing robustness and data size, providing a promising yet simple solution for melting point prediction.

Keywords: heat recovery, melting point, machine learning, multilayer perceptron

NONMENCLATURE

Abbreviations	
AI	Artificial intelligence
DFT	Density functional theory
FP	Fingerprint
HTF	Heat transfer fluid
LOOCV	Leave-one-out cross-validation
MAE	Mean absolute error
MD	Molecular dynamics
ML	Machine learning
PCM	Phase change material
RF	Random forest
SMILES	Simplified Molecular Input Line Entry System
SVR	Support vector regression
XGBoost	Extreme gradient boosting
Symbols	
R^2	Coefficient of determination

1. INTRODUCTION

Heat recovery materials, such as heat transfer fluids (HTFs), containment materials, sensible storage materials, and phase change materials (PCMs), play an important role in energy systems converting or supplying heat and cold, which are the largest form of end-use energy [1–3]. Artificial intelligence (AI) approaches can significantly accelerate the process of identifying new materials with desirable properties, reducing reliance on resource-intensive experimental methods and enabling the discovery of materials that may be overlooked by traditional approaches [4]. In the realm of heat recovery materials, machine learning and deep learning predictions have been used for composite materials with improved thermal conductivity [5], enthalpy of fusion of pure compounds [6], melting points for CHNO compounds [7] and drug development [8].

Reliable algorithms to predict the melting points of heat recovery materials are the cornerstone of data-driven material discovery for next-generation high-energy-density heat recovery systems. However, accurate and high-quality datasets are often scarce due to costly and time-consuming experimental

This is a paper for the 17th International Conference on Applied Energy (ICAE2025), December 8-12, 2025, Bangkok, Thailand.
† Contributed equally to this work.

measurements [5]. Density functional theory (DFT) and molecular dynamics (MD) simulations can help increase the scale of the dataset [9], though suffering from concerns about extensive computing power and a lack of high-quality simulation results. At the same time, state-of-the-art deep learning models typically involve huge parameters, which increases the risk of overfitting when applied to limited data.

In this study, we explore both classical machine learning models (ridge regression, random forest, XGBoost, support vector machines) and multilayer perceptron (MLP) models, including a transfer learning strategy where an MLP is pretrained on a larger dataset and subsequently fine-tuned on a smaller, high-quality dataset. The objective is to obtain a reliable and generalizable model for melting point prediction while mitigating overfitting and improving predictive accuracy.

2. METHODS

2.1 Data Sources and Curation

Two datasets were employed in this study: a moderate-scale dataset from Yaws' Handbook of Thermodynamic and Physical Properties [10] (hereafter referred to as the "Yaws dataset") and a smaller, high-quality dataset from the Thermal Energy System's Laboratory of Nanyang Technological University, developed by us in a previous work [3] (hereafter referred to as the "PCM dataset", highlighting the inclusion of phase change temperatures, but its usage is far beyond just PCMs). The Yaws dataset consists of around 5000 substances and their melting points, covering a wide range of various types of materials. The PCM dataset offers more reliable results thanks to rounds of cross-validation from various sources.

These datasets differ substantially in both size and scope, and were subjected to careful preprocessing before model development. Comparison of melting point distributions for the PCM dataset (blue) and the Yaws dataset (pink) is illustrated in Fig. 1. The Yaws dataset covers a wide temperature range from approximately -250 °C to over 400 °C (mean = -22.2 °C, standard deviation = 81.7 °C). Many compounds cluster near room temperature ranges. In contrast, the PCM dataset falls between –270 °C and 50 °C (mean = –103.2 °C, standard deviation = 80.3 °C), with most materials' melting points below –50 °C, aligning with their intended use in heat recovery. This shows a clear distribution shift that the PCM dataset's compounds occupy a smaller and colder subrange of the broader Yaws dataset's distribution.

2.2 Feature Selection

Melting points are influenced by both elemental composition and structural organization. Elemental factors such as heteroatom types, hydrogen-bond donors and acceptors, and polarity largely determine intermolecular forces and crystal stability [11,12]. Large-scale cheminformatics analyses further show that heteroatoms and hydrogen bonding potential dominate feature importance rankings, while more subtle topological descriptors provide secondary refinements [11]. Effective melting point prediction, therefore, requires feature representations that capture both elemental identity and local structural motifs.

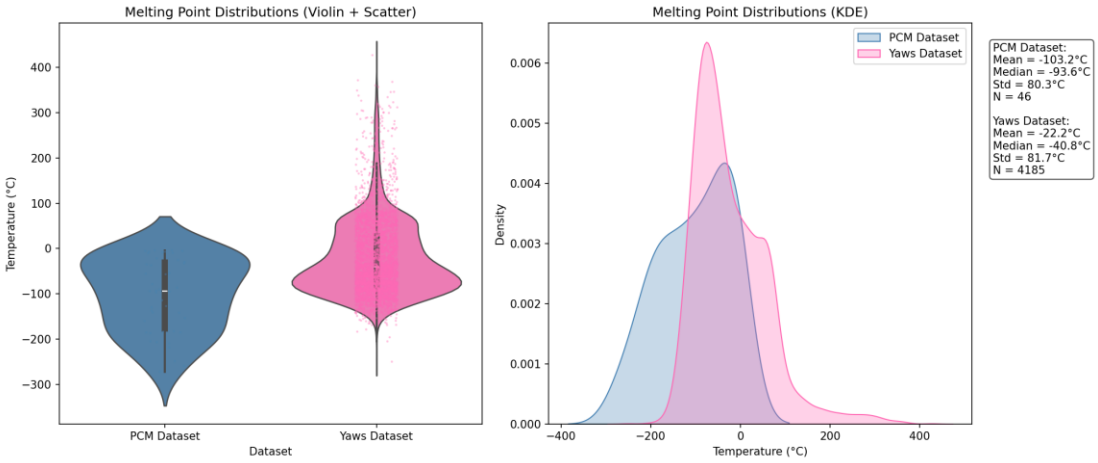

Fig. 1 Comparison of melting point distributions for the PCM datasets. (Left) The violin plots with individual data points illustrate the narrower range and lower median of the PCM dataset relative to the broader Yaws dataset. (Right) Kernel density estimates (KDE) highlight the distribution shift between the two datasets. Summary statistics (mean, median, standard deviation, and sample size) are shown for both datasets.

Material property predictions solely based on chemical formulas have been proven applicable and reliable [13]. In this study, we considered two complementary molecular representations, along with their combination, based on chemical formulas. First, Morgan fingerprints [14] were employed as 2048-bit binary vectors with radius 2, encoding the presence of key substructures such as rings, branching patterns, and functional groups. Second, we used SMILES embeddings from ChemBERTa [15,16], a transformer-based language model pretrained on large chemical corpora, which provides 384-dimensional dense embeddings that capture elemental composition, functional groups, and broader molecular context. Finally, to balance the strengths of both approaches, we constructed combined features by concatenating Morgan fingerprints with SMILES embeddings, integrating explicit structural information with contextualized representations learned from large-scale chemical language modeling.

2.3 Model Training and Evaluation

We implemented four classical machine learning models: ridge regression, random forest (RF), extreme gradient boosting (XGBoost), and support vector regression (SVR) as benchmarks for model performance comparison. Hyperparameters were tuned within reasonable ranges using grid search.

We constructed a feed-forward neural network using three types of input features: (i) Morgan fingerprints, (ii) SMILES embeddings from ChemBERTa, and (iii) their concatenation. The model consisted of feature encoding layers followed by fully connected layers with 512 hidden units, ReLU activation, and dropout regularization. Training was performed with the Adam optimizer with a learning rate of 3e-4, minimizing mean absolute error (MAE). For transfer learning, the feature encoding layers were frozen during fine-tuning, and only the regression head was updated using a reduced learning rate of 2e-4 to ensure stable convergence.

For the performance evaluation, we used 5-fold cross-validation and an 85/15 train–test split for the Yaws dataset. In contrast, the PCM dataset was too small for such partitioning, so leave-one-out cross-validation (LOOCV) was adopted. Model performance was reported in terms of MAE and the coefficient of determination (R^2). For neural network models, we further assessed robustness by averaging results over multiple random initializations and reporting the variance.

3. RESULTS AND DISCUSSION

3.1 Model performance

The neural network was evaluated on both the Yaws and PCM datasets and compared with classical machine learning methods (Table 1). The best results were obtained using combined fingerprint and SMILES embeddings, and all baseline comparisons were conducted under this feature setting.

Table 1 Performance of different models on the Yaws dataset (5-fold CV) and the PCM dataset (LOOCV).

Models	Yaws dataset (5-fold CV)		PCM dataset (LOOCV)	
	MAE (°C)	R^2	MAE (°C)	R^2
Ridge	58.44	-0.49	21.18	0.82
SVR	29.44	0.66	55.18	0.25
RF	18.11	0.87	37.94	0.57
XGB	18.43	0.88	33.36	0.61
MLP (non-pretrained)	27.57	0.70	38.72	0.49
MLP (pretrained+FT)	–	–	31.81	0.66

A transfer learning strategy was used to overcome the extremely limited size of the PCM dataset. The model was first pretrained on the larger Yaws dataset to capture general structure–property trends, and then fine-tuned to adapt to the narrower but more reliable PCM dataset domain.

Among the baseline models, ridge regression performed poorly on the large Yaws dataset, with an MAE of 58.4 °C and an R^2 below zero, confirming that linear models are too restrictive for chemically diverse compounds. On the PCM dataset, however, ridge regression is the best performing baseline, reaching an MAE of 21.2°C and an R^2 of 0.82. With only 46 samples available, the simplicity and strong regularization of linear models prevented overfitting and yielded stable predictions. This is probably due to the special composition of the PCM dataset, as the materials were selected deliberately for heat recovery.

Random forest and XGBoost performed better on the Yaws dataset, with MAEs around 18 °C and R^2

between 0.87 and 0.88, as their ensemble structures capture nonlinear interactions and benefit from large sample sizes. On the PCM dataset, their accuracy dropped substantially, with MAE between 33 and 38 °C and R^2 ranging from 0.57 to 0.61, and both are outperformed by the ridge regression model. The results indicate that, with so few training points, deep trees overfit, and ensemble averaging is insufficient to restore generalization.

SVR with RBF kernels achieved intermediate performance on Yaws, giving an MAE of 29.4°C and an R^2 of 0.66, comparable to the multilayer perceptron but worse than tree-based models. On the PCM dataset, SVR performs poorly, with an MAE of 55.2 °C and an R^2 of 0.25, reflecting its sensitivity to kernel scaling and regularization in extremely small training sets.

The MLP achieved MAE = 27.6 °C and R^2 of 0.70 on the Yaws dataset, better than the ridge regression but behind RF and XGB. On the PCM dataset, a non-pretrained model (training from scratch) performs weakly, with an MAE of 38.7 °C and an R^2 of 0.49. Transfer learning, however, improved accuracy, reducing the MAE to 31.8 °C and increasing R^2 to 0.66, demonstrating that pretraining on the broader Yaws dataset provided useful structural information for fine-tuning on the smaller dataset.

Overall, MLPs are not the strongest performers in either dataset, but they achieve competitive accuracy across both large and small datasets. Transfer learning enables the same framework to bridge the two domains. While ridge regression might have excelled in extremely low-data settings, MLPs offer a flexible compromise, balancing robustness and data size, providing a more reliable solution for melting point prediction.

3.2 Feature Contribution Analysis

The results of the feature combination experiment are shown in *Table 2* and *Fig. 2*. Using only SMILES embeddings (fingerprint (FP) fraction = 0.0) gives a moderate baseline performance (MAE = 38.2 °C, R^2 = 0.52), while relying solely on Morgan fingerprints (FP fraction = 1.0) performs even worse (MAE = 42.7 °C, R^2 = 0.49). The best accuracy is obtained when the model uses a mixture of features at FP fraction = 0.4, achieving MAE = 31.8 °C and R^2 = 0.66. Beyond this point, increasing the fingerprint fraction steadily degrades performance, with errors rising above 40 °C when fingerprints dominate. This pattern is consistent with the chemical determinants of melting point discussed earlier that multiple factors jointly affect the melting points.

*Table 2 Effect of fingerprint fraction on PCM LOOCV performance. The best case, observed at a fingerprint fraction of 0.4 with the lowest MAE, is highlighted in **bold**.*

FP fraction	LOOCV MAE (°C)	LOOCV R^2
0.0	38.18	0.521
0.1	35.36	0.588
0.2	39.07	0.511
0.3	42.76	0.499
0.4	**31.81**	**0.662**
0.5	36.99	0.555
0.6	34.68	0.619
0.7	41.41	0.523
0.8	42.00	0.499
0.9	40.63	0.489
1.0	42.72	0.490

Embeddings derived from ChemBERTa primarily capture elemental composition and functional group information, which prior studies identified as the dominant factors for melting point. In contrast, Morgan fingerprints emphasize local structures such as rings and branching. When embeddings are used alone, the model already learns much of the elemental signal, but adding a moderate proportion of fingerprint information improves predictions further by incorporating structural corrections. However, when fingerprints dominate, the model loses the broader compositional signal and accuracy declines.

In summary, these results confirm that elemental features captured by embeddings form the backbone of predictive performance, while structural fingerprints add complementary contributions.

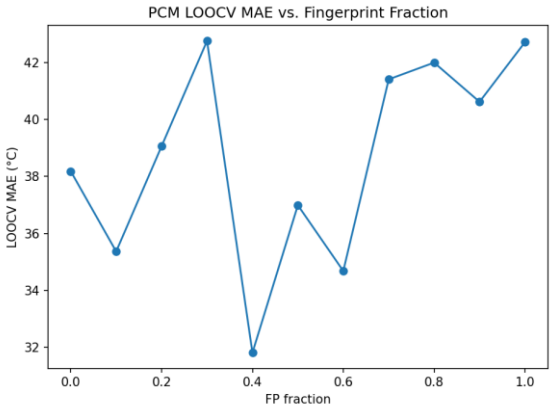

Fig. 2 Effect of feature composition on PCM LOOCV model performance. The best performance is observed when fingerprints contribute moderately (FP fraction = 0.4), whereas higher fingerprint fractions degrade accuracy

4. CONCLUSIONS

In this study, we systematically evaluated classical machine learning and deep learning approaches for predicting melting points—a critical yet challenging task in the field of heat recovery materials due to data scarcity and measurement variability.

Our findings reveal that while traditional ensemble models like random forest and XGBoost excel on large, diverse datasets such as the Yaws dataset, they suffer from overfitting when applied to small, specialized datasets like the PCM dataset, which contains only 46 high-quality entries. Ridge regression emerged as an effective baseline in low-data regimes, demonstrating the importance of model simplicity and regularization when data is extremely limited. In contrast, the MLP model, though not top-performing on either dataset individually, offers a balanced trade-off between flexibility and generalization. Crucially, we showed that transfer learning—pretraining an MLP on the larger Yaws dataset and fine-tuning on the smaller PCM dataset—significantly improves performance on the target domain, reducing MAE from 38.7 °C to 31.8 °C and increasing R^2 from 0.49 to 0.66. Furthermore, our feature ablation study highlights the complementary roles of molecular representations: ChemBERTa-derived SMILES embeddings capture essential elemental and functional group information that forms the foundation of predictive power, while Morgan fingerprints contribute valuable structural details. Optimal performance is achieved through a hybrid representation with a moderate contribution of fingerprint features (~40%), underscoring the value of combining contextualized embeddings with interpretable substructure descriptors.

This work demonstrates that thoughtful integration of transfer learning, hybrid feature engineering, and architecture selection can enhance predictive accuracy even in data-constrained scientific domains. These strategies pave the way for developing generalizable property predictors in materials informatics, particularly for niche applications such as HTSs and PCMs, where high-fidelity experimental and simulation data remain sparse. However, this study still faces challenges for real-world implementations, especially, the predicting error is not small enough. Some applications, for example, heat recovery and thermal energy storage systems working with air conditioners, require narrow temperature ranges for efficient operation. A few degrees of melting point prediction error may lead to significant changes in the energy consumption.

Therefore, further development of the models, including new algorithms, may be needed. Future efforts may also explore larger-scale pretraining on diverse chemical corpora and adaptation to multi-property prediction frameworks, moving closer to robust in silico screening tools for advanced heat recovery material design.

ACKNOWLEDGEMENT

The authors would like to thank the NGF-2024-16-024 grant, the Thermal Energy Systems Lab @ NTU, and the Aviation Green Propulsion and Airworthiness Lab for the support.

REFERENCE

[1] Ayachi F, Yang L, Dal F, Meneghetti A, Romagnoli A. Assessment of LNG Cold Energy utilization for Road Vehicles and Data-Centres cooling using Liquid Air. 10th Int Conf Appl Energy (ICAE2018), 22-25 August 2018, Hong Kong, China 2018;00:22–5.

[2] Yang L, Villalobos U, Akhmetov B, Onn KJ, Gil A, Tan WL, et al. Active TES With PCM for Refrigeration Applications. Encycl. Energy Storage, Elsevier; 2022. https://doi.org/https://doi.org/10.1016/B978-0-12-819723-3.00029-9.

[3] Yang L, Villalobos U, Akhmetov B, Gil A, Khor JO, Palacios A, et al. A comprehensive review on sub-zero temperature cold thermal energy storage materials, technologies, and applications: State of the art and recent developments. Appl Energy 2021;288:116555.

https://doi.org/10.1016/j.apenergy.2021.116555.

[4] Fu Y, Lin L, Liu C, Wang J. Assumption-free analysis for amplification-based quantitative nucleic acid detection. J Chem Phys 2024;160. https://doi.org/10.1063/5.0191132.

[5] Liu S, Han J, Shen Y, Khan SY, Ji W, Jin H, et al. The contribution of artificial intelligence to phase change materials in thermal energy storage: From prediction to optimization. Renew Energy 2025;238:121973. https://doi.org/10.1016/j.renene.2024.121973.

[6] Gharagheizi F, Gohar MRS, Vayeghan MG. A quantitative structure-property relationship for determination of enthalpy of fusion of pure compounds. J Therm Anal Calorim 2012;109:501–6. https://doi.org/10.1007/s10973-011-1727-y.

[7] Chen P, Liu H, Linghu Y, Zhang C, Wei X, Huang X. Deciphering melting behaviors of energetic compounds using interpretable Machine learning for melt-castable applications. Chem Eng J 2024;479:147392. https://doi.org/10.1016/j.cej.2023.147392.

[8] Xiouras C, Cameli F, Lunardon G, Kavousanakis ME, Vlachos DG, Stefanidis GD. Applications of Artificial Intelligence and Machine Learning Algorithms to Crystallization 2022. https://doi.org/10.1021/acs.chemrev.2c00141.

[9] Hong QJ. Melting temperature prediction via first principles and deep learning. Comput Mater Sci 2022;214:111684. https://doi.org/10.1016/j.commatsci.2022.111684.

[10] Yaws CL. Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds: Physical, Thermodynamic and Transport Properties for 5,000 Organic Chemical Compounds 2003.

[11] Tetko I V., Sushko Y, Novotarskyi S, Patiny L, Kondratov I, Petrenko AE, et al. How accurately can we predict the melting points of drug-like compounds? J Chem Inf Model 2014;54:3320–9. https://doi.org/10.1021/ci5005288.

[12] Katritzky AR, Mu L, Lobanov VS, Karelson M. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. J Phys Chem 1996;100:10400–7. https://doi.org/10.1021/jp953224q.

[13] Xue S-D, Hong Q-J. Materials Properties Prediction (MAPP): Empowering the prediction of material properties solely based on chemical formulas 2023:1–11.

[14] Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model 2010;50:742–54.

[15] Chithrananda S, Grand G, Ramsundar B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. ArXiv Prepr ArXiv201009885 2020.

[16] Ahmad W, Simon E, Chithrananda S, Grand G, Ramsundar B. Chemberta-2: Towards chemical foundation models. ArXiv Prepr ArXiv220901712 2022.