Materials Advances

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Mater. Adv.*, 2025, **6**, 6140

Machine learning-based predictions of melting temperature and enthalpy of fusion for protic organic salt phase change materials†

Saliha Saher, [a] Tu C. Le, *[b] Tamar L. Greaves, [c] Jennifer M. Pringle, [d] Douglas R. MacFarlane [a] and Karolina Matuszek *[a]

Protic organic salts have great potential to be used as phase change materials for thermal energy storage. However, tuning their melting temperatures and maximising their energy storage density (enthalpy of fusion) is a great challenge. The structures of the cation and anion play a crucial role in determining the thermal properties of protic organic salts. In this study, linear and non-linear machine learning models are used to predict the melting temperature (T_m) and enthalpy of fusion (ΔH_f) of 182 possible protic salts using thermal properties (T_m and ΔH_f) of 69 protic salts for training models. An additional feature of this study was the investigation of the prediction accuracy of models for salts with solid–solid phase transitions. It was found that the presence of solid–solid transition/s greatly impacted the ΔH_f predictions. The best linear models for ΔH_f were obtained for salts having no solid–solid transitions (R^2 of 0.82, standard error of estimation (SEE) of 4 kJ mol^{-1}). T_m predictions remained unaffected by the presence of solid–solid transitions. The best linear model for T_m prediction achieved R^2 of 0.63, and SEE of 28 °C. The non-linear models showed marginally lower performance compared to linear models. Experimental cross-validation demonstrated the acceptable predictive ability of linear models for both T_m and ΔH_f. This study opens new avenues for exploring the molecular origins of PCM properties and advancing the development of efficient energy storage materials.

Received 16th May 2025,
Accepted 23rd July 2025

DOI: 10.1039/d5ma00498e

rsc.li/materials-advances

1. Introduction

Despite the abundance of renewable energy, fossil fuels are still the backbone of the current energy system, continuing to exacerbate the global climate crisis and further destabilising energy security and the economy.[1,2] A key hurdle is the intermittency of renewable energy sources, which can be overcome by employing advanced energy storage methods. Thermal energy storage using phase change materials (PCMs) has recently emerged as a potential energy storage solution.[3] PCMs are the materials that can store and release heat during a reversible phase transition (usually solid to liquid) at a specific temperature. The two key properties of a PCM, crucial for optimal performance, are high enthalpy of fusion (ΔH_f), which is defined as the amount of energy absorbed during melting, and the melting temperature (T_m), which determines the operating temperature and, thus, the specific application.[4–6] Intermediate temperature PCMs, having melting temperatures between 100 °C and 220 °C, have the most potential for storing heat from a variety of abundant sources like solar thermal vacuum tubes, and industrial waste heat.[6–8] While there are plenty of traditional organic and inorganic PCMs with melting temperatures ranging from sub-ambient to high temperatures, there is a lack of efficient PCMs in the intermediate temperature range.[6]

Organic salts have recently emerged as potential intermediate temperature PCMs offering high thermal stability, low vapor pressure, and tuneable thermophysical properties.[9–11] Since, the structure of the constituent ions in organic salts drives their properties, an appropriate combination of cations and anions can yield an efficient high energy storage density PCM. However, a nearly infinite number of possible cation–anion combinations and a poor understanding of the underlying structure–property relationships pose a huge challenge to designing organic salt-based PCMs with the desired thermal properties. The identification of optimal PCMs through the

[a] *School of Chemistry, Monash University, Clayton, VIC 3800, Australia. E-mail: karolina.matuszek@monash.edu*
[b] *School of Engineering, STEM College, RMIT University, GPO Box 2476, Melbourne, Victoria, 3001, Australia. E-mail: tu.le@rmit.edu.au*
[c] *School of Science, STEM College, RMIT University, 124 La Trobe Street, Melbourne, VIC 3000, Australia*
[d] *Institute of Frontier Materials, Deakin University, 221 Burwood, VIC 3125, Australia*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5ma00498e

synthesis and testing of a vast number of possible salts is impractical, highlighting the need for advanced techniques to design efficient materials. Machine learning (ML) is an effective tool that has been successfully used to predict various attributes of ionic liquids including density,[12–14] viscosity,[14–17] surface tension,[14,18] melting temperature,[19–22] thermal conductivity,[17] refractive index,[15] heat capacity,[17,23] and toxicity.[24,25] However, ML remains largely underutilized in predicting the most important properties of PCMs *i.e.* T_m and ΔH_f. In 2009, Zhu *et al.*[26] first reported a quantitative structure–property relationship (QSPR) model to predict ΔH_f of imidazolium and quaternary ammonium based ionic liquids using six descriptors (the energy of lowest unoccupied molecular orbital, dipole moment, surface area, volume, shortest H-bond distance and cation–anion interaction energy). A good correlation (R^2 = 0.90 and standard deviation = 4.79 kJ mol^{-1}) was found between calculated and experimental values of ΔH_f.[26] Bai *et al.*[27] developed QSPR models for the prediction of ΔH_f for 40 ionic liquids (which consisted of four subsets; a mix of all 40 different ILs, 22 imidazolium cation based ILs, 10 halide anion containing ILs, and 9 containing imidazolium halide ILs). The models used different numbers of descriptors which were selected by stepwise addition of effective quantum chemical descriptors and removal of ineffective ones. The predicted and experimental ΔH_f for the studied ionic liquids by Bai *et al.*[27] showed good correlation (R^2 between 0.93 to 0.97) for all four models, while the imidazolium halide model showed good predictability and validity.[27] These promising reports suggest that ML methods can be applied for prediction of melting temperature and enthalpy of fusion of organic salt-based PCMs.

Beyond the realm of ionic liquids, machine learning in thermal energy storage is a rapidly expanding field, exploring different dimensions including prediction of eutectic and composite PCM compositions,[28,29] thermal performance prediction, optimisation of PCM-based thermal energy storage systems,[30–34] and thermophysical property prediction of PCMs.[35–38] A neural network and a linear model have been developed to predict the ΔH_f and particle size of micro-encapsulated paraffin wax to understand the influence of synthesis variables (*e.g.* paraffin wax/styrene mass ratio *etc*).[39] Wang *et al.*[28] constructed a back propagation artificial neural network (ANN) model to predict the composition and melting temperature of a eutectic mixture of KCl–NaF with high accuracy. Wang *et al.*[40] also combined particle swarm optimization and a backpropagation neural network for successful prediction of the ΔH_f of binary and ternary eutectic mixtures of inorganic molten salts, with R^2 values of 0.93 and 0.94 respectively. Kottala *et al.*[41] developed an ANN model to predict differential scanning calorimetry outcomes *i.e.* heat flow and temperature of composite PCM ($LiNO_3$ + NaCl) with various mass fractions of expanded graphite. The comparison between predicted and measured values showed a high performance of the model with an R^2 of 0.98. Pan *et al.*[42,43] developed ML models to predict the thermophysical properties (density, expansion coefficient, heat capacity, diffusion coefficient, thermal conductivity and viscosity) of molten $ZnCl_2$ and $ZnCl_2$–NaCl–KCl ternary eutectic mixtures, showing a good agreement with experimental values.

Considering the reasonable success of various ML methods[44–47] for various ionic liquid property prediction, along with diverse applications of ML in thermal energy storage, we decided to employ the technique to predict two key thermal properties *i.e.*, ΔH_f and T_m of protic organic salt PCMs. These predictive ML models will reduce the time and resources needed to explore efficient organic salt-based PCMs and will be highly beneficial in understanding the structure–property relationships at play in organic salt PCMs.

2. Materials and methods

2.1. Data

The experimental T_m and ΔH_f data of 52 (out of 69) protic organic salts used in this study were taken from published reports, most of which were performed in our laboratories.[48–52] They comprised fourteen different cations, namely guanidinium [Gdm], acetamidinium [aca], formamidinium [fa], pyrazolium [Pzy], imidazolium [Im], 1,2,4-triazolium [Tri], pyridinium [PyH], 2-hydroxypyridinium [2-OHPyH], 2-aminopyridinium [2-NH_2PyH], 2-hydroxy-3-aminopyridnium [2-OH-3-NH_2PyH], 2-methylpyridinium [2-MePyH], 3-methylpyridinium [3-MePyH], 4-methylpyridinium [4-MePyH] and 4-*tert*-butylpyridinium [4-*t*-butylPyH] paired with thirteen different anions, chloride (Cl^-), bromide (Br^-), iodide (I^-), nitrate $[NO_3]^-$, tetrafluoroborate $[BF_4]^-$, salicylate [Sal]$^-$, hydrogen sulfate $[HSO_4]^-$, methanesulfonate $[CH_3SO_3]^-$, ethanesulfonate $[C_2H_5SO_3]^-$, triflate $[CF_3SO_3]^-$, trifluoroacetate $[CF_3COO]^-$, benzenesulfonate $[C_6H_5SO_3]^-$, and *p*-tosylate $[p\text{-}CH_3C_6H_4SO_3]^-$. The chemical structures of the cations and anions used to construct the models are presented in Fig. 1.

The combination of fourteen cations and thirteen anions yields 182 possible salts. Since synthesising all 182 salts is not practical, machine learning methods are employed in this study to predict two key thermal properties, melting temperature and enthalpy of fusion, identify important structural features, and efficiently explore the chemical space. This approach significantly reduces the consumption of time and resources. It should be noted that to avoid any variation in the experimental values arising from the use of different differential scanning calorimetry methods (*i.e.* different heating rates and different instruments), data was taken initially only for protic organic salts that have been investigated as PCMs by our group. Additionally, 17 protic salts were synthesised to enrich the input data so that robust and well-representative models could be developed. After data collection, all ΔH_f were converted into standard units of kJ mol^{-1} instead of J g^{-1} as the former unit can better relate to molecular level structure–property relationships. The melting temperatures of the salts varied from 72 °C to 227 °C and enthalpies of fusion varied from 3.4 kJ mol^{-1} to 36 kJ mol^{-1}. The thermal properties of all 69 salts including 52 from the literature (Table S1, ESI†) and 17 newly synthesised salts (Table S2, ESI†) used in the training set are given in ESI.†

2.2. Synthesis of organic salts

All salts were synthesised by dissolving 10 mmoles of the base in 20 ml methanol followed by equimolar addition of the acid.

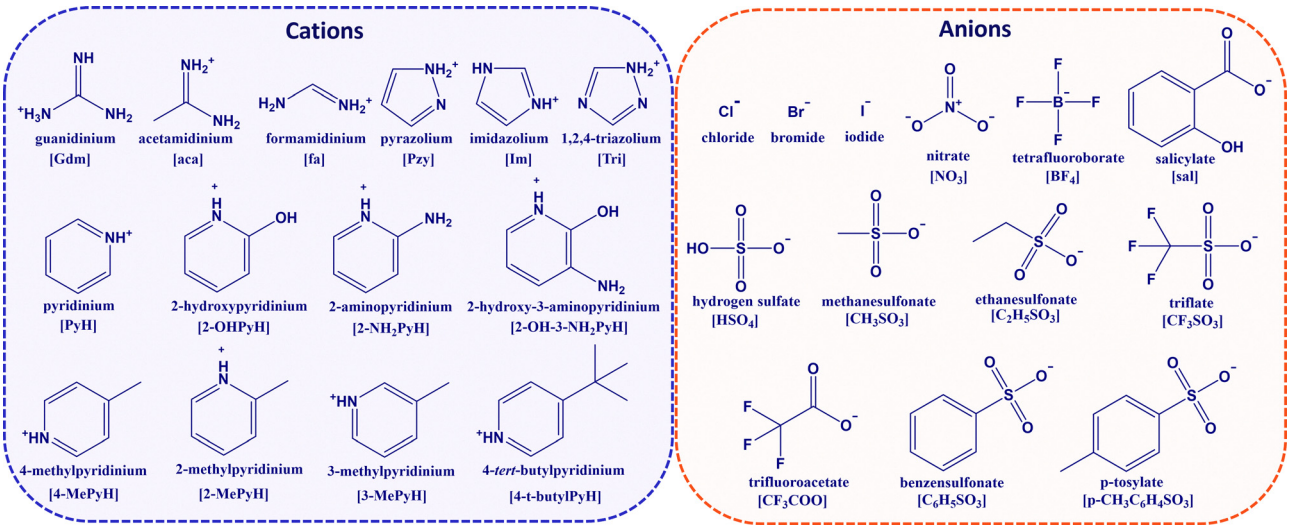

Fig. 1 Chemical structures, names and associated abbreviations of cations and anions used in the training dataset to construct the different machine learning models.

The reaction mixture was stirred for 2 hours at room temperature to ensure the completion of the reaction. The subsequent solution was concentrated by rotary evaporation followed by drying under vacuum (0.2 mbar, 40–60 °C) for 24–48 hours. The obtained solid salts were recrystallised with ethanol, dried under vacuum (0.2 mbar, 40–60 °C) for 24–48 hours. To minimise moisture absorption, thermal analyses were performed immediately after drying. Synthesised salts were characterised by nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). The detailed characterisation analysis is provided in ESI.†

2.3. Descriptors

Molecular descriptors are numbers that represent key structural/electronic features or physico-chemical properties of a molecule used to describe chemical behaviour or properties of molecules.[53] In this study, the molecular structures of the neutral acid (corresponding to the anion) and the neutral base (corresponding to the cation) for each organic salt were first drawn and geometry optimised using Avogadro software employing the integrated universal force field (UFF) *via* auto optimisation tool.[54,55] The MOL files of the acids and bases were then used as inputs for generating constitutional (0D), functional group-based (1D), and topological (2D) molecular descriptors in the Dragon software.[55] Structural (3D) descriptors were intentionally excluded to avoid potential inaccuracies resulting from imperfect geometry optimisation, and to reduce computational time. Although a large number of descriptors were initially generated, irrelevant and redundant features were later eliminated by sparse feature selection of machine learning models, retaining only those most relevant for modelling. A complete list of all generated descriptors is provided in the online data repository. The molecular descriptors correlate with constitutional properties (*e.g.*, the number of atoms of a particular type of element, the molar mass) and structural properties (*e.g.*, the numbers of particular functional groups or atom-centred fragments, molecular property-linked indicators such as H-donor, H-acceptor, and topological surface area).[53,56] The descriptors for cations and anions were generated independently to differentiate their role in property determination. Since this work was mainly focused on generating guidance for designing new organic salt based PCMs, only the chemically meaningful and unique descriptors were selected from the collection of generated descriptors.

2.4. Algorithms for regression models

The BioModeller program was used to derive the quantitative relationship between the input descriptors and the physical properties (ΔH_f and T_m) by applying both linear and non-linear algorithms: multiple linear regression with expectation maximization (MLREM) and Bayesian regularized artificial neural network with Laplacian prior (BRANNLP), respectively.[57–59] The neural networks consisted of three layers: input, hidden, and output layers. The number of nodes in the input layer corresponded to the number of descriptors, while the hidden layer contained 2 or 3 nodes, and the output layer had only a single node corresponding to the property of the organic salt being modelled. Importantly, MLREM and BRANNLP pruned out the irrelevant descriptors, leaving only the most relevant ones because of the sparse feature selection algorithms embedded. The robust QSPR models were developed using two datasets (an explanation of the differences between these datasets is provided in the following section) and the obtained models were used to independently interpolate the ΔH_f and T_m for a library of 182 organic salt PCMs by pairing all possible cations and anions considered in the study. The interpolated ΔH_f and T_m value by each model are provided in ESI† (Tables S5–S12). It should be noted that partitioning this small dataset into training and test sets (80 : 20) could not result in reasonable model development, therefore, all data points were used in the

training, and experimental validation was performed as a performance indicator of the models.

3. Results and discussion

3.1 Regression models for ΔH_f and T_m

Eight models were developed and assessed for predicting the T_m and ΔH_f of the protic organic salt PCMs. For each property, two linear models (MLREM) and two non-linear (BRANNLP) models were developed independently. In the first cycle, a training dataset of 69 organic salt PCMs (termed herein as dataset-a) was used to build both linear (MLREM) and non-linear (BRANNLP) models for T_m and ΔH_f separately. The 69 organic salt PCMs contained 22 salts exhibiting solid–solid phase transitions while the rest had only a single melting transition. During solid–solid phase transitions, the material absorbs and releases energy due to changes in crystal symmetry, molecular orientations or conformational changes without the material melting into the liquid state.[11,60,61] This is demonstrated by differential scanning calorimetry traces of [4-MePyH][CF_3SO_3] and [2MePyH]Cl in Fig. 2(a and b). The first salt exhibits only a melting transition during the heating cycle, while the second salt undergoes a solid–solid transition prior to melting. Although such materials with solid–solid transition/s can find other diverse applications,[9,61] they are undesirable in solid–liquid PCMs for thermal energy storage. This is primarily because solid–solid transitions involve enthalpy changes (ΔH_{s-s}) before melting, leading to significantly reduced energy uptake during melting.[11,60] Therefore, in the next round, the salts with solid–solid transitions along with five salts that were unstable during three cycles were excluded, leaving only 42 organic salt PCMs (termed herein as dataset-b), which were then used to develop linear (MLREM) and non-linear (BRANNLP) models. The results of all of the MLREM and BRANNLP models are given in Table 1. It is also worth noting that some of the salts in dataset-b may exhibit multiple phase transitions at lower temperatures, but these transitions may not have been observed within our selected temperature ranges. Such transitions could also result in a low observed value of ΔH_f.

The correlation coefficient/coefficient of determination (R^2) measures how well the model explains the variability of observed data around the mean (see eqn (1)). An R^2 score close to 1 indicates a good model. However, this can be misleading, and therefore it is often used in conjunction with the standard error of estimation (SEE), which quantifies the average deviation of predicted values from measured values (eqn (2)), expressed in the same units as the target variable.[62] In this context, SEE provides an absolute measure of error, making it a meaningful indicator of performance in this study. A lower SEE indicates better model performance.

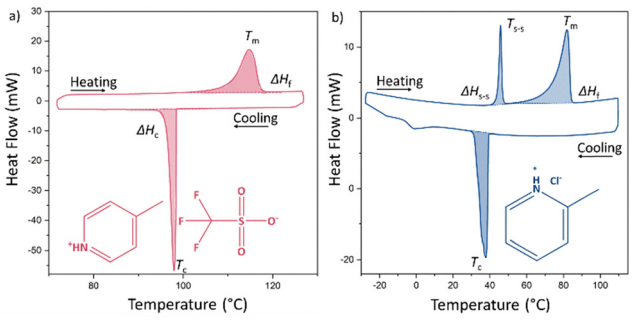

Fig. 2 Differential scanning calorimetry traces of (a) [4-MePyH][CF_3SO_3] indicating only one endothermic peak corresponding to melting of the salt, (b) [2-MePyH]Cl showing a solid–solid transition in addition to melting in the heating cycle.

Table 1 Statistical results of MLREM and BRANNLP models for protic organic salt PCMs. The acronyms TM and HF with MLREM and BRANNLP indicate the target property-melting temperature and enthalpy of fusion, respectively, while the number following the acronym represents the dataset size used to develop each model. The 69-point dataset includes salts with and without solid–solid transition, while 42-point dataset indicates salts with only one transition *i.e.* melting. The SEE units are °C for melting temperature and kJ mol^{-1} for enthalpy of fusion

Output	Data points	Model	Effective weights	R^2	SEE
T_m	69	MLREM_TM69	20	0.63	28
T_m	69	BRANNLP_TM69	13	0.51	25
ΔH_f	69	MLREM_HF69	18	0.65	5
ΔH_f	69	BRANNLP_HF69	8	0.43	5
T_m	42	MLREM_TM42	19	0.62	29
T_m	42	BRANNLP_TM42	12	0.52	22
ΔH_f	42	MLREM_HF42	15	0.82	4
ΔH_f	42	BRANNLP_HF42	11	0.70	3

$$R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \quad (1)$$

$$\mathrm{SEE} = \sqrt{\frac{\sum(y-\hat{y})^2}{n}} \quad (2)$$

here y, $\bar{y}$, $\hat{y}$ and n are the measured values, its mean, predicted value and number of observations in the dataset respectively. For the enthalpy of fusion, both MLREM_HF69 and BRANNLP_HF69 models exhibited low predictability for the first dataset (dataset-a, 69 data points). The MLREM model had an R^2 of 0.65 and SEE of 5 kJ mol^{-1} while the BRANNLP_HF69 model had a lower R^2 of 0.43 showing lower performance but a slightly lower SEE of 5 kJ mol^{-1}. The performance of the MLREM_HF42 model significantly improved for the second dataset (dataset-b) achieving R^2 of 0.82 and SEE reduced to 4 kJ mol^{-1}. The BRANNLP_HF42 model for this dataset also showed a performance with R^2 score of 0.70 and SSE of 3 kJ mol^{-1}. This shows that the relation between the descriptors and ΔH_f is well captured by the simple linear model. Additionally, a substantial improvement in the MLREM and BRANNLP model performances of the second data set suggests that the ΔH_f is less predictable for salts with solid–solid transitions. The parity plots of predicted *versus* measured values for the two MLREM and BRANNLP models for ΔH_f are shown in Fig. 3(a–d) to visualise the agreement between predicted and measured melting enthalpies. A good agreement is indicated by the closeness of data points along the diagonal line in the parity plot.

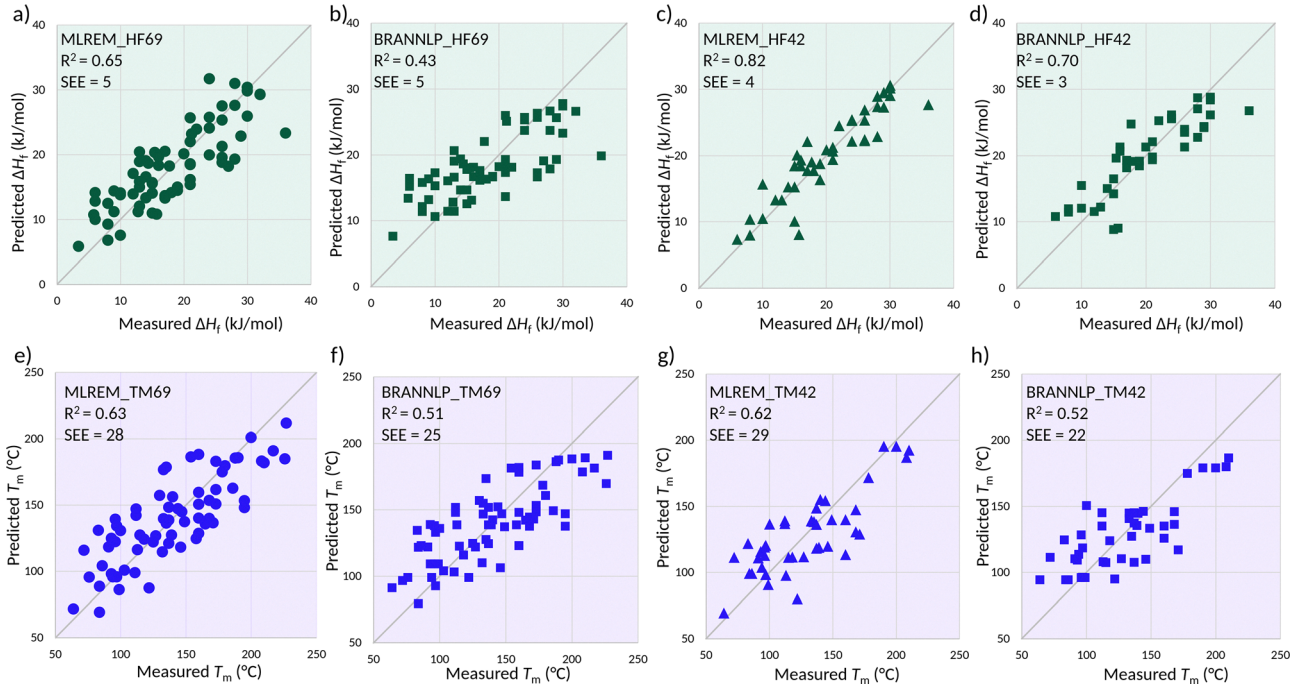

Fig. 3 Comparison of experimental *versus* predicted melting enthalpies (a)–(d) and melting temperatures (e)–(h) of protic organic salt PCMs according to MLREM and BRANNLP models. The panel ((a) and (b)) and ((e) and (f)) represent models developed using dataset-a (69 datapoints) while panels ((c) and (d)) and ((g) and (h)) represent models developed using dataset-b (42 data points).

Interestingly, the predictability of T_m by MLREM and BRANNLP was similar for both datasets. For the dataset-a, MLREM had an R^2 of 0.63 with a SEE of 28 °C. While the BRANNLP model showed a lower R^2 of 0.51 but lower SEE of 25 °C. For dataset-b, the values of R^2 remained mostly unchanged for both MLREM and BRANNLP models, however, there was a slight improvement in the SEE of the BRANNLP model. This can be seen in the parity plots in Fig. 3(e and h). Considering the small sample size; the R^2 value is reasonably acceptable for melting temperature prediction models, and previously R^2 values of 0.54–0.90 and SEE values between 25 °C and 45 °C have been reported.[19,63–66]

Overall, the presence of a solid–solid phase transition appears to have a strong influence on the accuracy of predicting the ΔH_f, while the T_m models largely remain unaffected. Therefore, future predictions for melting enthalpies in systems with such transitions should carefully consider the presence of solid–solid phase transitions. The linear models showed similar performance to the BRANNLP models for both T_m and ΔH_f, indicating that a simple linear relationship between descriptors and melting temperature or enthalpy of fusion can capture complex and hard-to-predict thermal properties like ΔH_f and T_m. The comparative lack of additional success of the BRANNLP model could be due to several factors including the limited size and wide diversity of the training dataset.

3.2. Descriptors

The analysis of the MLREM model and its coefficients for each descriptor provides insight into structure–property relationships in the organic salt PCMs. The coefficients represent how much the target property (ΔH_f or T_m) changes with a one-unit increase in the scaled value of a descriptor while holding all other descriptors constant. A positive coefficient indicates that an increase in the descriptor value correlates with an increase in ΔH_f or T_m, whereas a negative coefficient suggests the opposite trend. The magnitude of the coefficient reflects the relative importance of the descriptor in influencing the target property. The coefficients for the MLREM models for ΔH_f or T_m are shown in Fig. 4. All coefficients are based on scaled descriptor values to allow direct comparison across descriptors.

Although BRANNLP models do not provide explicit weights for individual descriptors, as their internal representations are distributed across multiple layers and neurons, they are still effective in selecting the most relevant descriptors.[67] This selective activation allows the models to capture complex, nonlinear structure–property relationships that may not be easily detected by linear models. As a result, BRANNLP models can support the screening and prioritisation of promising material candidates, even without directly reporting descriptor importance values.

To enhance interpretability, BRANNLP models were used in combination with MLREM models. While BRANNLP offers superior predictive accuracy for property estimation, MLREM provides insight into the relative influence of individual descriptors. This complementary approach allows both accurate prediction and a better understanding of the underlying factors governing the material properties.

A detailed discussion on descriptors used in each model is provided in the following section, while a summarised list of descriptors used in all models is also provided in ESI† Tables S3 and S4.

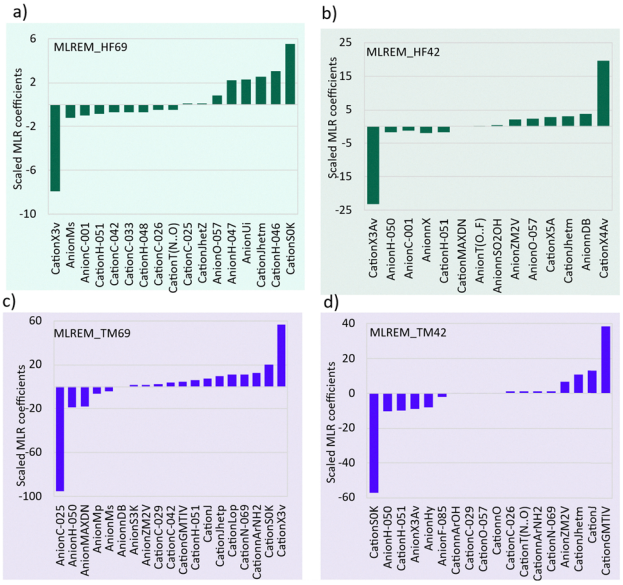

Fig. 4 Scaled regression coefficients of descriptors used in the MLREM models for enthalpy of fusion (a) and (b) and melting temperature (c) and (d).

3.2.1. Enthalpy of fusion descriptors. The analysis of descriptors for cations and anions in each model indicated that both the cation and the anion effectively contributed toward ΔH_f in the developed models as shown by the diversity of descriptors used in both MLREM models in Table 2. The MLREM_HF69 model for ΔH_f had 17 effective descriptors, out of which twelve were related to the cation. The selection of descriptors in this model indicates that the cation has a greater influence than the anion. The symmetry of the cation (CationS0K) had the highest positive correlation, which may be related to efficient crystal packing leading to higher enthalpy of fusion values.[68] Other features that significantly contributed positively to the enthalpy, include: (i) the number of saturated alkyl fragments without a halide group in the cation (CationH-046), (ii) molecular compactness, connectivity and distribution of mass in the cation (Cationjhetm), (iii) degree of unsaturation in the anion (AnionUi), (iv) number of hydrogen atoms in the saturated or unsaturated alkyl chain and (v) the number of phenol, enol or carboxylic hydroxyl group in the anion (AnionO-057), which can be related to the presence of H-bonding.[26,69–71]

On the other hand, CationX3v was the strongest negative contributor to enthalpy along with the mean topological state of the anion (AnionMs), the number of saturated fragments (more specifically methyl containing fragments (CH_3R)), the degree of substitution or saturation around the alpha carbon in the cation (CationH-051) and the number of haloalkyl fragments in the cation. In the other linear model MLREM_HF42 for enthalpy of fusion, 14 features were found to be effective, among which six were cationic. It seems that this model is based on a more balanced number of cationic and anionic features. The model captured the high order connectivity in the cation (CationX4Av and CationX5Av) as a positive contributor to the enthalpy of fusion other features, like the degree of unsaturation in the anion, and the presence of carbonyl/enol/phenolic hydroxyl group indicating potential H-bonding, were also listed as important features for high enthalpy of fusion. On the other hand, the presence of H attached to a heteroatom (AnionH-050), the presence of terminal methyl groups (AnionC-001) and the number of halogen atoms in the anion were found to have negative correlation.

Overall, MLREM models for enthalpy of fusion used different descriptors, except for four common ones (grouped in Table 2), each exhibiting the same type of correlation with ΔH_f across both models. The presence of aromatic hydroxyl groups in the anion (Anion-O57) and compact structure of the cations (CationJhetm) showed positive correlations, while terminal methyl groups (AnionC-001) and branching at the alpha carbon in the cation (CationH-051) displayed negative correlations.

Both BRANNLP models used different descriptors, with no overlap between them. For BRANNLP_HF69 model (dataset-a) seven descriptors were used as input, including four cationic and three anionic descriptors as shown in Table 3. The extent of electronegativity of the cation (CationMe), the connectivity pattern *e.g.* linear or branched (CationJ), the number of pyridines in the cation (CationnPyridines), the number of hydrogen atoms attached to the specific type of carbon atoms (sp^3, sp^2 and sp) in the cation (CationH-049), the number of unsubstituted benzenes (AnionnCbH) and the presence of carbonyl group (AnionO-058) in the anion were all found to influence the ΔH_f.

The BRANNLP_HF42 model (dataset-b) used nine descriptors including only two cationic and seven anionic descriptors (Table 3). The cationic features were mainly linked to the benzene-like aromaticity (CationBLI) and the presence of aromatic hydroxyl groups (CationnArOH). The anionic features included the amount of unsaturation (AnionnDB), the number of halogen atoms (AnionnX), the number of sulfonic groups ($AnionnSO_2OH$), the complexity and branching (AnionZM2V), the local connectivity of atoms, focusing on their first, second, and third neighbours (AnionX3Av) and the presence of a functional group containing carbon and a heteroatom like CF_3COO (AnionC-040).

3.2.2. Melting temperature descriptors. In the two MLREM models for T_m, each utilised 18 descriptors as shown in Table 4. Eight descriptors were common in both models (grouped in Table 4). Among the common descriptors five showed positive correlation with T_m, two showed negative correlation in both models while one descriptor (AnionZM2V) showed positive correlation in MLREM_TM69 and negative correlation in MLREM_TM42. The positively correlated descriptors were all related to cation including CationJ, CationC-029, CationGMTIV, CationN-069, $CationnArNH_2$ which likely indicate branching, halogenated fragments, topology and presence of aromatic NH_2 groups. The negatively correlated structural features include hydrogen atoms attached to a heteroatom and hydrogen attached to the alpha carbon.

The MLREM_TM69 model had 13 descriptors showing positive contributions and five descriptors with negative correlation. The strongest positive correlation of T_m was with the connectivity of atoms in the cation over a range of three bonds (CationX3V), which may be linked to the potential intermolecular interactions in the cation. The strongest negative

Table 2 Molecular descriptors used in training multiple linear regression with expectation maximization (MLREM) for predicting enthalpy of fusion (ΔH_f) of organic salts

Descriptors	Description	
AnionMs	Averaged electronic environments in molecules	MLREM_HF69
AnionH-047	Count of hydrogen–carbon bonds	
AnionUi	Presence of unsaturation in structure	
CationC-025	Count of R–CR–R structural units	
CationC-026	Number of halogenated alkyl fragments	
CationC-033	Number of R–CH$\cdots$X fragments	
CationC-042	Number of X–CH$\cdots$X fragments	
CationH-046	Number of H on sp^3 carbon without adjacent X	
CationH-048	Number of H on primary, secondary and tertiary carbons	
CationJhetZ	Connectivity index with *Z*-weighted distances	
CationS0K	Molecular structure symmetry	
CationT(N$\cdots$O)	Nitrogen–oxygen separation distances	
CationX3v	Direct count of three-bond connectivity	
AnionC-001	Number of CH_3R/CH_4 fragments	Common
AnionO-057	Number of aromatic hydroxyl groups	
CationH-051	Number of H attached to alpha-C	
CationJhetm	Measure of connectivity through mass weighted distance	
AnionnDB	Number of double bonds	MLREM_HF42
AnionnSO_2OH	Number of sulfonic acids	
AnionnX	Number of halogen atoms	
AnionZM2V	Describes molecular structure by counting atom connections	
AnionH-050	Count of hydrogen connected to heteroatom	
AnionT(O$\cdots$F)	Sum of distance between O and F atoms	
CationMAXDN	Measure of maximum negative structural changes	
CationX3Av	Average connectivity across three bonds	
CationX4Av	Average connectivity up to four bonds	
CationX5A	Measures average connectivity up to five bonds	

correlation was the number of tertiary alkyl fragments in the anion (AnionC-025), which can potentially affect the packing efficiency of the organic salt and decrease the melting temperature.

The MLREM_TM42 model for T_m had 12 descriptors showing positive correlation and among them the strongest influence was observed for the degree of connectivity of each atom and the degree of branching in the cation (CationGMTIV). Six features showed negative correlation, and among them CationS0k (representing the symmetry of the cation) was the strongest negative contributor. This is contrary to the general understanding that symmetric molecules/ion will pack efficiently and will result in a high melting temperature. This may be the reason that the model did not have high accuracy in melting temperature prediction (R^2 = 0.62, SEE = 29 °C). However, it should be noted that, in our first model, MLREM_TM69, CationS0k had a high positive correlation to the melting point.

In the two nonlinear models (BRANNLP_TM69 and BRANNLP_TM42) used for melting temperature prediction, largely different molecular descriptors were incorporated to

Table 3 Molecular descriptors used in training Bayesian regularized artificial neural network with Laplacian prior (BRANNLP) for predicting enthalpy of fusion (ΔH_f) of organic salts

Descriptors	Description	
AnionnCbH	Number of free benzene rings	BRANNLP_HF69
AnionO-058	Number of double bonded oxygen atoms	
AnionX1A	Measure of average connectivity	
CationH-049	Number of H on crowded carbon centres	
CationJ	Measure of connectivity through atomic distances	
CationMe	Reflects atomic electronegativity, carbon-based scale	
CationnPyridines	Number of pyridines	
AnionMs	Averaged electronic environments in molecules	BRANNLP_HF42
AnionnDB	Number of double bonds	
AnionnSO_2OH	Number of sulfonic acids	
AnionnX	Number of halogen atoms	
AnionZM2V	Describes molecular structure by counting atom connections	
AnionC-040	Halide containing functional groups	
AnionX3Av	Average of atom connectivity with three bonds	
CationBLI	Quantifies benzene-like structure in molecules.	
CationnArOH	Number of aromatic hydroxyls	

Table 4 Molecular descriptors used in training multiple linear regression with expectation maximization (MLREM) for predicting melting temperature (T_m) of organic salts

Descriptors	Description of descriptors	
AnionMs	Mean electrotopological state	MLREM_TM69
AnionC-025	Number of R–CR–R fragments	
CationLop	Lopping centric index	
AnionMAXDN	Maximal electrotopological negative variation	
AnionMp	Mean atomic polarizability	
AnionnDB	Number of double bonds	
AnionS3K	3-path Kier alpha-modified shape index	
CationC-042	Number of X–CH–X fragments	
CationJhetp	Balaban-type index from polarizability weighted distance matrix	
CationX3v	Valence connectivity index of order 3	
AnionH-050	H attached to heteroatom (−)	Common
AnionZM2V	Second Zagreb index by valence vertex degrees (−)(+)	
CationJ	Balaban distance connectivity index	
CationC-029	Number of R–CX–X fragments	
CationGMTIV	Gutman molecular topological index by valence vertex degrees	
CationH-051	Number of H attached to alpha-C (−)	
CationN-069	Number of Ar–NH_2/X–NH_2 fragments	
CationnArNH$_2$	Number of primary amines (aromatic)	
AnionF-085	F attached to $C_2(sp^2)$-$C_4(sp^2)$/$C_1(sp)$/$C_4(sp^3)$X	MLREM_TM42
AnionHy	Hydrophilic factor	
AnionX3Av	Average valence connectivity index of order 3	
CationC-026	Number of R–CX–R fragments	
CationJhetm	Balaban-type index from mass weighted distance matrix	
CationnArOH	Number of aromatic hydroxyls	
CationnO	Number of oxygen atoms	
CationO-057	Number of phenol/enol/carboxyl OH	
CationS0K	Kier symmetry index	
CationT(N···O)	Sum of topological distances between N···O	

capture structural features that influence the melting temperature except three common descriptors related to hydrogen atoms attached to heteroatom in anion (AnionH-050), polarity and rigidity in cation (CationC-040) and presence of polar groups *e.g.* –OH, –NH_2, –SO_3H (CationTPSA(Tot)). All descriptors used in BRANNLP models for T_m are listed in Table 5. In the BRANNLP_TM69 model only two descriptors, CationJ and AnionH-O50 (which indicate the connectivity in the cation and the number of hydrogen atoms attached to heteroatoms), were the same as those captured by the linear model MLREM_TM69. Several new descriptors related to atom connectivity (CationpiPC06), electrotopological structure (AnionMs), H-bonding and polar interactions (CationnHAcc, CationTPSA(Tot) and AnionH-050), number of halogen-containing functional groups

Table 5 Molecular descriptors used in training Bayesian regularized artificial neural network with Laplacian prior (BRANNLP) for predicting melting temperature (T_m) of organic salts

Descriptors	Description of descriptors	
CationJ	Balaban distance connectivity index	BRANNLP_TM69
AnionMs	Mean electrotopological state	
AnionALOGP	Ghose–Crippen octanol–water partition coeff. (logp)	
AnionTI1	First Mohar index TI1	
CationC-027	Number of R–CH–X fragments	
CationnHAcc	Number of acceptor atoms for H-bonds (N O F)	
CationpiPC06	Molecular multiple path count of order 06	
AnionH-050	H attached to heteroatom	Common
CationC-040	R–C(═X)–X/R–C#X/X–═C═X	
CationTPSA(Tot)	Topological polar surface area using N,O,S,P polar contributions	
AnionC-024	R–CH—R	BRANNLP_TM42
AnionRBF	Rotatable bond fraction	
CationH-050	H attached to heteroatom	
CationnHDon	Number of donor atoms for H-bonds (with N and O	
CationTI1	First Mohar index TI1	
CationX0	Connectivity index chi-0	
AnionZM2V	Second Zagreb index by valence vertex degrees	
CationLop	Lopping centric index	

(CationC-027, CationC-040) and hydrophobicity (AnionALOGP) were found to influence the melting point of the studied salts.

The BRANNLP_TM42 model exhibited a distinct feature selection, sharing only one descriptor, AnionZM2V, with the MLREM_TM42 model and another descriptor, CationTPSA(Tot), with the BRANNLP_TM69 model. New descriptors used in BRANNLP_TM42 model include connectivity and topological descriptors (CationTI1, CationX0, and CationLop), structural fragments (CationC-040 and AnionC-024), H-bonding and polar interactions (CationnHDon and CationH-050) and molecular flexibility descriptor in the anion (AnionRBF).

3.3. Experimental validation of models

To validate the accuracy of the prediction of ΔH_f and T_m by different models, seven new salts outside the training dataset were synthesised and their experimentally measured ΔH_f and T_m were compared with the model's predictions. While cross-validation (test *versus* training set) is often employed for statistical validation, the experimental validation presented here demonstrates the practical utility and reliability of the models beyond theoretical estimations. A model is considered reliable if the difference between the predicted and experimental values is within the standard error of estimation (SEE) of the model.

The synthesised organic salts were selected based on their structural similarity to those in the training set. Salts containing anions and cations within the model's chemical space were selected. For instance, the triazolium cation and chloride, benzenesulfonate and triflate anions were chosen because they appeared frequently in the training set, allowing us to assess whether the model accurately generalises to new but related compounds. Additionally, 2-methylpyridinium, 2-amino-3-hydroxypyridinium and pyridinium cations, along with ethanesulfonate and bromide anions were included to test the model's extrapolation capability. A comparison of predicted and measured enthalpies of fusion and melting temperatures for the seven new salts is given in Fig. 5 and Fig. 6 respectively.

The MLREM_HF42 model (R^2 = 0.82, SEE = 4 kJ mol^{-1}) had the lowest performance in experimental validation with only three salts ([Tri]Br, [4-*t*-BuPyH][$C_2H_5SO_3$], and [2-MePyH]Cl) showing enthalpies within acceptable range *i.e.* SEE. This model predicted around 5 kJ mol^{-1} lower enthalpies for [PyH][$C_2H_5SO_3$] and [2-MePyH][$C_2H_5SO_3$] and 5 and 16 kJ mol^{-1} higher enthalpies for [4-*t*-BuPyH][CF_3SO_3], and [2-NH_2-3-OHPyH][$C_6H_5SO_3$] respectively. Surprisingly, the BRANNLP_HF69 model, which had the lowest R^2 (0.43, SEE = 5 kJ mol^{-1}), showed a good agreement of predicted *versus* measured values with five salts having ΔH_f within the range of SEE and only two salts *i.e.*, [PyH][$C_2H_5SO_3$] and [2-NH_2-3-OHPyH][$C_6H_5SO_3$] showing enthalpies out of acceptable range. The predicted enthalpy for [PyH][$C_2H_5SO_3$] was 5 kJ mol^{-1} higher than measured and for [2-NH_2-3-OHPyH][$C_6H_5SO_3$], the enthalpy exceeded by 10 kJ mol^{-1} This highlights that R^2 should be used cautiously and not solely relied upon as a performance indicator of the models. It is important to mention that two salts [2-MePyH]Cl and [Tri]Br exhibit solid–solid phase transitions (2 kJ mol^{-1}) which could affect the ΔH_f but the experimental ΔH_f for these salts is in good agreement with the predicted ΔH_f by all four models. Overall, the four models failed to predict reliable ΔH_f for [2-NH_2-3-OHPyH][$C_6H_5SO_3$] with predicted ΔH_f significantly higher ΔH_f (10–16 kJ mol^{-1}) than measured, which exceeded the expected deviation given that the instrument error is 5%. This salt was found to supercool and did not fully crystallize even after cooling down to −50 °C. Similar thermal behaviour was observed for [2-MePyH][$C_2H_5SO_3$] which also has poor agreement between measured and predicted ΔH_f, except with the BRANNLP_HF69 model.

For the melting temperature, the comparison of experimental and predicted values for all four models (Fig. 6) showed that MLREM_TM69 was able to predict T_m with reasonable accuracy, with salts having T_m within SEE range except for [2MePyH][$C_2H_5SO_3$] which showed a difference of 60 °C. It may be argued that the SEE for MRLEM_TM69 model was higher (28 °C) than BRANNLP_TM69 (25 °C), thereby helping the T_m values fall within SEE range. However, the MLREM_TM69 (SEE = 28 °C) has shown better performance than MLREM_TM42, which has a slightly higher SEE of 29 °C. The BRANNLP_TM69

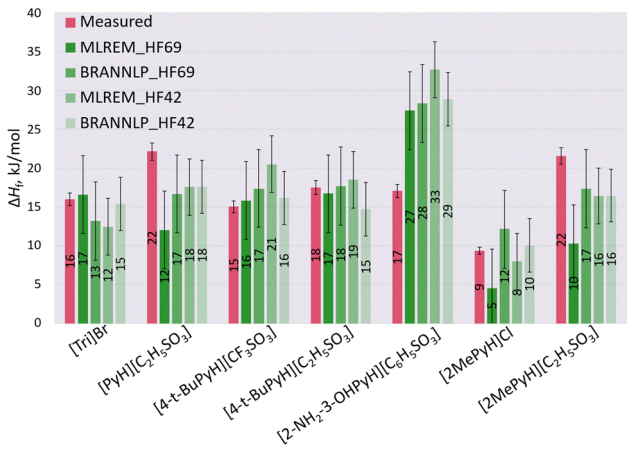

Fig. 5 Predicted *versus* experimental enthalpies of fusion (ΔH_f, kJ mol^{-1}) for selected organic salt PCMs using four different regression models, highlighting model accuracy in enthalpy estimation.

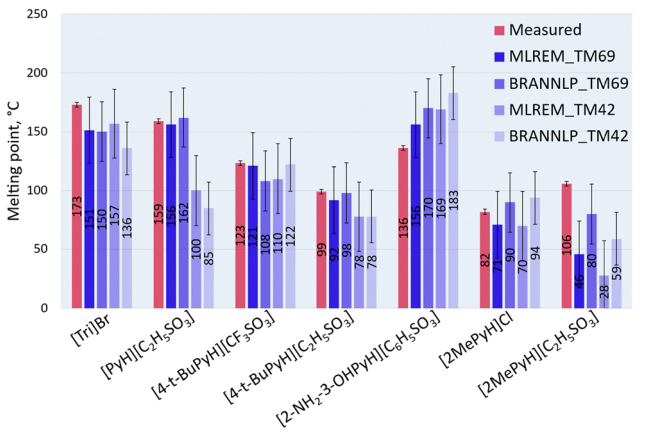

Fig. 6 Comparison of measured and predicted melting temperatures (T_m, °C) for selected organic salt PCMs across four regression models, demonstrating model performance in melting point prediction.

performed better than BRANNLP_TM42, with two salts, *i.e.* [2-NH_2-3-OHPyH][$C_6H_5SO_3$] and [2MePyH][$C_2H_5SO_3$] showing melting point outside the SEE range. It is evident from the bar plot in Fig. 5 that all models predicted higher T_m (20 to 47 °C higher) for [2-NH_2-3-OHPyH][$C_6H_5SO_3$] than the measured T_m and lower (by 16 to 37 °C) T_m for [Tri]Br.

Overall, while no model perfectly captured all experimental trends, BRANNLP_HF69 demonstrated the most consistent performance for enthalpies of fusion predictions, and the MLREM_TM69 showed best performance for predicting meting temperatures among other models. The experimental validation emphasises the limitations of relying solely on R^2 as an indicator of model reliability, particularly for compounds with atypical thermal behaviour.

4. Comparison with ML models in literature

This section presents a comparison between best performing ML models developed in our study for predicting T_m and ΔH_f with those reported in literature. While such comparisons are informative, they must be interpreted with caution and direct benchmarking of model's performance solely on the basis of statistical indicators such as R^2, standard error of estimation (SEE), or root mean square error (RMSE) can be potentially misleading. Nevertheless, a comparative analysis is provided here to contextualise our models within the broader landscape of predictive modelling for ionic liquids.

The best model for T_m prediction (in terms of statistical parameter R^2 and SEE) in our study was MLREM_TM6, which achieved an R^2 of 0.63 and an SEE of 28 °C. This places the model in a competitive range relative to established models in the literature. For instance, Venkatraman *et al.*[21] reported a partial least squares regression (PLSR) model trained on a dataset of 2212 ILs, which yielded an R^2 of 0.50 and an RMSE of 55 °C on the training set. Their non-linear models (*e.g.*, support vector regression, random forest, and k-nearest neighbours) exhibited improved performance, with R^2 values ranging from 0.64 to 0.67 and RMSE values between 46 °C and 49 °C. Low *et al.*[19] employed a *k*-nearest neighbours approach on the same dataset of 2212 IL to investigate the effect of different descriptor choice and best model achieved an R^2 of 0.76 with an SEE of 38 °C. Although Venkatraman *et al.*[21] and Low *et al.*[19] models show a moderate to strong correlation compared to our MLREM_TM69 model, the higher SEE indicates a trade-off between correlation strength and predictive precision, indicated by its significantly lower SEE (28 °C). It should be noted that Venkatraman *et al.*[21] and Low *et al.*,[19] used 2212 ILs, but 40% of ILs in this dataset had bromide and chloride anions, which may have introduced a bias in the models. The multiple linear regression (MLR) and multilayer perceptron neural network (MLPNN) models by Fatemi *et al.*[72] were quite accurate with R^2 of 0.91 and 0.97, respectively. However, the associated SEEs of 35 °C (MLR) and 22 °C (MLPNN) suggest that, despite higher correlation coefficients, the SEE values are comparable to the MLRM_TM69 model presented here. Overall, while non-linear models often yield higher R^2 values, the balance of accuracy and generalisability offered by our model affirms its utility, particularly in early-stage efforts.

In contrast to T_m, predictive modelling for ΔH_f remains relatively underexplored in the literature. Among the limited number of studies, Zhou *et al.*[26] developed a multiple linear regression (MLR) model based on 44 ionic liquids, predominantly imidazolium-based, and reported an R^2 of 0.90 with a standard deviation of approximately 5 kJ mol^{-1}. Similarly, Bai *et al.*[27] constructed four QSPR models with training sets comprising 9 to 30 ILs, achieving R^2 values between 0.93 and 0.97, and standard deviations of 3 kJ mol^{-1} for both. In our study, the MLREM_HF42 model emerged as the best-performing model for ΔH_f prediction, achieving an R^2 of 0.82 and SEE of 4 kJ mol^{-1}. This performance is highly competitive when compared to prior studies, particularly considering the slightly larger and broader diversity of the input dataset in the model training. While the absolute R^2 values are slightly lower than those reported by Zhou *et al.*[26] and Bai *et al.*,[27] the greater dataset size and structural diversity in this work support a higher degree of model generalisability and practical relevance. This trade-off between marginally lower fit and broader applicability reflects a more realistic and deployable model, especially for screening new or less conventional ILs with unknown fusion enthalpies.

5. Conclusions

This work explores simple linear (MLREM) and non-linear (BRANNLP) machine learning methods to predict two of the most important properties for phase change materials, *i.e.* the enthalpy of fusion and melting temperature. Two distinct datasets were used, one consisting of salts that do not exhibit solid–solid phase transitions, and the other including salts both with and without solid–solid transitions. The performance of MLREM and BRANNLP models for enthalpy of fusion predictions varied substantially for the two datasets, the best model was the linear MLREM_HF42 model with R^2 = 0.82 and SEE = 4 kJ mol^{-1} obtained for the dataset of PCMs without solid–solid phase transitions. The performance of both linear and non-linear models for predicting melting temperatures remained approximately unchanged for the two datasets. The best model for melting temperature prediction was MLREM_TM69 with R^2 = 0.63 and SEE = 28 °C for datasets containing PCMs both with and without solid–solid phase transitions. The models in this study did not incorporate total enthalpy (enthalpy of solid–solid transitions + enthalpy of fusion) in their predictions, as our focus was specifically on understanding enthalpy of fusion. Including total enthalpy would have masked the effects of solid–solid transitions or lowered enthalpy contributions, making it difficult to isolate the factors influencing melting behaviour.

Although R^2 was used as a performance indicator, this study highlights the importance of incorporating additional metrics like SEE and experimental validation to provide a comprehensive assessment of the model's performance. Experimental

validation of the enthalpy of fusion models showed poor performance for the highest R^2 model *i.e.* MLREM_HF42 (R^2 = 0.82) and better performance for the lowest R^2 model *i.e.* BRANNLP_HF69 (R^2 = 0.43). Overall, both the enthalpy of fusion and melting temperature models demonstrated moderate prediction accuracy in experimental validation.

This study represents the first attempt to utilise machine learning for predicting the properties of protic organic salt PCMs for thermal energy storage. With the availability of more data, and the guidance of this work, more accurate and robust models can be developed in the future.

Author contributions

S. S and K. M conceived the project. S. Saher collected and curated the data, conducted experiments, conducted the formal analysis, and wrote the original draft. T. C. L developed the machine learning models, contributed to methodology development, and assisted with formal analysis. T. L. G contributed to formal analysis, reviewed and assisted with proofreading the manuscript. D. R. M, K. M, J. M. P supervised the project, acquired funding, and provided resources and reviewed the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including raw figures file and machine learning data used to train models are available at **https://figshare.com/s/89aa6ffb3f493575b31f**.

Acknowledgements

S. S. gratefully acknowledge Higher Education Commission Pakistan. K. M. acknowledges financing from the Australian Research Council (ARC) through its Industry Fellowship Program (IE230100257).

References

1 A. Rahman, S. M. W. Murad, A. K. M. Mohsin and X. Wang, *J. Cleaner Prod.*, 2024, **452**, 142113.
2 W. J. Ripple, C. Wolf, J. W. Gregg, J. Rockström, M. E. Mann, N. Oreskes, T. M. Lenton, S. Rahmstorf, T. M. Newsome, C. Xu, J.-C. Svenning, C. C. Pereira, B. E. Law and T. W. Crowther, *BioScience*, 2024, **74**, 812–824.
3 H. M. Ali, T.-u Rehman, M. Arıcı, Z. Said, B. Duraković, H. I. Mohammed, R. Kumar, M. K. Rathod, O. Buyukdagli and M. Teggar, *Prog. Energy Combust. Sci.*, 2024, **100**, 101109.
4 K. Pielichowska and K. Pielichowski, *Prog. Mater. Sci.*, 2014, **65**, 67–123.
5 S. L. Piper, M. Kar, D. R. MacFarlane, K. Matuszek and J. M. Pringle, *Green Chem.*, 2022, **24**, 102–117.
6 K. Matuszek, M. Kar, J. M. Pringle and D. R. MacFarlane, *Chem. Rev.*, 2022, **123**, 491–514.
7 R. A. Lawag and H. M. Ali, *J. Energy Storage*, 2022, **55**, 105602.
8 Z. L. Yang, R. Walvekar, W. P. Wong, R. K. Sharma, S. Dharaskar and M. Khalid, *J. Energy Storage*, 2024, **87**, 111329.
9 K. Matuszek, S. L. Piper, A. Brzęczek-Szafran, B. Roy, S. Saher, J. M. Pringle and D. R. MacFarlane, *Adv. Mater.*, 2024, **36**, 2313023.
10 M. Mokhtarpour, A. Rostami, H. Shekaari, A. Zarghami and S. Faraji, *Phys. Chem. Chem. Phys.*, 2024, **26**, 13839–13849.
11 K. Matuszek, C. Hatton, M. Kar, J. M. Pringle and D. R. MacFarlane, *J. Non-Cryst. Solids: X*, 2022, **15**, 100108.
12 A. Barati-Harooni, A. Najafi-Marghmaleki, M. Arabloo and A. H. Mohammadi, *J. Mol. Liq.*, 2016, **224**, 954–964.
13 A. Barati-Harooni, A. Najafi-Marghmaleki and A. H. Mohammadi, *J. Mol. Liq.*, 2017, **231**, 462–473.
14 K. Golzar, H. Modarress and S. Amjad-Iranagh, *Int. J. Greenhouse Gas Control*, 2016, **53**, 187–197.
15 M. Mesbah, E. Soroush and M. R. Kakroudi, *J. Mol. Liq.*, 2017, **225**, 778–787.
16 C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana and X. Chen, *Fluid Phase Equilib.*, 2011, **300**, 95–104.
17 Y. Huang, X. Zhang, Y. Zhao, S. Zeng, H. Dong and S. Zhang, *Phys. Chem. Chem. Phys.*, 2015, **17**, 26918–26929.
18 S. Atashrouz, H. Mirshekar, A. Hemmati-Sarapardeh, M. K. Moraveji and B. Nasernejad, *Korean J. Chem. Eng.*, 2017, **34**, 425–439.
19 K. Low, R. Kobayashi and E. I. Izgorodina, *J. Chem. Phys.*, 2020, **153**, 104101.
20 J. A. Cerecedo-Cordoba, J. J. Gonzalez Barbosa, J. Frausto Solis and N. V. Gallardo-Rivas, *J. Chem. Inf. Model.*, 2019, **59**, 3144–3153.
21 V. Venkatraman, S. Evjen, H. K. Knuutila, A. Fiksdahl and B. K. Alsberg, *J. Mol. Liq.*, 2018, **264**, 318–326.
22 G. Carrera and J. Aires-de-Sousa, *Green Chem.*, 2005, **7**, 20–27.
23 W. Beckner, C. Ashraf, J. Lee, D. A. Beck and J. Pfaendtner, *J. Phys. Chem. B*, 2020, **124**, 8347–8357.
24 S. Ma, M. Lv, X. Zhang, H. Zhai and W. Lv, *Chemom. Intell. Lab. Syst.*, 2015, **144**, 138–147.
25 L. Cao, P. Zhu, Y. Zhao and J. Zhao, *J. Hazard. Mater.*, 2018, **352**, 17–26.
26 J. Zhu, L. Bai, B. Chen and W. Fei, *Chem. Eng. J.*, 2009, **147**, 58–62.
27 L. Bai, J. Zhu and B. Chen, *Fluid Phase Equilib.*, 2011, **312**, 7–13.
28 Y. Wang, C. Ling, H. Yin, W. Liu, Z. Tang and Z. Li, *Sol. Energy*, 2020, **204**, 667–672.
29 Q. Wu, Z. Wang, X. Hu, T. Zheng, Z. Yang, F. He, J. Li and J. Wang, *Acta Mater.*, 2020, **182**, 278–286.
30 D. K. Bhamare, P. Saikia, M. K. Rathod, D. Rakshit and J. Banerjee, *Build. Environ.*, 2021, **199**, 107927.

31 X. Xiao, Q. Hu, H. Jiao, Y. Wang and A. Badiei, *Sustainability*, 2023, **15**, 11365.
32 M. Abdolahimoghadam and M. Rahimi, *J. Energy Storage*, 2025, **106**, 114815.
33 G. Ren, A. Chuttar and D. Banerjee, *Int. J. Heat Mass Transfer*, 2022, **189**, 122628.
34 Y. Zhou, S. Zheng and G. Zhang, *Energy*, 2020, **192**, 116608.
35 K. R. Kumar, K. R. Balasubramanian, G. P. Kumar, C. Bharat Kumar and M. M. Cheepu, *Int. J. Thermophys.*, 2022, **43**, 145.
36 M. Goud and F. Raval, *J. Energy Storage*, 2022, **56**, 106163.
37 S. R. Naqvi, R. Tariq, Z. Hameed, I. Ali, S. A. Taqvi, M. Naqvi, M. Niazi, T. Noor and W. Farooq, *Fuel*, 2018, **233**, 529–538.
38 H. Bi, C. Wang, Q. Lin, X. Jiang, C. Jiang and L. Bao, *Energy*, 2020, **213**, 118790.
39 L. Sánchez, P. Sánchez, A. De Lucas, M. Carmona and J. F. Rodríguez, *Macromol. Symp.*, 2010, **287**, 162–167.
40 X.-m Wang, Y.-d Tao, G.-c Dong, S.-y Wang, Q. Miao, H.-l Ding, J. Lv, Q. Wu, Y. Jin and L.-h Tan, *Sol. Energy Mater. Sol. Cells*, 2025, **281**, 113328.
41 R. K. Kottala, B. K. Ramaraj, J. BS, M. G. Vempally and M. Lakshmanan, *Energy Sources, Part A*, 2022, 1–24.
42 G. Pan, P. Chen, H. Yan and Y. Lu, *Comput. Mater. Sci.*, 2020, **185**, 109955.
43 G. Pan, J. Ding, Y. Du, D.-J. Lee and Y. Lu, *Comput. Mater. Sci.*, 2021, **187**, 110055.
44 D. V. Duong, H.-V. Tran, S. K. Pathirannahalage, S. J. Brown, M. Hassett, D. Yalcin, N. Meftahi, A. J. Christofferson, T. L. Greaves and T. C. Le, *J. Chem. Phys.*, 2022, **156**, 154503.
45 S. J. Brown, D. Yalcin, S. Pandiancherri, T. C. Le, I. Orhan, K. Hearn, Q. Han, C. J. Drummond and T. L. Greaves, *J. Mol. Liq.*, 2022, **367**, 120453.
46 T. L. Greaves, K. S. Schaffarczyk McHale, R. F. Burkart-Radke, J. B. Harper and T. C. Le, *Phys. Chem. Chem. Phys.*, 2021, **23**, 2742–2752.
47 D. Yalcin, T. C. Le, C. J. Drummond and T. L. Greaves, *J. Phys. Chem. B*, 2019, **123**, 4085–4097.
48 K. Matuszek, R. Vijayaraghavan, M. Kar and D. R. MacFarlane, *Cryst. Growth Des.*, 2019, **20**, 1285–1291.
49 K. Matuszek, R. Vijayaraghavan, C. M. Forsyth, S. Mahadevan, M. Kar and D. R. MacFarlane, *ChemSusChem*, 2020, **13**, 159–164.
50 S. L. Piper, C. M. Forsyth, M. Kar, D. R. MacFarlane, K. Matuszek and J. M. Pringle, *Mater. Adv.*, 2021, **2**, 7650–7661.
51 K. Matuszek, R. Vijayaraghavan, M. Kar, S. Mahadevan and D. R. MacFarlane, *ChemSusChem*, 2021, **14**, 2757–2762.
52 S. Saher, S. L. Piper, C. M. Forsyth, M. Kar, D. R. MacFarlane, J. M. Pringle and K. Matuszek, *Mater. Adv.*, 2024, **5**, 2991–3000.
53 R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, John Wiley & Sons, 2008.
54 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminf.*, 2012, **4**, 17.
55 A. Mauri, V. Consonni, M. Pavan and R. Todeschini, *MATCH*, 2006, **56**, 237–248.
56 T. Le, V. C. Epa, F. R. Burden and D. A. Winkler, *Chem. Rev.*, 2012, **112**, 2889–2919.
57 F. R. Burden and D. A. Winkler, *J. Chem. Doc.*, 1999, **39**, 236–242.
58 F. R. Burden and D. A. Winkler, *J. Med. Chem.*, 1999, **42**, 3183–3187.
59 D. A. Winkler and F. R. Burden, *Mol. Simul.*, 2000, **24**, 243–258.
60 D. R. MacFarlane and M. Forsyth, *Adv. Mater.*, 2001, **13**, 957–966.
61 C. R. Raj, S. Suresh, R. R. Bhavsar and V. K. Singh, *J. Therm. Anal. Calorim.*, 2020, **139**, 3023–3049.
62 D. L. Alexander, A. Tropsha and D. A. Winkler, *J. Chem. Inf. Model.*, 2015, **55**, 1316–1322.
63 F. Gharagheizi, P. Ilani-Kashkouli and A. H. Mohammadi, *Fluid Phase Equilib.*, 2012, **329**, 1–7.
64 N. Farahani, F. Gharagheizi, S. A. Mirkhani and K. Tumba, *Thermochim. Acta*, 2012, **549**, 17–34.
65 J. O. Valderrama, *Ind. Eng. Chem. Res.*, 2014, **53**, 1004–1014.
66 K. Paduszyński, K. Kłębowski and M. Królikowska, *J. Mol. Liq.*, 2021, **344**, 117631.
67 F. R. Burden and D. A. Winkler, *QSAR Comb. Sci.*, 2009, **28**(10), 1092–1097.
68 R. Pinal, *Org. Biomol. Chem.*, 2004, **2**, 2692–2699.
69 H. Zhang, W. Xu, J. Liu, M. Li and B. Yang, *J. Mol. Liq.*, 2019, **282**, 474–483.
70 K. Matuszek, R. Vijayaraghavan, M. Kar and D. R. MacFarlane, *Cryst. Growth Des.*, 2019, **20**, 1285–1291.
71 C. Yin, Z.-X. Fei, J. Sun, L. Weng, X. Wang, K.-K. Yang and L.-Y. Shi, *Chem. Eng. J.*, 2023, **468**, 143495.
72 M. H. Fatemi and P. Izadian, *J. Theor. Comput. Chem.*, 2012, **11**, 127–141.