Polymer Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Polym. Chem.*, 2025, **16**, 3459

Polymer composites informatics for flammability, thermal, mechanical and electrical property predictions†

Huan Tran, *[a] Chiho Kim,[a] Rishi Gurnani,[a] Oliver Hvidsten,[a] Justin DeSimpliciis,[a] Rampi Ramprasad,[a] Karim Gadelrab,[b] Charles Tuffile,[b] Nicola Molinari,[b] Daniil Kitchaev[b] and Mordechai Kornbluth[b]

Received 11th December 2024,
Accepted 1st July 2025

DOI: 10.1039/d4py01417k

rsc.li/polymers

Polymer composite performance depends significantly on the polymer matrix, additives, processing conditions, and measurement setups. Traditional physics-based optimization methods for these parameters can be slow, labor-intensive, and costly, as they require physical manufacturing and testing. Here, we introduce a first step in extending Polymer Informatics, an AI-based approach proven effective for neat polymer design, into the realm of polymer composites. We curate a comprehensive database of commercially available polymer composites, develop a scheme for machine-readable data representation, and train machine-learning models for 15 flame-resistant, mechanical, thermal, and electrical properties, validating them on entirely unseen data. Future advancements are planned to drive the AI-assisted design of functional and sustainable polymer composites.

1 Introduction

Composites are materials created by combining two or more physically and chemically distinct phases to achieve desired properties or performance enhancements.[1–3] In polymer composites, as illustrated in Fig. 1, the main constituent phases include a matrix of base polymer, co-polymer, or polymer blend, and additional components such as reinforcement fibers, fillers, flame retardants, or functional additives.[2–6] A natural example of a polymer composite is wood, containing cellulose fibers embedded in lignin, a complex organic polymer.[7,8] In this arrangement, the continuous lignin matrix carries and distributes applied loads among the cellulose fibers, giving wood its mechanical strength. By combining diverse polymer matrices, reinforcement fibers, and functional additives[3,5,9–20] synthetic polymer composites may simultaneously meet multiple application-specific requirements, such as lightweight, high strength, corrosion resistance, durability under extreme conditions, and cost-effectiveness. As highlighted in Fig. 1, synthetic polymer composites are widely utilized across industries, including aerospace,[6] automotive,[21] and energy storage and conversion.[22,23]

Designing polymer composites, *i.e.*, rationally identifying formulations that meet predefined criteria for specific applications, is traditionally challenging, costly, and time-intensive, as candidates must be physically synthesized and tested.[1]

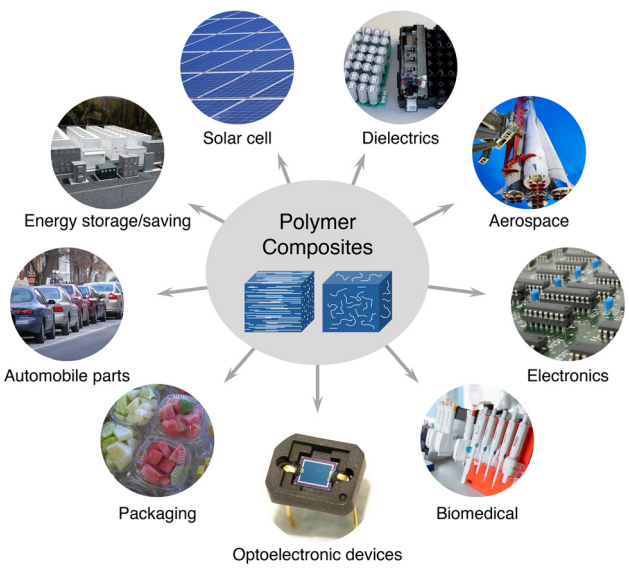

Fig. 1 (Center panel) polymer composites, formed by implanting reinforcement fibers, fillers, or functional additives in a polymer matrix, and (surrounding panels) their applications in different sectors of human life.

[a]*Matmerize Inc., Atlanta, GA 30332, USA. E-mail: huan.tran@matmerize.com*
[b]*Robert Bosch LLC, Watertown, MA 02472, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4py01417k

Because of the inherent complexity of these materials, physics-based evaluation methods like molecular dynamics simulations and finite-element analysis are highly intricate, while quantum mechanical approaches such as density functional theory remain largely out of reach. Empirical models and rules, such as group contribution, the "rule of mixtures",[24–27] the "Cox-Merz rule",[28] and the "Halpin–Tsai equations",[29,30] provide practical alternatives within specific domains but come with their own limitations.[31] Accordingly, a new robust evaluation method is essential to support experiments in polymer composite design.

Since the 2010s, machine-learning (ML) techniques have emerged as valuable complements to traditional approaches in materials science.[32–40] In the field of polymer composites, ML has been used to accelerate simulations[41] and predict physical properties[42] such as conductivity,[43,44] tensile strength,[45–47] fracture behavior,[48] and ductility.[45,46] The training data of these models are predominantly experimental in nature,[45,46,48] while some of them were generated using finite element method.[43,44] The data volume is typically small, ranging from less than ten[47] to a few dozens,[45,46,49] and up to a few hundreds at most.[48] Apparently, data shortage is a major challenges in the future of accelerated design of polymeric materials.[39]

This work aims to develop a set of robust ML models for polymer composites. To this end, we compiled and curated a database of over 5000 polymer composites, fabricated in laboratories and/or industry, with multiple measured properties. Using this database, six multi-task ML models were trained and deployed to predict 15 properties in 4 groups, including flame resistance, mechanical, thermal, and electrical characteristics. The developed models demonstrate good performance on the validation data curated separately and kept unseen to the entire process. We believe that ML, when combined with sufficiently large and diverse datasets and suitable representations, offers a pathway toward the accelerated design of polymer composites.

2 ML models for polymer composites

2.1 Data survey and curation

Polymer composites are manufactured by carefully controlling the selection of the polymer matrix, additives, their compositions, and processing conditions. However, recorded information is often incomplete. Data on polymer composites typically comes from two main sources: research articles and technical datasheets or brochures from manufacturers. Generally, research articles supply more detailed information than technical datasheets. As illustrated in Fig. 2(a), one study[50] examined composites with 50% ethylene-vinyl acetate (EVA) matrix combined with some specific compositions of magnesium hydroxide (MH), aluminum trihydroxide (ATH), and nanoclay (modified montmorillonite), all considered potential flame retardants. Manufacturing, processing, and measurement details, along with measured flammability-related characteristics, can be found in this reference.[50]

Data provided in technical datasheets of commercialized polymer composites are generally less detailed. Fig. 2(b) shows

(a)

Sample	EVA	← (wt%) →	MH	ATH	Nanoclay	LOI	Tensile strength (MPa)	Elongation at break (%)
EVA	100		0	0	0	19.5	18.4	708
EVA/MH	50		50	0	0	27.5	8.5	375
EVA/ATH	50		0	50	0	33.5	9.1	390
EVA/MH/C1	50		49	0	1	34.5	9.4	396
EVA/MH/C2	50		48	0	2	31.0	9.2	393
EVA/MH/C4	50		46	0	4	30.5	9.2	391
EVA/MH/C6	50		44	0	6	25.5	7.9	383
EVA/ATH/C1	50		0	49	1	27.0	8.4	386
EVA/ATH/C2	50		0	48	2	28.0	8.9	391
EVA/ATH/C4	50		0	46	4	26.0	8.8	389
EVA/ATH/C6	50		0	44	6	25.5	8.8	387

(b)

Ultramid® B3G7 R02
PA6-GF33
BASF

Mechanical Properties	Value	Unit	Test Standard
ISO Data			
Tensile Modulus	10000	MPa	ISO 527
Stress at Break	180	MPa	ISO 527
Strain at Break	3	%	ISO 527
Impact Strength (Charpy), +23°C	90	kJ/m²	ISO 179/1eU
Notched Impact Strength (Charpy), +23°C	15	kJ/m²	ISO 179/1eA
Flexural Modulus (23°C)	8400	MPa	ISO 178

Fig. 2 Two sources of polymer composites data curated for this work are (a) research articles and (b) technical datasheets/brochures provided by the manufacturers/distributors of commercialized products. Panel (a) was adapted from ref. 50 with permission while panel (b) was taken from a product brochure obtained from https://www.albis.com.

a top part of the brochure of "Ultramid® B3G7 R02", a product of BASF. This material is labeled as PA6-GF33, implying that it consists of Nylon 6 (PA6) as the polymer matrix and 33% of glass fibers (GF). Such conventions are fairly standard across the polymer composite industry,[21] although interpretations are not always straightforward. In case of "ALCOM® PA66 910/1.3 CF/GF30", a product of MOCOM Compounds Corporation, the label PA66-(CF + GF)30 implies that it contains PA66 polymer matrix and a total of 30% of glass fibers and carbon fibers (CF), but their separate compositions are unknown. Likewise, in the label of (ABS + PA6)-GF8 used for "Terblend® N NG-02 EF" (supplied by INEOS Styrolution), the polymer matrix is a blend of ABS and PA6, but their compositions are also unavailable. In another example, the label of PA6-GF30 FR used for "ALTECH PA6 A 2030/140 GF30 FR" (also provided by MOCOM) indicates that this material contains some flame retardants (FR), but does not provide their identity and compositions.

Such information incompleteness is expected to impede the targeted models in certain ways, for example, by introducing some level of uncertainty in the model's inferences. Nevertheless, if the database is large enough, the undesirable effects of missing data might be partially neutralized and diminished. On the other hand, data extracted from technical datasheets is critically important for our users, as it pertains to materials that are currently available on the market and can be readily purchased in a large quantity.

Our polymer composite database, curated from the two major sources and summarized in Table 1, contains 15 datasets for 15 flame-resistant, mechanical, thermal, and electrical properties. The flame-resistant datasets were curated from hundreds of research articles while the mechanical, thermal, and electrical datasets were extracted from about 10 000 technical datasheets, manually collected for about 5000 commercialized polymer composites. The reported properties were measured under some widely recognized standards, *e.g.*, ASTM E1354 (Cone calorimeter) and ASTM E662 (smoke chamber) for the flammability properties and ISO 527-1/-2 for the mechanical properties. Therefore, testing/measurement conditions are consistent across different sources for the same property/group of properties of the polymer composites.

As discussed above, the description of the materials, needed for the inputs of the ML models, is generally more complete in the research articles than in technical datasheets. The identity and the composition of the polymer matrix and additives are available in the flame-resistant datasets. However, such information is not always available in the mechanical, thermal, and electrical datasets. In some entries, the compositions of polymer matrix blend and the additives may be missing. Notably, for those involving flame retardants, no information on their identity and composition is available. A snapshot of the flame-resistant, mechanical, thermal, and electrical datasets is given in Fig. 3 while more information on the polymer matrices, the additives, and the flame retardants can be found in ESI.†

2.2 Methods

In this work, ML models for 15 properties (summarized in Table 1) were trained using Gaussian Process Regression (GPR) and deep learning (DL) algorithms, with 5-fold cross-validation as part of the training process. The entire workflow was carried out using the PolymRize™ platform. Model performance was also tested on completely unseen data, as described in Section 2.4.

Traditionally, each ML model is trained independently on a single dataset in a procedure known as single-task (ST) learning. On the other hand, multi-task (MT) learning combines multiple related datasets to train a single model, leveraging potential correlations among material properties rooted in physical and chemical laws. Technically, these datasets are stacked together and indicated using an additional selector vector appended to the standard descriptors. The combined dataset can be used for any learning algorithm. In this work,

Table 1 Summary of the datasets, including time to ignition TTI, peak heat release rate PHRR, averaged heat release rate AHRR, total heat release THR, optical smoke density D_s, maximum optical smoke density D_{max}, tensile modulus E, stress at break σ_{break}, glass transition temperature T_g, melting temperature T_m, longitudinal coefficient of thermal expansion α_{long}, transverse coefficient of thermal expansion α_{tran}, relative permittivity at 1 MHz $\varepsilon_{1\ MHz}$, relative permittivity at 100 Hz $\varepsilon_{100\ Hz}$, and breakdown electric strength E_{bd}, collected, cleaned, and used herein

Class	Property	Standard	Unit	Data range	Data size
Flame resistant	TTI	ASTM E1354	S	3.0–281.3	527
	PHRR	ASTM E1354	kW m^{-2}	12.9–1876	576
	AHRR	ASTM E1354	kW m^{-2}	58–750	100
	THR	ASTM E1354	MJ m^{-2}	2.5–609	316
	D_s	ASTM E662	—	0.1–857	474
	D_{max}	ASTM E662	—	1.0–964	124
Mechanical	E	ISO 527-1/-2	MPa	7.4–38 100	4098
	σ_{break}	ISO 527-1/-2	MPa	12–329	2738
Thermal	T_g	ISO 11357-1/-2	C	−109–337	608
	T_m	ISO 11357-1/-3	C	122–388	2044
	α_{long}	ISO 11359-1/-2	10^{-6} K^{-1}	−2.4–250	3373
	α_{tran}	ISO 11359-1/-2	10^{-6} K^{-1}	1.17–230	2889
Electrical	$\varepsilon_{100\ Hz}$	IEC 62631-2-1	—	2.5–15.0	813
	$\varepsilon_{1\ MHz}$	IEC 62631-2-1	—	2.5–7.0	797
	E_{bd}	IEC 60243-1	kV mm^{-1}	15–50	611

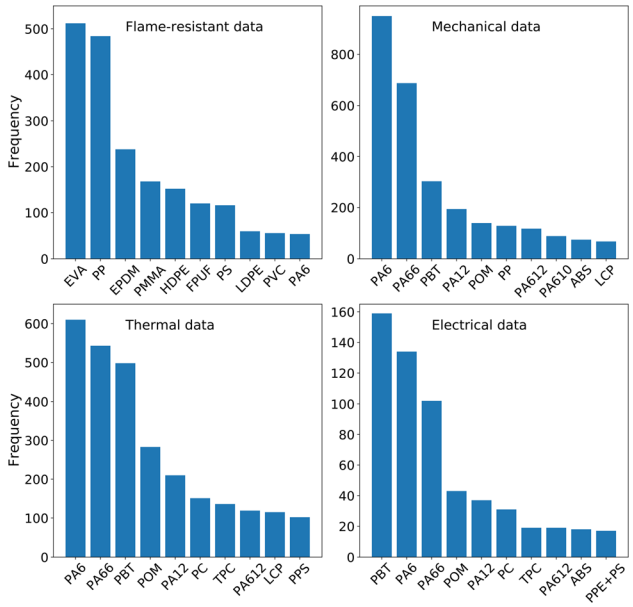

Fig. 3 Top ten base polymer matrices in four group of polymer composite datasets curated and used for this work.

MT learning is referred to as "physics-informed" (Pi) learning, as it uses augmentation data to implicitly convey these physics-containing correlations without requiring explicit mathematical expressions. Pi/MT approach is different from "physics-enforced" learning methods, which rely on directly encoding the correlations, given in terms of specific mathematical expressions, into the model. This study examines the Pi/MT approach against traditional ST learning for developing the targeted ML models (see Section 3 for details).

2.3 Descriptors

Table 2 summarizes the descriptors used to develop the models. Ideally, if SMILES strings[51] encoding the chemical structure of polymer repeat units are available, they can be converted into numerical descriptors.[36–38] However, many polymer matrices in our database lack well-defined SMILES strings, as they are often cross-linking and/or without sufficient information, making descriptor calculation infeasible. Therefore, the polymer matrices are represented by a categorical descriptor, cat_polym, taking their name (*e.g.*, PA6, ABS, PBT) as its value. The composition of glass fibers, carbon fibers, glass beads, and minerals is captured by the numerical descriptors num_gf, num_cf, num_gb, and num_md, respectively. The presence of impact modifiers and flame retardants is indicated by cat_impact and cat_fr1 (Yes/No values). Next, cat_condition specifies the sample state during standard tests as either dry (fully dried) or conditioned (ambient exposure at 23 °C and 50% humidity). Lastly, due to missing details in the thermal, mechanical, and electrical datasets, the density (num_density) is included as an augmentative descriptor, as it is consistently available across materials.

The flame-resistant models share several descriptors with the thermal, mechanical, and electrical models, including cat_polym, num_gf, num_cf, and cat_fr1. For cat_fr1 specifically, this descriptor specifies the identity of the first flame retardant, if present, while num_fr1 gives its composition. This numerical descriptor is unique to the flame-resistant models due to the absence of such data, as discussed above, in models of the other properties. Since materials in the flame-resistant datasets can contain up to four flame retardants, additional descriptors (cat_fr2, num_fr2, cat_fr3, num_fr3, cat_fr4, num_fr4) were included. Similarly, to account for up to two additional reinforcements and two additives beyond glass and carbon fibers, the

Table 2 Features used to develop the ML models

Feature	Description	Applicable to
cat_polym	Categorical, PA6, ABS, PBT, *etc.*	All models
num_gf	Numerical, weight fraction of glass fibers	All models
num_cf	Numerical, weight fraction of carbon fibers	All models
num_gb	Numerical, weight fraction of glass beads	Thermal, mechanical, & electrical models
num_md	Numerical, weight fraction of minerals	Thermal, mechanical, & electrical models
num_density	Numerical, material density (g cm^{-3})	Thermal, mechanical, & electrical models
cat_impact	Categorical, yes/no, if impact modifier included or not	Thermal, mechanical, & electrical models
cat_condition	Categorical, dry/conditioned, measurement condition	Thermal, mechanical, & electrical models
cat_rif1 – cat_rif2	Categorical, identity of other reinforcements if included	Flame-resistant models
num_rif1 – num_rif2	Numerical, weight fraction of other reinforcements if included	Flame-resistant models
cat_adv1 – cat_adv2	Categorical, identity of other additives if included	Flame-resistant models
num_adv1 – num_adv2	Numerical, weight fraction of other additives if included	Flame-resistant models
cat_fr1	Categorical, yes/no, if first flame retardant included	Thermal, mechanical, & electrical models
	Categorical, identity of first flame retardant if included	Flame-resistant models
num_fr1	Numerical, weight fraction of first flame retardant	Flame-resistant models
cat_fr2 – cat_fr4	Categorical, identity of other flame retardants if included	Flame-resistant models
num_fr2 – num_fr4	Numerical, weight fraction of other flame retardants	Flame-resistant models
num_cone_heatflux	Numerical, incoming heat flux (kW m^{-2}) in ASTM E1354 test	TTI, PHRR, AHRR, & THR models
num_cone_thickness	Numerical, thickness (mm) of the sample in ASTM E1354 test	TTI, PHRR, AHRR, & THR models
num_smoke_heatflux	Numerical, incoming heat flux (kW m^{-2}) in ASTM E662 test	D_s & D_{max} models
num_smoke_thickness	Numerical, thickness (mm) of the sample in ASTM E662 test	D_s & D_{max} models
cat_flaming	Categorical, true/false, flaming mode in ASTM E662 test	D_s & D_{max} models
num_smoke_time	Numerical, time (s) of the optical smoke density measurement	D_s & D_{max} models

descriptors cat_rif1, num_rif1, cat_rif2, num_rif2, cat_adv1, num_adv1, cat_adv2, and num_adv2 were used.

Beyond material descriptors, additional features are required for the specific tests measuring flame-resistant performances. Cone calorimeter tests, conducted under ASTM E1354, measure time to ignition (TTI), peak heat release rate (PHRR), average heat release rate (AHRR), and total heat release (THR). Two key parameters of the tests, *i.e.*, the incoming heat flux and the sample thickness, are described by num_cone_heatflux and num_cone_thickness. Likewise, smoke chamber tests, following ASTM E662, measure optical smoke density (D_s) and maximum optical smoke density (D_{max}) under flaming or non-flaming mode. Therefore, for D_s and D_{max} models, num_smoke_heatflux, num_smoke_thickness, cat_flaming (flaming *vs.* non-flaming mode), and num_smoke_time (measurement time) are included, as D_s is time-dependent.

This choice of descriptors may not be ideally comprehensive or complete, potentially omitting useful information if SMILES strings or polymer categories are available and usable. Nevertheless, for the curated data, this technical solution offers not only respectable model performance (discussed in Section 2.4) but also the convenient simplicity needed by the majority of model users.

2.4 Model performances and validations

For 15 properties summarized in Table 1, 15 ST models and 5 Pi/MT models were trained. Each Pi/MT model was trained on a set of properties that are intuitively/clearly correlated. For example, among 6 flame-resistant properties, TTI, PHRR, AHRR, and THR are typically measured simultaneously using a Cone calorimeter, thus they are clearly related and should be combined in a Pi/MT model. Likewise, D_s and D_{max} are measured simultaneously using a smoke chamber, thus another Pi/MT predictive model was developed for them. Starting from similar rationale, 3 other Pi/MT models were developed for the mechanical, thermal, and electrical properties.

As expected, the physics-informed MT models are systematically better than the corresponding ST models in multiple measures of performances, including the determination coefficient R^2, the absolute root-mean-square error aRMSE, and the relative root-mean-square error rRMSE, defined as the ratio between aRMSE and the whole range of the true data. While aRMSE cannot be compared across different datasets and models, rRMSE is more reliable for this purpose. These 3 performance metrics, computed on the training data, are summarized in Table 3. Among 15 models, 12 of them reach $R^2 > 0.9$, while other 2 models have $R^2 > 0.8$; rRMSE metric for all of them is about 5–6% and below. The electric strength model has a moderate $R^2 = 0.57$ and rRMSE $\simeq$ 12%. This result is reasonable and promising, given that our database suffers from unavoidable missing information and that the electric strength is related to and governed by multiple physics-based processes, spanning over multiple length and time scales, and thus understanding it is always highly challenging.[52–54] These 5 models, visualized in Fig. 4, are available in PolymRize™.[55]

Table 3 Summary of five deep-learning physics-informed MT models (separated by horizontal lines) developed for (1) time to ignition TTI, peak heat release rate PHRR, averaged heat release rate AHRR, and total heat release THR, (2) optical smoke density D_s and maximum optical smoke density D_{max}, (3) tensile modulus E and stress at break σ_{break}, (4) glass transition temperature T_g, melting temperature T_m, longitudinal coefficient of thermal expansion α_{long}, and transverse coefficient of thermal expansion α_{tran}, and (5) relative permittivity at 1 MHz $\varepsilon_{1\ MHz}$, relative permittivity at 100 Hz $\varepsilon_{100\ Hz}$, and breakdown electric strength E_{bd}

	Training			Validation		
Model	R^2	aRMSE	rRMSE	R^2	aRMSE	rRMSE
TTI	0.95	9.9	0.036	0.73	17.7	0.071
PHRR	0.94	86.1	0.046	0.74	154.7	0.124
AHRR	0.96	32.5	0.047	0.81	57.3	0.124
THR	0.97	17.9	0.029	0.34	35.05	0.172
D_s	0.99	18.4	0.021	0.78	116.2	0.142
D_{max}	0.99	25.1	0.026	0.89	69.7	0.105
E	0.97	944	0.025	0.98	624	0.030
σ_{break}	0.91	16.3	0.052	0.92	14.0	0.065
T_g	0.98	8.76	0.020	0.97	8.54	0.038
T_m	0.98	6.78	0.025	0.98	3.13	0.033
α_{long}	0.92	13.6	0.054	0.92	11.2	0.064
α_{tran}	0.83	14.3	0.062	0.52	13.7	0.121
$\varepsilon_{100\ Hz}$	0.97	0.65	0.052	0.81	1.3	0.096
$\varepsilon_{1\ MHz}$	0.85	0.25	0.055	0.48	0.41	0.140
E_{bd}	0.57	4.21	0.120	0.14	5.04	0.219

These deployed models were then validated on 15 completely unseen datasets curated independently. For each of them, the data were featurized and the targeted properties were predicted and compared with the ground truth. Predictions for time to ignition TTI, peak heat release rate PHRR, averaged heat release rate AHRR, total heat release THR, optical smoke density D_s, and maximum optical smoke density D_{max}, tensile modulus E, stress at break σ_{break}, glass transition temperature T_g, melting temperature T_m, longitudinal coefficient of thermal expansion α_{long}, transverse coefficient of thermal expansion α_{tran}, relative permittivity at 1 MHz $\varepsilon_{1\ MHz}$, relative permittivity at 100 Hz $\varepsilon_{100\ Hz}$, and breakdown electric strength E_{bd} on the unseen validation data are shown in Fig. 5. For all of the models, the predictions agree very well with the ground truth and aRMSE that is comparable with that reported in Table 3. In summary, all 5 MT models for 15 flame-resistant, mechanical, thermal, and electrical properties can reasonably predict the unseen data, suggesting that the training data of these models are sufficiently big and diverse to represent the common cases of polymer composites.

3 Physics-informed MT learning approach

The advantage of the physics-informed MT models over their ST counterparts is desirable and expected when the correlations among the training datasets are strong. Fig. 6 provides

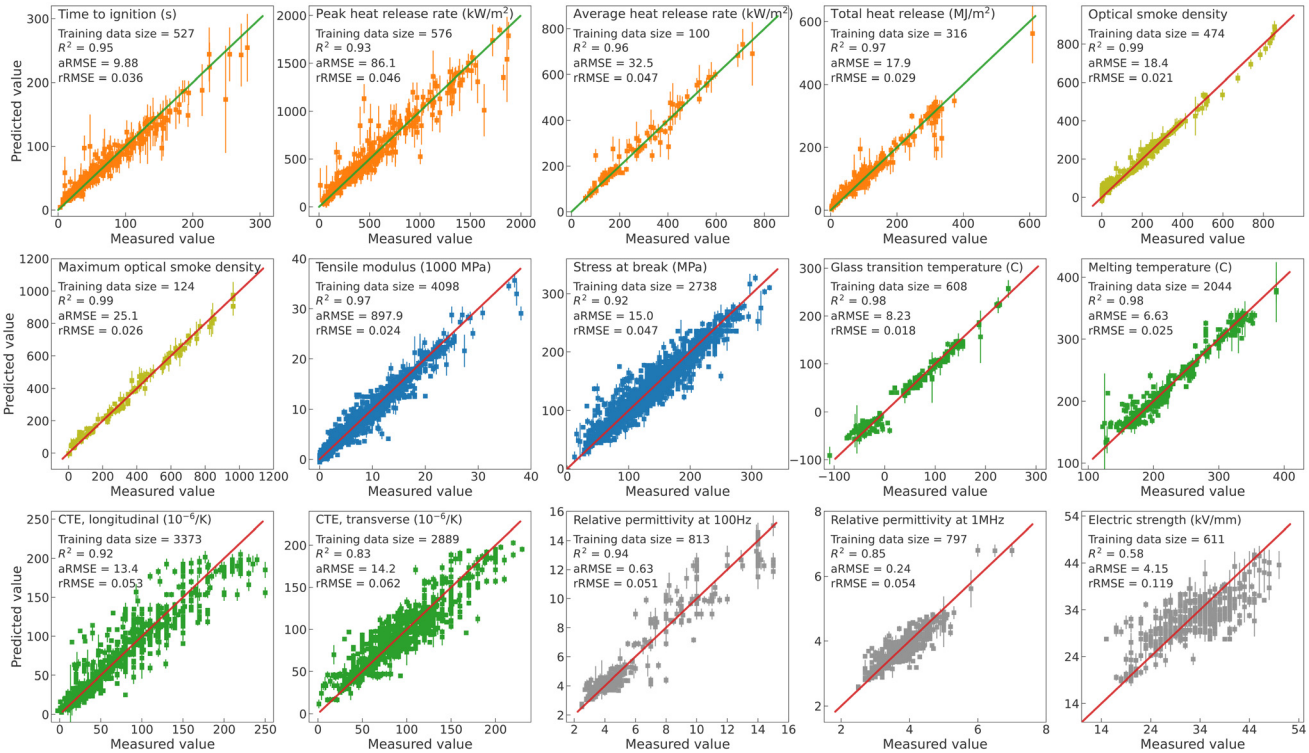

Fig. 4 Visualization of 5 physics-informed MT models developed (and deployed in PolymRize™) for 15 properties of polymer composites. For each of them, R^2 and aRMSE are provided. Each of 5 physics-informed MT models is marked by a distinct color.

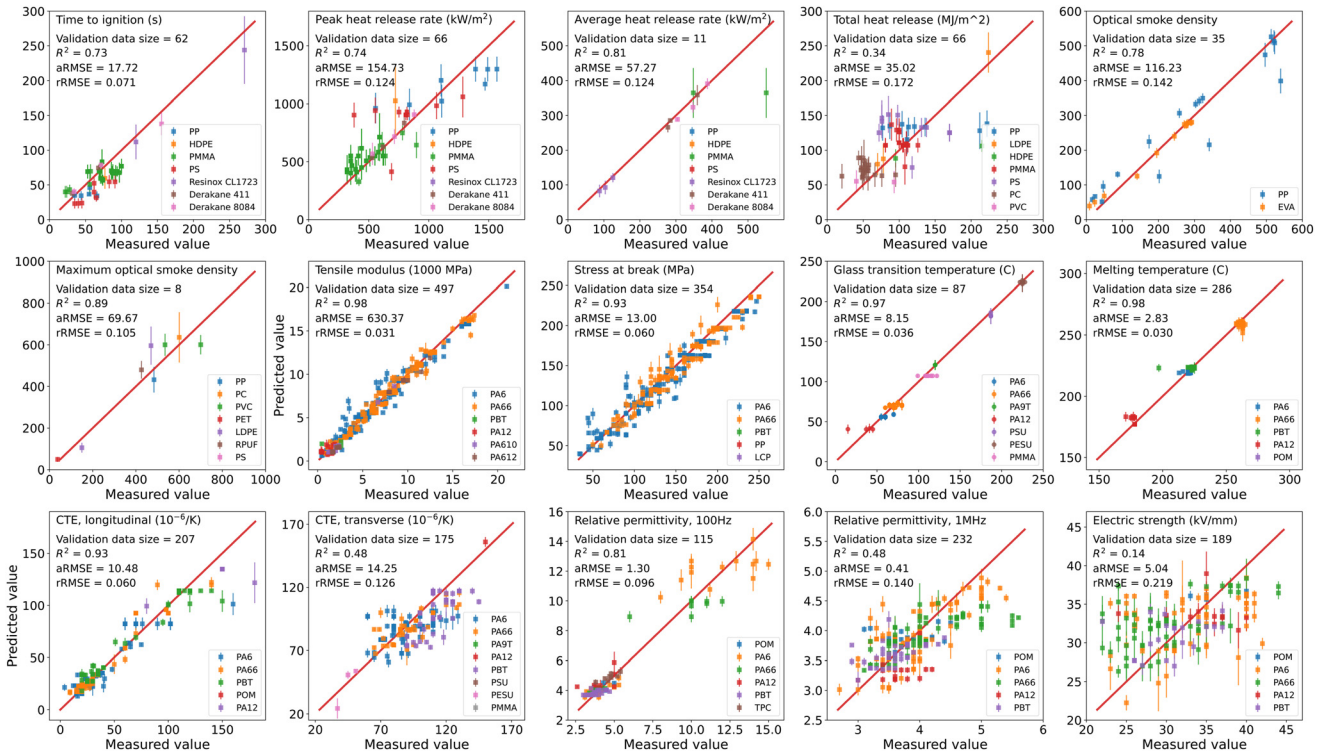

Fig. 5 Predictions of the developed models on the unseen validation datasets curated independently. For each of 15 properties, the base polymer matrix of the validating materials are distinguished by colors.

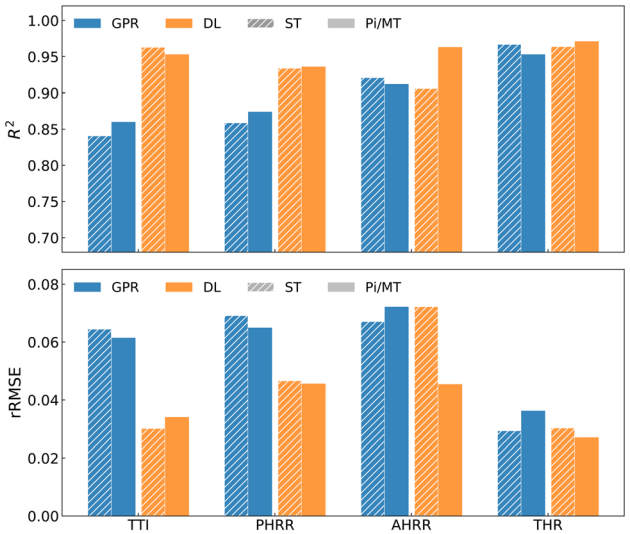

Fig. 6 Coefficient of determination R^2 and relative RMSE of the models trained for the time to ignition TTI, the peak heat release rate PHRR, the averaged heat release rate AHRR, and the total heat release THR. For both GPR and DT, ST and physics-informed MT models are indicated by crossed and solid patterns, respectively.

a summary of the models trained for time to ignition TTI, peak heat release rate PHRR, averaged heat release rate AHRR, and total heat release THR while similar information on all 15 properties can be found in ESI.† Between two learning algorithms, GPR is relatively less effective with TTI and PHRR than DL, for which R^2 is consistently higher than 90%. The physics-informed MT model trained using DL is clearly better, leveraging R^2 of all four properties above 95% while making them more comparable. Similarly, rRMSE is significantly reduced and becomes more balance across TTI, PHRR, AHRR, and THR.

The main rationale of the Pi/MT approach is that by deliberately generating, producing, supplying, and thus, "informing" the training process with data of related properties, the target ML models can be improved.[39] There is, in principle, no limit in the nature and the volume of the augmented data. Moreover, the expected correlations among the datasets are not required to be materialized into any solid mathematical expression. With these two major advantages, the physics-informed MT approach is expected to be widely used in the research area of polymer composites.[39]

4 Forward-looking perspectives and conclusions

Polymer composite data are often scarce and incomplete, posing significant challenges for developing ML predictive models. Addressing these challenges could unlock opportunities for accelerated property predictions and polymer composite design, specifically for extremes. By deploying five Pi/MT models on the largest datasets of their kind and demonstrating predictive performance across 15 widely used properties, this work marks an initial step toward that future.

From the ML perspective, the physics-informed MT learning approach consistently outperformed traditional ST learning, where each model is independently developed for a single property. Prior studies[56,57] suggest that MT architectures can capture hidden correlations among related properties. This work supports that theory. Nevertheless, small data size and large data noise, both of which are common in practice, can suppress the correlations and limit the MT learning efficiency. Addressing these issues remains open for future works.

Manual data curation, as performed here, is unsustainable given the abundance of polymer composite data. Advances in natural language processing, including large language models, named entity recognition, normalization, relation extraction, and co-referencing, may soon offer scalable solutions. Additionally, representing base polymers by name or label, as done in this study, is suboptimal. Future improvements could involve acquiring SMILES strings[51] for all polymers and extending chemical fingerprinting schemes[36,37] to better handle cross-linking polymers and other complex classes, further advancing model performance.

Author contributions

HT designed the project, developed and evaluated the models, and wrote the manuscript. OH, KG, CT, NM, DK, and MK collected, cleaned, and contributed to processing data. CK, RG, and JD developed platform, implemented Pi/MT, deployed, and tested the models. RR helped with designing the project and writing the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Data availability

The data used for this work was collected from public sources cited in the manuscript. Derivatives and data analysis results that support the discussions and conclusions of this work are available in the main text and the ESI.†

Acknowledgements

HT, RG, and OH acknowledge financial supports from National Science Foundation through the SBIR Phase I Grant #2322108. HT thanks Office of Naval Research for financial supports through the SBIR Phase I Contract #N68335-24-C-0121 and Paul Armistead for valuable comments on the manuscript. The authors thank Sydney Balcom (Matmerize, Inc.) for technical assistance.

References

1 E. J. Barbero, *Introduction to composite materials design*, CRC press, 2010.
2 L. Tong, A. P. Mouritz and M. Bannister, *3D fibre reinforced polymer composites*, Elsevier, 2002.
3 T. A. Vilgis, G. Heinrich and M. Klüppel, in *Reinforcement of polymer nano-composites: theory, experiments and applications*, Cambridge University Press, 2009.
4 B. T. Astrom, *Manufacturing of polymer composites*, Routledge, 2018.
5 S. Yuan, F. Shen, C. K. Chua and K. Zhou, *Prog. Polym. Sci.*, 2019, **91**, 141–168.
6 P. E. Irving and C. Soutis, *Polymer composites in the aerospace industry*, Woodhead Publishing, 2019.
7 K. V. Sarkanen and C. H. Ludwig, *Lignins: occurrence, formation, structure and reactions*, Wiley-Interscience, New York, 1971.
8 J. Ralph, K. Lundquist, G. Brunow, F. Lu, H. Kim, P. F. Schatz, J. M. Marita, R. D. Hatfield, S. A. Ralph, J. H. Christensen, *et al.*, *Phytochem. Rev.*, 2004, **3**, 29–60.
9 L. Xie and Y. Zhu, *Polym. Compos.*, 2018, **39**, 2985–2996.
10 X. He and Y. Wang, *Ind. Eng. Chem. Res.*, 2021, **60**, 1137–1154.
11 F. Zhang, Y. Feng and W. Feng, *Mater. Sci. Eng., R*, 2020, **142**, 100580.
12 J. Hu, Y. Huang, Y. Yao, G. Pan, J. Sun, X. Zeng, R. Sun, J.-B. Xu, B. Song and C.-P. Wong, *ACS Appl. Mater. Interfaces*, 2017, **9**, 13544–13553.
13 G. Liu, Y. Xiong and L. Zhou, *Compos. Commun.*, 2021, **27**, 100907.
14 S. Thomas, K. Joseph, S. K. Malhotra, K. Goda and M. S. Sreekala, in *Polymer composites, macro-and microcomposites*, John Wiley & Sons, 2012, vol. 1.
15 K. Friedrich, S. Fakirov and Z. Zhang, in *Polymer composites: from nano-to macro-scale*, Springer Science & Business Media, 2005.
16 C. Baillie and R. Jayasinghe, *Green composites: polymer composites and the environment*, Elsevier, 2004.
17 S. Maiti, M. R. Islam, M. A. Uddin, S. Afroj, S. J. Eichhorn and N. Karim, *Adv. Sustainable Syst.*, 2022, **6**, 2200258.
18 M. H. Islam, S. Afroj, M. A. Uddin, D. V. Andreeva, K. S. Novoselov and N. Karim, *Adv. Funct. Mater.*, 2022, **32**, 2205723.
19 R. Hsissou, R. Seghiri, Z. Benzekri, M. Hilali, M. Rafik and A. Elharfi, *Compos. Struct.*, 2021, **262**, 113640.
20 S. Kangishwar, N. Radhika, A. A. Sheik, A. Chavali and S. Hariharan, *Polym. Bull.*, 2023, **80**, 47–87.
21 A. Elmarakbi, *Advanced composite materials for automotive applications: Structural integrity and crashworthiness*, John Wiley & Sons, 2013.
22 N. S. Grundish, J. B. Goodenough and H. Khani, *Curr. Opin. Electrochem.*, 2021, **30**, 100828.
23 M.-X. Zhu, Q.-C. Yu, H.-G. Song, T.-X. Chen and J.-M. Chen, *ACS Appl. Energy Mater.*, 2021, **4**, 1449–1458.
24 K. Chawla, *Rev. Bras. Fis.*, 1974, **4**, 411–418.
25 M. W. Tham, M. N. Fazita, H. Abdul Khalil, N. Z. Mahmud Zuhudi, M. Jaafar, S. Rizal and M. M. Haafiz, *J. Reinf. Plast. Compos.*, 2019, **38**, 211–248.
26 L. Hart-Smith, *Wt. EngineeringEng.*, 1992, **52**, 29–45.
27 L. Hart-Smith, *Compos. Sci. Technol.*, 2002, **62**, 1515–1544.
28 W. Cox and E. Merz, *J. Polym. Sci.*, 1958, **28**, 619–622.
29 J. H. Affdl and J. Kardos, *Polym. Eng. Sci.*, 1976, **16**, 344–352.
30 D. Zhou, X. Tong, H. Liu, S. Lv, T. Srivatsan and X. Gao, *AIP Adv.*, 2024, **14**, 015157.
31 R. Song, A. H. Muliana and A. Palazotto, *Compos. Struct.*, 2016, **148**, 207–223.
32 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, *npj Comput. Mater.*, 2017, **3**, 54.
33 G. Pilania, C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, **3**, 2810.
34 T. D. Huan, A. Mannodi-Kanakkithodi and R. Ramprasad, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 014106.
35 A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman and R. Ramprasad, *Sci. Rep.*, 2016, **6**, 20952.
36 C. Kim, A. Chandrasekaran, T. D. Huan, D. Das and R. Ramprasad, *J. Phys. Chem. C*, 2018, **122**, 17575–17585.
37 H. Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J. P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton and R. Ramprasad, *J. Appl. Phys.*, 2020, **128**, 171104.
38 L. Chen, G. Pilania, R. Batra, T. D. Huan, C. Kim, C. Kuenneth and R. Ramprasad, *Mater. Sci. Eng., R*, 2021, **144**, 100595.
39 H. Tran, R. Gurnani, C. Kim, G. Pilania, H.-K. Kwon, R. Lively and R. Ramprasad, *Nat. Rev. Mater.*, 2024, **9**, 866–886.
40 R. Gurnani, S. Shukla, D. Kamal, C. Wu, J. Hao, C. Kuenneth, P. Aklujkar, A. Khomane, R. Daniels, A. A. Deshmukh, *et al.*, *Nat. Commun.*, 2024, **15**, 6107.
41 S. Cassola, M. Duhovic, T. Schmidt and D. May, *Composites, Part B*, 2022, **246**, 110208.
42 A. Sharma, T. Mukhopadhyay, S. M. Rangappa, S. Siengchin and V. Kushvaha, *Arch. Comput. Methods Eng.*, 2022, **29**, 3341–3385.
43 M. A. Matos, S. T. Pinho and V. L. Tagarielli, *Carbon*, 2019, **146**, 265–275.
44 M. Matos, S. Pinho and V. Tagarielli, *Scr. Mater.*, 2019, **166**, 117–121.
45 S. Altarazi, M. Ammouri and A. Hijazi, *Comput. Mater. Sci.*, 2018, **153**, 1–9.
46 B. R. S. Reddy, M. Premasudha, B. B. Panigrahi, K.-K. Cho and N. G. S. Reddy, *Polym. Compos.*, 2020, **41**, 3208–3217.
47 G. S. Zeng, C. Hu, S. Zou, L. Zhang and G. Sun, *Polym. Compos.*, 2019, **40**, 3923–3928.
48 V. Kushvaha, S. A. Kumar, P. Madhushri and A. Sharma, *J. Compos. Mater.*, 2020, **54**, 3099–3108.
49 M. Ataeefard, Y. Mohammadi and M. R. Saeb, *Polym. Sci., Ser. A*, 2019, **61**, 667–680.
50 Y.-Y. Yen, H.-T. Wang and W.-J. Guo, *Polym. Degrad. Stab.*, 2012, **97**, 863–869.

51 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
52 T. D. Huan, S. Boggs, G. Teyssedre, C. Laurent, M. Cakmak, S. Kumar and R. Ramprasad, *Prog. Mater. Sci.*, 2016, **83**, 236.
53 A. Mannodi-Kanakkithodi, G. Treich, T. D. Huan, R. Ma, M. Tefferi, Y. Cao, G. Sotzing and R. Ramprasad, *Adv. Mater.*, 2016, **28**, 6277–6291.
54 V. Sharma, C. C. Wang, R. G. Lorenzini, R. Ma, Q. Zhu, D. W. Sinkovits, G. Pilania, A. R. Oganov, S. Kumar, G. A. Sotzing, S. A. Boggs and R. Ramprasad, *Nat. Commun.*, 2014, **5**, 4845.
55 Matmerize, Inc., PolymRize, **https://polymrize.matmerize.com/**.
56 C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim and R. Ramprasad, *Patterns*, 2021, **2**, 100238.
57 R. Gurnani, C. Kuenneth, A. Toland and R. Ramprasad, *Chem. Mater.*, 2023, **35**, 1560–1567.