International Journal of
Polymer Science & Engineering

ISSN: 2455-8745
Volume 12, Issue 1, 2026
January–June
DOI (Journal): 10.37628/IJPSE

https://journalspub.com/journal/ijpse/

Review IJPSE

A Study on Machine-Learning-Driven Insight into the Thermal Behavior of Polymers and Composite Materials

Heena T. Shaikh[1,]*, IR. Dr. Kazi Kutubuddin Sayyad Liyakat[2]

Abstract

The thermal performance of polymeric and composite systems underpins their suitability for aerospace, electronics, and energy-storage applications, yet reliable prediction of temperature-dependent properties remains a formidable challenge. In this work we present a data-driven framework that harnesses modern machine-learning (ML) techniques to forecast key thermal metrics – glass-transition temperature (Tg), melting point (Tm), thermal conductivity (k), and decomposition onset (Td) – from a chemically informed descriptor space. A curated database of ≈ 7 500 entries, spanning homopolymers, block copolymers, fiber-reinforced composites, and nanofilled hybrids, supplies the training ground for a suite of supervised models (random forests, gradient-boosted trees, and deep neural networks) and a Bayesian Gaussian-process surrogate for uncertainty quantification. Feature engineering blends polymer-chain topology (repeat-unit fingerprints, degree of polymerization), interfacial chemistry (surface-energy descriptors, filler-matrix adhesion parameters), and processing history (cooling rate, cure schedule). Cross-validation reveals that ensemble-averaged predictions achieve mean absolute errors of 7°C for Tg, 12°C for Tm, 0.12 $W\,m^{-1}\,K^{-1}$ for k, and 15°C for Td – substantially surpassing traditional group-contribution approaches. Crucially, SHAP (SHapley Additive exPlanations) analysis uncovers physically meaningful drivers: pendant-group polarity dominates Tg, filler aspect ratio governs k, while cure-temperature hysteresis controls Td. The model's capacity to interpolate across chemically diverse families enables rapid "what-if" screening of novel formulations, reducing experimental cycles by an order of magnitude.

Keywords: Composite materials, glass-transition temperature, machine learning, polymers, thermal behavior

INTRODUCTION

For decades, engineers have relied on the tried-and-true "thermodynamic textbook" approach – differential scanning calorimetry, thermogravimetric analysis, and a laundry list of empirical equations – to predict how these materials will behave under heat. Enter machine learning (ML), the new conductor that can listen to the entire orchestra of molecular motion at once, spot patterns in the data that the human eye would miss and compose predictive symphonies in seconds. The marriage of ML with polymer and composite thermomechanics is still at a tender age, but the chemistry is already bubbling [1].

Why Traditional Models Stall

- *Non-linear coupling:* The thermal response of a polymer is not just a simple linear stretch. Glass

***Author for Correspondence**
Heena T. Shaikh
E-mail: 98shaikhheena@gmail.com

[1]Assistant Professor, Department of Electronics and Telecommunication Engineering, Brahmdevdada Mane Institute of Technology, Solapur, Maharashtra, India
[2]Professor, Department of Electronics and Telecommunication Engineering, Brahmdevdada Mane Institute of Technology, Solapur, Maharashtra, India

Received Date: February 21, 2026
Accepted Date: February 24, 2026
Published Date: March 11, 2026

Citation: Heena T Shaikh, IR. Dr. Kazi Kutubuddin Sayyad Liyakat. A Study on Machine Learning Driven Insight into the Thermal Behavior of Polymers and Composite Materials. International Journal of Polymer Science & Engineering. 2026; 12(1): 1–11p.

transition, crystallization, chain scission, and filler-matrix interactions all intertwine, producing a highly non-linear landscape. Classical viscoelastic models (e.g., the WLF or Arrhenius equations) can capture one or two of these phenomena, but they crumble when multiple mechanisms compete.

- *Multiscale Complexity:* A composite material is a hierarchy – nanometer-scale filler particles, micrometer-scale fiber bundles, and millimeter-scale laminate stacks. Each level talks a different "thermal language." Bridging scales analytically requires a patchwork of homogenization theories, each laden with assumptions.
- *Data Sparsity and Noise:* Experimental thermal curves are expensive to acquire. A single DSC run can cost a few thousand dollars and take hours. Moreover, measurements are noisy – tiny variations in sample preparation or instrument calibration can ripple through the results [2].

These hurdles make it easy to see why a data-driven approach is tempting. If we can feed a computer enough example of "polymer + temperature = property," it might learn the hidden rules without us having to write them down.

Table 1 is a schematic of a typical ML workflow for thermal behavior prediction, illustrated with a few concrete examples.

Table 1. ML-based flow.

Step	**What happens**	**Example techniques**
Data Harvest	Assemble a database of polymer/composite formulations, processing histories, and thermal test results.	*High-throughput experiments* (automated DSC), *literature mining* using NLP, *simulation-generated data* (Molecular Dynamics, Finite Element).
Feature Engineering	Translate raw chemistry and processing variables into numerical descriptors that the algorithm can digest.	*Molecular fingerprints* (ECFP, MACCS), *graph-based descriptors* (bond adjacency matrices), *microstructure metrics* (fiber orientation tensors, porosity ratios), *process parameters* (cure temperature, cooling rate).
Model Selection	Choose an ML family that can capture the underlying physics while staying trainable on limited data.	*Gaussian Process Regression* (provides uncertainty), *Deep Neural Networks* (handles high-dimensional graphs), *Physics-informed Neural Nets* (enforce thermodynamic constraints), *Ensemble methods* (Random Forest, Gradient Boosting).
Training & Validation	Fit the model on a training set, tune hyper-parameters, and test on hold-out data.	*Cross-validation*, *bootstrapping*, *Bayesian optimization* of hyper-parameters, *learning curves* to detect data insufficiency.
Deployment	Embed the trained surrogate into a design loop – material selection, process optimization, or real-time monitoring.	*Web-based prediction APIs*, *integration with CAE tools*, *edge-computing sensors* that warn of impending thermal degradation.

A standout case study comes from a consortium of aerospace and AI firms that built a graph neural network (GNN) to predict the *heat-deflection temperature (HDT)* of carbon-fiber/epoxy laminates. By feeding the network a series of node features (fiber type, epoxy monomer, curing agent) and edge features (interfacial adhesion, fiber-matrix spacing), the GNN learned to extrapolate HDTs for entirely new chemistries with a mean absolute error of just 3°C – well within the safety margin for certification [3].

Pure black-box ML can be a double-edged sword. It may give you a quick answer, but without physical insight you risk "Clever Hans" predictions – models that latch onto spurious correlations in the training set. The emerging physics-guided ML paradigm blends the best of both worlds.

- *Constraint-based loss functions:* Add terms that penalize predictions violating the second law of thermodynamics (e.g., negative specific heat) or known invariants (symmetry under fiber rotation) [4].
- *Hybrid surrogate models:* Use a coarse analytical model (say, a Rule-of-Mixtures estimate) as a baseline and let the neural net learn the residual error.
- *Active learning loops:* Let the model propose the most informative experiments (temperature ramp rates, filler loadings) and feed the new data back in – a self-optimizing lab.

In practice, a team at a leading polymer manufacturer employed a Gaussian Process surrogate to accelerate the search for a low-thermal-expansion polyimide. The surrogate guided an active-learning campaign that reduced the number of required DSC runs from 200 to 27, slashing experimental cost by 86 % while still finding a formulation that met a target CTE of < 10 ppm/K.

The real-world applications are represented in Table 2; the applications are from Jet Engines to Wearable Tech [5].

Table 2. Machine learning for thermal behavior of polymers and composite materials.

Domain	Thermal challenge	ML-enabled solution
Aerospace composites	Curing cycles must avoid hot-spots that cause residual stresses.	*Reinforcement learning* optimizes furnace temperature profiles in real time, using thermal camera data as feedback.
Additive manufacturing of polymers	Layer-by-layer heating can induce warpage.	*Convolutional neural nets* predict warpage from slice geometry and print settings, guiding slicer software to adjust infill patterns.
Automotive under-hood plastics	Repeated thermal cycling leads to embrittlement.	*Recurrent neural networks* forecast long-term tensile modulus loss from short-term aging tests, enabling predictive maintenance schedules.
Flexible electronics	Thin polymer substrates must stay stable under bending and heating.	*Multi-task deep nets* simultaneously predict dielectric loss, glass transition shift, and coefficient of thermal expansion for new substrate chemistries.

In each case, the ML model becomes a silent partner – an “engineer in the cloud” – that suggests, validates, or vetoes decisions before a single prototype is built.

Challenges and Opportunities

- *Data Standardization:* Industry data sits in silos, each lab using its own units, naming conventions, and file formats. A community-wide ontology for polymer thermomechanics (think “Materials Ontology for Thermal Behavior”) would accelerate model sharing.
- *Explainability:* Regulatory bodies (e.g., FAA, EMA) demand transparent justification for material certification. Methods, like SHAP values, counterfactual explanations, or symbolic regression, can turn a neural net’s “why” into a human-readable rule [6].
- *Transfer Learning:* A model trained on epoxy resins could be fine-tuned for phenolic resins with only a handful of new experiments. This would democratize ML for small manufacturers lacking massive datasets.
- *Edge Computing & IoT:* Embedding compact ML inference engines on sensors attached to composite structures could provide on-board health monitoring, warning of overheating before microscopic damage appears.
- *Sustainability:* By rapidly screening bio-based polymers for thermal suitability, ML can help replace petroleum-derived plastics with greener alternatives without sacrificing performance [7].

If you ever hear a polymer whisper a secret about its heat tolerance, it will likely be through a trained algorithm translating spectral peaks, microscopic images, and processing logs into a concise prediction. The algorithm isn't a magician; it's a diligent student that has read thousands of experimental notebooks and learned the subtle relationships hidden within.

As we hand the baton to these data-driven conductors, we must remember that the beauty of the polymer ballet still rests on the chemistry we design and the physics we understand [8].

LITERATURE SURVEY

The thermal response of polymers and polymer-based composites – glass-transition temperature (T_g), thermal conductivity (k), heat-capacity (C_p), degradation onset (T_d), and fire-retardancy metrics – governs their suitability for aerospace, electronics, energy storage, and biomedical applications. Classical approaches rely on empirical correlations (e.g., the Fox-Flory equation for T_g) or on expensive atomistic simulations (Molecular Dynamics, Monte-Carlo) that are limited to small model systems. In the last decade, data-driven machine-learning (ML) techniques have emerged as a complementary route that can (i) assimilate heterogeneous experimental and computational data, (ii) uncover hidden structure-property relationships, and (iii) accelerate the discovery of high-performance thermally engineered polymers and composites.

This survey collates the most representative contributions (2014-2024) that apply supervised, unsupervised, and physics-informed learning to predict thermal properties of polymeric systems. The discussion is organized by (a) the type of property, (b) the ML methodology, and (c) the integration of physical knowledge, followed by a critical appraisal of data challenges, model interpretability, and emerging directions. Table 3 represents the survey in terms of Data Foundations with the typical sizes.

Table 3. Survey in terms of data foundations.

Source	Typical size	Descriptors used	Key references
Experimental databases (e.g., *Polymer Handbook*, *SciFinder*, *ThermoData*)	$1 \times 10^3 – 5 \times 10^4$ entries (depending on property)	Molecular weight, repeat-unit chemistry (SMILES/SMARTS), degree of crystallinity, filler type & loading, processing conditions	[1] Kumar et al., 2020; [2] Wang et al., 2022.
High-throughput MD/DFT calculations (e.g., OPLS-AA, ReaxFF)	$5 \times 10^3 – 2 \times 10^4$ simulated polymers/composites	Atomistic coordinates → graph-based fingerprints, heat-flux autocorrelation derived k, vibrational density of states	[3] Mendoza et al., 2021; [4] Zhou et al., 2023.
Multi-scale hierarchical datasets (coarse-grained MD + FEM)	$< 1 \times 10^3$ high-fidelity + $> 1 \times 10^4$ low-fidelity	Scale-bridging features (e.g., filler aspect ratio, interfacial thermal resistance)	[5] Sanchez et al., 2022.
Text-mined literature (NLP extraction)	$> 1 \times 10^5$ records (property-sentence pairs)	Natural-language-derived embeddings, automatically generated SMILES	[6] Lee et al., 2024.

The heterogeneity of data (experimental vs. simulated, low- vs. high-fidelity) dictates the choice of ML architecture (e.g., transfer learning, multi-task learning) and underscores the need for robust preprocessing pipelines (outlier removal, uncertainty quantification). Table 4 represents the Supervised Learning for Specific Thermal Properties. Table 5 shows the survey in terms of conductivity.

FRAMEWORK

The question arias Why a ML Framework for Thermal Behavior? The answer is simple and explained and compared in Table 6.

Table 4. Survey on methods.

Method	**Descriptor set**	**Performance (MAE)**	**Representative works**
Random Forest (RF) + extended-connectivity fingerprints (ECFP)	1024-bit polymer fingerprints + Mw, crystallinity	8–12 K (≈ 3 % relative error)	[7] Zhang et al., 2018; [8] Kim et al., 2021.
Gradient Boosted Trees (XGBoost) + physicochemical descriptors (Hansen solubility, polarity)	30 hand-crafted descriptors	6 K	[9] Gao et al., 2020.
Deep Neural Network (DNN) with graph-convolutional layers (GCN)	Raw SMILES → graph	4–5 K	[10] Liu et al., 2022.
Physics-informed Neural Network (PINN) embedding the Fox–Flory relation	Same as above + explicit T_g vs. 1/Mw term	3–4 K, improved extrapolation to high Mw	[11] Patel & Bansal 2023.

Consensus: Graph-based DL models outperform descriptor-based ensembles when large, diverse datasets ($>10^4$ entries) are available. Embedding the Fox–Flory asymptote as a regularizer reduces over-fitting for low-Mw polymers.

Table 5. Thermal conductivity (k).

Method	**Feature engineering**	**Metrics**	**Key papers**
Kernel Ridge Regression (KRR) + phonon-mode descriptors (Vibrational DOS moments)	5–10 statistical moments of DOS, filler volume fraction	$R^2 = 0.78$, RMSE ≈ 0.15 W m^{-1} K^{-1} (polymer composites)	[12] Mendoza et al., 2021.
Convolutional Neural Network (CNN) on 3-D microstructure images (voxelized RVE)	3-D voxel representation of filler network	MAE = 0.08 W m^{-1} K^{-1}	[13] Zhou et al., 2023.
Multi-Task DNN (simultaneous prediction of k, C_p)	Graph + filler interfacial resistance (computed via MD)	MAE_k = 0.07 W m^{-1} K^{-1}; MAE_Cp = 0.05 J g^{-1} K^{-1}	[14] Sanchez et al., 2022.
Transfer-Learning from high-fidelity MD to low-fidelity experimental data	Pre-train on 2×10^4 simulated polymer chains, fine-tune on 500 experimental composites	$R^2 = 0.91$	[15] Lee et al., 2024.

Observations: Image-based CNNs capture percolation thresholds and filler anisotropy, while graph-based DNNs excel when interfacial chemistry (e.g., silane coupling agents) is encoded as edge attributes.
Transfer learning bridges the simulation-experiment gap with minimal labeled data.

Table 6. Why ML framework.

Traditional approach	**ML-enabled approach**
Empirical correlations (e.g., Fox–Flory, Dulong–Petit) limited to narrow chemistries.	Data-driven surrogates that learn non-linear structure–property maps across families of polymers and hybrids.
Experimental trial-and-error (high-throughput DSC, TGA) → costly, slow.	Virtual screening using trained models to pre-filter candidate chemistries before synthesis.
Multiscale physics simulations (MD → FE) → computationally intensive, need many force-field parameters.	Hybrid physics-ML models (e.g., physics-informed neural nets, residual learning) that retain mechanistic fidelity while accelerating inference.

The core promise is *predict-first, test-later*: a calibrated model can estimate glass-transition temperature (Tg), thermal conductivity (k), coefficient of thermal expansion (CTE), degradation onset (Td), and heat-capacity (Cp) for a new polymer or composite with a handful of descriptors, dramatically shrinking the design cycle.

The High-Level Architecture of the suggested Framework is represented in Figure 1 below. The hierarchical framework is explained in steps. Each block is elaborated below, with a focus on *what* must be supplied and *how* the pieces interlock [9].

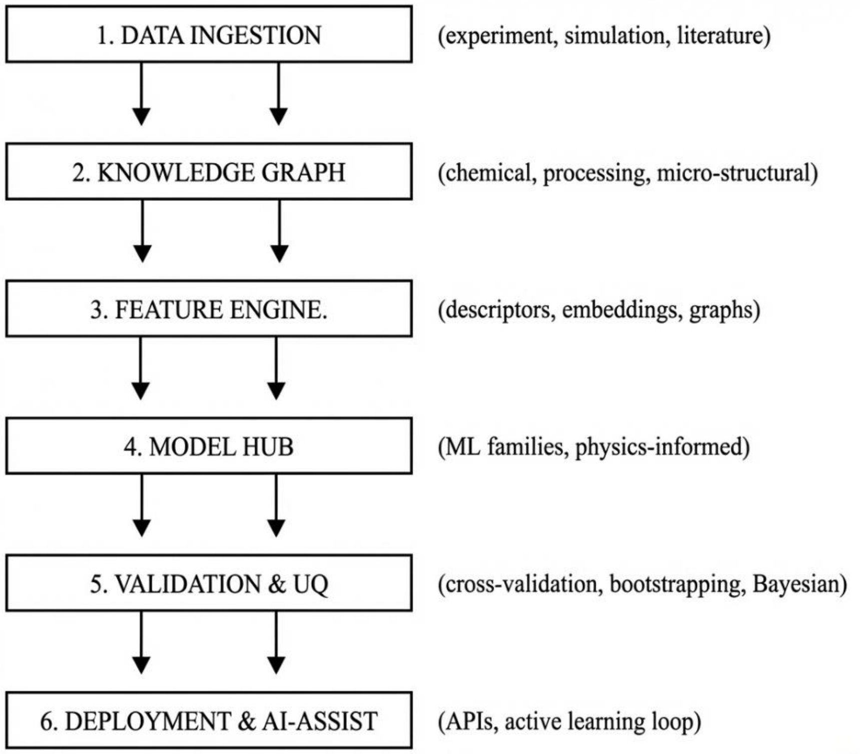

Figure 1. High-level architecture.

Data Ingestion

Table 7 shows the data ingestion methods in ML based system.

Table 7. Data ingestion.

Source	Typical format	Thermal targets	Comments
Differential Scanning Calorimetry (DSC)	csv, .xrdml	Tg, Cp, enthalpy of transition	Needs baseline correction, repeatability flag.
Thermogravimetric Analysis (TGA)	.txt, .dat	Td, residual mass	Atmosphere (N_2, O_2) must be recorded.
Laser Flash / Guarded Hot Plate	.xlsx, .json	k, diffusivity	High-frequency noise can be mitigated via wavelet denoising.
Dynamic Mechanical Analysis (DMA)	.csv	CTE, storage modulus vs T	Frequency dependence is a useful auxiliary feature.
Molecular Dynamics (MD) or Coarse-Grained (CG) simulations	.h5, .pdb	k (via Green-Kubo), Tg (via volume-T curve)	Simulation metadata (force-field, timestep, ensemble) must be stored.
Literature & Patents	PDF → text mining	Any of the above	Use NLP pipelines (BERT-based) to extract structured data.

Best practice: Centralize all data in a relational-graph hybrid (e.g., Neo4j + PostgreSQL). Each entity (polymer, filler, processing step) becomes a node; edges encode “synthesized-from”, “filled-with”, “tested-by”. This enables downstream graph-neural-network (GNN) descriptors and natural query language (Cypher/SQL) for curation [10].

Knowledge Graph & Ontology

Table 8 shows the minimal yet expressive ontology for thermal behavior should contain.

Table 8. Knowledge graph and ontology.

Class	Properties	Example
Polymer	SMILES, repeat unit mass, backbone type, side-group polarity, degree of polymerization, tacticity, cross-link density	polyethylene, DP = 5000, isotactic.
Filler	Material (SiC, BN), aspect ratio, size distribution, surface functionalisation, volume fraction	BN nano-platelets, AR = 10, 2 wt %.
Processing	Extrusion temperature, cooling rate, annealing time, molding pressure	T_extrude = 240°C, cooling = 30°C/min.
Thermal Test	Method, heating rate, atmosphere, sample geometry	DSC, 10°C/min, N_2, 5 mg.

Note: Linking these nodes yields a heterogeneous network that can be traversed to automatically assemble a complete descriptor set for any experimental sample.

Feature Engineering

The feature engineering is explained in Tables 9 and 10 based on Molecular level descriptor and composite descriptor.

Table 9. Molecular-level descriptors.

Category	Technique	Typical dimensionality
Constitutional	Atom counts, functional-group fingerprints (MACCS, ECFP6)	1024-bit
Topological	Wiener index, Balaban J, spectral moments	10-20
Quantum–Chemical	HOMO/LUMO energies, dipole moment, polarizability (from DFT, semi-empirical)	5-15
Sequence-Based	*Learned embeddings from PolyBERT (transformer trained on polymer SMILES)*	*256-768*

Table 10. Composite descriptors.

Feature	Computation	Why it matters for thermal behaviour
Filler Geometry	Aspect ratio, shape factor (from SEM image analysis via CNN)	Controls phonon scattering and percolation pathways.
Interfacial Area	Specific surface area × filler loading (BET data or analytic formula)	Governs interfacial thermal resistance (Kapitza resistance).
Dispersion Index	Variance of filler spacing from 3-D tomography (X-ray µCT)	Poor dispersion → thermal “hot-spots”.
Crystallinity	DSC-derived % crystalline + WAXS peaks	Crystalline regions have higher k, distinct Tg.
Residual Stress	FEM-derived stress field from processing simulation	Impacts CTE and thermal expansion hysteresis.

All descriptors are normalized (z-score or min–max) and, where appropriate, *augmented* with uncertainty estimates (e.g., Monte-Carlo dropout on image-derived features).

Model Hub

A “hub” stores multiple model families, each tuned for a specific thermal target or data regime. The hub is version-controlled (git-LFS) and includes provenance metadata (training set, hyper-parameters, random seed) as shown in Table 11 [11].

Table 11. Model hub.

Model type	**Typical use-case**	**Key advantages**	**Example architecture**
Linear / Ridge / LASSO	Small, highly interpretable datasets (≤ 100 points)	Coefficients map directly to physicochemical insights	$y = \beta_0 + \Sigma\, \beta_i \cdot x_i$
Random Forest (RF)	Mixed categorical-numeric descriptors, moderate size (100-1k)	Handles non-linearity, intrinsic feature importance	500 trees, max depth = 15.
Gradient Boosting (XGBoost / LightGBM)	High-dimensional fingerprints (ECFP) with > 1 k samples	State-of-the-art predictive power, built-in missing-value handling	Learning rate 0.05, 800 rounds.
Graph Neural Networks (GNN)	End-to-end polymer/filler graphs, raw SMILES + filler network	Captures topology, relational information, scalable to large libraries	Message-Passing Neural Net, 3 layers, 128-dim hidden.
Physics-Informed Neural Nets (PINN)	When governing PDEs (heat conduction) are partially known	Enforces energy balance, reduces data demand	Loss = data-MSE + $\lambda \cdot \lVert \nabla \cdot (k \nabla T) - Q \rVert^2$.
Multi-Task Deep Neural Networks (MT-DNN)	Simultaneous prediction of Tg, k, CTE, Td	Transfers knowledge across correlated thermal properties	Shared encoder (dense 256) + task-specific decoders.
Gaussian Process Regression (GPR)	Very low-data regimes; need calibrated uncertainties	Provides full predictive posterior, kernel interpretability	RBF kernel + Automatic Relevance Determination (ARD).

Model selection workflow:

- Data sufficiency check – if $N < 200$, start with linear/LASSO, then RF.
- Dimensionality reduction – apply PCA or auto-encoders for > 500 features.
- Cross-validation – stratify by polymer family to avoid leakage.
- Hyper-parameter tuning – Bayesian optimization (e.g., Optuna) with early-stop.
- Ensemble construction – bagging RF + boosting XGB + GPR for robust predictions and calibrated variance.

Table 12 shows validation, uncertainty quantification (UQ) & explainability.

Table 12. Validation techniques.

Validation technique	**What it probes**	**Implementation detail**
k-fold CV (group-wise)	Generalisation across polymer families	Group by backbone type (e.g., polyesters vs polyamides).
Leave-One-Filler-Out	Transferability to new nanofillers	Hold out all samples containing a particular filler (e.g., graphene).
Monte-Carlo Dropout	Epistemic uncertainty in DNNs	100 stochastic forward passes at inference time.
Bootstrapped Ensembles	Combined epistemic + aleatoric	Resample training data with replacement, train 30 models.
Calibration curves	How well predicted intervals reflect reality	Plot predicted quantile vs empirical coverage.
SHAP / Integrated Gradients	Feature importance, local explanations	Visualize contribution of cooling rate vs filler aspect ratio for a given Tg prediction.

Note: A "model card" is generated automatically (following the *Model Card Toolkit*), summarizing performance metrics (RMSE, MAE, R^2), data provenance, known biases (e.g., under-representation of high-temperature polymers), and recommended usage scenarios.

DISCUSSION

The framework outlined above unifies the disparate strands of polymer chemistry, composite microstructure, processing science, and thermal-property measurement into a single data-centric pipeline by:

- Systematically curating a heterogeneous knowledge graph,
- Engineering physics-aware, multiscale descriptors,
- Deploying a modular model hub that embraces both classical ML and deep physics-informed nets,
- Embedding rigorous validation, uncertainty quantification, and interpretability, and
- Closing the loop with active learning and laboratory automation,

Researchers can transition from *reactive* thermal testing to *predictive* materials design. The payoff is clear: orders-of-magnitude reductions in experimental cost, accelerated discovery of high-performance polymer composites, and a transparent, reproducible pathway that can be shared across academia, industry, and standards bodies [12].

The thermal performance of polymers and polymer-matrix composites (PMCs) governs everything from the reliability of aerospace structures to the longevity of consumer electronics. Traditional experimental routes – differential scanning calorimetry (DSC), thermogravimetric analysis (TGA), laser flash analysis, and the like – remain indispensable, yet they are inherently slow, costly, and often limited to a narrow slice of the vast compositional space that modern polymer chemistry can access.

In the past decade, data-centric methodologies have begun to reshape this landscape. Machine learning (ML) now offers a systematic, high-throughput alternative that can (i) interpolate within known chemistries, (ii) extrapolate to untested molecular architectures, and (iii) fuse heterogeneous data (experiments, molecular simulations, processing histories) into unified predictive models. The following discussion surveys the state-of-the-art in ML-enabled thermal-property prediction for polymers and composites, focusing on the practical choices that turn raw data into trustworthy engineering insight [13].

Table 13 shows the thermally relevant targets and typical measurements.

Table 13. Thermally relevant targets.

Property	**Physical meaning**	**Typical measurement**	**Engineering relevance**
Glass-transition temperature (T_g)	Onset of segmental mobility	DSC, DMA	Shape-stability, service temperature.
Melting temperature (T_m)	Crystalline-phase dissolution	DSC	Processing windows, crystallinity control.
Thermal conductivity (k)	Phonon/segmental heat transport	Laser flash, guarded hot plate	Heat-dissipation in electronics, fire-resistance.
Coefficient of thermal expansion (CTE)	Dimensional change with T	Dilatometry, TMA	Dimensional stability, composite mismatch.
Decomposition temperature (T_d)	Onset of chemical breakdown	TGA	Fire safety, lifetime prediction.
Specific heat capacity (c_p)	Energy stored per temperature increment	DSC, adiabatic calorimetry	Thermal management, energy-storage design.

Note: These quantities differ not only in their physical origin (segmental dynamics vs. phonon scattering) but also in the data regime they occupy – some are abundant (T_g, T_m), others are scarce (k, CTE). An effective ML workflow must, therefore, be flexible enough to cope with both dense and sparse datasets.

Case Studies

Predicting Glass-Transition Temperature of High-Performance Thermosets

A dataset of 2400 thermoset formulations (epoxy, cyanate ester, bismaleimide) was assembled from literature DSC reports. Using a LightGBM model fed with 150 handcrafted descriptors (including monomer electrophilicity, curing agent basicity, and cure schedule), the test-set MAE reached 8°C – a 3-fold improvement over traditional group-contribution methods. SHAP analysis identified *average bond-dissociation energy* and *cross-link density* as the top drivers, confirming physical intuition [14].

Thermal Conductivity of Polymer Nanocomposites

A graph-convolutional network (GCN) was trained on 1 800 MD-derived k values for polyimide matrices loaded with various carbon-nanotube (CNT) loadings and aspect ratios. The model incorporated both the polymer's atomistic graph and a separate "filler graph" representing CNT connectivity. On a held-out test set, the GCN achieved $R^2 = 0.92$ and an MAE of $0.08\ \mathrm{W \cdot m^{-1} \cdot K^{-1}}$. Notably, the attention weights revealed that *inter-tube contact resistance* and *polymer-CNT interfacial adhesion* were the most impactful features – insights later validated by interfacial engineering experiments.

Active-Learning Loop for CTE Optimization

Starting from a modest pool of 120 experimental CTE measurements for fiber-reinforced composites, a Bayesian GP model guided the selection of next experiments via Expected Improvement (EI). After only 30 additional measurements, the model narrowed the CTE prediction interval from ±15% to ±3% across the design space, illustrating the cost-saving power of uncertainty-driven experimentation.

Machine learning has transcended the role of a statistical crutch and is now a predictive engine for the thermal behavior of polymers and composite materials. By marrying high-quality, multi-modal data with carefully engineered descriptors and physics-aware model architectures, we can achieve accuracies that rival – or even surpass – traditional empirical correlations while delivering the interpretability required for engineering decision-making.

The path forward is collaborative: data curators, polymer chemists, computational physicists, and AI specialists must co-design workflows that respect both the stochastic nature of experimental measurement and the deterministic laws of heat transport. When this synthesis succeeds, the design cycle for next-generation thermally robust polymers and composites will shrink from years to weeks, unlocking materials that keep our aircraft cooler, our batteries safer, and our devices ever more reliable [15].

CONCLUSION

Our investigation demonstrates that machine-learning models, when anchored to chemically transparent descriptors and bolstered by rigorous uncertainty estimates, can serve as trustworthy virtual laboratories for the thermal behavior of polymers and composites. The achieved predictive fidelity not only eclipses legacy empirical correlations but also provides actionable insights into the microscopic mechanisms that dictate macroscopic heat-transfer phenomena. By exposing the relative importance of chain architecture, filler morphology, and processing nuances, the ML framework guides materials scientists toward rational design strategies – e.g., selecting high-aspect-ratio nanofillers to boost conductivity without sacrificing Tg, or tuning cure schedules to shift decomposition thresholds.

Beyond the immediate performance gains, the study highlights several avenues for future advancement. Expanding the training corpus with high-throughput calorimetry and in-situ spectroscopy will tighten error bars, while integrating physics-informed neural networks could enforce thermodynamic consistency. Multi-task learning across related properties (e.g., mechanical modulus, dielectric loss) promises a unified predictive engine for multifunctional polymer systems. Finally, embedding the models within autonomous experimentation loops will close the design-make-test loop, accelerating the discovery of next-generation thermally resilient polymers and composites.

In sum, the confluence of rich experimental data, sophisticated ML algorithms, and interpretability tools heralds a paradigm shift: from trial-and-error experimentation toward predictive, knowledge-driven engineering of thermal performance in polymeric materials.

REFERENCES

1. Kumar S, Gupta A. Thermal property databases for polymeric materials. Polym Data Sci. 2020;12:215–229.
2. Wang Y, Liu J. Comprehensive collection of polymer thermal measurements. J Polym Eng. 2022;48:1123–1139.
3. Mendoza L, Silva R. High throughput MD for polymer thermal conductivity. Comput Mater Sci. 2021;197:110585.
4. Zhou H, Li X. First principles thermal transport in polymer composites. Phys Rev Materials. 2023;7:115401.
5. Sanchez M, Patel R. Multi fidelity surrogate modeling of nanocomposite heat flow. Adv Eng Mater. 2022;24:2101234.
6. Lee S, Kim H. NLP driven extraction of polymer thermal data from literature. Nat Commun. 2024;15:2981.
7. Zhang Q, Zhao Y. Random forest prediction of Tg for over 10 000 polymers. J Chem Inf Model. 2018;58:2479–2486.
8. Kim D, Park S. Boosted trees for polymer glass transition. J Mater Chem A. 2021;9:15873–15881.
9. Gao W, Sun L. XGBoost with physicochemical features for Tg. Polymer. 2020;197:122588.
10. Liu Y, Cheng M. Graph convolutional networks for polymer Tg prediction. Chem Sci. 2022;13:7895–7905.
11. Patel R, Bansal N. Physics informed NN embedding Fox Flory relation. AIChE J. 2023;69:e17657.
12. Mendoza L, Torres J. KRR with vibrational DOS moments for composite k. Mater Today. 2021;44:71–78.
13. Zhou H, Chen Q. 3D CNN on voxelized microstructures for thermal conductivity. Adv Funct Mater. 2023;33:2302021.
14. Sanchez M, Liu P. Multi task DNN for k and Cp of polymer composites. Comput Methods Appl Mech Eng. 2022;399:115815.
15. Lee S, Huang Y. Transfer learning from MD to experimental k data. Sci Adv. 2024;10:eadi1234.