Contents lists available at ScienceDirect

Next Materials

journal homepage: www.sciencedirect.com/journal/next-materials

Dielectric constant prediction in polymers: A chemical structure based approach

S. Sharifi^a, S. Bonardd^b, L.A. Miccio^{b,c,*}

^a Department of Physics, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran

^b Centro de Física de Materiales (CFM) (CSIC-UPV/EHU), Paseo Manuel de Lardizábal 5, San Sebastián 20018, Spain

^c Institute of Materials Science and Technology (INTEMA), National Research Council (CONICET), Colón 10850, Mar del Plata, Buenos Aires 7600, Argentina

outperform a more computationally expensive neural network (ANN). Additionally, we implemented an extension of this approach to also handle SMILES-based polymer representations, allowing approximated predictions for molecular structures without available crystallographic data. This study highlights the potential of data-driven approaches for accelerating the discovery of novel dielectric polymers, providing a computational tool that can complement experimental efforts in materials design.

1. Introduction

The dielectric constant is perhaps the most fundamental electric property of materials, since it characterizes their response under the influence of an external electric field. This property reflects the material's ability to be polarized through spontaneous (electronic cloud and chemical bond distortions) or permanent polarization phenomena, the later attributed to the reorientation and align of molecular dipoles. The dielectric constant is directly related to a broad spectrum of technological applications, including energy storage, microelectronics, and high-voltage insulation, each demanding precise control to optimize performance [1–5]. Among the available dielectric materials, polymers have emerged as indispensable in modern technology, offering tuneable permittivity while also providing mechanical flexibility and processability. Polymers also present other advantages, like notably higher dielectric strength values, lightweight and synthesis versatility. In addition, polymer and nanocomposite engineering have led to

substantial enhancements in their dielectric properties, with researchers continually refining material formulations to improve efficiency, reliability, and sustainability [6-9]. High-permittivity polymers are particularly crucial in capacitive energy storage applications, where enhanced charge storage capacity is directly translated into improved device performance [4,10]. Recent developments in polymeric materials have led to substantial increases in dielectric constant, with reported values well-above 5.0, that recently has been selected as the benchmark value to classified high-dielectric polymer [11]. Breakdown strengths exceeding 770 MV/m have also been achieved, highlighting the effectiveness of innovative molecular design strategies. To attain such high performance, several structural and compositional modifications have been explored [12]: like the incorporation of polar groups into polymer backbones to increase the dielectric constant through orientational polarization effects [13]; or high-temperature polyolefins with phenyl pendant groups that also exhibit increased permittivity due to their rigid bicyclic backbone (which in turn facilitates dipole alignment) [14].

https://doi.org/10.1016/j.nxmate.2025.100795

Received 19 April 2025; Received in revised form 21 May 2025; Accepted 3 June 2025 Available online 7 June 2025

2949-8228/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



Research article





^{*} Corresponding author at: Institute of Materials Science and Technology (INTEMA), National Research Council (CONICET), Colón 10850, Mar del Plata, Buenos Aires 7600, Argentina.

E-mail address: luisalejandro.miccios@ehu.es (L.A. Miccio).

Similarly, polyurea- and polythiourea-based systems employ functional groups with high dipolar moments, achieving both high dielectric constants (~5.7) [15] and robust breakdown strengths (>700 MV/m) [16]. Another effective strategy involves the integration of nanocomposites, like the inclusion of ultralow concentrations of inorganic nanodots as fillers within ferroelectric polymer matrices to enhance dielectric response by inducing interfacial dipoles and altering polymer chain conformations [7]. Ferroelectric polymer nanocomposites have also demonstrated exceptional energy densities (~26.0 J/cm³) and breakdown strengths exceeding 600 MV/m, making them promising candidates for high-energy-density capacitor applications [7]. Furthermore, molecularly interpenetrating interfaces have emerged as a novel approach to enhancing both dielectric properties and mechanical robustness. These interfaces introduce distinctive intermolecular interactions that reinforce the polymer network while maintaining high dielectric performance [17]. Such strategies are particularly relevant for applications requiring both flexibility and high charge storage capacity, such as wearable and flexible electronics [3,18].

In contrast to high-permittivity polymers, low-permittivity polymers serve as critical components in microelectronic devices, particularly as interlayer dielectrics in integrated circuits [8,19,20]. These materials play a key role in minimizing parasitic capacitance, reducing power consumption, and mitigating signal delay (essential factors in the continued miniaturization of devices) [16]. Despite their importance, research on low-permittivity polymer development remains relatively underrepresented compared to high-permittivity materials. This disparity highlights the need for systematic investigations into the long-term stability, mechanical properties, and scalability of these polymer candidates. While certain polymer architectures, such as porous polymer films and fluorinated materials have demonstrated promising dielectric properties, challenges remain in balancing permittivity with mechanical and processing reduction compatibility. Processing-property correlations indicate that achieving uniform porosity and stable molecular architectures is critical to maintaining the integrity of low-permittivity films. The fabrication of bilayer polymer films with nanometer-scale control has been suggested as a potential route for optimizing performance while ensuring compatibility with modern lithographic techniques [19,21–24].

Regarding other applications, polymer-based insulators are widely employed in high-voltage power transmission systems, where they must exhibit exceptional dielectric stability and breakdown strength to minimize energy losses and ensure operational reliability [25,26]. Cross-linked polyethylene has been the industry standard for insulation in high-voltage direct current and alternating current cable systems due to its excellent dielectric properties, mechanical durability, and thermal stability [27-30]. However, the crosslinked structure of polyethylene renders it non-recyclable, posing significant environmental concerns as large-scale power grids expand. Recent efforts have focused on designing alternative insulating polymers that maintain desirable properties while offering recyclability. Strategies such as controlled cross-linking, optimized crystallinity, and interfacial engineering have been explored to develop high-performance, sustainable dielectric insulators. All-organic flexible dielectric films with engineered cross-linking networks have been reported to achieve breakdown strengths up to 770 MV/m, demonstrating their potential as next-generation high-voltage insulators.

Despite these experimental advances, accurately predicting the dielectric permittivity of polymers remains a formidable challenge. The complex interplay of molecular architecture, processing conditions, and external factors such as temperature and frequency introduce significant variability in dielectric behaviour. As a result, traditional structure–property relationships fall short in capturing these intricacies, imposing the adoption of data-driven approaches [31]. In this regard, machine learning has emerged as a powerful tool for predicting polymer properties, offering the potential to accelerate materials discovery and/or optimization [32–36]. Regarding dielectric properties of

polymer materials, initial studies have focused on frequency- and temperature-dependent dielectric constant predictions [37], using polymer fingerprinting techniques and statistical learning models. Subsequent work extended the paradigm to the frequency domain [37], and began to experiment with connectivity-aware graph or language models such as GCNs [38] and multitask transformers [39]. Application-oriented studies have meanwhile targeted specialised niches like low-k telecommunication polymers [40] or high-permittivity epoxy resins for capacitors [41], highlighting both the versatility of ML and the field's growing practical relevance. By combining high-throughput computational screening with these neural network-based modelling approaches, researchers extracted meaningful structure-property correlations and generated reliable permittivity predictions under diverse conditions [42,43]. Despite this progress, two critical gaps remain unresolved: (i) no existing approach can process both fully characterised crystalline polymers provided as CIF files and virtual candidates represented only by SMILES strings (i.e. a chemical structure representation only approach); and (ii) state-of-the-art models still depend heavily on global composition features or expensive DFT-derived tensors, limiting true high-throughput screening. Finally, the scarcity of consistent, high-quality data on dielectric permittivity is another critical bottleneck.

In this work, we address these data and complexity challenges by approximating the polymer structure using crystallographic information inputs (CIF) and high-quality dielectric constant values derived from first-principles calculations [44]. To tap into the rich structural details captured by these CIF files, we employ a custom, automated pipeline that reads, filters, and processes each polymer structure. Our approach extracts global features (such as unit cell parameters, volume, among others) as well as atom-wise descriptors (composition), which are subsequently tokenized and embedded. These embedded tokens capture local compositional and site-level information, thereby enhancing the model's ability to learn subtle structure-property correlations. We feed the resulting comprehensive feature representations into a machine learning model which concatenates the aggregated global features with averaged atom embeddings to predict dielectric permittivity. Ultimately, our approach can assist in the discovery of novel polymers with precisely tuned dielectric properties, accelerating progress in a range of applications from next-generation electronics to high-performance power cables and energy storage devices.

2. Methods

In this section we outline the key steps of our methodology, from data loading and feature extraction to neural network architecture and training procedure. In addition, we have included an expansion for SMILES-based inputs, so that any new chemical structures can be sketched and tested easily. All pipelines are schematically shown in Fig. 1.

2.1. Data description and preprocessing

All polymer structures and associated dielectric constant data used in this study were obtained from a curated collection of nearly 1000 polymeric compounds for which structural information is provided in CIF format, along with experimental or computationally derived dielectric constant values [44]. The dataset arises from a density functional perturbation theory approach that computes the dielectric constant (each sample was treated as a solid-state crystal in the calculations). This approach yields the static dielectric constant, which effectively corresponds to zero (or near-zero) frequency. Some of these zero-frequency calculated values were compared to measured dielectric constants at low frequencies and room temperature [44]. When discrepancies were found to be large, the authors re-optimized the structures until the computed properties converged more closely to the experimental results. In addition, the calculations were done on



Fig. 1. Schematic pipeline for the machine learning model predicting dielectric constants from CIF files and SMILES. **left (model training):** set up the computing environment and define filtering parameters. CIF files are read and processed to extract unit cell properties, atomic features, and metadata, while applying necessary filters. The processed features and labels are then stored, and outliers are removed. Atom vocabulary is built, and a custom dataset is created, followed by dataset splitting into training and validation sets. The Atom-Aware MLP model is then defined and trained using batch processing, optimization, and early stopping techniques. Training and validation loss are recorded, followed by visualization of model performance using scatter plots, hexbin plots, and error metrics. Finally, the trained model and vocabulary are saved for future use. **right (extension to SMILES inputs)**: set up the computing environment, and convert the input SMILES string into an ASE atoms object. The molecular structure is then embedded and optimized before extracting atomic symbols and coordinates. A bounding box and unit cell are defined, which is subsequently expanded into a supercell and saved as a CIF file. The unit cell parameters are extracted, and atomic features such as atom count and average atomic number are computed. The Atom-Aware MLP model is defined, and a decision is made to either load a pre-trained model or create a new one. The model is then used to predict dielectric constants, store the predictions, and visualize the generated structures.

geometries optimized at T = 0 K, with no explicit thermal expansion or finite-temperature effects. Therefore, the reported dielectric constants do not include temperature-dependent phenomena such as increased molecular mobility, thermally activated defects, or changes in phonon (lattice vibrational) occupations that would arise at higher temperatures. Therefore, while a polymer's dielectric constant at room temperature and at a certain applied frequency can differ from the static-limit computed value, knowing the latter is extremely valuable for screening since it provides a baseline measure of a polymer's intrinsic polarizability.

Each CIF file in the dataset includes detailed atomistic information, unit cell parameters, and metadata describing the material's composition. The files contain extra descriptors in the form of labels, source, or classification notes. Our script iterates over all files in directory and attempts to parse each file into an ASE (atomic simulation environment) object [45].

2.1.1. Unit cell parameters

For each valid CIF file, the code extracts the unit cell parameters from the CIF text (Cell lengths, Cell angles, and the Unit cell volume). If any of these parameters are missing from the CIF text, they are computed directly from the ASE Atoms object.

2.1.2. Atomic features and tokenization

The script reads the atomic positions and symbols to determine the number of atoms in the structure and the average atomic number of these atoms. Both quantities are appended to the list of aggregated features, giving a total feature vector of length nine. In addition, while reading each CIF, the script retrieves atom site labels (C1, C2, etc) and their corresponding atom types. Each atom is then represented by a "token" combining the type and its label (C_C1, C_22, etc).

2.1.3. Metadata

We also scan each CIF file for the lines containing "# Label" (compound label), "# Source" (string describing data origin), and "# Class" (a string describing a classification or grouping, when provided). These metadata fields are stored alongside the extracted features and dielectric constant. All entries are aggregated into lists of 9-dimensional feature tensors, target dielectric constants, compound identifiers, per-sample atom tokens, and additional metadata from each CIF file.

2.1.4. Outliers

After reading each CIF file, the script searches for the "Dielectric constant, total" field in the text. Any structure whose dielectric constant exceeded a selected threshold (9.0 in this work) is excluded. In addition, we also use a z-score threshold (2.5) for any remaining extreme values. Specifically, the script calculates the mean and standard deviation of the dielectric constants and removes any samples that deviate by more than the selected threshold. The final dataset includes nearly 900 samples. These steps help removing any potential outliers that might unduly skew model training.

2.1.5. Vocabulary construction

The atomic tokens extracted from all CIF files are used to build a vocabulary (mapping each unique token to an integer index). This is done by iterating over every token and assigning a unique index to each distinct token. The resulting vocabulary is used to convert atom tokens into embedding indices during model training.

2.2. Modelling

We used a simple fully connected network with two hidden layers to process the aggregated features and the atomic token embeddings. This network was compared with a Random Forest (RF) approach, which we will show its able to outperform the previous (both in accuracy and computational resources). The pipeline has been constructed as follows:

2.2.1. Inputs

We implemented a custom class that stores the aggregated 9-dimensional feature vectors, the target labels (dielectric constants), the compound labels (string identifiers), and the raw atom tokens. Within the method, if a vocabulary has been set, each sample's list of tokens is converted to a tensor of integer indices. This enables straightforward embedding lookups in the subsequent neural network. Then, a custom function is used to handle the variable-length token lists. The function returns a batch tensor of features, a batch tensor of labels, a list of compound label strings, and a list of token index tensors. Therefore, this design allows accommodating varying numbers of atoms in each structure without forcing truncation or padding. Before training, the dataset is randomly split into training and validation subsets (this random split ensures an unbiased evaluation of the model during training).

2.2.2. Embedding layer

Each unique atom token is mapped to a learnable embedding vector (default dimension 16). For each sample, the embeddings of all tokens (atoms) are averaged to form a single vector representing that sample's composition and local environment features. The averaged embedding vector is concatenated with the 9-dimensional aggregated feature vector, yielding a combined input.

2.2.3. Fully connected layers (neural network only)

The concatenated vector is passed through two hidden layers (each of size 32 by default) using Leaky ReLU activations (which in turn avoid dead neurons). The final linear layer outputs a single prediction for the dielectric constant.

2.2.4. Training and hyperparameter optimization

We trained several models using mean squared error (MSE) as loss function. We also employed a grid hyperparameter optimization using Adam optimizer, learning rates ranging from 0.001 to 0.0001, batch sizes ranging from 16 to 512 samples, and dropout probability of either 0 or 0.1. During each training, we calculated the validation loss at the end of each epoch, and employed an early stopping mechanism for halting training if validation loss does not improve (by at least a value $\delta = 10^{-5}$) for a set number of epochs (50–300 depending on learning rate and batch size). When triggered, the model reverted to the best-performing weights seen so far. At the conclusion of training, we plot both the training and validation loss curves across epochs to diagnose overfitting or underfitting and to verify the stability of training. A similar procedure was implemented for the RF models [46], using a grid search optimization on the number of estimators and max depth.

2.3. Extension to SMILES-based inputs

While our primary workflow relies on CIF inputs, our approach can be extended to handle SMILES (Simplified Molecular-Input Line-Entry System[47]) representations of polymer or monomer units. This extension allows users who may not have crystallographic data on hand to provide a SMILES string and still obtain a dielectric constant prediction. Below, we detail the steps involved:

2.3.1. SMILES parsing and 3D geometry

We used RDKit to parse SMILES strings and to generate an initial 3D geometry. Then, a simple force field optimization was employed for refining the geometry, producing 3D Cartesian coordinates for each atom in the monomer.

2.3.2. Conversion to an ASE atoms object

Once the 3D structure of the monomer is computed, we extracted each atom's element symbol and position from the RDKit conformer. These atomic symbols and coordinates were then used to create an ase. Atoms object (this step bridges the SMILES-centric RDKit representation with the ASE-based workflow).

2.3.3. Definition of a periodic unit cell

Unlike small molecules that require no periodic extension, polymeric materials typically span a repeating unit structure. To approximate polymer properties, our method computes the bounding box of the monomer and adds a pre-defined margin (typically 3 Å) along each dimension. This enlarged bounding box is used to define a 3D cell. For a more realistic representation of the repeating polymer environment, we can replicate the initial cell via an expansion matrix. For instance, a $2 \times 2 \times 2$ supercell can be built using make_supercell in ASE, or by manually extending one dimension to simulate a chain-like environment. These manipulations ensure that the final structure better reflects the structure of polymers. The margin size and expansion matrix can be tuned to capture a more realistic polymer environment (a small margin or single replication may underestimate inter-chain effects, while very large expansions can significantly increase computational load).

2.3.4. CIF generation and feature extraction

The resulting 3D structure was written to a temporary CIF file using ASE's write method. This file was immediately reprocessed with the same feature-extraction code we applied to standard CIF inputs. Thus, unit cell parameters (a,b,c, α , β , γ), volume, number of atoms, and average atomic number are automatically extracted, hence generating a 9-dimensional feature vector suitable for our models.

2.3.5. Integration with the existing models

After generating the features, the new sample is passed through our model. A predicted dielectric constant is computed in precisely the same manner as for data originating from CIF files.

This approach offers a flexible interface in which users can supply either standard CIF files (for well-characterized polymers) or SMILES strings (for less-characterized polymers or any hypothetical designs). The method can be therefore adapted to other polymer classes, such as branched polymers, or co-polymers, provided that an appropriate representation exists.

3. Results and discussion

In this section we present and discuss the key findings of our study. We start by showing the results of ANN and random forest training processes, and follow by analysing their corresponding strengths and weaknesses. We then present the extension of the method to other samples by using a second dataset, comprised of compounds whose both dielectric constant and SMILES strings are available. Finally, and with the aim of providing a deeper interpretation of the models' behaviour, we focus on the chemical structures of the samples with the largest deviations.

3.1. ANN and RF models

We trained several ANN models using a grid hyperparameter optimization approach (see methods section and Section 1 in supplementary information file). Fig. 2 shows hexbin plots comparing predicted versus true dielectric constant values for both training (a) and validation (b) sets for a single ANN model. The colour intensity represents the density of data points, with darker regions indicating higher data concentrations, and the solid line indicates the Predicted = True values. The plots show a strong linear trend along the diagonal, although some scatter samples are observed for higher dielectric constant values (starting at approximately 5-6), where the ANN model tends to systematically underestimate the target values. In terms of average metrics, training achieved a mean average percentage error (MAPE) of 9.1 % and a root median squared error (RMSE) of 0.57, while validation achieved MAPE = 10.8 % and RMSE = 0.67. It is important to stress that these results were obtained with only the above-mentioned structural information as input, without needing to rely on any experimental measurements like polarizability or other potentially related properties. However, and



Fig. 2. Hexbin plots comparing predicted versus true dielectric constant values. (a) Training set results, showing a strong correlation along the diagonal, indicating good model fit. (b) Validation set results, where the model maintains reasonable predictive performance but with slightly increased scatter. The color intensity represents the density of data points, with darker regions indicating higher data concentrations. The solid line represents the predicted = true values.

despite the observed overall metrics, it is also important to acknowledge that the ANN models present the above-mentioned biases at large dielectric constant values that render them nearly useless in this range. This issue has been observed throughout all trained ANN models (as shown in the hexbin plots in Figure SI 1), which evidences a systematic deviation of the approach that will be addressed below.

On the other hand, RF models were able to achieve strong performances with only a fraction of the computational resources. In this case we followed a similar training procedure, for which more information (grid hyperparameter optimization and the selected configuration) can be found in Section 2 of the supplementary information. Fig. 3 shows the predicted versus true dielectric constant results in the validation set for the selected model (see shadowed row in Table SI 2 and the corresponding scatter plot in Figure SI 2), which present a MAPE= 9.63 %, R^2 = 0.691, and RMSE= 0.583.

As reflected by the lower obtained overall deviations, in this case the previous bias is not nearly as pronounced, and even more extreme samples are reasonably well predicted (see that the concentration of values above 6–7 appear relatively closer to the diagonal than in the previous approach). This suggests that the extracted features, i.e. the computed unit cell parameters and atomic composition-based descriptors, effectively capture the dominant factors influencing permittivity (thereby favouring decision-tree-based algorithms that rely on well-defined feature importance [46]).

Despite the strong predictive capability demonstrated, some limitations should also be acknowledged in this approach. On the one hand, the available information in the dataset is mainly composed by mediumpermittivity polymers, with comparatively few high- and lowpermittivity materials (which are in turn essential for industrial applications). This imbalance might lead to reduced prediction reliability in certain regions of permittivity space (like in the case of very high permittivity values for the ANN models). On the other hand, since the dataset is derived from crystallographic information files, the approximations made for more amorphous polymers (which are prevalent in all practical applications), could present slightly larger deviations and biases. As a general trend, the larger deviations tend to cluster around a small set of chemically atypical backbones that are under-represented in the training distribution. Our inspection of the highest-deviations indicates three main recurring chemical structures: samples with organometallic main chains (Sn, Ti), sample with highly fluorinated and chlorinated segments (PVDF-\beta, perfluoro-acetone alt-ethylene), and some samples with π -conjugated sulphur heterocycles linked through donor-acceptor spacers (C4H2S with CO/CS/NH). In the case of organometallic samples, classical atom-type fingerprints encode tin and titanium merely as "heavy atoms", so the distinctive d-orbital polarizability that drives the true permittivity is flattened, leading to conservative estimates. On the other hand, in case of highly fluorinated and chlorinated polymers, the strong C-F/C-Cl dipoles raise the dielectric constant, but the model's mixed atomic-charge descriptors were calibrated mainly on hydrogenated analogues; it therefore misses the extra dipolar



Fig. 3. Hexbin plot comparing predicted versus true dielectric constant values in the validation set results. The color intensity represents the density of data points, with darker regions indicating higher data concentrations. The solid line represents the predicted = true values.

contribution and tends to under-estimate fluoropolymers. Finally, in the third case the samples share an electron-rich C_4H_2S ring but differ in how strongly, and at what periodicity, electron-withdrawing (CO, CS) or electron-donating (NH) spacers are inserted. Their dielectric constant is extraordinarily sensitive to this subtle electronic push–pull balance.

However, an important strength of this work lies precisely in its robust performance when predicting dielectric constants in the low to medium-high permittivity range. This capability aligns well with the primary objective of our study: identifying foundational or "base" polymeric structures and their correlation with dielectric constants. Such structural insights can significantly guide the rational design and synthesis of new polymeric materials aimed at achieving enhanced dielectric properties. By systematically exploring and optimizing these "base" structures, future efforts can effectively develop novel polymers with elevated permittivity suitable for advanced applications in energy storage, microelectronics, and high-voltage insulation.

3.2. Extension to external data and SMILES-based predictions

To assess model robustness and to further increase its applicability. we extended our pipeline to use it for inputting any chemical representations in the form of SMILES strings. Among many others, this approach has been widely employed in the past in the study of the structure dependence of other properties, and has shown good accuracy, applicability and simplicity [48-50]. As shown in Fig. 4 for some selected examples, our external test dataset consists of experimentally characterized polymers with available SMILES representations (see also the chemical representations for each of the strings corresponding to their SMILES), which we convert into 3D structures assuming that the monomer is enclosed in a unit cell (see methods section for more details). When using these approximated inputs, and with the exception of a few examples that we analyse in detail below, the RF model exhibited a good agreement between the predicted and true values, thus reinforcing the hypothesis that many critical permittivity features can be captured from structural and composition information alone.

Fig. 5 shows the scatter plot of the obtained prediction result where, to further interpret the model's performance, we examined the polymers with the highest deviations between predicted and true dielectric constants. In addition, a summary of external test samples with their individual metrics is also presented in the Table SI 3. These cases provide valuable insights into the strengths and limitations of our feature selection approach. Hence, several trends among the polymers with the largest deviations can be observed:

3.3. Underestimated dielectric constant in highly polarizable systems

Poly (methyl α -chloroacrylate) exhibited a large deviation (17.90 %), with a predicted dielectric constant of 2.79 compared to the true value of 3.40 (see label a in Fig. 5). One possible rationalization for this result is related to the presence of the highly electronegative chlorine atoms, which contribute to strong local dipole interactions and polarization effects. The model likely underestimates these contributions, therefore indicating that either additional molecular descriptors capturing halogen-related polarization effects, or more chlorinated data samples during training could improve accuracy. On the other hand, nylon 6 also showed significant underestimation (12.82 %, label b in Fig. 5). This polymer displays strong intermolecular hydrogen bonding and chain alignment effects, which influence permittivity apparently beyond what is captured by our unit cell descriptors alone. As in the previous case, this suggests that future models could benefit from including features related to hydrogen bonding potential, chain mobility, and molecular mobility.

3.4. Overestimated dielectric constant

Poly (diallyl phenyl phosphonate) and Poly (diallyl phthalate) both

S. Sharifi et al.

Next Materials 8 (2025) 100795



Fig. 4. SMILES, 2D chemical structure and approximated 3D structure for selected compounds in the test set.



Fig. 5. Scatter plot of predicted vs true dielectric constant values. The dashed line represents the predicted = true values, and the shadowed area represents a 0.5 absolute deviation range. Labels a to g represent the samples with the largest observed deiviations: a) poly (methyl α -chloroacrylate), b) nylon 6, c) PMMA, d) poly (vinyl acetate), e)nylon 12, f) nylon 11 and g) poly (ether ketone).

presented overestimated values (positive deviations of 8.72 % and 8.96 %, respectively). These materials contain rigid phenyl rings, which can reduce chain mobility and polarization response, leading to lower effective dielectric constants than the model predicted. Additionally, it is important to consider the density factor of polymeric materials, particularly when interpreting prediction deviations. A reduced density in such systems can decrease the dipolar density per unit volume,

thereby significantly lowering the observed dielectric constant. Consequently, even in systems where dipole mobility is relatively high, a lower density of dipoles per unit volume can notably diminish the dielectric response, leading to larger discrepancies between predicted and actual permittivity values.

Polycarbonate (3.80 % deviation) is another case where the model overestimated the dielectric constant. Nylon-based polymers with

longer carbon chains also presented overestimated dielectric constants (nylon 11 and nylon 12). In particular, nylon 11 was predicted 3.87 (true dielectric constant 3.30), and nylon 12 was predicted 3.74 (true 3.60). While nylons possess strong intermolecular hydrogen bonding networks that contribute to dielectric response through interchain interactions, in these cases these are diluted by the longer carbon chains when compared to the shorter nylon 6. Furthermore, it is worth noting increasing complexity arising from the odd-even effect in polyamides, where the number of methylene groups significantly impacts their physical and dielectric properties [51]. In our dataset, the model slightly overestimates the dielectric constant for these longer-chain nylons. Other chemistries, like Poly(amide-imide), showed minimal deviation (0.72 %) suggesting that its combination of aromatic and amide functionalities is well captured by the model. Other examples include Poly (p-hydroxybenzoate) and Poly (hexamethylene adipamide), that also showed small deviations (3.63 % and 5.62 %, respectively), thus reinforcing the hypothesis that hydrogen bonding induced variability is reasonably captured.

As a final remark, it is important to emphasize that the DFT-derived labels used for training predominantly capture the electronic contribution to the static permittivity at 0 K. This is justified by the fact that, at such low temperatures, the glass-transition temperatures of most polymers considered here result in minimal contributions (i.e., the α -relaxation is largely frozen). Additionally, transport processes such as ionic conductivity and interfacial phenomena (e.g., Maxwell-Wagner-Sillars effects) are significantly suppressed due to the absence of these longsegmental motions. On the other hand, if present, β -relaxations ($\Delta \epsilon \beta$) would represent the primary thermally activated correction. However, the activation energies typically associated with these relaxations hinder their occurrence at such low temperatures. Furthermore, the reported β-relaxation strengths for common engineering polymers generally range from 0.1 to 0.8 (rarely exceeding 1.0) [52] which is comparable to the uncertainty of the present model. Values above 1.0 could be associated with special cases of high-dielectric polymer materials that have been excluded from the training process of the model. Therefore, the adopted training-validation strategy remains valid for exploring structure-property relationships and yields meaningful performance estimates for chemical representation inputs (e.g., SMILES).

These insights not only enhance our understanding of the structure-property relationships governing polymer dielectric behaviour but also provide clear directions for improving the predictive approach in future iterations.

4. Conclusions

This study demonstrates the effectiveness of machine learning-based approaches for predicting the dielectric constant of polymers using only structural descriptors. By employing both artificial neural networks and random forest models, we also explored the predictive capabilities of different architectures in terms of both accuracy and resource economy. In the conditions employed in our study, the obtained results indicated that the RF model outperforms the ANN, particularly in capturing extreme dielectric constant values while maintaining lower computational costs. The ANN model exhibited a systematic underestimation bias at high dielectric constants, likely due to the limited representation of high-permittivity polymers in the training set. The generalization of the approach to SMILES-based predictions further demonstrated the flexibility of the model, allowing for dielectric constant estimations even when other data is unavailable. The developed computational approach provides a valuable tool for accelerating the identification of nextgeneration dielectric polymers for applications in energy storage, microelectronics, and high-voltage insulation technologies.

Declaration of Competing Interest

relationships which may be considered as potential competing interests: L. A. Miccio reports equipment, drugs, or supplies was provided by NVIDIA Corp. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU used for this research.

Associated content

Additional information on the ANN and RF grid optimization, and test results can be found in the supplementary information file.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.nxmate.2025.100795.

Data availability

The data that supports the findings of this study are available within the article, at https://github.com/lamiccio/Polymer_CIFs.git and in the supplementary information file.

References

- H. Li, Y. Zhou, Y. Liu, L. Li, Y. Liu, et al., Dielectric polymers for High-Temperature capacitive energy storage, Chem. Soc. Rev. (2021), https://doi.org/10.1039/ d0cs00265i
- [2] Q. Li, L. Chen, M.R. Gadinski, S. Zhang, et al., Flexible High-Temperature dielectric materials from polymer nanocomposites, Nature (2015), https://doi.org/10.1038/ nature14647.
- [3] X. Wu, X. Chen, Q.M. Zhang, D.Q. Tan, et al., Advanced dielectric polymers for energy storage, Energy Storage Mater. (2021), https://doi.org/10.1016/j. ensm.2021.10.010.
- [4] L. Zhu, Exploring strategies for high dielectric constant and low loss polymer dielectrics, J. Phys. Chem. Lett. (2014), https://doi.org/10.1021/jz501831q.
- [5] H. Luo, X. Zhou, C. Ellingford, Y. Zhang, et al., Interface design for high energy density polymer nanocomposites, Chem. Soc. Rev. (2019), https://doi.org/ 10.1039/c9cs00043g.
- [6] D. Ai, H. Li, Y. Zhou, L. Ren, et al., Tuning nanofillers in In situ prepared polyimide nanocomposites for High-Temperature capacitive energy storage, Adv. Energy Mater. (2020), https://doi.org/10.1002/aenm.201903881.
- [7] L. Li, C. Jingsai, C. Yunyun, H. Ting, L. Yang, Z. Yao, Z. Guanghui, Z. Yan, X. C, D. Lijie, et al., Significant improvements in dielectric constant and energy density of ferroelectric polymer nanocomposites enabled by ultralow contents of n anofillers, Adv. Mater. (2021), https://doi.org/10.1002/adma.202102392.
- [8] Y. Thakur, T. Zhang, C. Iacob, T. Yang, J. Bernholc, et al., Enhancement of the dielectric response in polymer nanocomposites with low dielectric constant fillers, Nanoscale (2017), https://doi.org/10.1039/c7nr01932g.
- [9] Prateek, V.K. Thakur, R. Gupta, Recent progress on ferroelectric polymer-based nanocomposites for high energy density capacitors: synthesis, dielectric properties, and future aspects, Chem. Rev. (2016), https://doi.org/10.1021/acs. chemrev.5b00495.
- [10] A.F. Baldwin, R. Ma, T.D. Huan, Y. Cao, R. Ramprasad, G.A. Sotzing, Effect of incorporating aromatic and chiral groups on the dielectric p roperties of Poly (Dimethyltin Esters), Macromol. Rapid Commun. 35 (2014) 2082–2088, https:// doi.org/10.1002/marc.201400507.
- [11] E. Baer, L. Zhu, 50th anniversary perspective: dielectric phenomena in polymers and multilayered dielectric films, Macromolecules 50 (2017) 2239–2256, https:// doi.org/10.1021/acs.macromol.6b02669.
- [12] Q.-K. Feng, S.-L. Zhong, J.-Y. Pei, Y. Zhao, D.-L. Zhang, D.-F. Liu, Y.-X. Zhang, Z.-M. Dang, Recent progress and future prospects on All-Organic polymer dielectrics for energy storage capacitors, Chem. Rev. 122 (2022) 3820–3878, https://doi.org/ 10.1021/acs.chemrev.1c00793.
- [13] Y. Ningning, C. Jianxiong, R. Fuhao, D. Xiaorong, H. Lijun, M. Jie, Designing highenergy-density all-organic flexible dielectric films fr om molecular structure to polymer cross-linking networks, J. Power Sources (2024), https://doi.org/ 10.1016/j.jpowsour.2024.234265.
- [14] H. Jing, M. Irene, M. Madhubanti, S. Harikrishna, K. Ashish, Y. O, R. R, S. Gregory, C. Yang, High-Temperature High-k polyolefin by rational molecular design, Proc. Natl. Acad. Sci. (2024), https://doi.org/10.1073/pnas.2415388121.
- [15] Y. Thakur, M. Lin, S. Wu, Z. Cheng, D.-Y. Jeong, Q.M. Zhang, Tailoring the dipole properties in dielectric polymers to realize high energy density with high

The authors declare the following financial interests/personal

S. Sharifi et al.

breakdown strength and low dielectric loss, J. Appl. Phys. 117 (2015), https://doi.org/10.1063/1.4915942.

- [16] R. Ma, A.F. Baldwin, C. Wang, I. Offenbach, M. Cakmak, R. Ramprasad, G. A. Sotzing, Rationally designed polyimides for High-Energy density capacitor appli cations, ACS Appl. Mater. Interfaces 6 (2014) 10445–10451, https://doi.org/ 10.1021/am502002v.
- [17] S. Liang, Z. Fengyuan, L. Li, L. Jiajie, D. Jiufeng, P. Zizhao, N. Yujuan, C. Jiaxin, L. Yuqi, L. Yani, et al., Superior capacitive energy storage enabled by molecularly interpenetra ting interfaces in layered polymers, Adv. Mater. (2024), https://doi. org/10.1002/adma.202412561.
- [18] Q. Li, F. Yao, Y. Liu, G. Zhang, et al., High-temperature dielectric materials for electrical energy storage, Annu. Rev. Mater. Res. (2018), https://doi.org/10.1146/ annurev-matsci-070317-124435.
- [19] Z. Hu, X. Liu, T. Ren, H.A.M. Saeed, Q. Wang, et al., Research progress of low dielectric constant polymer materials, J. Polym. Eng. (2022), https://doi.org/ 10.1515/polyeng-2021-0338.
- [20] S. Wang, C. Yang, X. Li, H. Jia, S. Liu, X. Liu, T. Minari, Q. Sun, Polymer-Based dielectrics with high permittivity and low dielectric loss for flexible electronics, J. Mater. Chem. C. 10 (2022) 6196–6221, https://doi.org/10.1039/D2TC00193D.
- [21] Y. Cheng, Q.-S. Ji, G. Zhang, X. Zhang, Z. Liu, H. Gong, Z. Zhang, Synergistic enhancement of dielectric polymers through fluorine incorporation for improved energy storage, reduced loss, and enhanced processability, Adv. Funct. Mater. (2024), https://doi.org/10.1002/adfm.202406219.
- [22] B. Chu, X. Zhou, K. Ren, B. Neese, M. Lin, et al., A dielectric polymer with high electric energy density and fast discharge speed, Science 80 (2006), https://doi. org/10.1126/science.1127798.
- [23] L.A. Miccio, M.M. Kummali, P.E. Montemartini, P.A. Oyanguren, G.A. Schwartz, N. Alegra, J. Colmenero, Determining concentration depth profiles in fluorinated networks by means of electric force microscopy, J. Chem. Phys. 135 (2011), https://doi.org/10.1063/1.3624574.
- [24] L.A. Miccio, J. Otegui, M.E. Penoff, P.E. Montemartini, G.A. Schwartz, Fluorinated networks dynamics studied by means of broadband dielectric spectroscopy, J. Appl. Polym. Sci. 132 (2015), https://doi.org/10.1002/app.42690.
- [25] J.-B. Ping, Q.-K. Feng, Y.-X. Zhang, X.-J. Wang, L. Huang, S.-L. Zhong, Z.-M. Dang, A bilayer High-Temperature dielectric film with superior breakdown strength and energy storage density, NanoMicro Lett. 15 (2023) 154, https://doi.org/10.1007/ s40820-023-01121-6.
- [26] D. Ghosh, D. Khastgir, Degradation and stability of polymeric High-voltage insulators and prediction of their service life through environmental and accelerated aging processes, ACS Omega 3 (2018) 11317–11330, https://doi.org/ 10.1021/acsomega.8b01560.
- [27] J.-G. Gao, L.-W. Liu, W.-F. Sun, Dielectric characteristics of crosslinked polyethylene modified by grafting Polar-Group molecules, Polym. (Basel) 15 (2023), https://doi.org/10.3390/polym15010231.
- [28] A. Nazrin, T.M. Kuan, D.-E.A. Mansour, R.A. Farade, A.M. Ariffin, M.S.A. Rahman, N.I.Bin Abdul Wahab, Innovative approaches for augmenting dielectric properties in Cross-Linked polyethylene (XLPE): a review, Heliyon 10 (2024), https://doi. org/10.1016/j.heliyon.2024.e34737.
- [29] Mecheri, Y.; Boukezzi, L.; Boubakeur, A.; Lallouani, M. Dielectric and Mechanical Behavior of Cross-Linked Polyethylene under Thermal Aging. In Proceedings of the 2000 Annual Report Conference on Electrical Insulation and Dielectric Phenomena (Cat. No.00CH37132); 2000; Vol. 2, pp. 560–563 vol.2.
- [30] E. Linde, L. Verardi, D. Fabiani, U.W. Gedde, Dielectric spectroscopy as a condition monitoring technique for cable insulation based on crosslinked polyethylene, Polym. Test. 44 (2015) 135–142, https://doi.org/10.1016/j. polymertestine.2015.04.004.
- [31] M.-X. Zhu, T. Deng, L. Dong, J.-M. Chen, Z.-M. Dang, Review of machine Learning-Driven design of Polymer-Based dielectrics, IET Nanodielectrics 5 (2022) 24–38, https://doi.org/10.1049/nde2.12029.
- [32] M.J. Uddin, J. Fan, Interpretable machine learning framework to predict the glass transition temperature of polymers, 2024, 16, 1049, Polym 16 (2024) 1049, https://doi.org/10.3390/POLYM16081049.
- [33] L.A. Miccio, C. Borredon, G.A. Schwartz, A glimpse inside materials: polymer structure – glass transition temperature relationship as observed by a trained

artificial intelligence, Comput. Mater. Sci. 236 (2024) 112863, https://doi.org/ 10.1016/J.COMMATSCI.2024.112863.

- [34] J. Fiosina, P. Sievers, M. Drache, S. Beuermann, Polymer reaction engineering meets explainable machine learning, Comput. Chem. Eng. 177 (2023) 108356, https://doi.org/10.1016/j.compchemeng.2023.108356.
- [35] A. Tayyebi, A.S. Alshami, Z. Rabiei, X. Yu, N. Ismail, M.J. Talukder, J. Power, Prediction of organic compound aqueous solubility using machine learning: a comparison study of Descriptor-Based and Fingerprints-Based models, J. Chemin.. 15 (2023) 99, https://doi.org/10.1186/s13321-023-00752-6.
- [36] J.G. Ethier, R.K. Casukhela, J.J. Latimer, M.D. Jacobsen, B. Rasin, M.K. Gupta, L. A. Baldwin, R.A. Vaia, Predicting phase behavior of linear polymers in solution using machine learning, Macromolecules 55 (2022) 2691–2702, https://doi.org/10.1021/ACS.MACROMOL.2C00245.
- [37] C. Lihua, K. Chiho, B. Rohit, P.L. Jordan, W. Chao, L. Zongze, A.D. Ajinkya, W. Yifei, T. H, V. P, et al., Frequency-Dependent dielectric constant prediction of polymers using m achine learning, npj Comput. Mater. (2020), https://doi.org/ 10.1038/s41524-020-0333-6.
- [38] Zeng, M.; Kumar, J.N.; Zeng, Z.; Savitha, R.; Chandrasekhar, V.R.; Hippalgaonkar, K. Graph Convolutional Neural Networks for Polymers Property Prediction 2018.
- [39] C. Kuenneth, W. Schertzer, R. Ramprasad, Copolymer informatics with multitask deep neural networks, Macromolecules 54 (2021) 5957–5961, https://doi.org/ 10.1021/acs.macromol.1c00728.
- [40] J. Liang, J. Liang, J. Liang, J. Liang, S. Xu, S. Xu, S. Xu, L. Hu, L. Hu, X. Zhang, et al., Machine-Learning-assisted low dielectric constant polymer discovery, Mater. Chem. Front (2021), https://doi.org/10.1039/d0qm01093f.
- [41] Y. Shiraki, Y. Kawanami, K. Shinmei, H. Kaneko, Machine learning model for predicting dielectric constant of epoxy resin with additional data selection and design of monomer structures for low dielectric constant, ACS Appl. Polym. Mater. 7 (2025) 2809–2818, https://doi.org/10.1021/acsapm.4c03279.
- [42] C. Qian, K. Sun, W. Bao, Recent advance on machine learning of <scp>MXenes</ Scp> for energy storage and conversion, Int. J. Energy Res (2022), https://doi.org/ 10.1002/er.7833.
- [43] H. Lei, X. Li, J. Wang, Y. Song, G. Tian, et al., DFT and molecular dynamic simulation for the dielectric property analysis of polyimides, Chem. Phys. Lett. (2022), https://doi.org/10.1016/j.cplett.2021.139131.
- [44] T.D. Huan, A. Mannodi-Kanakkithodi, C. Kim, V. Sharma, G. Pilania, R. Ramprasad, A polymer dataset for accelerated property prediction and design, Sci. Data 3 (2016) 160012, https://doi.org/10.1038/sdata.2016.12.
- [45] A. Hjorth Larsen, J. Jørgen Mortensen, J. Blomqvist, I.E. Castelli, R. Christensen, M. Dułak, J. Friis, M.N. Groves, B. Hammer, C. Hargus, et al., The atomic simulation Environment-a python library for working with atoms, J. Phys. Condens. Matter 29 (2017) 273002, https://doi.org/10.1088/1361-648X/aa680e.
- [46] L. Breiman, L. Breiman, Random For. (2001), https://doi.org/10.1023/a: 1010933404324.
- [47] D. Weininger, SMILES, a chemical language and information system. 1. introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31–36, https://doi.org/10.1021/ci00057a005.
- [48] L.A. Miccio, Understanding polymers through transfer learning and explainable AI, Appl. Sci. 14 (2024), https://doi.org/10.3390/app142210413.
 [49] C. Borredon, L.A. Miccio, G.A. Schwartz, Transfer Learning-Driven artificial
- [49] C. Borredon, L.A. Miccio, G.A. Schwartz, Transfer Learning-Driven artificial intelligence model for glass transition temperature estimation of molecular glass formers mixtures, Comput. Mater. Sci. 238 (2024) 112931, https://doi.org/ 10.1016/J.COMMATSCI.2024.112931.
- [50] U.M. Casado, F.I. Altuna, L.A. Miccio, Towards sustainable material design: a comparative analysis of latent space representations in AI models, Sustainability 16 (2024).
- [51] R.A. Pérez-Camargo, J. Torres, A.J. Müller, Understanding Even-Odd effects in linear Semi-Crystalline polymers: polyethers, polycarbonates, polyesters, and polyamides, Polymers 324 (2025) 128233, https://doi.org/10.1016/j. polymer.2025.128233.
- [52] F. Kremer, A. Schönhals. A. Broadband Dielectric Spectroscopy, 2003. ISBN 9783540434078.