ACS APPLIED MATERIALS & INTERFACES

www.acsami.org

Forum Article

An Informatics Approach for Designing Conducting Polymers

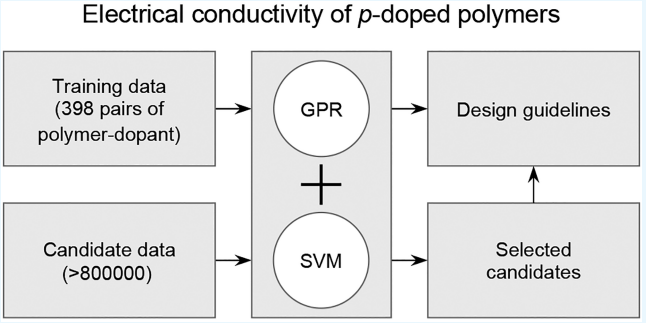
Harikrishna Sahu, Hongmo Li, Lihua Chen, Arunkumar Chitteth Rajan, Chiho Kim, Natalie Stingelin, and Rampi Ramprasad*

Cite This: https://doi.org/10.1021/acsami.1c04017

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Doping conjugated polymers, which are potential candidates for the next generation of organic electronics, is an effective strategy for manipulating their electrical conductivity. However, selecting a suitable polymer−dopant combination is exceptionally challenging because of the vastness of the chemical, configurational, and morphological spaces one needs to search. In this work, high-performance surrogate models, trained on available experimentally measured data, are developed to predict the p-type electrical conductivity and are used to screen a large candidate hypothetical data set of more than 800 000 polymer−dopant combinations. Promising candidates are identified for synthesis and device fabrication. Additionally, new design guidelines are extracted that verify and extend knowledge on important molecular fragments that correlate to high conductivity. Conductivity prediction models are also deployed at www.polymergenome.org for broader open-access community use.



KEYWORDS: *conducting polymer, organic electronics, design guidelines, virtual screening, machine learning*

1. INTRODUCTION

Conjugated organic polymers are garnering interest as potential candidates for a wide variety of uses, ranging from electronics to biomedical applications.[1−4] One attribute of importance in such applications is high electrical conductivity (σ), and the optimal value of σ can be realized with appropriate dopants mediating electron or hole conduction.[5−7] A few thiophene-based polymers exhibit conductivities up to 1×10^3 S/cm[8−10] and are expected to be useful as hole transport materials. Nevertheless, given the breadth of the polymer and dopant chemical spaces, identifying viable candidates using just prior experience, chemical insight, and ingenuity is nontrivial.

σ is proportional to the product of the hole/electron density and the charge-carrier mobility.[11,12] Doping generates charge carriers by extracting or transferring electrons to the frontier orbitals of the polymer, thus increasing the density of charge carriers. It can be accomplished using a wide range of dopants, including alkali metals, halides, small organic molecules, and metal complexes. The diverse chemical nature and size of these dopants lead to irregular and significant changes in the physical properties of the polymer as well as their solid-state microstructure.[10,13,14] In addition to the nature of the dopant, the chemical structure of the polymer significantly influences σ. It is negatively correlated with the reorganization energy associated with charge carriers and positively correlated with the electronic coupling between frontier orbitals of adjacent conjugated segments in the π-stacking direction.[15] It is worth noting that the electronic coupling strongly depends on the polymer packing and its general molecular arrangement.

Machine learning (ML) is a well-known approach for selecting and designing materials by inferring hidden correlations from previous measurements/calculations.[16−19] Such informatics and related computational efforts have contributed in the past to accelerating the rational design of application-specific polymers.[20−26] Within the context of electrical conductivity, a few attempts have been made to establish machine learning models for σ of inorganic materials[27] and properties related to σ for organic materials.[28−32] Notably, Atahan-Evrenk et al.[29] proposed a surrogate model for the intramolecular reorganization energy using graph- and geometry-based features. Models are also developed for predicting electron transfer coupling, which is a critical factor in determining electron transfer rates, to reduce computational cost.[30−32] Although these studies attempted to build models for properties related to the σ of polymers, none of them aimed to directly estimate σ.

In this work, we have developed ML models to predict σ of p-doped polymers using a data set of 398 polymer-dopant combinations measured experimentally at room temperature. The schematic workflow for this study is depicted in Figure 1.

Special Issue: Artificial Intelligence/Machine Learning for Design and Development of Applied Materials

Received: March 2, 2021
Accepted: May 17, 2021

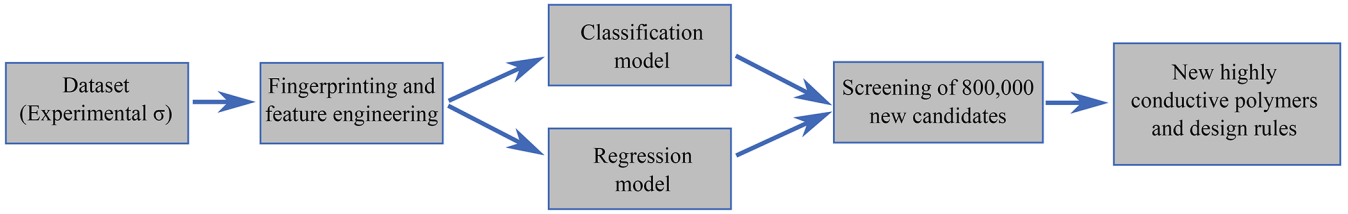


Figure 1. Machine-learning workflow for the electrical conductivity (σ) prediction and designing of doped polymers.

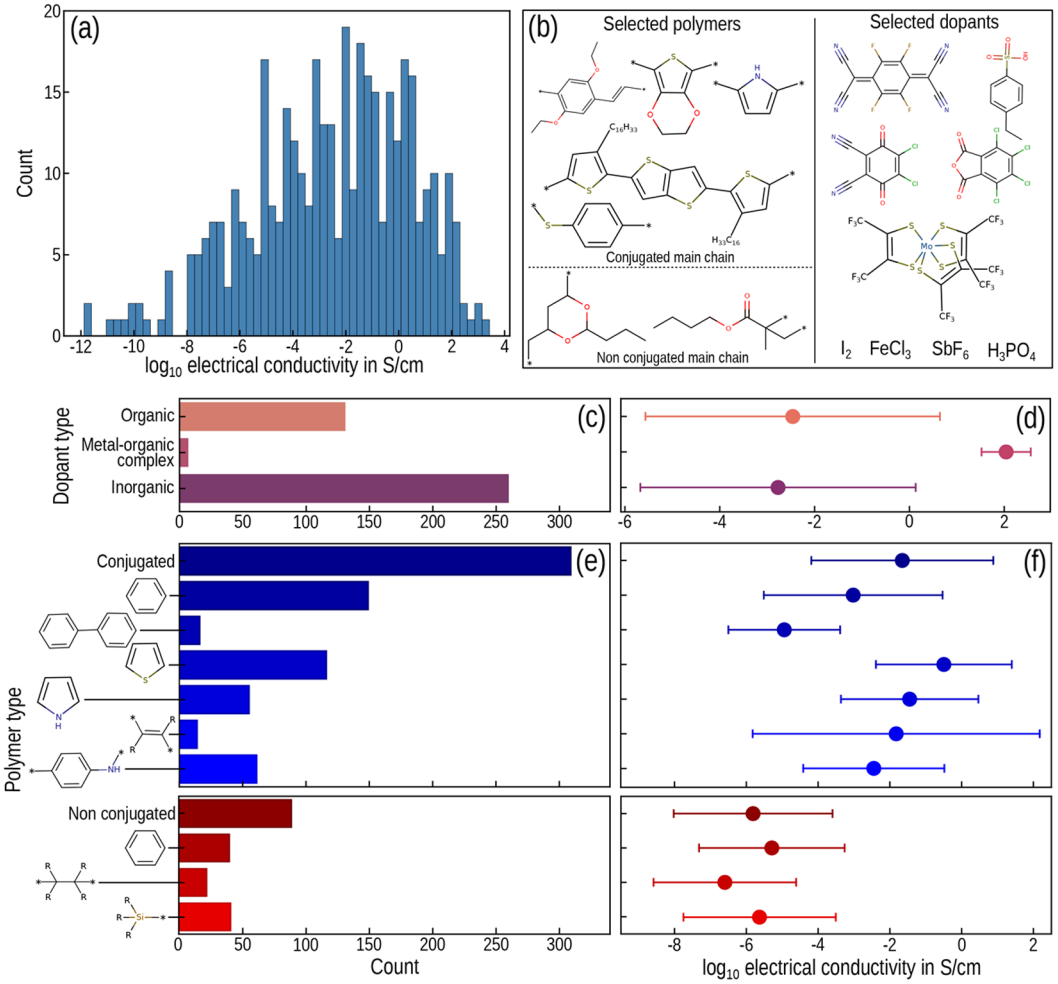


Figure 2. Experimental data set for the electrical conductivity (σ) of doped polymers. (a) Histogram for the experimental σ of 398 unique polymer−dopant combinations. (b) Chemical structures of exemplary polymers and dopants present in the data set. (c) Types of dopants and the number of samples for each case. (d) Average σ for each dopant type, along with the standard deviation as an error bar. (e) Types of polymers and the number of samples for each case. (f) Average σ for each polymer type, along with the standard deviation as an error bar.

Polymers and dopants are represented using chemical, electronic, and geometrical features by considering a three-level hierarchical fingerprinting scheme discussed previously.[33] Gaussian process regression (GPR) and support vector machine (SVM) algorithms were used to map the polymer−dopant representation with the corresponding σ values. The resulting ML models are able to accurately and rapidly predict σ of a new polymer−dopant combination. These predictive models were then used to screen more than 800 000 new candidates, and 500 promising ones are proposed for experimental verification. By analyzing models and screening results, important molecular fragments that control σ were identified, verifying and extending the list of known ones. Finally, we propose new guidelines for designing potential polymer−dopant combinations in terms of molecular fragments and properties.

2. DATA SET AND MACHINE LEARNING MODELS

2.1. Data Set. Our room-temperature electrical conductivity data set for p-type doping was collected from literature[10,34−55] and from our own laboratory. There were 226 different homopolymers and 65 distinct dopants in the data set, forming 398 distinct polymer−dopant combinations. If more than one measurement was available for a combination, the mean value of all measurements was considered. A histogram of the experimental data set shown in Figure 2a

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

indicates that there are an adequate number (88) of samples with high σ (>10^0 S/cm). The data set covered a wide variety of chemical structures of polymers as well as dopants. Polymers and dopants were constructed from {C, H, N, Br, Cl, P, O, Si, S, F} and {C, H, N, Sb, B, F, P, Mg, Fe, Cu, S, Cl, Br, Li, I, O, Na, Mo, Zn} sets of elements, respectively. The chemical structures of a few representative polymers and dopants are depicted in Figure 2b. Figure 2c,d shows that various types of dopants, such as small organic molecules, inorganic molecules, and metal−organic complexes, are in the dopant set, and the average conductivity of polymers doped with metal−organic complexes was relatively high in comparison to other dopants. As shown in Figure 2e, in addition to 309 π-conjugated polymer-dopant blends, there are 89 systems comprising nonconjugated polymers. In conjugated polymers, benzene and thiophene rings were overrepresented compared to other conjugated segments. Nonconjugated polymers were mostly silicon-based polymers, alkanes including polyethylene derivatives as well as materials comprising benzene moieties in their backbone. Figure 2f shows the average conductivity value and the corresponding standard deviation for each polymer type. As shown in the figure, thiophene-based conjugated polymers have higher σ compared to others. It is worth emphasizing that although nonconjugated polymers are known to have low σ, they are included in the data set to extend the range of conductivity and polymer space, which is required for efficient training of ML models. Additionally, this will enable our models for screening low conducting polymers, a critical factor for dielectric materials.

2.2. Fingerprints and Feature Engineering. The numerical representation of polymers and dopants is the first step toward building a high-performing machine-learning model. The key features of polymers were captured by three hierarchical levels of descriptors, i.e., (1) atomic-level fragments, (2) block-level fragments, and (3) chain-level features.[33] Dopants are small molecules and include both organic and inorganic compounds, and there is a need to consider both geometrical and electronic aspects in the feature space. Therefore, in addition to geometrical fingerprints, maximum and minimum partial charge, van der Waals surface area (VSA) descriptors, and Kier and Hall's EState Descriptors[56] were also calculated for dopants, as implemented in the RDKit package.[57] Out of a total of 456 features obtained for each sample, important ones were identified by several feature selection approaches such as forward selection and the least absolute shrinkage and selection operator (LASSO).[58] The forward selection approach was started by selecting one feature and appending the most important one to the selected featured list during each iteration. In LASSO, coefficients of less important features shrink to zero by adding a penalty term to the loss function, and they were removed from a model. The down-selected feature representations of polymer−dopant candidates along with their respective experimental conductivity were used for model training.

2.3. Machine-Learning Models. Classification and regression are the two primary classes of supervised machine learning problems. Classification is a process of finding a suitable mapping function from the input variables to categorical output variables. However, in a regression model, a mapping function is established to predict continuous output variables. As the working principles of these two approaches are different and a distinct set of features are selected for each case, we applied both strategies to build surrogate models.

We trained a regression model using the GPR algorithm combined with a radial basis function (RBF) as a kernel. There are potentially many functions that can fit the training data, and GPR finds a distribution of functions, in which the mean of this probability distribution is the most probable prediction. For the classification model, the data set was divided into three classes based on the values of σ. Samples with conductivity value less than 1×10^{-4} S/cm and higher than 1×10^{-1} S/cm were assigned to Low and High classes, respectively. The remaining samples were assigned to the Medium class. The surrogate models were trained using SVM combined with RBF as a kernel. In SVM, hyperplanes are defined in such a way that they best divide a data set into classes.

Models were trained by varying sampling techniques. First, models were built by removing a particular type of polymer from the training set (Figure S1), and then the performance of each model was tested for the removed polymer type. Second, the model performance was tested for a polymer type by adding 50% polymers of the same kind to the training set. As the presence of a polymer type leads to improved performance, we considered the random and stratified random samplings for the regression and classification models, respectively. The hyperparameters of the models were optimized by following a 5-fold cross-validated grid-search approach. As mentioned in section 2.2, we used different feature selection techniques to identify important ones. Learning curves were generated for each case to check the performance and stability of the models and are depicted in Figures S2 and S3. The forward selection approach and LASSO led to the best results for classification and regression models, respectively.

3. RESULTS AND DISCUSSION

First, the performance of each model was verified, and obtained errors were compared with expected experimental deviations. Second, the models were used for screening new candidates, and promising ones were selected for experimental verification. Third, models and screening results were analyzed to extract useful guidelines for designing highly conducting organic materials.

3.1. Model Performance and Validation. Each electrical conductivity model was trained using 90% of the data set, and the rest of the data set was used for testing. The performance of the classification model is reported in Figure 3a in terms of precision, recall, and f1-score values. The confusion matrix is shown in Figure S4, which is a table with different combinations of predicted and actual values. For train and test sets, the f1 scores of the High class were 0.82 and 0.81 with precision values of 0.73 and 0.76, respectively.

The performance of the regression model was estimated for an unseen data set by two metrics: (a) root-mean-square error (RMSE) and (b) the coefficient of determination (R^2). The parity plot for predicted and true conductivity values is depicted in Figure 3b. The R^2 for the train and test sets were 0.86 and 0.81, respectively. The RMSE for the train (test) set was 1.12 (1.37) S/cm, mostly originated from experimental anomalies that arise from a change in processing method, measurement technique, solvent, etc. To compare experimental uncertainties with model predicted errors, we considered 175 unique polymer/dopant combinations with multiple reported σ, a total of 1269 measurements. A histogram for deviations from the respective mean for a polymer/dopant combination is depicted in Figure S5. It appears that the measured uncertainties follow a normal distribution with a

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

(a)

	Precision	Recall	F1 Score
Low	0.95 (0.70)	0.82 (0.88)	0.88 (0.78)
Medium	0.79 (0.85)	0.71 (0.65)	0.74 (0.73)
High	0.73 (0.76)	0.92 (0.87)	0.82 (0.81)

(b)

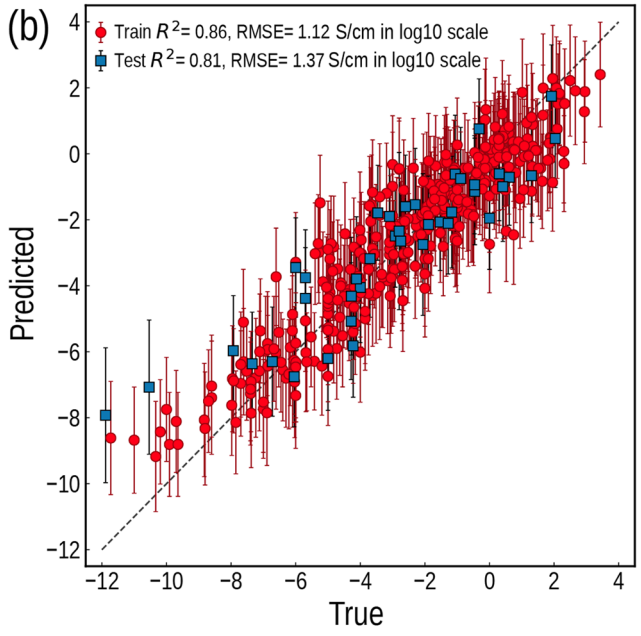


Figure 3. Performance of machine learning models for electrical conductivity (σ). (a) Precision, recall, and f1-score values of each class for the classification model. Values outside and inside the parentheses are for train and test sets, respectively. (b) Parity plot for the regression model. Conductivity values are provided in $\log_{10}$ scale. The uncertainty associated with each data point is shown as an error bar. For both the models, 90% of data was used for training, and 10% was used as a test data set.

standard deviation of 1.53 $\log_{10}$ σ (S/cm) with a peak at the mean value. Thus, intrinsic uncertainties associated with conductivity measurement are close to the model RMSE value, indicating that our model is capable of capturing the intrinsic noise associated with σ measurements around the data set. It is worth noting that to minimize the effect of experimental variation, the feature space must contain experimental conditions that govern the solid-state microstructure of materials.

Even though the prediction of conductivity was quite challenging for doped polymers, the models showed satisfactory results that can be used for further analysis. These models are deployed on the Polymer Genome platform (www.polymergenome.org) to provide easy access to the polymer community. The user can predict the conductivity at room temperature for any polymer combined with 20 selected dopants present in the training set.

3.2. Screening of Candidate Materials. We established a candidate set of unique polymer−dopant combinations, different than those present in the training set. The chemical identities of 12 904 previously synthesized homopolymers were collected from the literature.[22,59−61] Each of these polymers was combined with 65 dopants present in the training set, and a data set of 838 760 candidates formed our candidate set. Note that synthesis routes for both the polymer and the dopant are well-known; however, the electrical conductivity values of the polymer−dopant combinations are unknown.

All relevant features were computed for the candidates. To get an overview of the chemical space, we performed principal component analysis (PCA) of chemical features of both candidate and training sets. The scatter plot of the first two components (Figure 4a) indicates that the training set covered only a portion of the candidate space, and there was an opportunity to discover suitable candidates in the unexplored chemical space.

Candidates were screened using both regression and classification models. They were considered highly conducting materials if the predicted class was High by the classification model and predicted $\sigma > 10^{\circ}$ S/cm by the regression model. Note that polymer-dopant combinations were rejected if they are associated with high uncertainty ($>1 \times 10^2$ S/cm) in predicting σ. These conditions led to selections of a total of 4197 candidates. As shown in Figure 4a, selected candidates expanded the polymer space of the training set. Table S1 provides a list of 500 most-suitable polymer−dopant combinations, involving 192 unique polymers and 22 dopants, with predicted σ and associated uncertainty. These cases are proposed for synthesis and device fabrication.

A few promising candidates are shown in Figure 4b, and among them, the predicted conductivities of poly(3-cyanothiophene) and poly(3-heptafluoropropyl thiophene) in combination with $(Mo(tfd)_3)$ and $Mo(tfd\text{-}COCF_3)_3$, respectively, are encouraging ($\sim 1 \times 10^4$ S/cm). Figure 4c depicts the change in average σ of a polymer type when combined with 6 selected dopants, providing guidance for selecting suitable polymer−dopant combinations. The minimum and maximum values for each polymer type and dopant combination are also provided. As can be seen, for a specific dopant, thiophene-based polymers often show higher σ compared to other polymer types. Polyphenylamine derivatives are far superior to benzene-based polymers as lone pair electrons contribute to the π-conjugation. For several polymers, the average σ is high when they are blended with Mo complexes, probably due to high doping efficiency and increased polaron delocalization.[10]

3.3. Guidelines for Designing Organic Materials. Design guidelines provide insights and heuristics for rational pathways to designing polymers that meet target property requirements. In this work, we propose a set of design rules for choosing molecular fragments that correlate to high electrical conductivity.

3.3.1. SHAP Analysis. The Shapley additive explanations (SHAP) approach has recently received considerable attention due to its profound ability to explain statistical models.[62−64] It evaluates the importance of having a particular value for a feature compared to the prediction for that feature's baseline value. In this work, SHAP values were approximately computed with the kernel explainer for both classification and regression models. Feature importance and density plots for classification and regression models are depicted in Figures S9 and S10, respectively. SHAP analysis of the classification model renders more useful insights as it explains feature importance for the desired class (i.e., High class) compared to the overall performance in the regression model. The density

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

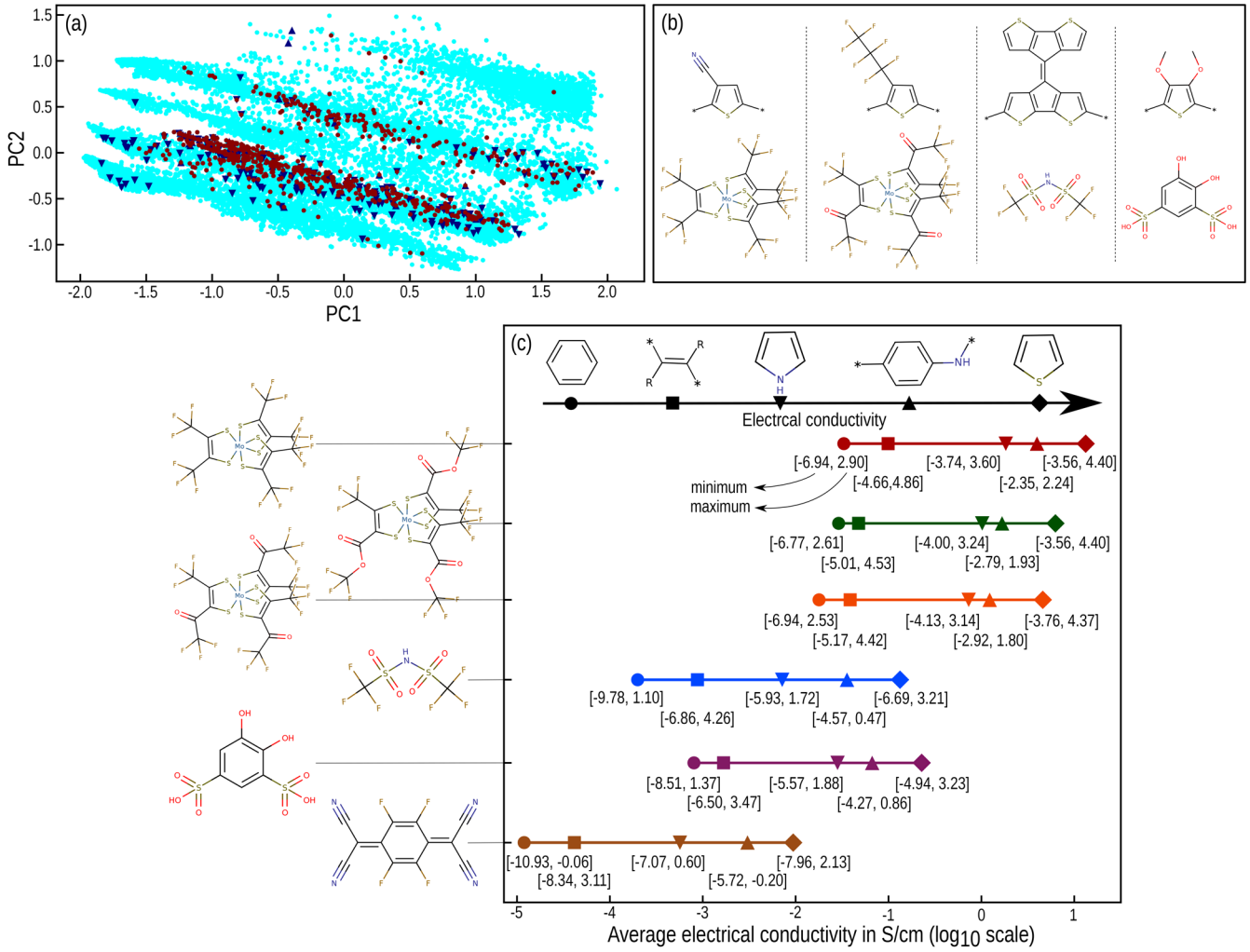


Figure 4. Chemical space of training, candidate and selected polymer data sets, and preferred polymer−dopant combinations. (a) Chemical space of polymers in the training set (blue; triangle up if $\sigma > 10^{\circ}$ S/cm, otherwise triangle down), 12 904 candidate polymers (cyan rounds), and 663 unique polymers present in 4197 selected candidates (dark red rounds) illustrated using the first two principal components (PC1 and PC2). (b) Chemical structures of a few polymers and dopants selected in the machine-learning assisted virtual-screening. (c) The average σ of different polymer types combined with a few selected dopants. The minimum and maximum predicted values of σ for a particular polymer type and dopant combination are also provided in a square bracket.

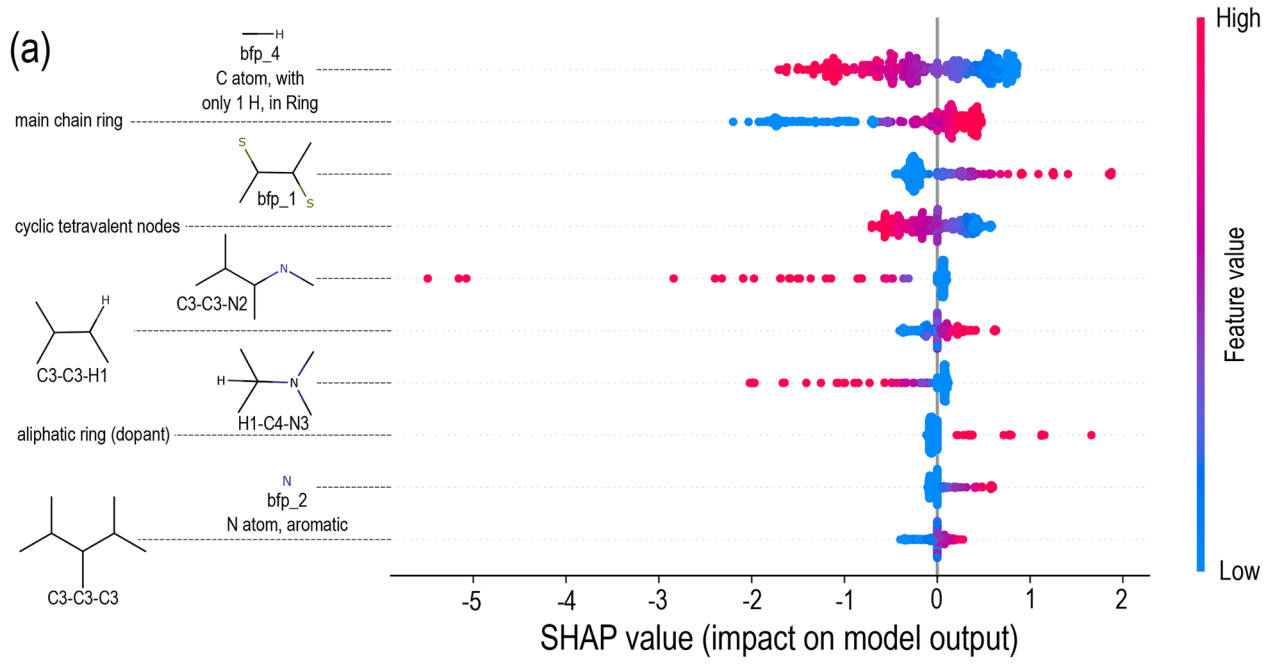
plot for the most important features of the classification model is shown in Figure 5a. Characteristic features for determining σ were extracted from SHAP analysis of both the models; they are depicted in Figure 5b.

As shown in the figure, the number of non-hydrogen ring atoms in the main chain (denoted as *main_chain_ring*) and carbon atoms with only 1 H in a ring (*bfp_4*) are two crucial descriptors for both the models. As expected, an increase in *main_chain_ring* of polymers leads to a rise in σ. The presence of adjacent thiophene/pyrrole rings are captured by *bfp_1*/*bfp_5*, and has a positive impact on conductivity. However, the presence of a benzene ring negatively affects hole conduction as it is often associated with a large dihedral angle in the main chain, leading to a weak π-conjugation.[65,66] A rise in the *percentage of non-hydrogen atoms* for a polymer chain indicates an increase in the degree of unsaturation and number of heteroatoms, and thus it is positively correlated with σ. *Number of cyclic tetravalent nodes in a repeat unit*, *N3−C4−H1*, *C4−C4−H1*, and *C3−C3−N2* are negatively correlated as they block π-conjugation, and a low value is preferred. In dopants, an increase in the number of aliphatic rings improves σ, probably because of enhanced solubility. Overall, a feature is positively correlated with the conductivity if it extends π-conjugation or improves solubility, and vice versa.

3.3.2. Importance of Molecular Fragments. For designing conducting polymers, it is worth estimating the importance of molecular fragments commonly used in the polymer research community. The *z*-score is often used as a numerical representation of the fragment importance that describes the relationship between a fragment's occurrence to the average, measured in standard deviation units.[67−69] The details of the *z*-score calculation are provided in the Supporting Information.

In this work, *z*-scores for building blocks of polymers were computed, ignoring the presence of dopants. For *z*-score calculation, a total of 12 976 unique polymers present in train and candidate sets were considered. Out of them, a subset of 680 polymers were labeled as highly conducting polymers by combining good candidates obtained from high-throughput screening (section 3.2) and samples with conductivity values $>10^{0}$ S/cm in the training set. A database of fragments is

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX



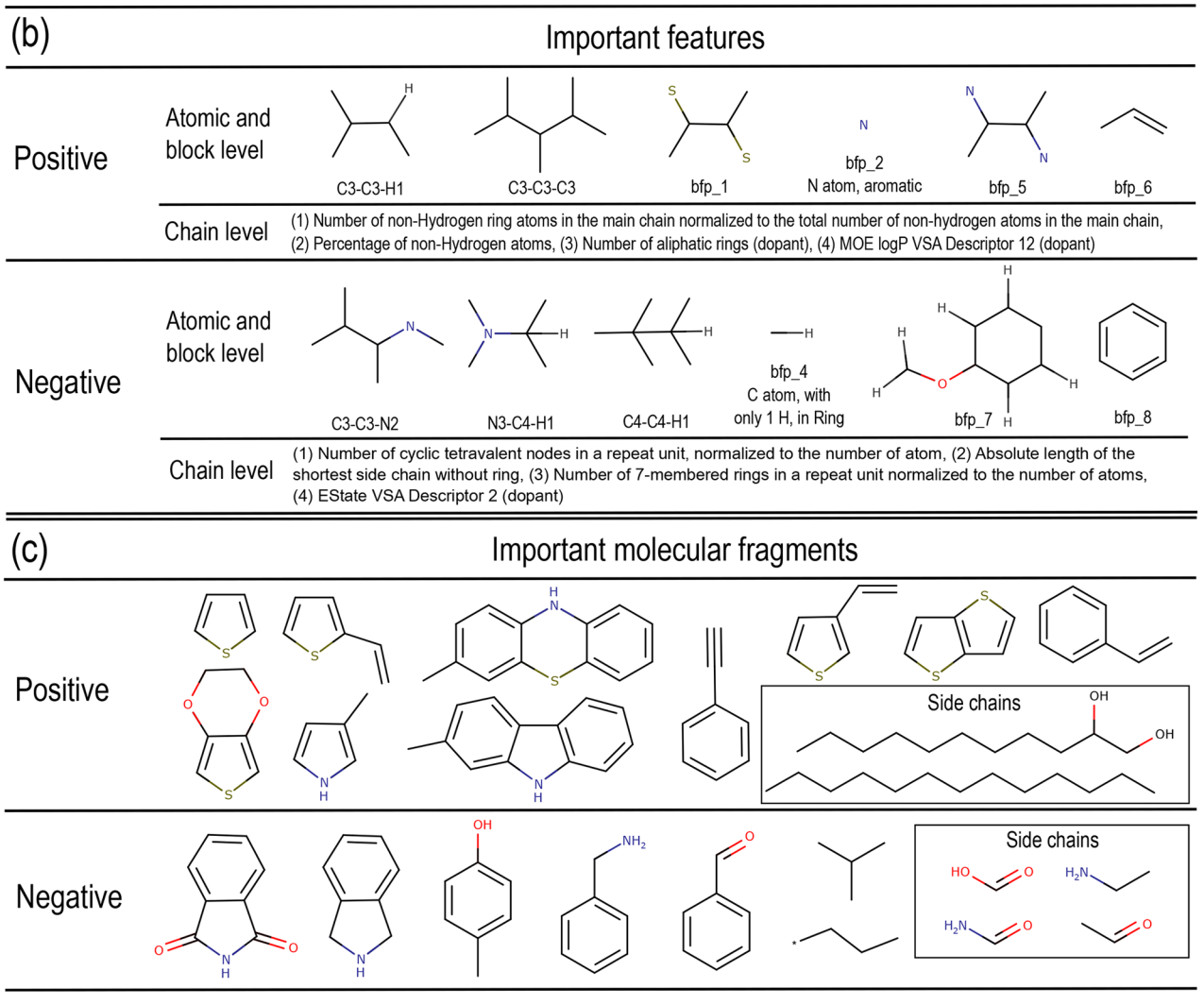


Figure 5. Key molecular fragments and properties controlling electrical conductivity (σ). (a) Density plots of the top contributing descriptors determined by the mean absolute SHAP values for classifying High class in classification model, (b) crucial features obtained from the Shapley values, and (c) molecular fragments with z-score higher and lower than 10.

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

established from the polymer data set, using the breaking of retrosynthetically interesting chemical substructures (BRICS) fragmentation[70] as implemented in the RDKit package. Using this method, each polymer was decomposed into such synthetically meaningful fragments, and the resulting list was reduced to a set of 3331 unique fragments (after fingerprint comparison). Note that in the z-score calculation, only one fragment is considered at a time, ignoring the presence of other moieties in a polymer. When a molecular fragment is combined with a wide variety of moieties in a sufficiently large number of polymers, the impact of other fragments is minimized, offering a reliable score for the desired one.

We computed the z-score of each fragment, and the important ones are shown in Figure 5c. As can be seen, thiophene and its derivatives are mostly overrepresented in highly conducting polymers, along with pyrrole and other conjugated systems. Phenylacetylene and styrene moieties favor high σ in comparison to benzene and benzaldehyde building blocks. A good z-score of long aliphatic side chains indicates their importance for solubility and materials processing. The z-scores of amide and carboxyl functional groups suggest that their presence is not favorable for high conductivity. The presence of a tertiary carbon atom in a polymer-chain decreases π-conjugation, leading to a negative z-score. An extended list of fragments with their z-scores is provided in the Tables S2 and S3. In short, to design efficient conducting polymers, we strongly recommend incorporating fragments with positive z-scores while avoiding negative ones.

4. CONCLUSION

We have developed reliable classification and regression models to predict the electrical conductivity of p-doped polymers at room temperature using support vector machine (SVM) and Gaussian process regression (GPR) algorithms, respectively. These models were then used to screen more than 800 000 candidates, and 500 promising polymer−dopant combinations were selected for experimental verification. Additionally, key molecular fragments in determining σ were identified. In future work, we plan to extend the feature space further to cover the effect of solid-state microstructure and environmental conditions such as temperature, processing, and measurement techniques.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsami.1c04017.

> Additional information on model optimization and validation, including learning curves, parity plots, and confusion matrix; predicted electrical conductivity of promising candidates; correlations between features and electrical conductivity; important molecular fragments that control the electrical conductivity (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Rampi Ramprasad − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0003-4630-1565; Email: rampi.ramprasad@mse.gatech.edu

Authors

Harikrishna Sahu − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0001-5458-9488

Hongmo Li − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Lihua Chen − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0002-9852-8211

Arunkumar Chitteth Rajan − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Chiho Kim − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Natalie Stingelin − *Department of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsami.1c04017

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the National Science Foundation (Grant 1729737).

■ REFERENCES

(1) Prunet, G.; Pawula, F.; Fleury, G.; Cloutet, E.; Robinson, A.; Hadziioannou, G.; Pakdel, A. A Review on Conductive Polymers and their Hybrids for Flexible and Wearable Thermoelectric Applications. *Mater. Today Phys.* **2021**, *18*, 100402.

(2) Lee, Y.; Park, J.; Son, J.; Woo, H. Y.; Kwak, J. Degenerately Doped Semi-Crystalline Polymers for High Performance Thermoelectrics. *Adv. Funct. Mater.* **2021**, *31*, 2006900.

(3) Guo, X.; Facchetti, A. The Journey of Conducting Polymers from Discovery to Application. *Nat. Mater.* **2020**, *19*, 922−928.

(4) Wang, Y.; Yu, Y.; Liao, H.; Zhou, Y.; McCulloch, I.; Yue, W. The Chemistry and Applications of Heteroisoindigo Units as Enabling Links for Semiconducting Materials. *Acc. Chem. Res.* **2020**, *53*, 2855−2868.

(5) Lu, Y.; Yu, Z.-D.; Liu, Y.; Ding, Y.-F.; Yang, C.-Y.; Yao, Z.-F.; Wang, Z.-Y.; You, H.-Y.; Cheng, X.-F.; Tang, B.; Wang, J.-Y.; Pei, J. The Critical Role of Dopant Cations in Electrical Conductivity and Thermoelectric Performance of n-Doped Polymers. *J. Am. Chem. Soc.* **2020**, *142*, 15340−15348.

(6) Zhao, W.; Ding, J.; Zou, Y.; Di, C.-a.; Zhu, D. Chemical Doping of Organic Semiconductors for Thermoelectric Applications. *Chem. Soc. Rev.* **2020**, *49*, 7210−7228.

(7) Ma, Z.; Shi, W.; Yan, K.; Pan, L.; Yu, G. Doping Engineering of Conductive Polymer Hydrogels and their Application in Advanced Sensor Technologies. *Chem. Sci.* **2019**, *10*, 6232−6244.

(8) Neusser, D.; Malacrida, C.; Kern, M.; Gross, Y. M.; van Slageren, J.; Ludwigs, S. High Conductivities of Disordered P3HT Films by an Electrochemical Doping Strategy. *Chem. Mater.* **2020**, *32*, 6003−6013.

(9) Ferchichi, K.; Bourguiga, R.; Lmimouni, K.; Pecqueur, S. Concentration-Control in All-Solution Processed Semiconducting Polymer Doping and High Conductivity Performances. *Synth. Met.* **2020**, *262*, 116352.

(10) Liang, Z.; Zhang, Y.; Souri, M.; Luo, X.; Boehm, A.; Li, R.; Zhang, Y.; Wang, T.; Kim, D.-Y.; Mei, J.; Marder, S. R.; Graham, K. R.

Influence of Dopant Size and Electron Affinity on the Electrical Conductivity and Thermoelectric Properties of a Series of Conjugated Polymers. *J. Mater. Chem. A* **2018**, *6*, 16495−16505.
(11) Russ, B.; Glaudell, A.; Urban, J. J.; Chabinyc, M. L.; Segalman, R. A. Organic Thermoelectric Materials for Energy Harvesting and Temperature Control. *Nat. Rev. Mater.* **2016**, *1*, 16050.
(12) Kroon, R.; Mengistie, D. A.; Kiefer, D.; Hynynen, J.; Ryan, J. D.; Yu, L.; Müller, C. Thermoelectric Plastics: From Design to Synthesis, Processing and Structure-Property Relationships. *Chem. Soc. Rev.* **2016**, *45*, 6147−6164.
(13) Haneef, H. F.; Zeidell, A. M.; Jurchescu, O. D. Charge Carrier Traps in Organic Semiconductors: A Review on the Underlying Physics and Impact on Electronic Devices. *J. Mater. Chem. C* **2020**, *8*, 759−787.
(14) Yee, P. Y.; Scholes, D. T.; Schwartz, B. J.; Tolbert, S. H. Dopant-Induced Ordering of Amorphous Regions in Regiorandom P3HT. *J. Phys. Chem. Lett.* **2019**, *10*, 4929−4934.
(15) Oberhofer, H.; Reuter, K.; Blumberger, J. Charge Transport in Molecular Materials: An Assessment of Computational Methods. *Chem. Rev.* **2017**, *117*, 10319−10357.
(16) Batra, R.; Song, L.; Ramprasad, R. Emerging Materials Intelligence Ecosystems Propelled by Machine Learning. *Nat. Rev. Mater.* **2020**, DOI: 10.1038/s41578-020-00255-y.
(17) Cole, J. M. A Design-to-Device Pipeline for Data-Driven Materials Discovery. *Acc. Chem. Res.* **2020**, *53*, 599−610.
(18) Roch, L. M.; Saikin, S. K.; Häse, F.; Friederich, P.; Goldsmith, R. H.; León, S.; Aspuru-Guzik, A. From Absorption Spectra to Charge Transfer in Nanoaggregates of Oligomers with Machine Learning. *ACS Nano* **2020**, *14*, 6589−6598.
(19) Sahu, H.; Rao, W.; Troisi, A.; Ma, H. Toward Predicting Efficiency of Organic Solar Cells via Machine Learning and Improved Descriptors. *Adv. Energy Mater.* **2018**, *8*, 1801032.
(20) Chen, L.; Pilania, G.; Batra, R.; Huan, T. D.; Kim, C.; Kuenneth, C.; Ramprasad, R. Polymer Informatics: Current Status and Critical Next Steps. *Mater. Sci. Eng., R* **2021**, *144*, 100595.
(21) Chandrasekaran, A.; Kim, C.; Ramprasad, R. *Machine Learning Meets Quantum Physics*; Springer International Publishing, 2020.
(22) Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122*, 17575−17585.
(23) Mannodi-Kanakkithodi, A.; Chandrasekaran, A.; Kim, C.; Huan, T. D.; Pilania, G.; Botu, V.; Ramprasad, R. Scoping the Polymer Genome: A Roadmap for Rational Polymer Dielectrics Design and Beyond. *Mater. Today* **2018**, *21*, 785−796.
(24) Ma, R.; Sharma, V.; Baldwin, A. F.; Tefferi, M.; Offenbach, I.; Cakmak, M.; Weiss, R.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Rational Design and Synthesis of Polythioureas as Capacitor Dielectrics. *J. Mater. Chem. A* **2015**, *3*, 14845−14852.
(25) Huzayyin, A. A.; Boggs, S. A.; Ramprasad, R. Density Functional Analysis of Chemical Impurities in Dielectric Polyethylene. *IEEE Trans. Dielectr. Electr. Insul.* **2010**, *17*, 926−930.
(26) Baldwin, A. F.; Huan, T. D.; Ma, R.; Mannodi-Kanakkithodi, A.; Tefferi, M.; Katz, N.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Rational Design of Organotin Polyesters. *Macromolecules* **2015**, *48*, 2422−2428.
(27) Juneja, R.; Singh, A. K. Unraveling the Role of Bonding Chemistry in Connecting Electronic and Thermal Transport by Machine Learning. *J. Mater. Chem. A* **2020**, *8*, 8716−8721.
(28) Krämer, M.; Dohmen, P. M.; Xie, W.; Holub, D.; Christensen, A. S.; Elstner, M. Charge and Exciton Transfer Simulations Using Machine-Learned Hamiltonians. *J. Chem. Theory Comput.* **2020**, *16*, 4061−4070.
(29) Atahan-Evrenk, S.; Atalay, F. B. Prediction of Intramolecular Reorganization Energy Using Machine Learning. *J. Phys. Chem. A* **2019**, *123*, 7855−7863.
(30) Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine Learning for Predicting Electron Transfer Coupling. *J. Phys. Chem. A* **2019**, *123*, 7792−7802.
(31) Lederer, J.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine Learning-Based Charge Transport Computation for Pentacene. *Adv. Theory Simul.* **2019**, *2*, 1800136.
(32) Miller, E. D.; Jones, M. L.; Henry, M. M.; Stanfill, B.; Jankowski, E. Machine Learning Predictions of Electronic Couplings for Charge Transport Calculations of P3HT. *AIChE J.* **2019**, *65*, e16760.
(33) Doan Tran, H.; Kim, C.; Chen, L.; Chandrasekaran, A.; Batra, R.; Venkatram, S.; Kamal, D.; Lightstone, J. P.; Gurnani, R.; Shetty, P.; Ramprasad, M.; Laws, J.; Shelton, M.; Ramprasad, R. Machine-Learning Predictions of Polymer Properties with Polymer Genome. *J. Appl. Phys.* **2020**, *128*, 171104.
(34) Kiefer, D.; Kroon, R.; Hofmann, A. I.; Sun, H.; Liu, X.; Giovannitti, A.; Stegerer, D.; Cano, A.; Hynynen, J.; Yu, L.; Zhang, Y.; Nai, D.; Harrelson, T. F.; Sommer, M.; Moulé, A. J.; Kemerink, M.; Marder, S. R.; McCulloch, I.; Fahlman, M.; Fabiano, S.; Müller, C. Double Doping of Conjugated Polymers with Monomer Molecular Dopants. *Nat. Mater.* **2019**, *18*, 149−155.
(35) Hofmann, A. I.; Kroon, R.; Yu, L.; Müller, C. Highly Stable Doping of a Polar Polythiophene Through Co-processing With Sulfonic Acids and Bistriflimide. *J. Mater. Chem. C* **2018**, *6*, 6905−6910.
(36) Kroon, R.; Kiefer, D.; Stegerer, D.; Yu, L.; Sommer, M.; Müller, C. Polar Side Chains Enhance Processability, Electrical Conductivity, and Thermal Stability of a Molecularly *p*-Doped Polythiophene. *Adv. Mater.* **2017**, *29*, 1700930.
(37) Jacobs, I. E.; Aasen, E. W.; Oliveira, J. L.; Fonseca, T. N.; Roehling, J. D.; Li, J.; Zhang, G.; Augustine, M. P.; Mascal, M.; Moulé, A. J. Comparison of Solution-Mixed and Sequentially Processed P3HT:F4TCNQ Films: Effect of Doping-Induced Aggregation on Film Morphology. *J. Mater. Chem. C* **2016**, *4*, 3454−3466.
(38) Glaudell, A. M.; Cochran, J. E.; Patel, S. N.; Chabinyc, M. L. Impact of the Doping Method on Conductivity and Thermopower in Semiconducting Polythiophenes. *Adv. Energy Mater.* **2015**, *5*, 1401072.
(39) Méndez, H.; Heimel, G.; Opitz, A.; Sauer, K.; Barkowski, P.; Oehzelt, M.; Soeda, J.; Okamoto, T.; Takeya, J.; Arlin, J.-B.; Balandier, J.-Y.; Geerts, Y.; Koch, N.; Salzmann, I. Doping of Organic Semiconductors: Impact of Dopant Strength and Electronic Coupling. *Angew. Chem., Int. Ed.* **2013**, *52*, 7751−7755.
(40) Singh, A.; Chandra, A. Graphite Oxide/β-Ni$(OH)_2$ Composites for Application in Supercapacitors. *AIP Conf. Proc.* **2012**, *1538*, 253−256.
(41) Goto, T.; Awano, H.; Takahashi, T.; Yonetake, K.; Sukumaran, S. K. Effect of Processing Temperature on Thermal Doping of Polyaniline Without Shear. *Polym. Adv. Technol.* **2011**, *22*, 1286−1291.
(42) Madl, C. M.; Kariuki, P. N.; Gendron, J.; Piper, L. F. J.; Jones, W. E. Vapor Phase Polymerization of Poly(3,4-ethylenedioxythiophene) on Flexible Substrates for Enhanced Transparent Electrodes. *Synth. Met.* **2011**, *161*, 1159−1165.
(43) Bodugöz-Sentürk, H.; Güven, O. Enhancement of Conductivity in Polyaniline-[Poly(Vinylidene Chloride)-co-(Vinyl Acetate)] Blends by Irradiation. *Radiat. Phys. Chem.* **2011**, *80*, 153−158.
(44) Singh, R.; Singh, R. K.; Kumar, J.; Kant, R.; Kumar, V. The Origin of DC Electrical Conduction and Dielectric Relaxation in Pristine and Doped Poly(3-hexylthiophene) Films. *J. Polym. Sci., Part B: Polym. Phys.* **2010**, *48*, 1047−1053.
(45) Somboonsub, B.; Srisuwan, S.; Invernale, M. A.; Thongyai, S.; Praserthdam, P.; Scola, D. A.; Sotzing, G. A. Comparison of the Thermally Stable Conducting Polymers PEDOT, PANi, and PPy using Sulfonated Poly(imide) Templates. *Polymer* **2010**, *51*, 4472−4476.
(46) Yamamoto, T.; Usui, M.; Ootsuka, H.; Iijima, T.; Fukumoto, H.; Sakai, Y.; Aramaki, S.; Yamamoto, H. M.; Yagi, T.; Tajima, H.; Okada, T.; Fukuda, T.; Emoto, A.; Ushijima, H.; Hasegawa, M.; Ohtsu, H. π-Conjugated Polymers Consisting of Isothianaphthene and Dialkoxy-p-phenylene Units: Synthesis, Self-Assembly, and

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX

Chemical and Physical Properties. *Macromol. Chem. Phys.* **2010**, *211*, 2138−2147.
(47) Massoumi, B.; Aghili, H.; Entezami, A. Investigation of Electrochemical Copolymerization of 1-Naphthylamineaniline in the Presence of Various Organic Sulfonic Acids. *J. Chin. Chem. Soc.* **2009**, *56*, 741−747.
(48) Ak, M.; Toppare, L. Synthesis of Star-shaped Pyrrole and Thiophene Functionalized Monomers and Optoelectrochemical Properties of Corresponding Copolymers. *Mater. Chem. Phys.* **2009**, *114*, 789−794.
(49) Li, Y.; Wang, J.; Tang, J.; Liu, Y.; He, Y. Conductive Performances of Solid Polymer Electrolyte Films based on PVB/$LiClO_4$ Plasticized by PEG200, PEG400 and PEG600. *J. Power Sources* **2009**, *187*, 305−311.
(50) Idris, N. K.; Aziz, N. A. N.; Ramli, S.; Isa, M. I. N.; Senin, H. B.; Carini, G.; Abdullah, J. B.; Bradley, D. A. Chitosan Based Film Electrolytes Doped Oleic Acid: An Electrical Study. *AIP Conf. Proc.* **2007**, *1017*, 315−320.
(51) Ayad, M. M.; Rehab, A. F.; El-Hallag, I. S.; Amer, W. A. Preparation and Characterization of Polyaniline Films in the Presence of N-Phenyl-1,4-Phenylenediamine. *Eur. Polym. J.* **2007**, *43*, 2540−2549.
(52) Kumar, A.; Hussain, A. 120MeV Si_9+ Ion Irradiation Effects on Poly(3-methylthiophene) Conducting Polymer. *Nucl. Instrum. Methods Phys. Res., Sect. B* **2006**, *251*, 451−456.
(53) Ou, R.; Cui, G.; Gerhardt, R. A.; Samuels, R. J. Effect of Stretching on the Structure and Electrical Conductivity of Doped and Undoped Poly(phenylene vinylene) Thin Films. *Electrochim. Acta* **2006**, *51*, 1728−1735.
(54) Galal, A. Characterization of Conducting Poly(3-methylthiophene) Films Prepared Under Sono-Electrochemical Conditions. *J. Appl. Polym. Sci.* **2006**, *102*, 2416−2425.
(55) Xu, J.-C.; Liu, W.-M.; Li, H.-L. Titanium Dioxide Doped Polyaniline. *Mater. Sci. Eng., C* **2005**, *25*, 444−447.
(56) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press, 1999.
(57) *RDKit: Cheminformatics and Machine Learning Software*; Open-source, 2018. http://www.rdkit.org/.
(58) Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Statist. Soc. B* **1996**, *58*, 267−288.
(59) Huan, T. D.; Mannodi-Kanakkithodi, A.; Kim, C.; Sharma, V.; Pilania, G.; Ramprasad, R. A Polymer Dataset for Accelerated Property Prediction and Design. *Sci. Data* **2016**, *3*, 160012.
(60) Bicerano, J. *Prediction of Polymer Properties*; CRC Press, 2002.
(61) Mark, J. *Polymer Data Handbook*; Oxford University Press, 1999.
(62) Yang, Z.-Y.; Dong, J.; Yang, Z.-J.; Lu, A.-P.; Hou, T.-J.; Cao, D.-S. Structural Analysis and Identification of False Positive Hits in Luciferase-Based Assays. *J. Chem. Inf. Model.* **2020**, *60*, 2031−2043.
(63) Xie, Y. R.; Castro, D. C.; Bell, S. E.; Rubakhin, S. S.; Sweedler, J. V. Single-Cell Classification Using Mass Spectrometry through Interpretable Machine Learning. *Anal. Chem.* **2020**, *92*, 9338−9347.
(64) Ye, W.-L.; Shen, C.; Xiong, G.-L.; Ding, J.-J.; Lu, A.-P.; Hou, T.-J.; Cao, D.-S. Improving Docking-Based Virtual Screening Ability by Integrating Multiple Energy Auxiliary Terms from Molecular Docking Scoring. *J. Chem. Inf. Model.* **2020**, *60*, 4216−4230.
(65) Dharmapurikar, S. S.; Chithiravel, S.; Mane, M. V.; Deshmukh, G.; Krishnamoorthy, K. Dihedral Angle Control to Improve the Charge Transport Properties of Conjugated Polymers in Organic Field Effect Transistors. *Chem. Phys. Lett.* **2018**, *695*, 51−58.
(66) Sahu, H.; Shukla, R.; Goswami, J.; Gaur, P.; Panda, A. N. Alternating Phenylene and Furan/Pyrrole/Thiophene Units-Based Oligomers: A Computational Study of the Structures and Optoelectronic Properties. *Chem. Phys. Lett.* **2018**, *692*, 152−159.
(67) Sahu, H.; Yang, F.; Ye, X.; Ma, J.; Fang, W.; Ma, H. Designing Promising Molecules for Organic Solar Cells via Machine Learning Assisted Virtual Screening. *J. Mater. Chem. A* **2019**, *7*, 17480−17488.
(68) Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A. L.; Blood-Forsythe, M. A.; Seress, L. R.; Román-Salgado, C.; Trepte, K.; Atahan-Evrenk, S.; Er, S.; Shrestha, S.; Mondal, R.; Sokolov, A.; Bao, Z.; Aspuru-Guzik, A. Lead Candidates for High-Performance Organic Photovoltaics from High-Throughput Quantum Chemistry − the Harvard Clean Energy Project. *Energy Environ. Sci.* **2014**, *7*, 698−704.
(69) Lopez, S. A.; Sanchez-Lengeling, B.; de Goes Soares, J.; Aspuru-Guzik, A. Design Principles and Top Non-Fullerene Acceptor Candidates for Organic Photovoltaics. *Joule* **2017**, *1*, 857−870.
(70) Degen, J.; Wegscheid-Gerlach, C.; Zaliani, A.; Rarey, M. On the Art of Compiling and Using 'Drug-Like' Chemical Fragment Spaces. *ChemMedChem* **2008**, *3*, 1503−1507.

https://doi.org/10.1021/acsami.1c04017
ACS Appl. Mater. Interfaces XXXX, XXX, XXX−XXX