Nano World Journal

https://doi.org/10.17756/nwj.2023-s5-015

Review Article

Open Access

Artificial Intelligence in Nanomaterials and Polymers: A Detailed Review Approach on Nano-informatics, Polymer Informatics, and Informational Materials Science

Neelesh Ashok[1,2], Prabaharan Poornachandran[1], K.P. Soman[2] and V.G. Sujadevi[1*]

[1]*Center for Internet Studies and Artificial Intelligence, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India*
[2]*Department of Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore, Tamil Nadu, India*

***Correspondence to:**
V.G. Sujadevi
Center for Internet Studies and Artificial Intelligence,
Amrita Vishwa Vidyapeetham,
Kollam, Kerala, India.
E-mail: sujap@am.amrita.edu

Received: October 11, 2023
Accepted: December 19, 2023
Published: December 22, 2023

Citation: Ashok N, Poornachandran P, Soman KP, Sujadevi VG. 2023. Artificial Intelligence in Nanomaterials and Polymers: A Detailed Review Approach on Nano-informatics, Polymer Informatics, and Informational Materials Science. *NanoWorld J* 9(S5): S76-S80.

Published by United Scientific Group

Abstract

This paper envisages a detailed review about progress in nano-informatics, polymer-informatics, and material informatics. The incorporation of engineered nanomaterials (ENMs) enable novel materials in which matter can be controlled at nano-range (1 to 100 nm). Numerous areas of human life, research and engineering are influenced by artificial intelligence (AI)-based data driven approaches. Nano reinforcements play a vital role in materials development. Nano-informatics is a subset of materials informatics or informational materials technology and a powerful tool for predicting and characterization of nanocomposites. This article reviews a brief about data repositories and nano-informatic centric approach for property description, activation energies, etc., using molecular representation. Simplified Molecular Input Line Entry System (SMILES and SELFIES) serve as a basis for molecular string representation in chemical language and generative models. Another area of novel research is in polymers with nanomaterials in which AI and ML (machine learning) perspectives have taken a role for efficient creation, design, development, and discovery of novel materials known as polymer informatics. Surrogate models are trained on the available data to predict and screen properties of potential polymer candidates for a given target property. This review article also outlines the data-driven repositories to address tremendous physico-chemical variety of polymers at various sizes. It is being investigated how to deal with inverse problems by recommending polymers that use cutting-edge AI techniques to achieve application goals. Efficiency gains, quicker research, and higher output may come about as different elements of the developing polymer informatics ecosystem mature and combine. This article provides an overall picture of present developments in data-driven experimental polymer chemistry with the aid of artificial intelligence.

Keywords

Nano-informatics, Molecular representations, Polymer informatics, Artificial intelligence, Materials science, Materials informatics, Chemical language models

Introduction

The applications of computational modeling and methodology for the prediction of nanoparticle property including data base and multi-physics-based modeling using AI has larger potential to enhance the nanomaterial into large scale of applications. A platform for utilizing current polymer data for effective functional polymer design, polymer informatics is an interdisciplinary area of research that combines polymer science, AI, data science, and ML [1,2]. The group of organic or synthetic compounds viz. polymers are of many macromolecules, which are very massive molecules that are multiples numerous similar chemical blocks known as monomers. Various natural and artificial materials, as well as a major portion of the constituents in living things, are composed of polymers.

When it comes to food packaging, water purification, clothes, housing, health care, electronics, and transportation, polymers play a crucial role in daily life. One or more of the characteristics of polymers—economical, easy to manufacture, enhanced strength-to-mass ratio, and low embodied energy to synthesize and process them—make them ideal for a variety of applications [3, 4].

Amongst various methods for nano-informatics, data-based quantitative structure–property relationship (QSPR) and quantitative structure–activity relationship (QSAR) models assist materials researchers and scientists to attain a better insight of the physicochemical properties that produce the desired functionality of the reinforcements [5].

This article reviews recent advancements in tools and data repositories in nano-informatics and informational materials science in general, specifically with polymer-informatics. The applications of AI and ML databases with molecular representation using SMILES and SELFIES is also envisaged in this review article.

Traditional technologies are facing significant challenges as a result of the massive number of organic molecules currently in existence and the ever-increasing demand for nano materials. QSAR modelling approach or otherwise nano-QSAR that utilizes ML and AI approach utilizes a set of predictors for various characteristics of nanomaterials known as nano-descriptors. In nano-informatics, the former mentioned datacentric approach plays a vital role. The predicted property includes physiochemical interactions, biological activity, etc. The various nano-descriptors viz. sequential modelling approach (example by structure-activity prediction networks - SAPNet, predicted response modelling like predictor of cellular uptake etc. are shown in figure 1.

Data-centric informatics-based methods are becoming helpful to predict and determine material properties subjected to adverse conditions or compute using traditional methods. Predictions are typically interpolative, by fingerprinting a material numerically first and then applying a mapping between the fingerprint and the property of interest. Predictions may also be extrapolative and further extending into novel material spaces that provide predictions on uncertainties that are properly taken into account.

Polymer informatics is developing as a viable instrument to assure ground-breaking discoveries in the field of polymer science, thanks to the quick advancements in high-throughput computing, AL, and ML applications.

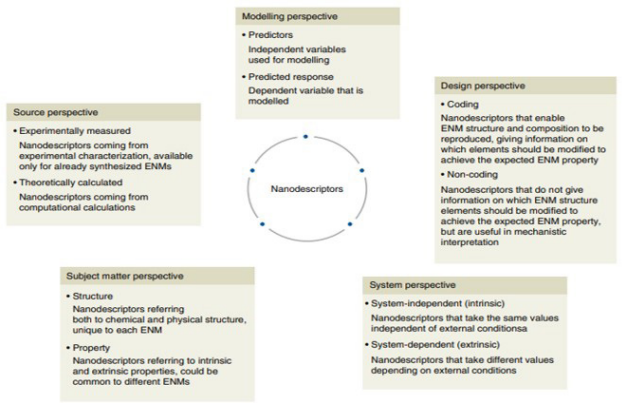

Figure 1: Modeling, design, and characteristics of nanomaterials and nano-descriptors.

Review About Informatics in Materials Sciences

The constant process of discovering new materials has marked the advancement of humanity. The current generation of materials discovery, which this article refers to as Materials 4.0, is the fourth paradigm of materials science research. It briefly describes the infrastructure that has been put in place for Materials 4.0 and provides several descriptions of materials discovery and lifecycle prediction assessment using this methodology [2]. Human needs and aspirations have traditionally driven material development, and this is expected to continue in the near future. By 2050, there will likely be 10 billion people on the planet, which will lead to an increase in demand for professional healthcare, individualized consumer goods, clean, high-efficiency energy, and clean energy. Data-driven or ML technologies have created new possibilities for the materials discovery and logical design of materials as huge amounts of data produced by contemporary experimental and computational techniques become more widely accessible [6]. Polymers are very adaptable materials that are widely used. Researchers are increasingly utilizing polymer informatics and data science to build novel materials and analyze the links between their structure and properties. A developing field is polymer informatics. Despite the abundance of helpful tools and databases, many of them remain underestimated. Here, a few of the databases and technologies that are readily available in the field of polymer informatics are described in detail [7]. The field of polymer informatics is expanding quickly based on cheminformatics, materials informatics, and data science platforms. ML algorithms can now predict a variety of macromolecular characteristics with a respectable degree of accuracy, making data-driven analyses, forecasts, and recommendations for experimental polymer research more useful. In order to properly handle the hierarchical connections of polymer systems, it is essential to develop suitable treatments for higher-order structures and experimental techniques [8].

As algorithms and computation get more sophisticated, the time and cost associated with creating new materials will decrease. Several kinds of materials, including metals and oxides, have found acceptance in the newly growing subject of materials informatics. However, there are significant obstacles that must be overcome in the specific context of polymer research before sophisticated ML strategies for developing novel materials may be used. These difficulties mostly have to do with how polymeric systems and their attributes are reported [9]. When it comes to their chemistry, structure, and uses, polymers are incredibly diverse. However, because the chemical space of polymers is nearly unlimited, it is difficult to locate the optimum polymer with the correct mix of qualities for a specific application [10-14]. Surrogate models that can instantly anticipate the characteristics of polymers that were built using machine learning methods on existing property data might help to alleviate this difficult search dilemma. In materials informatics, several polymer characteristics are concurrently learned and predicted using a multi-task learning technique (Figure 2) [8, 15].

Molecular Representation – Case Study

This case study covers molecular representation and pre-

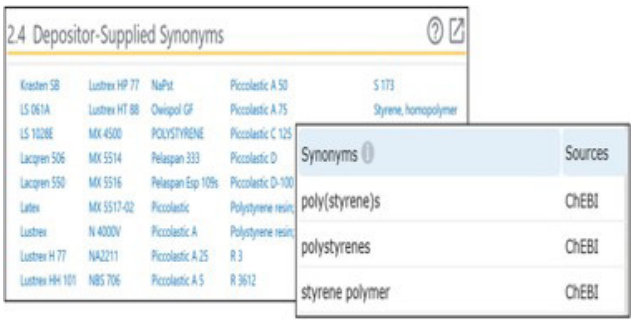

Figure 2: Databases for polymer-informatics.

diction whether a molecule is based on atomic and activation energy function. The comparative study of different activation functions and to classify the presence of molecules in humans or not is carried out in this work.

This study mainly envisages the applications of SMILES and SELFIES in representing the chemical structure of a molecule. Various deep learning and ML approaches are detailed in this work. Table 1 above gives an overall picture about the accuracy prediction using various methods. The dataset used in this work is ZINC15 dataset.

The output labels are the subset extracted from the ZINC15 dataset that is non-human metabolites and endogenous. The subset of this large data is evaluated by extracting the molecules which come under non-human metabolites (Primary metabolites - also known as metabolites, not reported in humans) and endogenous (Primary metabolites observed in humans) [16-23].

Open-source data are the foundation of data-driven research, but they are difficult for individuals to access. In this work, a benchmark database called PI1M is created, which stands for 1 million polymers, to provide data resources for ML research in polymer informatics. 12,000 manually collected polymers from PolyInfo, the largest polymer collection currently in existence, are used to train a generative model. Then 12,000 polymers are created using the model [1, 17].

A unique representation for polymers termed "polymer embedding" is first provided in order to carry out several polymer informatics regression tasks, including density, glass transition temperature, melting temperature, and dielectric constants. They conclude that the PI1M database encompasses a similar chemical space to PolyInfo while significantly filling places where PolyInfo data are scarce by comparing the polymer embedding trained with PolyInfo data to that trained with PI1M data. It is envisaged that the PI1M benchmark database will be beneficial for further polymer informatics research [17, 24]. Lack of big FAIR (findable, accessible, interoperable, and reusable) databases is a major obstacle to the implementation of polymer informatics.

In order to obtain grass transition temperatures with the least amount of human input, natural language processing software in conjunction with specifically created software modules were utilised and finally it was able to extract over 250 glass transition temperatures. All of the results' information is

Table 1: Accuracy prediction for various models.

Technique	Activation function	Number of layers	Accuracy
Deep learning	RELU	1D-CNN 1 layer	93.73
		1D-CNN 2 layer	95.46
	ELU	1D-CNN 1 layer	92.44
		1D-CNN 2 layer	94.81
	SELU	1D-CNN 1 layer	93.3
		1D-CNN 2 layer	93.3
Machine learning	Random forest classifier		97.84

available on the Polymers Property Predictor and Databases website, which may be viewed without charge. They learned throughout this process that it was difficult to recognise the names of the polymers in the literature since polymers are referred to by several names, including common names, sample names, labels, etc. [25]. In order to overcome this problem, researchers later looked into named entity recognition. They are concentrating on making it possible for user-submitted data to grow our databases.

The lack of uniformity in the representation of polymer names results in irregular polymer indexing, which is a significant barrier to the widespread use of datasets linked to polymers. Additionally, this limits the extensive application of materials informatics for innovation across a broad spectrum of polymer science and polymeric-based material classes. The present strategy, which uses several distinct chemical identities, hasn't been found to be effective in solving the issue and is unclear to researchers. ChemProps is a multi-algorithm-based mapping technique that is suggested by this research to address the polymer indexing [4] problem. It is simple to update. The RESTful API enables simple system integration and lightweight data exchange. The weight factor assigned to each algorithm that generates scores for prospective chemical names is calibrated to maximise the lowest possible value of the score difference between the potential chemical names and the actual chemical name. The 160 training data points are checked ten times to prevent overfitting issues. The resulting set of weight factors achieves a test accuracy of 100% on the 54 test data points. The weight factors will alter as ChemProps advances [26].

Other polymer databases can reduce duplicate entries and provide a more accurate "search by SMILES" functionality by utilising ChemProps as an API call-based common name-to-SMILES converter. ChemProps is an excellent tool for automatically filling polymer properties because of its straightforward design [22, 27]. The ATHAS data bank contains several valuable databases of physical characteristics related to polymers, such as the glass transition temperature (Tg) and heat capacity difference at Tg for completely amorphous polymers, the equilibrium melting point and heat of fusion at 100% crystallinity, etc. Using the ATHAS data pool, ML found QSPRs between repeated polymeric structural unit fingerprints and each physical feature.

For the two types of artificial neural networks that are utilized as regression models, the optimal hidden-layer designs

were determined using the contour plots of the root mean square error for each physical feature. Both artificial neural network constructions' observed values and expected values had a good correlation, suggesting that the physical characteristics could be predicted using nothing more complex than the repeating polymeric structural components. Physical properties of poly(p-dioxanone) that have yet to be recorded in the ATHAS data bank were also predicted, demonstrating that the projected attributes corresponded with reported values from the literature within 25% of the practical temperature range. An innovative method for determining the heat capacity of polymers is given that combines ML with ATHAS analysis. Nano-informatics is a rapidly developing research entity both at academia as well as industries. Nano-informatics is facilitated with enormous potential to provide new mechanistic insights into ENM interactions with living systems. This also facilitate a transition to *in silico* risk assessment. The characterization of engineered nanomaterials is a challenge faced by nano-informaticians and experimentalists. In characterizations using conventional experimental approach, novel parameters and measurement methodology with transparent procedures are implemented. Thus, enabling novel reproducibility of the work with better accuracy. This also facilitates requirements for reporting formats that include metadata, storage tools and data gathering [28]. However, there are some challenges in developing nano-descriptors like component relationships in multi-component nanomaterials. The interpretability of nano-descriptors can be enhanced by nano-QSARs with adverse outcome pathways [29].

Conclusion

Informatics of materials science or materials informatics in general is a vibrant and rapidly booming area of research not only in engineering, but various other fields like medical, high performance property predictions for adverse applications, etc. Informatics on nanomaterials and polymeric materials provide more mechanistic insights into the domain materials science. The advancements of AI and ML approaches otherwise data centric approach provide a better perspective in current scenario. This review article provides an overall snapshot of concepts in the progress of materials informatics, particularly database repositories for property prediction. The complete representation of nano and polymer informatics which relates structure with property prediction. The various physiochemical interactions and biological activity as well mechanical and fundamental properties are related with perspective on databases. A case study about the molecular representation as a classification problem whether the molecule is present or not in a human is also outlined in this review article. The case study basically focus on utility of molecular representation viz. SMILES and SELFIES in the dataset. In this research article, various studies on nano-descriptors and nano-informatics and some challenges involved in developing comprehensive models for predictive nano-informatics is also explored in detail. Nano-informatics is expected to enrich the exploration of frontiers in materials science and promote the integration of information as well as utilization of technical knowhow regarding nano reinforced materials for material discovery.

Acknowledgments

None.

Conflict of Interest

None.

References

1. Audus DJ, de Pablo JJ. 2017. Polymer informatics: opportunities and challenges. *ACS Macro Lett* 6(10): 1078-1082. https://doi.org/10.1021/acsmacrolett.7b00228
2. Chen L, Pilania G, Batra R, Huan TD, Kim C, et al. 2021. Polymer informatics: current status and critical next steps. *Mater Sci Eng R Rep* 144: 100595. https://doi.org/10.1016/j.mser.2020.100595
3. Sha W, Li Y, Tang S, Tian J, Zhao Y, et al. 2021. Machine learning in polymer informatics. *InfoMat* 3(4): 353-361. https://doi.org/10.1002/inf2.12167
4. Adams N. 2010. Polymer Informatics. In Meier M, Webster D (eds) Polymer Libraries. Advances in Polymer Science. Springer, Berlin, pp 107-149.
5. Wu S, Yamada H, Hayashi Y, Zamengo M, Yoshida R. 2020. Potentials and challenges of polymer informatics: exploiting machine learning for polymer design. *arXiv preprint* arXiv:2010.07683. https://doi.org/10.48550/arXiv.2010.07683
6. Adams N, Murray-Rust P. 2008. Engineering polymer informatics: towards the computer-aided design of polymers. *Macromol Rapid Commun* 29(8): 615-632. https://doi.org/10.1002/marc.200700832
7. Zeng Y, Cao H, Ouyang Q, Qian Q. 2021. Multi-task learning and data augmentation for negative thermal expansion materials property prediction. *Mater Today Commun* 27: 102314. https://doi.org/10.1016/j.mtcomm.2021.102314
8. Wang Y, Zhang M, Lin A, Iyer A, Prasad AS, et al. 2020. Mining structure–property relationships in polymer nanocomposites using data driven finite element analysis and multi-task convolutional neural networks. *Mol Syst Des Eng* 5(5): 962-975. https://doi.org/10.1039/D0ME00020E
9. Yamaguchi T, Yamashita Y. 2021. Quality prediction for multi-grade batch process using sparse flexible clustered multi-task learning. *Comput Chem Eng* 150: 107320. https://doi.org/10.1016/j.compchemeng.2021.107320
10. Cassola S, Duhovic M, Schmidt T, May D. 2022. Machine learning for polymer composites process simulation–a review. *Compos B Eng* 27: 110208. https://doi.org/10.1016/j.compositesb.2022.110208
11. Ramprasad R, Batra R, Pilania G, Mannodi-Kanakkithodi A, Kim C. 2017. Machine learning in materials informatics: recent applications and prospects. *NPJ Comput Mater* 3(1): 54. https://doi.org/10.1038/s41524-017-0056-5
12. Jackson NE, Webb MA, de Pablo JJ. 2019. Recent advances in machine learning towards multiscale soft materials design. *Curr Opin Chem Eng* 23: 106-114. https://doi.org/10.1016/j.coche.2019.03.005
13. Jose R, Ramakrishna S. 2018. Materials 4.0: materials big data enabled materials discovery. *Appl Mater Today* 10: 127-132. https://doi.org/10.1016/j.apmt.2017.12.015
14. Zhou T, Song Z, Sundmacher K. 2019. Big data creates new opportunities for materials research: a review on methods and applications of machine learning for materials design. *Engineering* 5(6): 1017-1026. https://doi.org/10.1016/j.eng.2019.02.011
15. Cencer MM, Moore JS, Assary RS. 2022. Machine learning for polymeric materials: an introduction. *Polym Int* 71(5): 537-542. https://doi.org/10.1002/pi.6345

16. Hatakeyama-Sato K. 2023. Recent advances and challenges in experiment-oriented polymer informatics. *Polym J* 55(2): 117-131. https://doi.org/10.1038/s41428-022-00734-9
17. Kuenneth C, Rajan AC, Tran H, Chen L, Kim C, et al. 2021. Polymer informatics with multi-task learning. *Patterns* 2(4): 100238. https://doi.org/10.1016/j.patter.2021.100238
18. Ma R, Luo T. 2020. PI1M: a benchmark database for polymer informatics. *J Chem Inf Model* 60(10): 4684-4690. https://doi.org/10.1021/acs.jcim.0c00726
19. Tchoua R, Hong Z, Audus D, Patel S, Ward L, et al. 2020. Developing databases for polymer informatics. *Bull Am Phys Soc* 65. https://doi.org/10.1021/acs.jcim.0c00726
20. Hu B, Lin A, Brinson LC. 2021. ChemProps: a RESTful API enabled database for composite polymer name standardization. *J Cheminformatics* 13(1): 1-3. https://doi.org/10.1186/s13321-021-00502-6
21. Chandrasekaran A, Kim C, Venkatram S, Ramprasad R. 2020. A deep learning solvent-selection paradigm powered by a massive solvent/nonsolvent database for polymers. *Macromolecules* 53(12): 4764-4769. https://doi.org/10.1021/acs.macromol.0c00251
22. Ishikiriyama K. 2022. Polymer informatics based on the quantitative structure-property relationship using a machine-learning framework for the physical properties of polymers in the ATHAS data bank. *Thermochim Acta* 708: 179135. https://doi.org/10.1016/j.tca.2021.179135
23. Ma R, Liu Z, Zhang Q, Liu Z, Luo T. 2019. Evaluating polymer representations via quantifying structure–property relationships. *J Chem Inf Model* 59(7): 3110-3119. https://doi.org/10.1021/acs.jcim.9b00358
24. Xu C, Wang Y, Farimani AB. 2023. TransPolymer: a transformer-based language model for polymer property predictions. *NPJ Comput Mater* 9(1): 64. https://doi.org/10.1038/s41524-023-01016-5
25. Hayashi Y, Shiomi J, Morikawa J, Yoshida R. 2022. RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. *NPJ Comput Mater* 8(1): 222. https://doi.org/10.1038/s41524-022-00906-4
26. Venkatraman V, Alsberg BK. 2018. Designing high-refractive index polymers using materials informatics. *Polymers* 10(1): 103. https://doi.org/10.3390/polym10010103
27. Kim C, Chandrasekaran A, Huan TD, Das D, Ramprasad R. 2018. Polymer genome: a data-powered polymer informatics platform for property predictions. *J Phys Chem C* 122(31): 17575-17585. https://doi.org/10.1021/acs.jpcc.8b02913
28. Nagasawa S, Al-Naamani E, Saeki A. 2018. Computer-aided screening of conjugated polymers for organic solar cell: classification by random forest. *J Phys Chem C* 9(10): 2639-2646. https://doi.org/10.1021/acs.jpclett.8b00635
29. Hatakeyama-Sato K, Tezuka T, Nishikitani Y, Nishide H, Oyaizu K. 2019. Synthesis of lithium-ion conducting polymers designed by machine learning-based prediction and screening. *Chem Lett* 48(2): 130-132. https://doi.org/10.1246/cl.180847