

Chapter 2

Data Mining in Material Science

Moganapriya Chinnasamy¹, Rajasekar Rathanasamy^{2*}, Samir Kumar Pal¹, Manoj Kumar Kathiresan², Sathish Kumar Palaniappan¹

¹Department of Mining Engineering, Indian Institute of Technology Kharagpur, West Bengal State, India

²Department of Mechanical Engineering, Kongu Engineering College, Erode, Tamil Nadu State, India

* rajasekar.cr@gmail.com

Abstract

The invention of novel materials is the most propelling factor promoting the growth of contemporary civilization and technological innovation; nevertheless, previous materials research relied mostly on random techniques, which is arduous and labor-intensive. With the introduction of big data, which brings a deep upheaval in human society and considerably advances science, artificial intelligence, machine learning, and deep learning methods have recently made remarkable progress in materials science research. However, there are few comprehensive generalizations and descriptions of its applications in materials research. In this chapter, a brief summary of the evolution of materials science research is presented, followed by an emphasis on the key principles and basic processes of AI technique.

Keywords

Artificial intelligence, Machine learning, Deep learning, Materials

Contents

Data Mining in Material Science.....	24
1. Introduction.....	25
2. Machine learning and materials science.....	28
3. ML algorithms in materials science.....	31

4. Steps in machine learning for materials science.....	33
4.1 Experience	33
4.2 Task.....	34
4.3 Classification	34
4.4 Regression.....	34
4.5 Clustering.....	34
4.6 Dimension reduction and conception	34
4.7 Efficient searching.....	35
4.8 Performance measure.....	35
4.9 Model particulars	35
4.10 Supervised model.....	36
Conclusion.....	38
References	39

1. Introduction

Artificial intelligence (AI) is a branch of computer science that aims to understand and build intelligent systems. Researchers have seen remarkable development in AI in recent decades, with some systems already achieving human-level and potentially super-human competence for a variety of activities such as voice recognition, picture analysis, machine interpretation, and games. There is a widespread conviction that AI will fundamentally alter many aspects of our culture and economy. Self-driving vehicles, for example, which incorporate actual visual identification and control, are on the verge of becoming a reality. This incredible development is causing a major transition in AI exploration from theoretical pursuit to an abundant larger sector with significant industry and governmental funding.

Considering the remarkable breakthroughs in AI, the research fraternity as a whole has taken notice and is investigating the application of AI for discoveries [1-11]. The materials science sector, in particular, has begun to use AI tools to speed material discovery. The current tendency is to apply machine learning (ML) approaches because many recent AI successes, particularly those involving superhuman powers are anchored in ML and deep learning (DL) approaches. Even though the conditions that facilitate DL are indisputable, particularly for visualization, language recognition, linguistic conversion, and independent driving, the constraints are soundly identified. In general, DL methodologies greatly rely on the accessibility of massive quantities of references or labeled data that is frequently

unavailable. The present state of deep learning was likened to System 1. It comprises human's usual data processing because it is quick, easy, and a form of automatic pattern identification. Perception, encompassing vision, and hearing, is a component of System 1. The perceptual capacities are extremely advanced and a significant portion of the human intellectual cortex is dedicated to observation as depicted in figure 1.

Humans as well have an additional logical System 2, which is sluggish and requires significant thought and reasoning to solve difficult issues beyond reflexive reactions unlike System 1. Such complicated jobs are unsuitable for pure ML and DL. Nonetheless, AI spans a wide range of approaches, including hunting, cognitive, preparation, and data illustration. These strategies played a variety of parts throughout the AI growth stages. They are positioned to become more significant and play a growing role in supplementing pure ML techniques as well as further extending and propelling AI research to address increasingly complicated issues. These challenges include those involving scientific discovery, as well as jobs that need critical thought and reasoning and are carried out by people utilizing System 2.

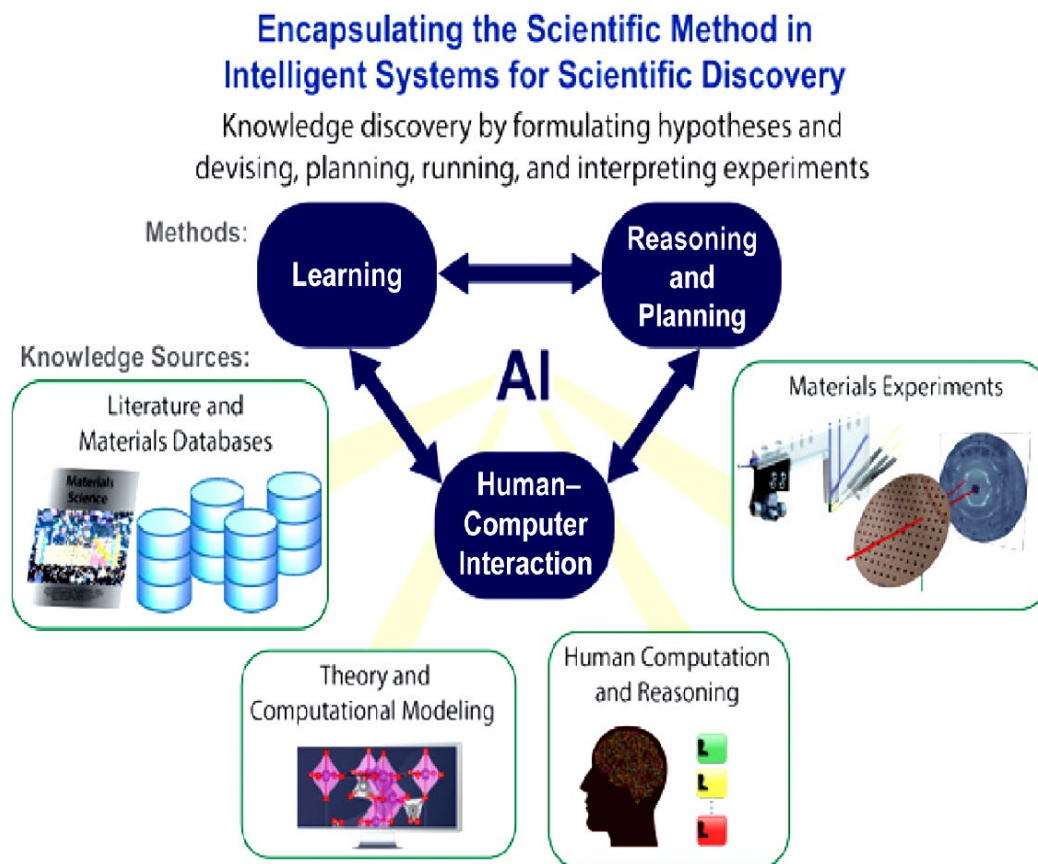


Figure 1 Structure of AI [4]

Before data-driven methodologies, material discovery evolved through three paradigms [12]. The first framework is based on experiments using trial approaches, the next is on the conversion of investigational data into a theoretical model, and the last is based on the computational simulation of theoretic representations. Individually model emerged in response to the outcomes of the preceding models. Recently, the fast expansion of existing data created through the preceding three models, along with the advent of the extremely effective algorithm, has permitted the use of AI, kicking off the quarter and the most current model of material discoveries. The main benefit of data-assisted and driven discovery is significantly reducing the necessary computing power, hence removing the limitations of the compositional transmission area. In recent years, artificial intelligence [10, 13-18] and virtual screening [19-27] approaches have been effectively applied to subfields of material and chemical science to identify novel organic compounds and constituents with extraordinary functions. Every area, due to its structural features, necessitates specialized approaches for simulated screening. Furthermore, to employ AI approaches in various subfields, a substantial quantity of great reliability and eminent data [28] must first be produced.

Two-dimensional materials (2D) are developing a sub-field of materials research. 2D materials, with their outstanding and controllable characteristics, show great potential for semiconductor, energy, and health applications [29-32]. Few unique 2D materials were fruitfully produced [33] later the finding of graphene [34], which has a meek 2D assembly of carbon atoms, however enticing and complicated physics. Despite a small number of experimentally produced compounds, novel 2D material depositories established on realistic considerable imitations have lately emerged [19, 35, 36]. The silicon repositories were created using two consecutive ways to build novel 2D materials: layer exfoliation through 3D bulk assemblies and combined atom interchange in 2D configurations. The initial method relies on the screening of exfoliated covered resources through 3D material databases [37-39]. This technique produces novel 2D material with assemblies that may utilize 2D patterns. The subsequent method encompasses exchanging single or additional atoms of an identified 2D pattern with molecules of extra chemical constituent. The following approach produces novel 2D constituents with identical crystal patterns as the original but with dissimilar chemical components. Regardless of the dual methodologies, the resource-intensive computation constrained the chemical search area for new 2D materials to the point where the quantity of 2D materials that were considered with density functional theory (DFT) calculations, the workhorse computational technique [40], has been reduced to a few thousand in recent years. Even though the DFT-calculated 2D materials records contain a few thousand materials, they are excellent equitable information repositories for data-driven algorithms. Amongst the primary aims, machine

learning is a promising way of exploring extraordinarily broad search spaces for possible chemical combinations in the order of particles, 2D and 3D materials. The procedure creates contestant resources systematically throughout an unparalleled chemical universe of molecules, recognizes probable stable resources, and forecasts important attributes of constant materials alternatives. They utilized this technique to create a computer-generated 2D Materials databank that comprises 316,505 probable constant 2D materials with projected attributes as a proof of concept [41].

2. Machine learning and materials science

“Machine learning” implementation approaches ML approaches which contain “supervised, unsupervised and reinforcement learning” and are an essential aspect of artificial intelligence. Each instance in supervised learning is made up of input and output, often referred to as a supervised signal. The labeling of data, comprising data grouping, information attributes, and feature idea placement, is one of the major aspects of supervised learning [42]. After training the algorithms with labeled information, factors of procedures are altered and predicted on the assessment outcomes of expected ones and predicted data and the process is repeated till the procedures congregate to optimal result, at which point the intelligent decision-making capability with the specific model is obtained. Unsupervised learning, unlike supervised learning, uses unlabelled data to search and derive probable links between samples. Clustering and dimensionality reduction are two common unsupervised learning methods. Clustering requires analysis of dispersal in feature areas of data measurements to categorize analogous information into single clusters and distinct information with dissimilar types. This is necessary because the categorization of data is unclear in advance. Furthermore, because higher dimensional information cliques are fairly uncommon in ML, difficulties such as distance computations and insufficient sample data are common, which is referred to as dimension catastrophe. The theoretical basis of reducing the dimension in ML is plotting the information sockets from novel higher dimensional to lower dimensional area and dimension reduction procedures primarily include factor analysis (FA), singular value decomposition (SVD), principal component analysis (PCA), and independent component analysis (ICA) [43-47].

Unlike “supervised and unsupervised learning, reinforcement learning” obtains knowledge statistics and updates the typical factors continuously based on input from the situation, it does not necessitate information to be supplied prior. In general, when the machine performs appropriately, a positive incentive is obtained; otherwise, a negative incentive is gathered. In this situation, the ML algorithms create several dynamic planning concepts, and eventually picked the action mode which maximizes the incentive. This demonstrates that the reinforcement learning technique requires less information and is easier to develop,

making it suitable for dealing with increasingly complex choice issues. Furthermore, as shown in figure 2 deep reinforcement learning that associates reinforcement through deep learning, is emerging as the fastest research hub in AI, particularly in automated driving, robots, language dispensation, and other domains [48-52]. Figure 3 depicts a framework of materials discovery which is used basically and designed based on ML approaches, with three primary processes listed: sample fabrication, algorithm model development, model verification, and materials prediction [53].

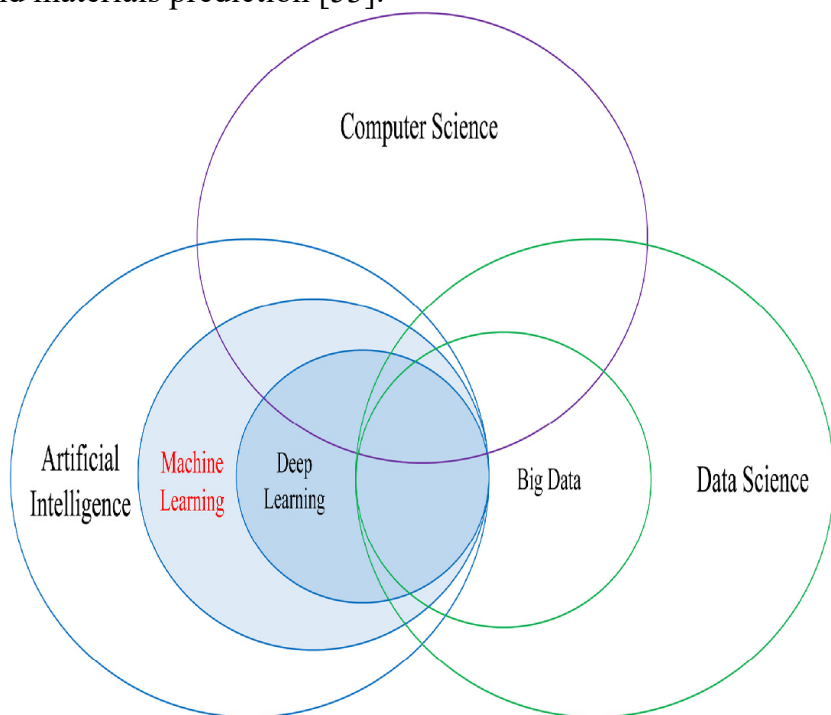
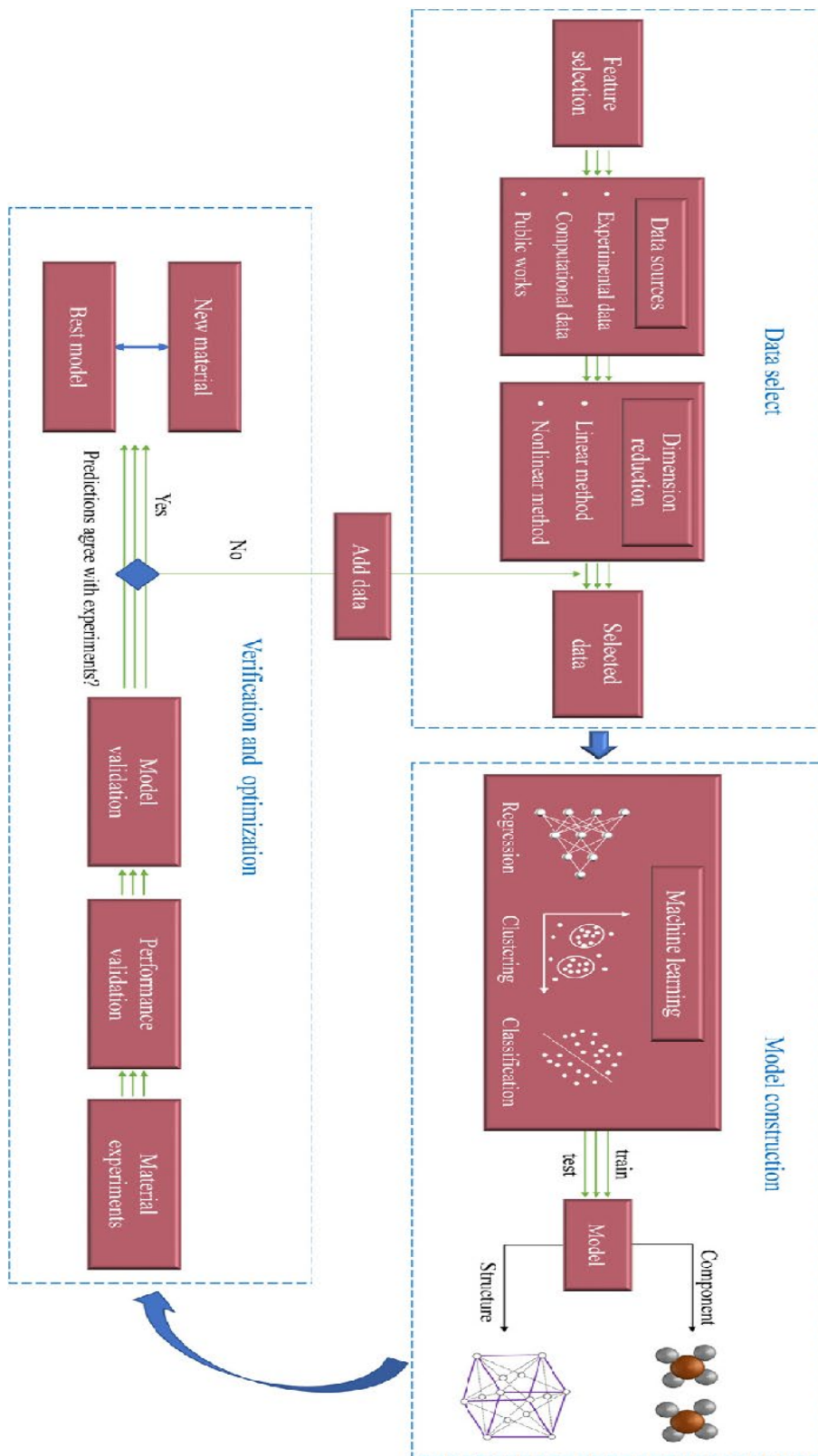


Figure 2 Interlinking of AI, ML, and DL [53]

Figure 3 Structural design of materials discovery and ML [53]



3. ML algorithms in materials science

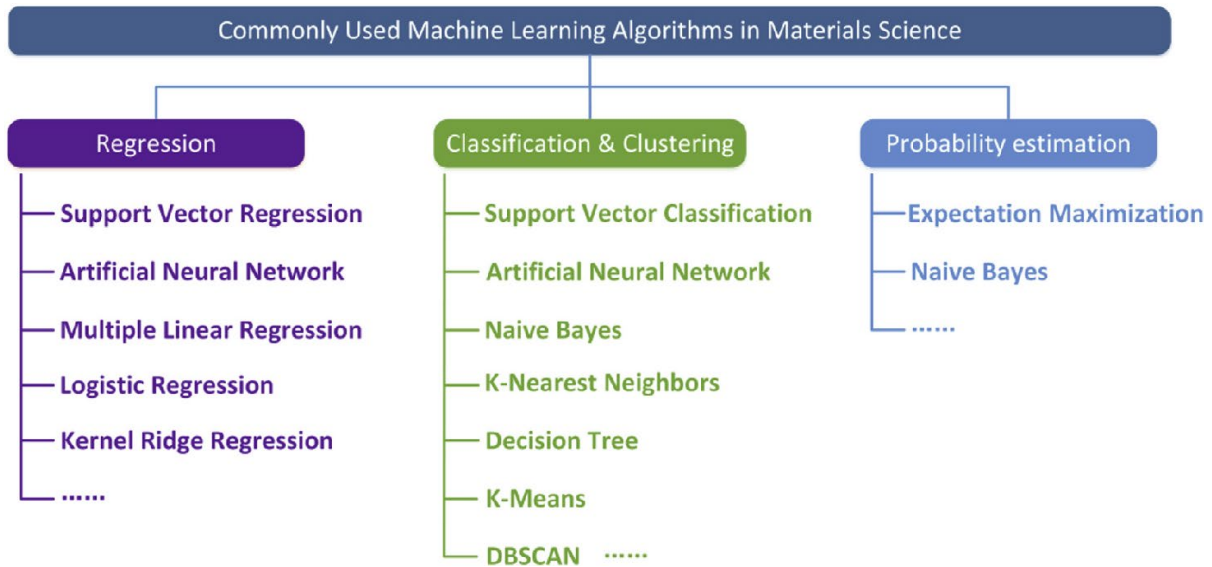


Figure 4 ML algorithm in materials science [54]

The critical stage in the development of a machine learning system is the selection of an efficient machine learning system since it has a significant impact on generalization ability and prediction accuracy [55]. Every single algorithm has its unique set of applications; hence no approach is ideal for all problems. As seen in Figure 4, in materials science, the most often employed machine learning algorithms fall into 4 groupings: “probability estimation, regression, clustering, and classification” [54]. Probability estimate procedures are mostly utilized for novel material detection whereas “regression, clustering, and classification” techniques are primarily employed for predicting the material characteristics at all levels. Furthermore, ML approaches are frequently integrated with intellectual optimization techniques [56], which are primarily employed for optimizing model characteristics. Additionally, these techniques may be used to tackle additional complex tasks [57], like spatial configuration and material property optimization [58, 59] as portrayed in Figures 5 and 6.

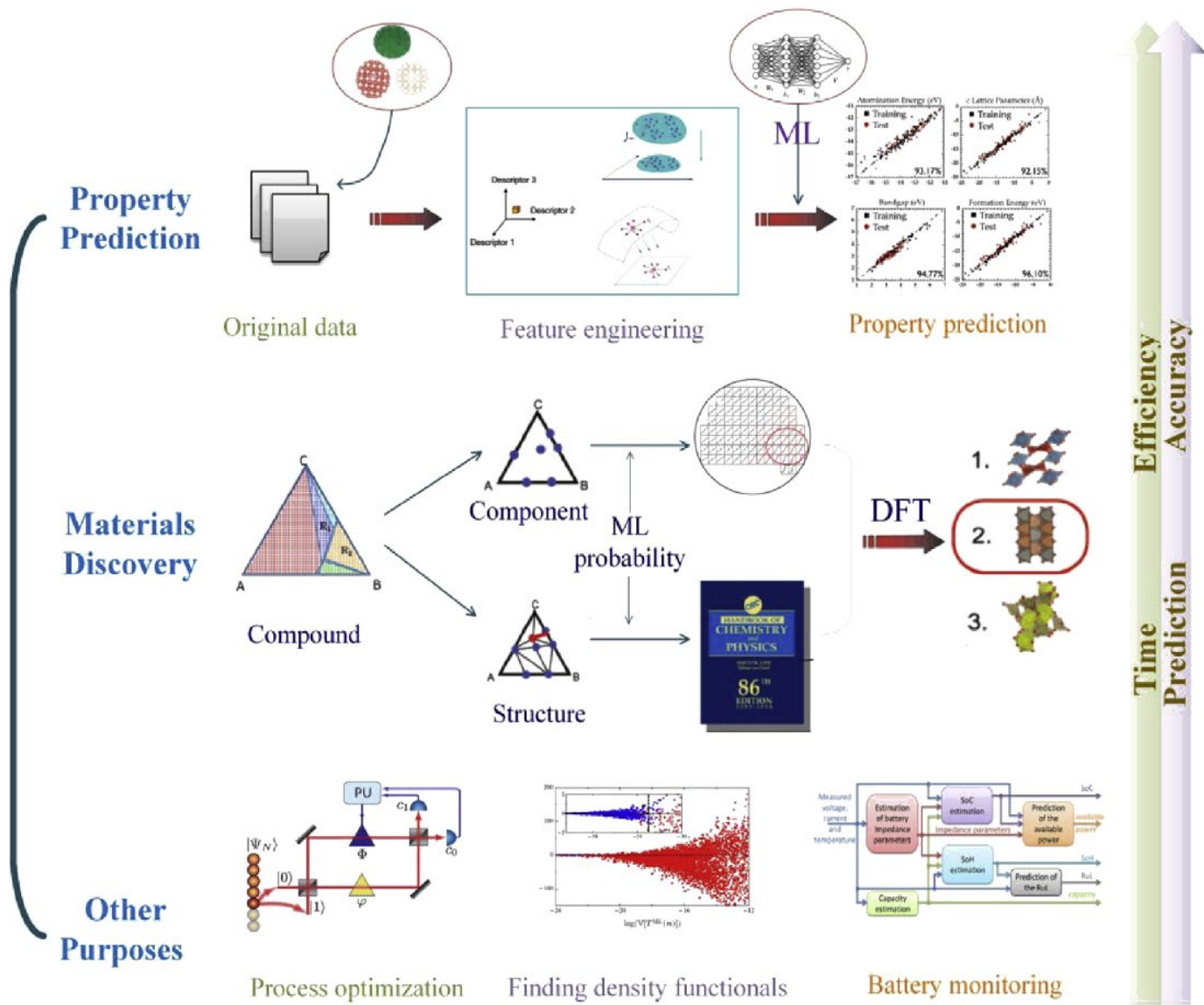


Figure 5 Application of ML in materials science [54]

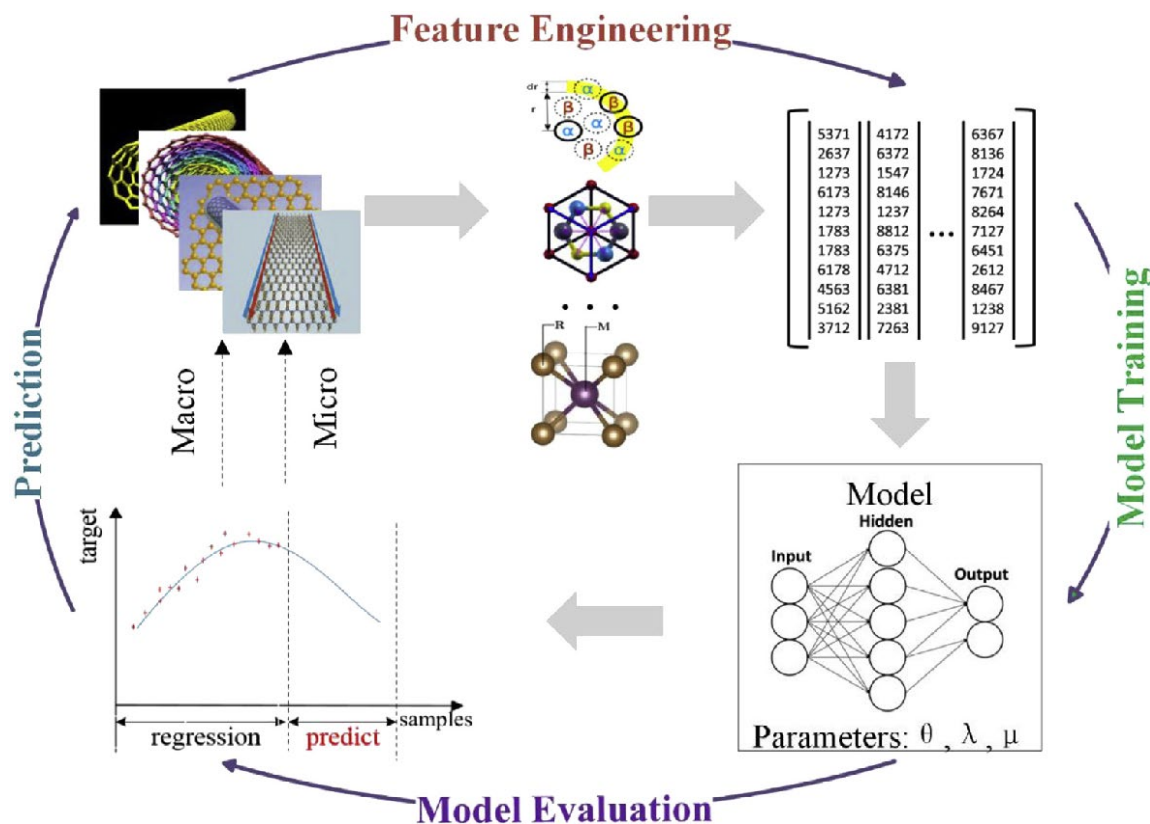


Figure 6 ML in material property forecast [54]

4. Steps in machine learning for materials science

Mitchell defines ML as a "mainframe program is supposed to acquire from experience E regard to a set of tasks T and performance measure P" [60] that is utilized as foundation background L in the following way as depicted in Figure 7.

4.1 Experience

Utmost ML algorithms may be classed as either unsupervised or supervised learning depending on the sort of samples that are provided throughout the training phase. Unsupervised learning uses a collection of instances containing just features to establish a few significant connections between the cases. A dataset of instances containing characteristics and associated labels that are "right" values linked with characteristics is used in supervised learning. Another type is reinforcement learning (RL), in which the representative knows through engaging with surroundings to acquire reward feedback. Conversely, RL confronts several obstacles, the most significant of which is the difficulty in creating interactive settings with rapid response. As a result, it is not yet commonly used in material discovery, and this evaluation confines itself to be guided and unsupervised.

4.2 Task

ML could be used for a variety of purposes. The task generally entails processing samples provided to an algorithm. The illustrations include features/descriptors that are typically organized as vectors to define the physicochemical characteristics, structural features, composition attributes, or formulation process conditions of the material.

4.3 Classification

The algorithm is concerned with determining which group or category a certain problem fits. It is accomplished through the training function f that transfers the characteristics vector to one of the unique modules. In its place of settling on a single group, f can instead provide a probability distribution across every group, with every element in the output vector representing the likelihood that the sample corresponds to the assured group. Material categorization challenges have been successfully solved using ML algorithms. A grouping model can forecast the effective formation of synthesized materials [61] and also the regions which exhibit any flow defect whenever given for specific synthesis parameters [62].

4.4 Regression

Regression is another typical job in which the algorithm attempts to study the function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that generates an uninterrupted y value or a series of data points represented as a vector y . Regression models are commonly used to estimate material attributes like specific heat of objects, wave equation temperature, and optical gap of materials [63-65].

4.5 Clustering

It is a type of unsupervised learning technique, which comes in handy with a huge volume of unlabelled data [66-68]. It groups the set into clusters wherein elements inside the same group are more "similar" to one another than to another. What is meant by "similar" varies depending on context and circumstances. Even if the dataset is unlabelled, grouping objects into clusters allows for valuable insight into the data. Clustering algorithms have been used in material discovery in materials textual data mining and micro-structural image processing.

4.6 Dimension reduction and conception

While teaching a network, it may depend on several variables [69]. In such cases, it is advantageous to decrease the measurement of characteristics by projecting into smaller dimensions while retaining quite as abundant data as feasible [70]. It will assist in improving computational efficiency, and model performance, reduce overfitting, and aid

in the discovery of insights for specific applications. Dimension reduction was utilized to enhance prediction outcomes in material discovery, such as summarizing entire long unique features into lesser information to obtain better performance [71, 72]. Dimension reduction was employed to discover the essential science of physical prototypes by removing less important aspects [73]. Furthermore, mapping higher dimensional information to 2D or 3D charts for conception is a significant dimension reduction function since it allows meaningful insights to be drawn from an understandable graphic. It has been used to depict the high-dimensional material design space in material discovery[74].

4.7 Efficient searching

Conducting replications and lab-scale studies to collect more data is arduous and expensive. Effectual searching strategies can assist in identifying the best relevant additional data points to annotate, reducing total data-collecting efforts. Effectual probing was applied to the design of Perovskite, multilayer films, carbide, and nitride, among other materials [74, 75]. Overall, determining which objective to tackle is the initial stage in applying new ML in material discovery. Though, the five work types are not necessarily independent of one another.

4.8 Performance measure

The performance metric is employed to evaluate the performance of an algorithm on a certain job. The accuracy and error rate can be used to assess the algorithm's success in classification [76]. The log probability may be determined that produce probability dissemination. Mean square error or various variants of error are commonly used in the regression. A task's dataset is often divided into three parts: training data, a validation set of data, and data for testing. The algorithms will be taught during training and then fine-tuned based on their efficiency. The test efficiency of the optimized learning method is generally tested depending on the capability to execute test data, which functions as a measure of how effectively the model can generalize. The clustering technique in unsupervised learning may be quantified using exterior and interior indexes. Exterior indices require previously defined clustering configuration, i.e. the actual labeling for data in the dataset. The outcome of internal indices is assessed using quantities and attributes intrinsic to the database [77, 78].

4.9 Model particulars

The procedure of ML to employ is perspective and critical to select the proper model for a suitable task to obtain optimal effectiveness without inadequate or underfitting. Numerous ML models are often employed in the domain of AI-aided material discovery

and are described in many papers. These approaches are introduced in supervised, unsupervised, and weakly supervised learning paradigms. Figure 1 depicts the entire architecture.

4.10 Supervised model

The supervised learning algorithm is initially qualified using labeled data containing N training instances. The i^{th} training example, for instance, consists of a combination of dual vectors for structures and tickets. The purpose of the model is to learn the given function. The model features are critical to be determined by the problem context. Figure 8 depicts a typical supervised learning approach in material discovery. It can create datasets through lab trials, simulations, or pre-existing datasets. There are two types of characteristics that are often employed in material discovery internal, and external information about the material classification. The intrinsic content is determined by the qualities of materials and chemicals utilized, but extrinsic data is determined by the surroundings wherein the materials reside. The two sorts of characteristics can be utilized as raw features directly. However, before feeding the raw characteristics into the prediction model, certain additional modifications are usually made to them. Normalization is the first pre-processing transformation that may be used to convert the quantities of attributes to a comparable basis. Another alternative pre-processing alteration is dimension reduction, which is employed if there are too many distinct features, particularly when compared to the number of labeled training samples. In this instance, it may be preferable to lower the dimension of the attributes to avoid overfitting. [14, 79]. In addition to the above-mentioned pre-processing methods, domain precise transition will be performed using the same preceding model. The preceding framework executes a few raw characteristics and transfers the output to the last forecasting framework. The translation of diverse chemical structures, for example, into fixed-length signature vectors can aid in improving forecasting ability in molecular property prediction tasks. Following these steps, the various treated and unrefined characteristics may be reduced to generate the complete characteristic given vector. Choosing the characteristics to employ is the utmost important stage in guaranteeing a successful model presentation [80]. Regardless of the highest excellent model, it will perform poorly if the features supplied to it are unsuitable. Following feature processing, several supervised models may be employed to forecast the finalized tasks. The most often employed supervised learning algorithms in material discovery were outlined and outcomes are typically a sequence of interesting qualities expressed vectorially [81].

Figure 7 Types of ML methods [81]

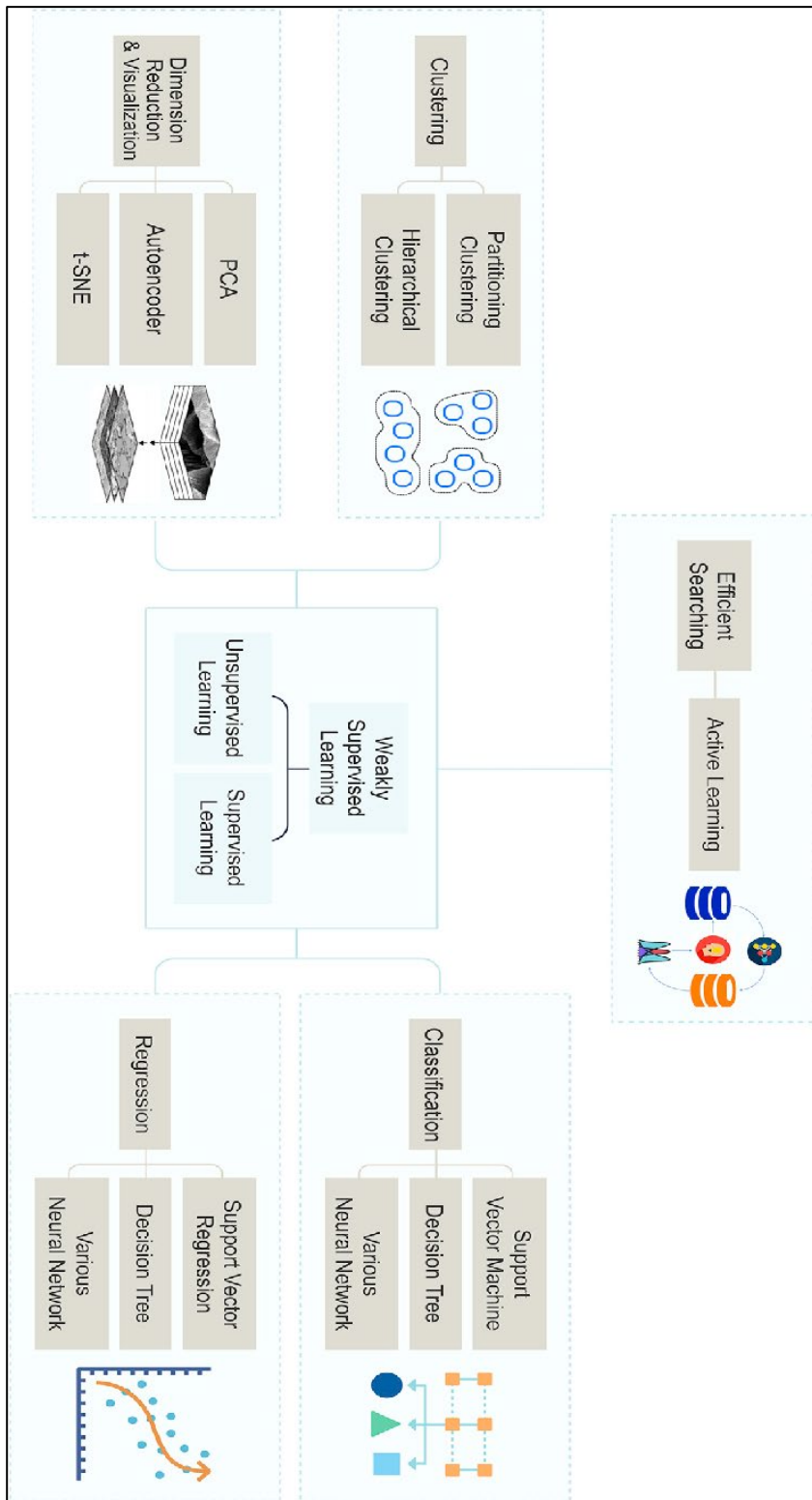
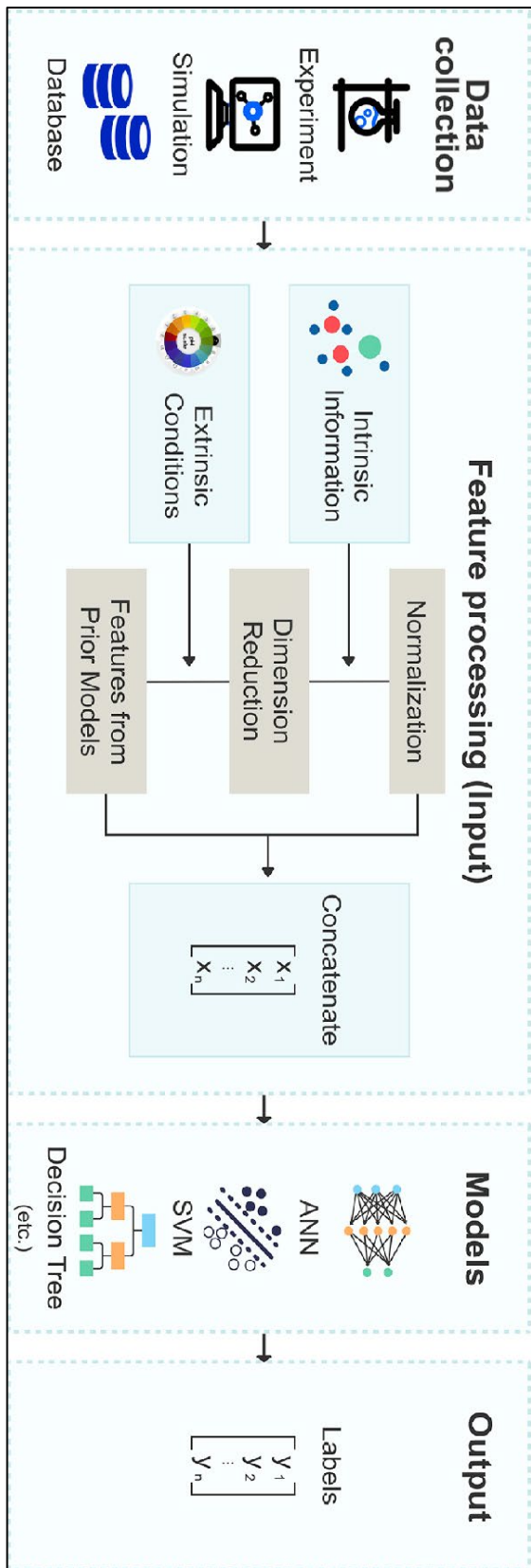


Figure 8. Supervised Learning in ML [81]



Conclusion

Machine learning, as a discipline of AI and one of the most popular categories of analytic techniques, is an important way for computers to learn. ML applications in materials science include novel material discovery, material property prediction, and other objectives spanning from the macro to micro level. The purposes of ML application in materials science are employed for quite diverse materials. A wide range of related studies shows that ML can be utilized to create precise and effective solutions for materials science. The areas in which ML can be used in materials science are becoming wider as theories and methodologies evolve.

References

- [1] J. Bohannon, The cyber scientist, *Science*. 357 (2017) 18-21.
<https://doi.org/10.1126/science.357.6346.18>
- [2] P.D. Luna, J. Wei, Y. Bengio, A.A. Guzik, E. Sargent, Use Machine Learning to Find Energy Materials, Nature Publishing Group, 2017. <https://doi.org/10.1038/d41586-017-07820-6>
- [3] Y. Gil, M. Greaves, J. Hendler, H. Hirsh, Amplify scientific discovery with artificial intelligence, *Science*. 346 (2014) 171-172. <https://doi.org/10.1126/science.1259439>
- [4] C.P. Gomes, B. Selman, J.M. Gregoire, Artificial intelligence for materials discovery, *MRS Bulletin*. 44 (2019) 538-544. <https://doi.org/10.1557/mrs.2019.158>
- [5] E.A. Hansen, S. Zilberstein, Monitoring and control of anytime algorithms: A dynamic programming approach, *Artif. Intell.* 126 (2001) 139-157.
[https://doi.org/10.1016/S0004-3702\(00\)00068-0](https://doi.org/10.1016/S0004-3702(00)00068-0)
- [6] A. Kelly, Think Twice: Review of Thinking, Fast, and Slow by Daniel Kahneman (2011), *Numeracy*. 10 (2017) 15. <https://doi.org/10.5038/1936-4660.10.2.15>
- [7] R.D. King, K.E. Whelan, F.M. Jones, P.G. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, S.G. Oliver, Functional genomic hypothesis generation and experimentation by a robot scientist, *Nature*. 427 (2004) 247-252. <https://doi.org/10.1038/nature02236>
- [8] J.R. Kitchin, Machine learning in catalysis, *Nat. Catal.* 1 (2018) 230-232.
<https://doi.org/10.1038/s41929-018-0056-y>
- [9] P. Nikolaev, D. Hooper, F. Webber, R. Rao, K. Decker, M. Krein, J. Poleski, R. Barto, B. Maruyama, Autonomy in materials research: A case study in carbon nanotube growth, *NPJ Comput. Mater.* 2 (2016) 1-6.
<https://doi.org/10.1038/npjcompumats.2016.31>

- [10] R. Ramprasad, R. Batra, G. Pilania, A.M. Kanakkithodi, C. Kim, Machine learning in materials informatics: Recent applications and prospects, *NPJ Comput. Mater.* 3 (2017) 1-13. <https://doi.org/10.1038/s41524-017-0056-5>
- [11] E. Smalley, AI-powered drug discovery captures pharma interest, *Nat. Biotechnol.* 35 (2017) 604-606. <https://doi.org/10.1038/nbt0717-604>
- [12] A. Agrawal, A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science, *Apl Mater.* 4 (2016) 053208. <https://doi.org/10.1063/1.4946894>
- [13] L. Himanen, A. Geurts, A.S. Foster, P. Rinke, Data-driven materials science: Status, challenges, and perspectives, *Adv. Sci.* 6 (2019) 1900808. <https://doi.org/10.1002/advs.201900808>
- [14] J. Noh, J. Kim, H.S. Stein, B.S. Lengeling, J.M. Gregoire, A.A. Guzik, Y. Jung, Inverse design of solid-state materials via a continuous representation, *Matter.* 1 (2019) 1370-1384. <https://doi.org/10.1016/j.matt.2019.08.017>
- [15] Q. Zhou, P. Tang, S. Liu, J. Pan, Q. Yan, S.C. Zhang, Learning atoms for materials discovery, *Proc. Natl. Acad. Sci.* 115 (2018) E6411-E6417. <https://doi.org/10.1073/pnas.1801181115>
- [16] C.C. Fischer, K.J. Tibbetts, D. Morgan, G. Ceder, Predicting crystal structure by merging data mining with quantum mechanics, *Nat. Mater.* 5 (2006) 641-646. <https://doi.org/10.1038/nmat1691>
- [17] J. Schmidt, M.R. Marques, S. Botti, M.A. Marques, Recent advances and applications of machine learning in solid-state materials science, *NPJ Comput. Mater.* 5 (2019) 1-36. <https://doi.org/10.1038/s41524-019-0221-0>
- [18] C. Moganapriya, R. Rajasekar, V.K. Gobinath, P. Prabhakaran, S.K. Jaganathan, A Frontier Statistical Approach Towards Online Tool Condition Monitoring and Optimization for Dry Turning Operation of SAE 1015 Steel, *Archiv. Metall. Mater.* 66 (2021) 901-909.
- [19] J. Zhou, L. Shen, M.D. Costa, K.A. Persson, S.P. Ong, P. Huck, Y. Lu, X. Ma, Y. Chen, H. Tang, 2DMatPedia, an open computational database of two-dimensional materials from top-down and bottom-up approaches, *Sci. Data.* 6 (2019) 1-10. <https://doi.org/10.1038/s41597-018-0005-2>
- [20] I.E. Castelli, K.S. Thygesen, K.W. Jacobsen, the Calculated optical absorption of different perovskite phases, *J. Mater. Chem. A.* 3 (2015) 12343-12349. <https://doi.org/10.1039/C5TA01586C>

- [21] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, *Phys. Rev. B.* 89 (2014) 094104. <https://doi.org/10.1103/PhysRevB.89.094104>
- [22] A.K. Singh, K. Mathew, H.L. Zhuang, R.G. Hennig, Computational screening of 2D materials for photocatalysis, *J. Phys. Chem. Lett.* 6 (2015) 1087-1098. <https://doi.org/10.1021/jz502646d>
- [23] R.G. Bombarelli, J.A. Iparraguirre, T.D. Hirzel, D. Duvenaud, D. Maclaurin, M.A.B. Forsythe, H.S. Chae, M. Einzinger, D.G. Ha, T. Wu, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, *Nat. Mater.* 15 (2016) 1120-1127. <https://doi.org/10.1038/nmat4717>
- [24] S. Curtarolo, G.L. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, *Nat. Mater.* 12 (2013) 191-201. <https://doi.org/10.1038/nmat3568>
- [25] L. Yu, A. Zunger, Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials, *Phys. Rev. Lett.* 108 (2012) 068701. <https://doi.org/10.1103/PhysRevLett.108.068701>
- [26] J. Hachmann, R.O. Amaya, A. Jinich, A.L. Appleton, M.A.B. Forsythe, L.R. Seress, C.R. Salgado, K. Trepte, S.A. Evrenk, S. Er, Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry-the Harvard Clean Energy Project, *Energy Environ. Sci.* 7 (2014) 698-704. <https://doi.org/10.1039/C3EE42756K>
- [27] E.O.P. Knapp, C. Suh, R.G. Bombarelli, J.A. Iparraguirre, A.A. Guzik, What is high-throughput virtual screening? A perspective from organic materials discovery, *Annual Rev. Mater. Res.* 45 (2015) 195-216. <https://doi.org/10.1146/annurev-matsci-070214-020823>
- [28] M.C. Sorkun, A. Khetan, S. Er, AqSolDB, a curated reference set of aqueous solubility and 2D descriptors for a diverse set of compounds, *Sci. Data.* 6 (2019) 1-8. <https://doi.org/10.1038/s41597-019-0151-1>
- [29] S. Li, L. Ma, M. Zhou, Y. Li, Y. Xia, X. Fan, C. Cheng, H. Luo, New opportunities for emerging 2D materials in bioelectronics and biosensors, *Curr. Opin. Biomed. Eng.* 13 (2020) 32-41. <https://doi.org/10.1016/j.cobme.2019.08.016>

- [30] N. Briggs, S. Subramanian, Z. Lin, X. Li, X. Zhang, K. Zhang, K. Xiao, D. Geohegan, R. Wallace, L.Q. Chen, A roadmap for electronic grade 2D materials, *2D Mater.* 6 (2019) 022001. <https://doi.org/10.1088/2053-1583/aaf836>
- [31] C. Moganapriya, R. Rajasekar, T. Mohanraj, V. Gobinath, P.S. Kumar, C. Poongodi, Dry machining performance studies on TiAlSiN coated inserts in turning of AISI 420 martensitic stainless steel and multi-criteria decision making using Taguchi-DEAR Approach, *Silicon.* 14 (2021) 4183-4196. <https://doi.org/10.1007/s12633-021-01202-4>
- [32] C. Moganapriya, R. Rajasekar, K. Ponappa, P.S. Kumar, S.K. Pal, J.S. Kumar, Effect of coating on tool inserts and cutting fluid flow rate on the machining performance of AISI 1015 steel, *Materials Testing.* 60 (2018) 1202-1208. <https://doi.org/10.3139/120.111271>
- [33] G.R. Bhimanapati, Z. Lin, V. Meunier, Y. Jung, J. Cha, S. Das, D. Xiao, Y. Son, M.S. Strano, V.R. Cooper, Recent advances in two-dimensional materials beyond graphene, *ACS Nano.* 9 (2015) 11509-11539. <https://doi.org/10.1021/acsnano.5b05556>
- [34] K.S. Novoselov, A.K. Geim, S.V. Morozov, D.E. Jiang, Y. Zhang, S.V. Dubonos, I.V. Grigorieva, A.A. Firsov, Electric field effect in atomically thin carbon films. *Science.* 306 (2004) 666-669. <https://doi.org/10.1126/science.1102896>
- [35] S. Hastrup, M. Strange, M. Pandey, T. Deilmann, P.S. Schmidt, N.F. Hinsche, M.N. Gjerding, D. Torelli, P.M. Larsen, A.C.R. Jensen, The computational 2D materials database: High-throughput modeling and discovery of atomically thin crystals, *2D Mater.* 5 (2018) 042002. <https://doi.org/10.1088/2053-1583/aacfc1>
- [36] M. Ashton, J. Paul, S.B. Sinnott, R.G. Hennig, Topology-scaling identification of layered solids and stable exfoliated 2D materials, *Phys. Rev. Lett.* 118 (2017) 106101. <https://doi.org/10.1103/PhysRevLett.118.106101>
- [37] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1 (2013) 011002. <https://doi.org/10.1063/1.4812323>
- [38] G. Bergerhoff, I. Brown, F. Allen, Crystallographic databases, International Union of Crystallography, Chester. 360 (1987) 77-95.
- [39] S. Gražulis, A. Daškevič, A. Merkys, D. Chateigner, L. Lutterotti, M. Quiros, N.R. Serebryanaya, P. Moeck, R.T. Downs, A.L. Bail, Crystallography Open Database (COD): An open-access collection of crystal structures and platform for world-wide

- collaboration, *Nucleic Acids Res.* 40 (2012) D420-D427.
<https://doi.org/10.1093/nar/gkr900>
- [40] A. Jain, Y. Shin, K.A. Persson, Computational predictions of energy materials using density functional theory, *Nat. Rev. Mater.* 1 (2016) 1-13.
<https://doi.org/10.1038/natrevmats.2015.4>
- [41] M.C. Sorkun, S. Astruc, J. Koelman, S. Er, An artificial intelligence-aided virtual screening recipe for two-dimensional materials discovery, *NPJ Comput. Mater.* 6 (2020) 1-10. <https://doi.org/10.1038/s41524-020-00375-7>
- [42] P.R. Regonia, C.M. Pelicano, R. Tani, A. Ishizumi, H. Yanagi, K. Ikeda, Predicting the band gap of ZnO quantum dots via supervised machine learning models, *Optik.* 207 (2020) 164469. <https://doi.org/10.1016/j.ijleo.2020.164469>
- [43] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science.* 313 (2006) 504-507. <https://doi.org/10.1126/science.1127647>
- [44] M. Becker, J. Lippel, A. Stuhlsatz, T. Zielke, Robust dimensionality reduction for data visualization with deep neural networks, *Graphical Models.* 108 (2020) 101060. <https://doi.org/10.1016/j.gmod.2020.101060>
- [45] C. Moganapriya, R. Rajasekar, P. Sathish Kumar, T. Mohanraj, V. Gobinath, J. Saravanakumar, Achieving machining effectiveness for AISI 1015 structural steel through coated inserts and grey-fuzzy coupled Taguchi optimization approach, *Structural and Multidisciplinary Optimization.* 63 (2021) 1169-1186.
<https://doi.org/10.1007/s00158-020-02751-9>
- [46] C. Moganapriya, M. Vigneshwaran, G. Abbas, A. Ragavendran, V.H. Ragavendra, R. Rajasekar, Technical performance of nano-layered CNC cutting tool inserts-An extensive review, *Mater. Today: Proc.* 45 (2021) 663-669.
<https://doi.org/10.1016/j.matpr.2020.02.731>
- [47] G.C. Nayak, R. Rajasekar, C.K. Das, Effect of SiC coated MWCNTs on the thermal and mechanical properties of PEI/LCP blend, *Composites Part A: Appl. Sci. Manuf.* 41 (2010) 1662-1667. <https://doi.org/10.1016/j.compositesa.2010.08.003>
- [48] M.N. Rastgoo, B. Nakisa, F. Maire, A. Rakotonirainy, V. Chandran, Automatic driver stress level classification using multimodal deep learning, *Expert Systems with Applications.* 138 (2019) 112793. <https://doi.org/10.1016/j.eswa.2019.07.010>
- [49] L. Romeo, J. Loncarski, M. Paolanti, G. Bocchini, A. Mancini, E. Frontoni, Machine learning-based design support system for the prediction of heterogeneous machine

- parameters in industry 4.0, *Expert Systems with Applications*. 140 (2020) 112869.
<https://doi.org/10.1016/j.eswa.2019.112869>
- [50] M. Paolanti, L. Romeo, M. Martini, A. Mancini, E. Frontoni, P. Zingaretti, Robotic retail surveying by deep learning visual and textual data, *Robot. Auto. Syst.* 118 (2019) 179-188. <https://doi.org/10.1016/j.robot.2019.01.021>
- [51] P. Hähnel, J. Mareček, J. Monteil, F.O. Donncha, Using deep learning to extend the range of air pollution monitoring and forecasting, *J. Comput. Phys.* 408 (2020) 109278. <https://doi.org/10.1016/j.jcp.2020.109278>
- [52] R.R. Blázquez, M.M. Organero, Using multivariate outliers from smartphone sensor data to detect physical barriers while walking in urban areas, *Technol.* 8 (2020) 58. <https://doi.org/10.3390/technologies8040058>
- [53] Y. Juan, Y. Dai, Y. Yang, J. Zhang, Accelerating materials discovery using machine learning, *J. Mater. Sci. Technol.* 79 (2021) 178-190.
<https://doi.org/10.1016/j.jmst.2020.12.010>
- [54] Y. Liu, T. Zhao, W. Ju, S. Shi, Materials discovery and design using machine learning, *J. Materiomics* 3 (2017) 159-177. <https://doi.org/10.1016/j.jmat.2017.08.002>
- [55] C.M. Bishop, N.M. Nasrabadi, *Pattern recognition and machine learning*, Springer, 2006.
- [56] S. Fang, M. Wang, W. Qi, F. Zheng, Hybrid genetic algorithms and support vector regression in forecasting atmospheric corrosion of metallic materials, *Comput. Mater. Sci.* 44 (2008) 647-655. <https://doi.org/10.1016/j.commatsci.2008.05.010>
- [57] W. Paszkowicz, K.D.M. Harris, R.L. Johnston, Genetic algorithms: A universal tool for solving computational tasks in Materials Science Preface, *Comput. Mater. Sci.* 45 (2009) IX-X. <https://doi.org/10.1016/j.commatsci.2008.07.008>
- [58] C.E. Mohn, W. Kob, A genetic algorithm for the atomistic design and global optimisation of substitutionally disordered materials, *Comput. Mater. Sci.* 45 (2009) 111-117. <https://doi.org/10.1016/j.commatsci.2008.03.046>
- [59] X.J. Zhang, K.Z. Chen, X.A. Feng, Material selection using an improved genetic algorithm for material design of components made of a multiphase material, *Mater. Design.* 29 (2008) 972-981. <https://doi.org/10.1016/j.matdes.2007.03.026>
- [60] T.M. Mitchell, *Machine Learning*, McGraw-hill, New York, 1997.

- [61] C.W. Coley, R. Barzilay, T.S. Jaakkola, W.H. Green, K.F. Jensen, Prediction of organic reaction outcomes using machine learning, *ACS Central Sci.* 3 (2017) 434-443. <https://doi.org/10.1021/acscentsci.7b00064>
- [62] E.D. Cubuk, S.S. Schoenholz, J.M. Rieser, B.D. Malone, J. Rottler, D.J. Durian, E. Kaxiras, A.J. Liu, Identifying structural flow defects in disordered solids using machine-learning methods, *Phys. Rev. Lett.* 114 (2015) 108001. <https://doi.org/10.1103/PhysRevLett.114.108001>
- [63] Y. Dong, C. Wu, C. Zhang, Y. Liu, J. Cheng, J. Lin, Bandgap prediction by deep learning in configurationally hybridized graphene and boron nitride, *NPJ Comput. Mater.* 5 (2019) 1-8. <https://doi.org/10.1038/s41524-019-0165-4>
- [64] Y. Zhuo, A.M. Tehrani, A.O. Oliynyk, A.C. Duke, J. Brgoch, Identifying an efficient, thermally robust inorganic phosphor host via machine learning, *Nat. Commun.* 9 (2018) 1-10. <https://doi.org/10.1038/s41467-017-02088-w>
- [65] S.K. Kauwe, J. Graser, A. Vazquez, T.D. Sparks, Machine learning prediction of heat capacity for solid inorganics, *Integrating Materials and Manufacturing Innovation.* 7 (2018) 43-51. <https://doi.org/10.1007/s40192-018-0108-9>
- [66] L. Kaufman, P.J. Rousseeuw, *Finding groups in data: An introduction to cluster analysis*, John Wiley & Sons, 2009.
- [67] M.E. Celebi, *Partitional clustering algorithms*, Springer, 2014. <https://doi.org/10.1007/978-3-319-09259-1>
- [68] N. Grira, M. Crucianu, N. Boujemaa, Unsupervised and semi-supervised clustering: A brief survey, *A review of machine learning techniques for processing multimedia content.* 1 (2004) 9-16.
- [69] A.R. Kitahara, E.A. Holm, Microstructure cluster analysis with transfer learning and unsupervised learning, *Integrating Materials and Manufacturing Innovation.* 7 (2018) 148-156. <https://doi.org/10.1007/s40192-018-0116-9>
- [70] L.L.C. Kasun, Y. Yang, G.B. Huang, Z. Zhang, Dimension reduction with extreme learning machine, *IEEE Trans. Image Process.* 25 (2016) 3906-3918. <https://doi.org/10.1109/TIP.2016.2570569>
- [71] T. Xie, A.F. Lanord, Y. Wang, Y.S. Horn, J.C. Grossman, Graph dynamical networks for unsupervised learning of atomic scale dynamics in materials, *Nat. Commun.* 10 (2019) 1-9. <https://doi.org/10.1038/s41467-018-07882-8>
- [72] A. Mardt, L. Pasquali, H. Wu, F. Noé, VAMPnets for deep learning of molecular kinetics, *Nat. Commun.* 9 (2018) 1-11. <https://doi.org/10.1038/s41467-017-02088-w>

- [73] E.Y. Lee, B.M. Fulan, G.C. Wong, A.L. Ferguson, Mapping membrane activity in undiscovered peptide sequence space using machine learning, *Proc. Natl. Acad. Sci.* 113 (2016) 13588-13593. <https://doi.org/10.1073/pnas.1609893113>
- [74] K. Tran, Z.W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO₂ reduction and H₂ evolution, *Nat. Catal.* 1 (2018) 696-703. <https://doi.org/10.1038/s41929-018-0142-1>
- [75] A. Talapatra, S. Boluki, T. Duong, X. Qian, E. Dougherty, R. Arróyave, Autonomous efficient experiment design for materials discovery with Bayesian model averaging, *Phys. Rev. Mater.* 2 (2018) 113803. <https://doi.org/10.1103/PhysRevMaterials.2.113803>
- [76] B.W. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *Biochimica et Biophysica Acta (BBA)*. 405 (1975) 442-451. [https://doi.org/10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9)
- [77] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, G.C. Tseng, Evaluation and comparison of gene clustering methods in microarray analysis, *Bioinformatics*. 22 (2006) 2405-2412. <https://doi.org/10.1093/bioinformatics/btl406>
- [78] S. Dudoit, J. Fridlyand, A prediction-based resampling method for estimating the number of clusters in a dataset, *Genome Biology*. 3 (2002) 1-21. <https://doi.org/10.1186/gb-2002-3-7-research0036>
- [79] P.M. Shenai, Z. Xu, Y. Zhao, Applications of Principal Component Analysis (PCA) in Materials Science, in: P. Sanguansat (Eds.), *Principal Component Analysis*, 2012, pp. 25-40. <https://doi.org/10.5772/37523>
- [80] M. Ayyar, M.P. Mani, S.K. Jaganathan, R. Rathanasamy, Preparation, characterization and blood compatibility assessment of a novel electrospun nanocomposite comprising polyurethane and ayurvedic-indhulekha oil for tissue engineering applications, *Biomed. Tech.* 63 (2018) 245-253. <https://doi.org/10.1515/bmt-2017-0022>
- [81] J. Li, K. Lim, H. Yang, Z. Ren, S. Raghavan, P.Y. Chen, T. Buonassisi, X. Wang, AI applications through the whole life cycle of material discovery, *Matter*. 3 (2020) 393-432. <https://doi.org/10.1016/j.matt.2020.06.011>