A high throughput molecular screening for organic electronics via machine learning: present status and perspective

Akinori Saeki* and Kakaraparthi Kranthiraja

Department of Applied Chemistry, Graduate School of Engineering, Osaka University, Suita, Osaka 565-0871, Japan

E-mail: saeki@chem.eng.osaka-u.ac.jp

Published online December 10, 2019

Organic electronics such as organic field-effect transistors (OFET), organic light-emitting diodes (OLED), and organic photovoltaics (OPV) have flourished over the last three decades, largely due to the development of new conjugated materials. Their designs have evolved through incremental modification and stepwise inspiration by researchers; however, a complete survey of the large molecular space is experimentally intractable. Machine learning (ML), based on the rapidly growing field of artificial intelligence technology, offers high throughput material exploration that is more efficient than high-cost quantum chemical calculations. This review describes the present status and perspective of ML-based development (materials informatics) of organic electronics. Although the complexity of OFET, OLED, and OPV makes revealing their structure-property relationships difficult, a cooperative approach incorporating virtual ML, human consideration, and fast experimental screening may help to navigate growth and development in the organic electronics field. © 2019 The Japan Society of Applied Physics

1. Introduction

Data science that solely relies on massive data is often called the 4th paradigm of science, after the 3rd (computer science) and the 2nd (first-principle theory) paradigms.1–4 After the launch of the Material Genome Initiative in 2011, many projects aiming for data-driven material development have begun on a worldwide level. This has been aided by state-of-the-art artificial intelligence (AI) technology that allows for high accuracy image/voice recognition and efficient advertisement spots based on consumer attribute information.5 The AI algorithm used in social applications is compatible with pattern recognition of images/spectra taken in scientific research6,7 as well as in material design,8,9 failure analysis,10 and clustering of specific functionalities.10–12 Machine learning (ML) is a technology that makes a computer (algorithm) learn the relationship between inputs and outputs and predict an output from new inputs based on the learned data. The present ML development is greatly leveraged on the improvement of computer hardware and algorithms such as the artificial neural network (ANN), support-vector machine (SVM), kernel ridge regression (KRR), Bayesian network, and random forest (RF). Despite the unclear causality (i.e. back box) of ML models,13,14 they are expected to be an alternative tool for material exploration requiring less time and money investment than other methods.

Inorganic materials are prime subjects for theory- and data-driven screening,15–17 as they show a relatively close relationship between structure (crystal and band structures) and property (electronic, thermal, and mechanical properties). Moreover, websites with large inorganic material repositories such as Cambridge Crystallographic Data Centre and AFLOW, constitute an easily accessible entry point for ML studies. For instance, ML studies have been conducted on ion transport,18 redox potential,19,20 cathode,21,22 and dendrite formation23 of a lithium ion battery. Thermolectric materials readily fit the high throughput ML and quantum mechanical calculations,24–26 because their figure-of-merits, composed of Seebeck coefficient, electrical resistance, and thermal conductivity, can be calculated from the atomic and crystal structures or retrieved from the database. Inorganic photovoltaics,27 perovskite solar cells,28–36 and magneto-electric composites37–39 are other notable fields of material survey via such an approach. In contrast to inorganic materials that are mostly assumed to have a periodic boundary condition, organic materials have a finite, small unit of a molecule, leading to confinement of charge and energy, the anisotropic intermolecular boundary formed through weak van der Waals force, and hierarchical structures on the micro-meso-macro scale.38,39 Nevertheless, the computational and ML approaches have been successful in drug discovery40 and the prediction of redox potential,41 photoabsorption,42 energetics43 and solubility44 of organic molecules, in which these properties are linked to the involved atoms, functional groups, chemical bonds, and steric features of single molecules. An energy-structure function of metal-organic framework,45 exploration of intercalating organic agents to exfoliate inorganic two-dimensional materials,46 and organic-metal complex for catalysis47 are other examples.

Organic electronics based on π-conjugated materials progressed over three decades, from the incubation period after the pioneering works of photoconductive perylene-bromine complex in 1954,48 to conducting polymer of halogen-doped polyacetylene in 1977.49 They were implemented for thin-film devices of organic photovoltaics (OPV) in 198650 and organic light-emitting diodes (OLED) in 1987.51 Marcus theory, developed from 1956–196452,53 deals with electron transfer from donor (D) to acceptor (A) and density functional theory (DFT), developed from ~1965,54 are the premier theoretical bases for calculating a charge transfer/transport among two molecules as well as the electronic property of a single molecule (the highest occupied molecular orbital: HOMO, the lowest unoccupied molecular orbital: LUMO, bandgap, oscillator strength, etc.), respectively. Owing to the emergence of high-performing computers and user-friendly software, theoretical and experimental chemists
now have easy access to pre-examination and post-experimental analysis of a target molecule. Combining Marcus theory and DFT is powerful for calculating charge transport in molecular-based organic field-effect transistors (OFET) where the size of the molecule and its molecular packing can be precisely determined by experimental approach. The complexity is considerably increased, however, when a conjugated polymer has a size, conformational distribution and OLED/OPV devices comprising heterojunction of multiple components. In particular, the bulk heterojunction (BHJ) network of OPV, composed of binary and ternary p-type and n-type semiconducting materials, is formed via (BHJ) network of OPV, composed of binary and ternary p-type and n-type semiconducting materials, is formed via complicated interactions of solubility,\(^{55,56}\) crystallinity (intermolecular interaction),\(^{57,58}\) and miscibility of involved components.\(^ {59,60}\) This situation hampers the conventional theoretical calculation for predicting device performance (external quantum efficiency: EQE for OLED and power conversion efficiency: PCE for OPV) from the physicochemical properties of single molecule.

This review discusses how theoretical and ML calculations are used to explore conjugated materials for OFET, OLED, and OPV. A thorough exploration of organic chemicals and their morphologies in condensed matter would be too vast a scope for experimental (synthesis and characterization) and computational (DFT, molecular dynamics, and kinetic Monte Carlo) approaches, whereas the ML and smart approximation in the quantum theory can be used to address these issues.\(^{61–63}\)

2. Organic field-effect transistors

OFET is a primary element driving logical circuits, such as displays, in electronic devices, as well as a pilot device for evaluating the charge carrier mobility (\(\mu\)) of an organic semiconductor in the in-plane direction (parallel to the substrate).\(^ {64,65}\) In the early stage of OFET research in the 1990s–2000s, pentacene, an acene molecule with five-fused benzenes, was used extensively for OFET studies (Fig. 1). The pentacene OFETs fabricated by thermal evaporation in a vacuum chamber exhibited hole mobilities as high as 3–6 cm\(^2\) V\(^{-1}\) s\(^{-1}\) by optimizing film processing (e.g. substrate temperature and evaporation process) and a gate insulator to ensure ordered herringbone-packing and pure, large-grains.\(^ {56–68}\) OFETs using the rubrene molecule that yields a large, flexible single crystal reached hole mobility of 15–40 cm\(^2\) V\(^{-1}\) s\(^{-1}\).\(^ {69,70}\) Subsequently, the structural modification of OFET molecules was triggered by sulfur-containing heteroacene of [1]benzo[3,2-h][1]benzo[b]thiophene (BTBT),\(^ {71}\) of which alkylated derivatives (e.g. C\(_8\)-BTBT\(^ {72,73}\) and Ph-BTBT-10\(^ {74}\)) have been applied to a solution-processed OFETs with hole mobility over 10 cm\(^2\) V\(^{-1}\) s\(^{-1}\). This high score can be explained by band-like, coherent charge transport as reported in 3,11-didecyldinaphtho[2,3-b:2′,3′-d′]benzo[1,2-b:4,5-b′]dithiophene (C\(_{160}\)-DNBDT-NW).\(^ {75}\) In addition, n-type characteristics have been observed in \(\pi\)-conjugated molecules with electron-withdrawing fluorine\(^ {66,76,77}\) and imides.\(^ {78,79}\) In addition, ambipolar (both p- and n-type) semiconductors are the key materials in a complementary circuit.\(^ {80,81}\)

Classical Marcus\(^ {52}\) and Marcus–Levich–Jortner (MLJ)\(^ {82,83}\) formulas enable calculations of the charge transfer rate and charge carrier mobility. These calculations are based on molecular geometry as determined by X-ray diffraction, reorganization energy obtained from DFT optimization of charged and neutral molecules, and transfer integral among neighboring molecules (Fig. 1). MLJ includes the quantum description of the nonclassical degrees of freedom represented by a single effective mode of frequency and Huang–Rhys parameters.\(^ {84}\) Although the dependence of mobility on the temperature and electric field strength requires a macroscopic model that incorporates energetic and spatial disorders,\(^ {85–87}\) Marcus theory has served as a basis for explaining observed mobility,\(^ {88}\) and screening candidate molecules for synthesis.\(^ {89–92}\) The successful synthesis of new molecules achieving OFET hole mobility over 10 cm\(^2\) V\(^{-1}\) s\(^{-1}\) is a notable example of computational screening by the combination of DFT and Marcus theory.\(^ {93}\) Along these lines, DFT calculations are useful for exploring dielectric response,\(^ {94}\) antiaromaticity,\(^ {95}\) and air stability.\(^ {96}\)

The early work on ML of organic semiconductors appeared in 2013, in attempts to extract semiconductors from Schiff base molecules.\(^ {97}\) In 2018–2019, the number of papers using ML for screening organic semiconductor properties has increased. Day et al. reported the classification of N-heteroacene (pyrrole-based azaphenacene) molecules using a kernel-based ML algorithm and assessed their electron mobilities.\(^ {98}\) Despite differences in the arrangement of hydrogen bond functionality, they found the predicted crystal structures of the molecules could be classified into a small number of packing types. Musil et al. showed the prediction of stability and properties of polymorphs of pentacene and azapentacene molecules by using the Gaussian process regression ML schemes based on a kernel function.\(^ {99}\) They showed the estimation of DFT lattice energy with sub-kJ mol\(^{-1}\) accuracy, which reduced the computational cost for predicting charge carrier mobility by a factor of ten compared with conventional quantum chemical calculations (Fig. 2). Galiardi et al. performed KRR ML to predict a charge transfer integral, which was subsequently used to compute hole mobility and its anisotropy through off-lattice kinetic Monte Carlo simulations (Fig. 3).\(^ {100}\) Atahan-Evrenk and Atalay generated a molecular...
library of 5631 molecules and calculated their electronic data using DFT.\textsuperscript{101) The data were then examined using the KRR and deep neural net (DNN) regression models based on graph- and geometry-based descriptors [Fig. 4(a)]. They found a high determination coefficient of 0.92 for the prediction of intramolecular reorganization energy by using DNN [Fig. 4(b)]. These results suggest that the fast ML approach can skip time-consuming DFT calculations to obtain the transfer integral and reorganization energy necessary to estimate charge carrier mobility.

In addition to material exploration, ML is also used for recognition of signal patterns from OFET and electrochemical\textsuperscript{102,103) devices. For instance, signals from a single-walled carbon nanotube FET were subjected to ML to discern purine compounds.\textsuperscript{104) Similarly, ML analyzed signals from a silicon nanowire FET to detect a volatile organic compound.\textsuperscript{105) ML studies relating to OFET have become increasingly diverse, and a new application of ML may appeal to the interdisciplinary areas of biological, chemical, and material sciences.

3. Organic light-emitting diodes

As shown in Fig. 5(a), OLED is a multilayered device that transfers holes from an anode through a hole transport layer (HTL), electrons from a cathode through an electron injection layer (EIL) and an electron transport layer (ETL), and recombines holes and electrons at the emitting material layer (EML). The OLED device structure is more complicated than OFET, and the direction of charge transport is out-of-plane (perpendicular to the substrate), the opposite of in-plane OFET. The HTL, ETL, and EIL sandwich the EML and facilitate the charge transport and injection. The first thin-film OLED reported by Tang uses green-emitting tris(8-quinolinolato)aluminum(III) (abbreviated to Alq$_3$) as an EML.\textsuperscript{51) Using conjugated polymers [e.g. polyfluorene,\textsuperscript{106)
poly(9,9-dioctylfluorene-co-bithiophene), and poly(p-phenylene vinylene) (PPV) as an EML extended the electro-luminescence (EL) color to blue, greenish yellow, and orange/red, respectively. Owing to the spin statistics, singlet ($S_1$) and triplet ($T_1$) excited states are formed in a 1:3 ratio in the charge recombination within the EML [Fig. 5(b)]. Thus, only 25% of charges associated with $S_1$ (fluorescence) contribute to EL, while the remaining 75% of charges associated with $T_1$ are mostly lost through non-radiative paths. To address this issue, heavy metal (platinum and iridium) complexes are used, as they are efficient phosphorescence emitters (emission from $T_1$) and boost the EQE of OLED devices.

The remarkable progress in OLED was made by Adachi et al. in 2012, who reported a thermally-activated delayed fluorescence (TADF) molecule composed of D and A units [e.g. 4CzIPN shown in Fig. 5(b)]. The D-A design spatially separates the HOMO and LUMO, decreases the $S_1$-$T_1$ energy offset ($\Delta E_{ST}$), and allows thermal activation from $T_1$ to $S_1$ at room temperature (i.e. reverse intersystem crossing; RISC), leading to a theoretical internal quantum efficiency (IQE) up to 100%. The key to the efficient TADF molecule is the balance between the small $\Delta E_{ST}$ and the large oscillator strength ($f$) of $S_1$-$S_0$ transition, the latter of which is directly linked to the fluorescence quantum yield and is benefited by a large spatial overlap of HOMO and LUMO. These two factors are in a trade-off relationship. Kaji and Adachi et al. reported that DACT-II (a TADF molecule) exhibited an excellent score in both $\Delta E_{ST}$ and $f$ from the DFT calculation, resulting in a remarkable IQE of 100% and EQE of 29.4%. Penfold surveyed 31 representative TADF molecules by DFT and showed a positive correlation between $\Delta E_{ST}$ and HOMO-LUMO overlap, thus confirming the tradeoff of $\Delta E_{ST}$ and $f$. Despite the precise estimation of electronic properties relating to TADF by DFT, it still requires relatively large computational cost. In contrast, ML can greatly increase the number of molecules to be screened and cut down the cost. Levine and Shu utilized genetic algorithm and identified 3792 promising candidate fluorophores from $1.26 \times 10^6$ molecules (Fig. 6).
combined ML, DFT, and human expert votes for consideration of TADF molecules [Fig. 7(a)].115) First, they searched the $1.6 \times 10^6$ molecules using an ANN algorithm; next, they screened $4 \times 10^5$ molecules by DFT, using a single-parameter figure-of-merit of the delayed fluorescence rate constants ($k_{TADF}$) comprising $\Delta E_{ST}$ and $f$ as a general guide for efficient TADF molecules [Fig. 7(b)]. Last, they selected 4 molecules by human voting, synthesized them, and characterized their OLEDs. This work demonstrated a high EQE of 22%, showing a successful example of comprehensive, high-throughput research based on the combination of ML, theory, and the human brain.

It is interesting to show how device and material parameters impact the performance of OLED devices. Woon et al. have collected these parameters of 304 blue-emitting OLED data from the literature and performed RF ML with current efficiency (cd/A) at 1000 cd m$^{-2}$ as the output.116) From feature importance analysis, they showed the triplet energy of the ETL to be the most critical feature. These works encourage ML-oriented material screening and fast optimization of complicated device elements.

4. Organic photovoltaics
The first OPV breakthrough, after the bilayer device in 1986,50) was the mixture of p- and n-type semiconductors fabricated by the thermal evaporation of small molecules reported by Hiramoto et al. in 1991117) and by the solution process of PPV and soluble fullerene of [6,6]-phenyl-C$_{61}$-butyric acid methyl ester (PCBM), reported by Heeger et al. in 1995.118) The mixed phase of electronically heterogeneous semiconductors, called BHJ, significantly enlarges the surface area of the p/n heterojunction and improves the efficiencies of charge separation and transport.119–121) P-type conjugated polymers have evolved from the homopolymers such as PPV and regioregular poly(3-hexylthiophene) (P3HT) to D-A type copolymers that have a narrow bandgap (narrow bandgap polymer: NBP). They successfully boost the PCE from $\sim$4% for homopolymers122) to $\sim$11% for NBPs.123) The advent of the non-fullerene acceptor (NFA) enables the fine tuning of HOMO/LUMO and bandgap,124–126) further enhances the PCE by over 15% in a single junction OPV.127,128)
Like OFET and OLED, DFT-based screening is a primitive approach to the large molecular space of OPV. Lin et al. performed DFT calculations of the D-A repeating unit of conjugated polymers, where the bandgap and HOMO of ~40 monomer A coupled with ~40 monomer B (totally different 780 polymer units) were mapped on the Scharber plot. Imamura et al. used a Hückel approximation model to calculate the energy level with reduced time based on the DFT results of monomer units. Combination of DFT and the tight-binding model demonstrates a less expensive simulation for estimating the electronic properties of molecules.

ML-based exploration of OPV materials has been reported since 2011. Aspuru-Guzik et al. started the Harvard clean energy project, which constructed an automated, high-throughput in silico framework using the World Community Grid by IBM. Their website (likely closed, due to the group’s movement) provides a few millions of virtual molecules with their chemical structures, energy levels determined by DFT, and estimated OPV parameters. In 2011, Hutchison et al. examined over 9 × 10^4 molecules to solve the inverse design by genetic algorithm. In 2015, Li et al. performed screening of organic dyes for a dye-sensitized solar cell using six feature selection methods with SVM. In 2017, Aires-de-Sousa et al. reportedly showed a good mean absolute error of ~0.15 V to predict HOMO and LUMO by using RF ML, which was based on the DFT database of ~1 × 10^4 molecules. LUMO and optical bandgap energies were predicted by Jørgensen et al. in 2018, who used a grammar variational autoencoder to analyze the simplified molecular-input line-entry system (SMILES) string of ~4 × 10^3 molecules with their DFT database (Fig. 9).

In 2018, our group reported RF ML of polymer:PCBM OPV based on the ~1.2 × 10^3 experimental data manually taken from the literature. The RF model used the structural properties of molecular access system or extended connectivity fingerprint (ECFP) key encoded from the SMILES and the material properties (HOMO, bandgap, and weight-averaged molecular weight: $M_w$) as the inputs. This RF model showed improved prediction accuracy of PCE compared to the ANN model (Fig. 10). This RF model was used to screen the conjugated polymers, and demonstrating that it was capable of suggesting optimal alkyl chains for the specific polymer. This was a notable step toward efficient material design, because conventional DFT calculations cannot distinguish the effect of insulating alkyl chains on bulk properties except for a possible effect of steric hindrance. The fast experimental screening of semiconductors by the electrode-less Xe-flash time-resolved microwave

The chemical structure of a typical NBP (poly[2,6-(4,4-bis-(2-ethylhexyl)-4 H-cyclopenta[2,1-b; 3,4-b’] dithiophene)-alt-4,7-(2,1,3-benzothiadiazole)]: PCPDTBT) for OPV is shown in Fig. 8. It consists of a D unit, A unit, and alkyl chains. The D and A units are the primary building blocks that constitute the electronic nature of the conjugated backbone. Although the alkyl side chains do not significantly impact the properties of the polymer backbone, optimization of their length and structure (linear or branch) is very important, because they largely alter the solubility, crystallinity, orientation (face-on or edge-on) of polymers and miscibility with n-type semiconductors [e.g. PCBM and 5,5,11,11-tetrakis(4-hexylphenyl)-dithieno[2,3-b:2′,3′-d′]s-indaceno[1,2-b:5,6-b′]dithiophene (ITIC) as an A-D-A type NFA, see Fig. 8 for their chemical structures]. The hierarchical structure in BHJ films, i.e. the crystallite and amorphous phases, interface of p- and n-type materials, and mesoscale bi-continuous network, greatly influence the performance of OPV devices. The complexity of the OPV arises not only from the freedom of molecular design (choice of D, A, and alkyl chains), but also from the active layer morphology partly associated with processing (e.g. solvent, additive, thermal annealing, and p/n blend ratio).

![Fig. 8.](Color online) Schematic of an OPV device. Representative materials: a DA NBP of PCPDTBT, a soluble fullerene of PCBM, and a NFA (ITIC) are shown.

![Fig. 9.](Color online) The grammar variational autoencoder maps strings to/from vector space in a number of steps. Reproduced with permission from Ref. 145, Copyright: 2018 American Institute of Physics.)
conductivity technique\textsuperscript{148} used in this study is also useful, because it can predict the PCE and explore the processing condition without tedious fabrication of devices. This RF-ML approach was further applied to polymer-NFA OPVs and showed a good correspondence between the ML prediction (PCE = 11.2\%) and the experiment (the maximum PCE = 11.0\%) for a new polymer and NFA.\textsuperscript{149} Recently, Lee examined the ML algorithm for a ternary-blend polymer-fullerene OPV based on 124 data manually constructed from the literature and showed that RF yielded the highest accuracy.\textsuperscript{150}

For the ML of OPV materials, the choice of input parameters (i.e. descriptors) and models (e.g. ANN, KRR, and RF) greatly varies in output accuracy. Troisi and Ma et al. used 13 parameters as the input, such as ionization potential, dipole, HOMO-HOMO offset, and LUMO-LUMO offset, which were obtained from DFT calculations. As a consequence, a good Pearson’s coefficient of 0.79 for the PCE prediction was obtained by using the gradient boosting model (Fig. 11).\textsuperscript{151} Troisi et al. examined applicability of electronic and structural parameters, type of fingerprints, and ML models (nearest neighbors and KRR) to a polymer: fullerene OPV.\textsuperscript{152} They indicated that the inclusion of both electronic and structural parameters with KRR and Morgan fingerprint (similar to ECFP) showed the highest Pearson’s correlation coefficient.

BHJ morphology is visualized through atomic force microscopy (AFM) and transmission electron microscopy; however, researchers usually quantify them only by surface roughness, cross section, and particle size. Persson et al. developed an automated image analysis (called GTFiber, available from the website) that can distinguish the size, direction, and crystallinity of the polymer fibers in AFM images (Fig. 12).\textsuperscript{153} Salvador and Brabec et al. reported a thermodynamic figure-of-merit regarding the polymer:fullerene blend by combining the solubility parameter predicted from ANN and Flory-Huggings theory.\textsuperscript{154} Fink and Ameri et al. used ANN and DFT to determine the Hildebrand solubility parameters and analyzed the ternary blend morphology of a polymer:PCBM:sensitizer reconstructed from energy-filtered TEM and resonant soft X-ray scattering measurements.\textsuperscript{155}

The general problem for material scientists is the large amount of time and effort needed to find the optimal parameters through one-variable-at-a-time experimentation.

Fig. 10. (Color online) (a) Scheme of converting a chemical structure to digitized data (fingerprints are either MACCS or ECFP). (b) RF result (5-fold cross validation, CV) of polymer: fullerene (FA) acceptor OPV by using ECFP as the fingerprint. The horizontal and vertical axes represent the normalized experimental PCE and predicted PCE, respectively, and $r$ is the correlation coefficient. Reproduced with permission from Ref. 146, Copyright: 2018 American Chemical Society.

Fig. 11. (Color online) Theoretically predicted versus experimental PCE for the testing set (30 molecules) (a) for all data points using the leave-one-out cross-validation technique and (b) for the GB model. Inset shows probability density of prediction errors. Reproduced with permission from Ref. 151, Copyright: 2018 Wiley-VCH.

Fig. 12. (Color online) (a) TEM image of P3HT film. (b) Colored orientation maps obtained from the image analysis of the film. The size is $5 \times 5 \, \mu m^2$. Reproduced with permission from Ref. 153, Copyright: 2017 American Chemical Society.
Adutwum et al. mentioned in their perspective that design of experiments (DoE) is suited for multivariable analysis such as a complex system of OPVs. The DoE powered by ML may allow a researcher in the laboratory to reduce their time for screening variables, thus preserving the remaining time for a more productive purpose. In that sense, ML could be a ubiquitous tool aiding extremely high-throughput material screening, efficient planning of experiments, team management, and decision-making for next research targets.

Despite the wide use of ML in organic electronics, some issues have still remained such as unclear causality (depending on the algorithm), insufficient prediction accuracy (depending on the target) due to the complexity of the system, and the small number of dataset with a high quality. Thus, manual consideration by researchers and data collection via more efficient calculations and experiments are still important for the ML-oriented studies.

5. Concluding remarks

The recent progress and perspective of ML- and theory-driven exploration of organic semiconductors has been reviewed. DFT and Marcus theory are undoubtedly powerful for evaluating electronic properties (HOMO/LUMO, bandgap, reorganization energy, etc.) and charge transport properties (transfer integral) of OFET/OLED/OPV materials at the pre- and post-experiment stages. The ML approach enables prediction of these properties without high-cost quantum chemical calculations, based on the massive data. Attempts to predict device performance (e.g. PCE of OPV) from the experimental or DFT database have begun to be published, particularly in the past three years (2017–2019). For ML studies, the size and quality of data, choice of ML algorithm (e.g. ANN, KRR, SVM, RF), and the use of good descriptors (material and structural properties) are important for obtaining high prediction accuracy. In addition to material exploration, ML is utilized for pattern recognition of electric signals from OFET, image analysis of BHJ film, and experiment planning in reduced time. Therefore, ML can facilitate the progress of organic electronics and widen its horizon in the future.

Acknowledgments

This work was supported by the PRESTO program (Grant No. JPMIPR15N6) of the Japan Science and Technology Agency (JST), the Japan Society for the Promotion of Science (JSPS) through a KAKENHI Grant-in-Aid for Scientific Research (A) (Grant No. JP16H02285). K. K. acknowledges the JSPS postdoctoral fellowships for research in Japan (Grant No. JP18F18035).

1) S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, Nature Mater. 12, 191 (2013).
2) A. Agrawal and A. Choudhary, APL Mater. 4, 053208 (2016).
3) K. Takahashi and Y. Tanaka, Dalton Trans. 45, 10497 (2016).
4) K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, Nature 559, 547 (2018).
5) M. I. Jordan and T. M. Mitchell, Science 349, 255 (2015).
6) E. Strelcov, A. Belianinov, Y.-H. Hsieh, S. Jesse, A. P. Baddorff, Y.-H. Chu, and S. V. Kalinin, ACS Nano 8, 6446 (2014).
7) S. Kayohara, T. Miyata, K. Tsuda, and T. Mizoguchi, Sci. Rep. 8, 13548 (2018).
8) G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Chem. Mater. 22, 3762 (2010).
136) M. C. Scharber, D. Mühlbacher, M. Koppe, P. Denk, C. Waldauf, A. J. Heeger, and C. J. Brabec, Adv. Mater. 18, 789 (2006).
137) Y. Imamura, M. Tashiro, M. Katouda, and M. Hada, J. Phys. Chem. C 121, 28275 (2017).
138) M. Mesta, J. H. Chang, S. Shil, S. Thygesen, and J. M. G. Lastra, J. Phys. Chem. A 123, 4980 (2019).
139) J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, and A. Aspuru-Guzik, J. Phys. Chem. Lett. 2, 2241 (2011).
140) R. Olivares-Amaya, C. Amador-Bedolla, J. Hachmann, S. Atahan-Evrenk, R. S. Sánchez-Carrera, L. Vogt, and A. Aspuru-Guzik, Energy Environ. Sci. 4, 4849 (2011).
141) E. O. Pyzer-Knapp, K. Li, and A. Aspuru-Guzik, Adv. Funct. Mater. 25, 6495 (2015).
142) I. Y. Kanal, S. G. Owens, J. S. Bechtel, and G. R. Hutchison, J. Phys. Chem. Lett. 4, 1613 (2013).
143) H. Li, Z. Zhong, L. Li, R. Gao, J. Cui, T. Gao, L. H. Hu, Y. Lu, Z.-M. Su, and H. Li, J. Comput. Chem. 36, 1036 (2015).
144) F. Pereira, K. Xiao, D. A. R. S. Latino, C. Wu, Q. Zhang, and J. Aires-de-Sousa, J. Chem. Inf. Model. 57, 11 (2017).
145) P. B. Jørgensen, M. Mesta, S. Shil, J. M. G. Lastra, K. W. Jacobsen, K. S. Thygesen, and M. N. Schmidt, J. Chem. Phys. 148, 241735 (2018).
146) S. Nagasawa, E. Al-Naamani, and A. Saeki, J. Phys. Chem. Lett. 9, 2639 (2018).
147) D. Rogers and M. Hahn, J. Chem. Inf. Model. 50, 742 (2010).
148) A. Saeki, S. Yoshikawa, M. Touji, Y. Koizumi, M. Ide, C. Vijayakumar, and S. Seki, J. Am. Chem. Soc. 134, 19035 (2012).
149) Y.-C. Lin et al., J. Mater. Chem. A 7, 3072 (2019).
150) M.-H. Lee, Adv. Energy Mater. 9, 1900891 (2019).
151) H. Sahu, W. Rao, A. Troisi, and H. Ma, Adv. Energy Mater. 8, 1801032 (2018).
152) D. Padula, J. D. Simpson, and A. Troisi, Mater. Horiz. 6, 343 (2019).
153) N. E. Persson, J. Rafshoon, K. Naghshpour, T. Fast, P.-H. Chu, M. McBride, B. Risteen, M. Grover, and E. Reichmanis, ACS Appl. Mater. Interfaces 9, 36090 (2017).
154) J. D. Perea et al., J. Phys. Chem. C 121, 18153 (2017).
155) X. Du et al., Macromolecules 50, 2415 (2017).
156) B. Cao, L. A. Adutwum, A. Oliynyk, E. J. Luber, B. C. Olsen, A. Mar, and J. M. Buriak, ACS Nano 12, 7434 (2018).