

ARTICLE OPEN



Deep learning for development of organic optoelectronic devices: efficient prescreening of hosts and emitters in deep-blue fluorescent OLEDs

Minseok Jeong^{1,2}, Joonyoung F. Joung^{1,2}, Jinhyo Hwang^{1,2}, Minhi Han^{1,2}, Chang Woo Koh¹, Dong Hoon Choi¹✉ and Sungnam Park¹✉

The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, which are key factors in optoelectronic devices, must be accurately estimated for newly designed materials. Here, we developed a deep learning (DL) model that was trained with an experimental database containing the HOMO and LUMO energies of 3026 organic molecules in solvents or solids and was capable of predicting the HOMO and LUMO energies of molecules with the mean absolute errors of 0.058 eV. Additionally, we demonstrated that our DL model was efficiently used to virtually screen optimal host and emitter molecules for organic light-emitting diodes (OLEDs). Deep-blue fluorescent OLEDs, which were fabricated with emitter and host molecules selected via DL prediction, exhibited narrow emission (bandwidth = 36 nm) at 412 nm and an external quantum efficiency of 6.58%. Our DL-assisted virtual screening method can be further applied to the development of component materials in optoelectronics.

npj Computational Materials (2022)8:147; <https://doi.org/10.1038/s41524-022-00834-3>

INTRODUCTION

Since the molecular orbital theory was proposed in the 20th century, the concepts of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) have been used in various research areas^{1–8}. For example, Fukui et al. found that π -electrons in the HOMO played decisive roles in the reaction of aromatic hydrocarbons and extended this result to the concept of frontier molecular orbitals (i.e., the HOMO and LUMO)^{9,10}. In organic photovoltaics (OPVs), the power conversion efficiency (PCE) can be significantly increased by optimizing the frontier molecular orbital energies of the component materials^{11,12}. Shockley and Queisser suggested that the optimal bandgap in light-harvesting materials is approximately 1.3 eV, which is based on a compromise between the short-circuit current (J_{sc}) and open-circuit voltage (V_{oc}) made to maximize the PCE¹¹. In organic light-emitting diodes (OLEDs), the HOMO and LUMO energies are crucial factors when designing new component materials. For example, ideal host materials should have proper HOMO-LUMO energy gaps, which are required for sufficient spectral overlaps with emitters for efficient energy transfer¹³. An appropriate alignment of HOMO and LUMO energy levels of component materials in OLEDs is required to transport charge carriers to the emitting layer and trap them there, leading to a high exciton recombination yield^{14,15}.

To design and develop new materials, the molecular properties such as HOMO and LUMO energies need to be accurately estimated. The density functional theory (DFT) calculations have been extensively used in many research areas to calculate molecular properties^{16–21}. Recently, deep learning (DL) methods based on big-data have emerged as a promising solution for reliable estimation of molecular properties with substantially reduced computational costs^{22–39}. The DL methods based on the databases of DFT-calculated molecular properties including

HOMO and LUMO energies have been reported^{8,22,40–49}, but they have limitations in practical applications for the development of optimal molecules in various research areas because of the following reasons: (1) most databases obtained by DFT calculations contain relatively small molecules^{50–53}, (2) DFT-calculated values are different from actual experimental ones⁵⁴, and (3) molecule-environment interactions are not included in most currently available databases obtained by DFT-calculations. The molecular properties such as absorption and emission wavelengths, bandwidths, and HOMO and LUMO energies are significantly influenced by the molecule-environment interactions. For practical chemical applications, these limitations of previously-developed DL methods based on DFT-calculated database need to be overcome.

In this work, we built an experimental database containing the HOMO and LUMO energies of various molecules via direct data collection from the literature. We then developed a DL model trained using an experimental database to quickly and reliably predict HOMO and LUMO energies close to the experimental values. The architecture of DL model is based on our previously developed DL optical spectroscopy (DLOS) model, which has been proven to accurately predict optical and photophysical properties influenced by the local environments⁵⁵. When compared with the DFT calculations, our DL model exhibited better performance in terms of computational time and prediction error. Interestingly, our DL model was found to rationally recognize the effects of donor-acceptor structures, substituents, conjugation length, heteroatoms, and heavy atoms on the HOMO and LUMO energies. Finally, we demonstrated DL-assisted virtual screening for developing emitter and host molecules for ultra-deep-blue fluorescent OLEDs.

¹Department of Chemistry and Research Institute for Natural Science, Korea University, Seoul 02841, Korea. ²These authors contributed equally: Minseok Jeong, Joonyoung F. Joung, Jinhyo Hwang, Minhi Han. ✉email: dhchoi8803@korea.ac.kr; spark8@korea.ac.kr

RESULTS AND DISCUSSION

Experimental database

Our experimental database includes the HOMO and LUMO energies of 3026 organic molecules in solvent or solids, yielding 3362 molecule/solvent combinations, and our data collection procedure is described in the Methods in detail. Figure 1a shows the distributions of HOMO and LUMO energies of molecules in our experimental database. In our experimental database, the numbers of datapoints are 2990 and 3077 datapoints for the HOMO and LUMO energies, respectively. Figure 1b shows the solvents used for the measurement of the HOMO and LUMO energies of molecules. Dichloromethane is found to be the most frequently used solvent for cyclic voltammetry (CV). The experimental errors of the HOMO and LUMO energies are 0.089 and 0.112 eV, respectively (See the Methods for details).

In Fig. 1c, the molecular weight distribution in our experimental database is compared with that in the QM9 database which is one of DFT-calculated databases. Our experimental database has a much broader molecular weight distribution than the QM9 database and contains much larger molecules that have been developed for practical use in many research areas (Supplementary Fig. 1)^{56–59}.

In our experimental database of HOMO and LUMO energies, 598 molecule/solvent combinations are the same combinations in our previously reported experimental database of optical properties⁶⁰. Using the same combinations in these two databases, the bandgap (the difference of the HOMO and LUMO energies for a given molecule) is plotted against the electronic transition energies (i.e., absorption and emission energies) in Fig. 1d. The bandgap is shown to be correlated reasonably well with both the absorption and emission energies, and the bandgap is smaller than the absorption energy but is larger than the emission energy. Additionally, the LUMO energy is correlated with the absorption and emission energies, but the HOMO energy is not correlated with them, as shown in Supplementary Fig. 2.

Deep learning model

Our DL model is basically based on the graph convolutional network (GCN) and its architecture is schematically illustrated in Fig. 2a. For the inputs of our DL model, the molecule and solvent are represented as the molecular graph which can include up to 150 atoms (without hydrogen atoms) to cover most of practically used organic molecules. The detail description of the molecular graph and our DL model can be found in the Methods and elsewhere⁵⁵. Note that the solvent is included in our DL model because the molecular orbital energies are affected by the local environment of a given molecule as will be discussed in the next section in great detail. In this work, we used the transfer learning-based DL model to predict the HOMO and LUMO energies because the bandgap and LUMO energy were correlated with the absorption and emission energies (Fig. 1d) which were well predicted by our previously developed DLOS. Transfer learning methods are DL techniques that reuse a model developed for a task as a starting point for training a new model on a similar task^{61–66}. Here, we compared two learning methods, i.e., learning from scratch (LS) and transfer learning (TL) methods. In our TL method, our DL model was trained with the experimental database of HOMO and LUMO energies using the pre-trained parameters of our previously developed DLOS as initial conditions (See the Method for details)⁵⁵. The TL method has been shown to be more effective with small datasets. The performance of our DL model was examined by varying the training dataset size as shown in Fig. 2b, Supplementary Tables 1 and 2. The mean absolute errors (MAEs) of the TL method are always smaller than those of the LS method, which indicates that our current DL model benefits from the pre-trained parameters of our previously reported DLOS model. When the dataset size is 50, the MAE of the TL method is about 10.8% smaller than that of the LS method. When the dataset size is ~2000, the MAE of the TL method is only 4.8% smaller than that of the LS method. As the dataset size is increased, the MAEs of TL and LS methods are found to get closer to the experimental error of ~0.1 eV.

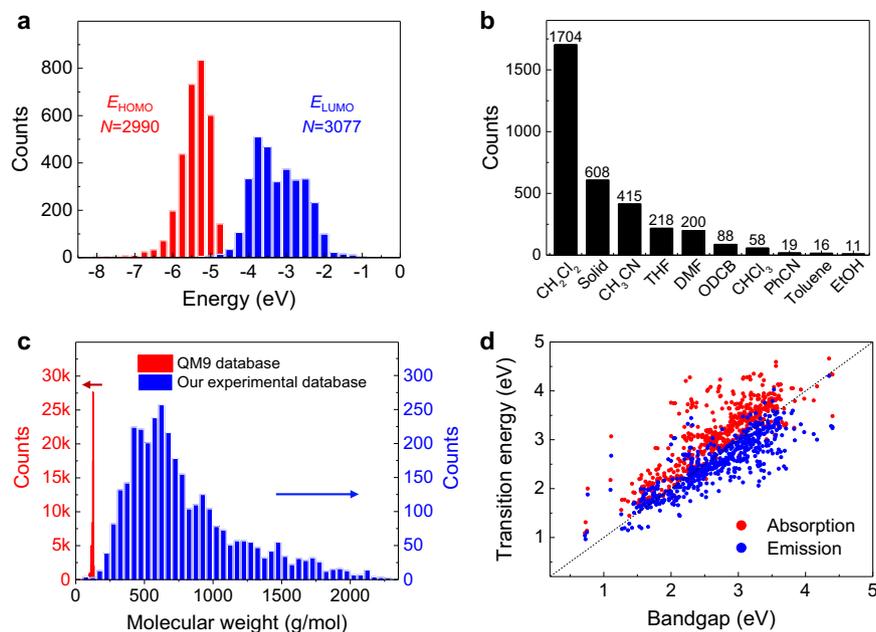


Fig. 1 Experimental database of HOMO and LUMO energies. **a** Distributions of HOMO and LUMO energies (E_{HOMO} and E_{LUMO}) of molecules in our experimental database. The number of datapoints (N) is indicated. **b** Histogram of solvents (CH_2Cl_2 : dichloromethane, CH_3CN : acetonitrile, THF: tetrahydrofuran, DMF: *N,N*-dimethylformamide, ODCB: 1,2-dichlorobenzene, PhCN: benzonitrile, and EtOH: ethanol). **c** Distributions of the molecular weights of molecules in the QM9 database (in red) and our experimental database (in blue). **d** Plot of the bandgap (E_g) vs. the transition energy from our experimental database.

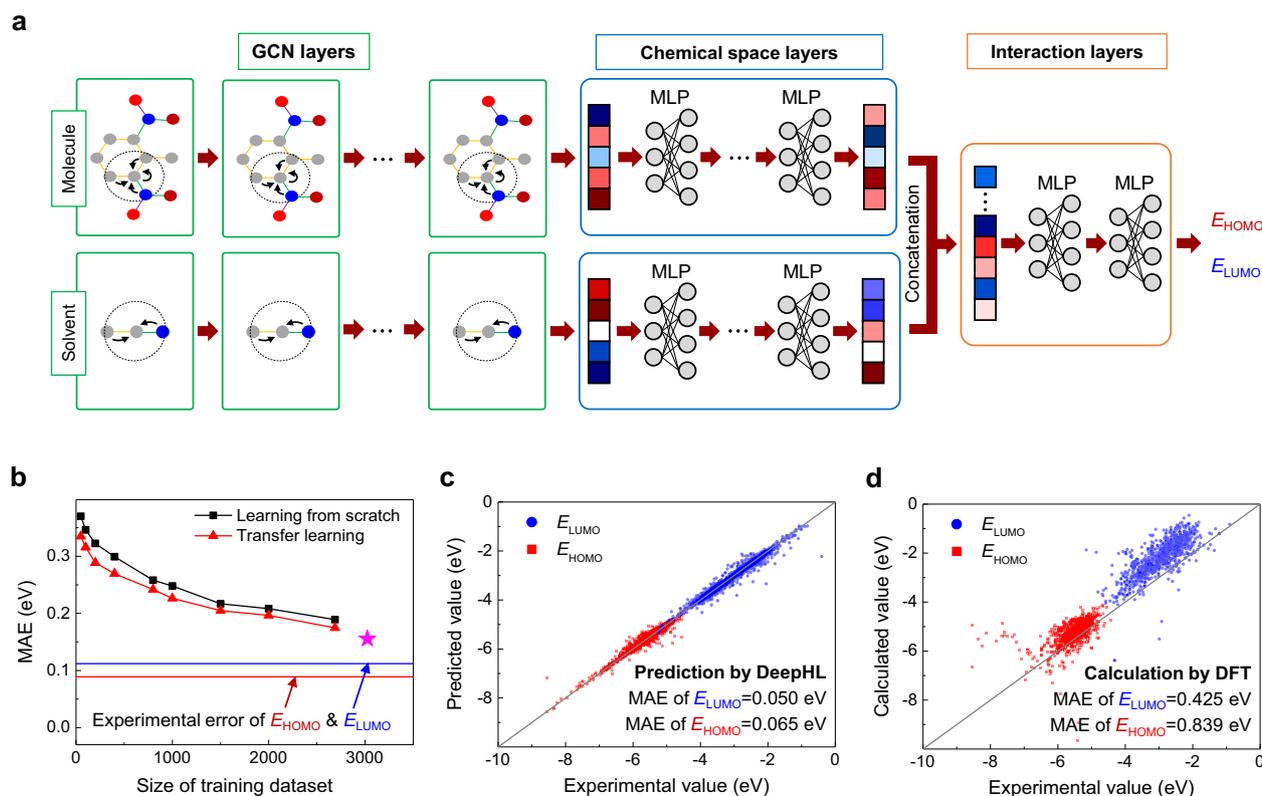


Fig. 2 Deep learning model. **a** Schematic illustration of our DL model used to predict the HOMO and LUMO energies (E_{HOMO} and E_{LUMO}) of 2,6-dinitrotoluene in acetonitrile. See the Methods for the detailed architecture of our DL model. **b** The mean absolute errors (MAEs) in the test dataset as a function of the training dataset size. The MAE of our DeepHL model is indicated by a magenta star. The solid lines represent the experimental errors of HOMO (red) and LUMO (blue) energies, respectively. **c** HOMO and LUMO energies of all molecules in our database predicted by our DL model (DeepHL). **d** HOMO and LUMO energies of 988 organic molecules, calculated using DFT methods (B3LYP/6-31 G(d)).

Our final DL model (referred to as ‘DeepHL’), which was re-trained with the combined training and validation datasets, shows high performance in predicting the HOMO and LUMO energies with relatively small MAEs, as shown in Fig. 2c. In addition, the distributions of the prediction error in the training and test datasets are presented in Supplementary Fig. 3. The MAEs of the HOMO and LUMO energies for the test dataset are 0.148 and 0.163 eV, respectively, which are very close to the experimental errors of HOMO and LUMO energies. For the total dataset, the MAEs for the HOMO and LUMO energies are 0.050 and 0.065 eV, respectively.

To compare the accuracy of DeepHL and theoretical methods, 988 organic molecules were randomly selected from our database, and their HOMO and LUMO energies were calculated using the density functional theory (DFT) calculations with the B3LYP functional and 6–31 G(d) basis set, as implemented in the Gaussian 16 software package⁶⁷. As shown in Fig. 2d, the MAEs of the DFT-calculated HOMO and LUMO energies are 0.425 and 0.839 eV, respectively, which are much larger than the DeepHL-prediction errors of the test dataset (0.148 and 0.163 eV for the HOMO and LUMO energies, respectively). Furthermore, DeepHL is superior to DFT calculations in terms of computation time, taking only 0.82 s to predict the HOMO and LUMO energies of 338 molecules (in the test dataset). Note that the HOMO and LUMO energies predicted by DFT calculations depend on DFT functionals and basis sets. In fact, it has been reported that DFT-calculated HOMO and LUMO energies deviate from the experimental values with a relatively large error (i.e., the smallest DFT-calculation error = 0.77 eV)⁵⁴. In this work, B3LYP/6-31 G(d) was used because it was commonly used to calculate large organic materials in

OLEDs and solar cells, and also used to build computational databases including PubChemQC⁶⁸.

It should be noted that several DL methods based on DFT-calculated HOMO and LUMO energies have recently been reported as summarized in Supplementary Table 3. The database size of DFT-calculated HOMO and LUMO energies ranges from ~7000 to ~133,000. The DL models trained with the QM9 database show superior accuracy as shown in Supplementary Table 3^{40–44,48,69}. However, they are not useful for practical applications in designing and developing new materials in many research areas because the QM9 database contains only small molecules in which the number of atoms is limited up to 9 including C, N, O, S, and halogens except hydrogen atoms⁵¹. For example, hexafluoropropane ($\text{C}_3\text{H}_2\text{F}_6$, molecular weight = 152 g/mol) is the heaviest molecule in the QM9 database (Supplementary Fig. 1). By contrast, our experimental database contains more practically used molecules than DFT-calculated databases (Supplementary Fig. 1). In addition, the DL models based on DFT-calculated databases might inherit the calculation errors as shown in Fig. 2d, and might not be able to predict the values close to the actual experimental values. In this respect, DeepHL based on the experimental database is substantially more useful and practical in developing new materials than the DL models trained with DFT-calculated databases.

Performance of our DeepHL

Trained with the experimental database, DeepHL can predict the HOMO and LUMO energies of various molecules widely used in optoelectronics. In this section, we will show how well our DL model (DeepHL) is trained and predicts HOMO and LUMO energies based on the different structural moieties of molecules in our experimental database. Figure 3 summarizes four different groups

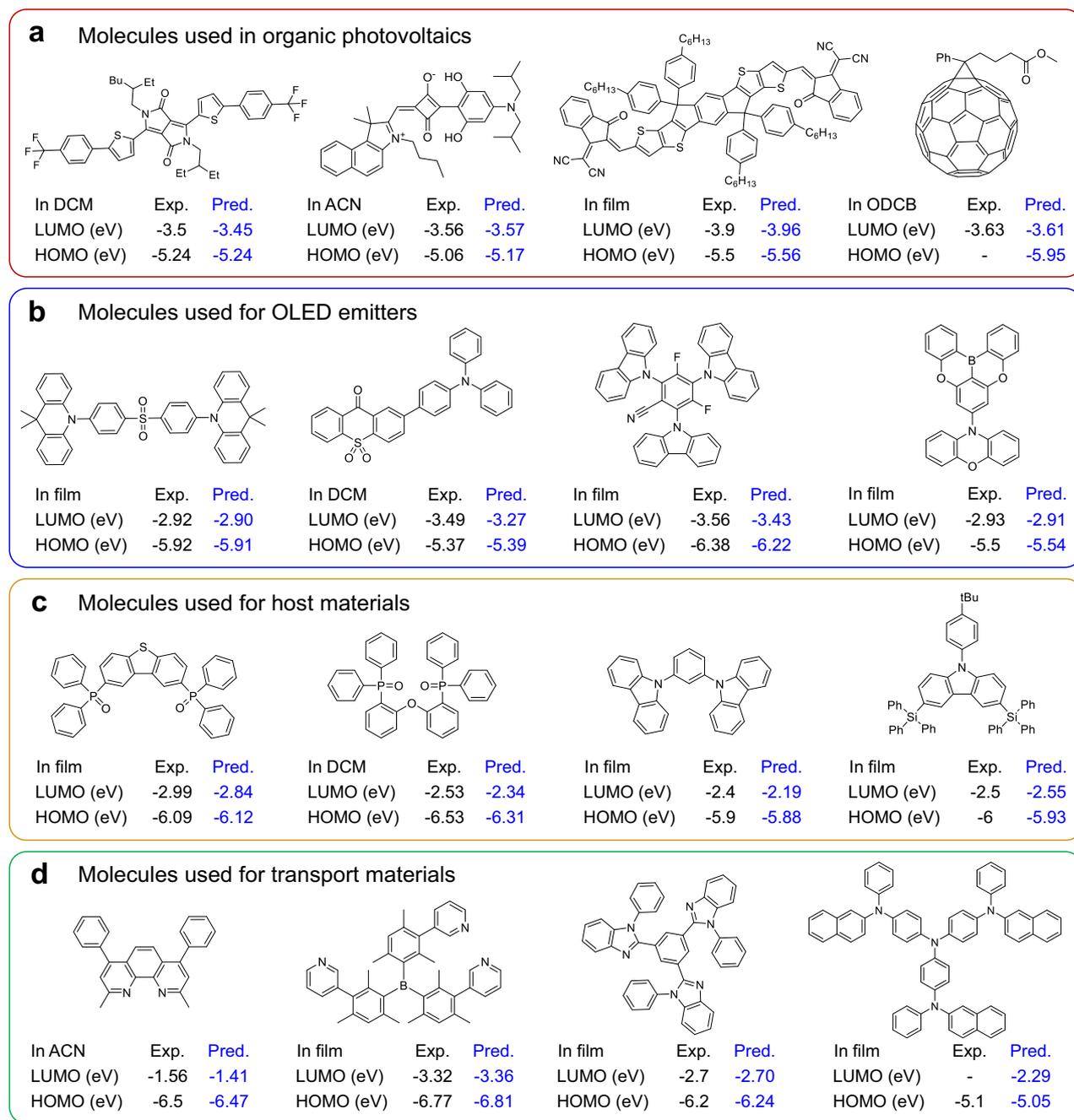


Fig. 3 Experimental and DeepHL-predicted HOMO and LUMO energies for the selected molecules in our experimental database. a Molecules used in organic photovoltaics. **b** Molecules used for OLED emitters. **c** Molecules used for host materials. **d** Molecules used for transport materials.

of molecules used in OPVs and OLEDs along with their experimental and DeepHL-predicted HOMO and LUMO energies. For example, the two molecules on the left in Fig. 3a were developed for electron donors in OPVs^{70,71}, and their HOMO energies were accurately predicted by DeepHL. Additionally, the ITIC and PC₆₀BM on the right in Fig. 3a, which are the electron acceptor, had their LUMO energies accurately predicted by DeepHL as well^{72,73}. Figures 3b and c show emitters with various core structures and host molecules commonly used in OLEDs, respectively^{74–81}, and their HOMO and LUMO energies were accurately predicted by DeepHL. Additionally, the molecules shown in Fig. 3d are frequently used as the electron and hole transport layers in the

OLEDs^{82–85}, and their HOMO and LUMO energies are shown to have been accurately predicted within the MAEs.

It is interesting to note that the high accuracy of DeepHL results from an understanding of the effects of molecular structures on the HOMO and LUMO energies, including donor–acceptor structures, functional groups, heteroatoms and heavy atoms, and conjugation length. Engineering of donor (D)–acceptor (A) structures is widely used as a design strategy to tune the HOMO and LUMO energies of given molecules. As shown in Fig. 4a, DeepHL can predict the HOMO and LUMO energies of a given D–A type molecule by identifying donor and acceptor moieties in the molecule. For example, three molecules containing triazine (TRZ)

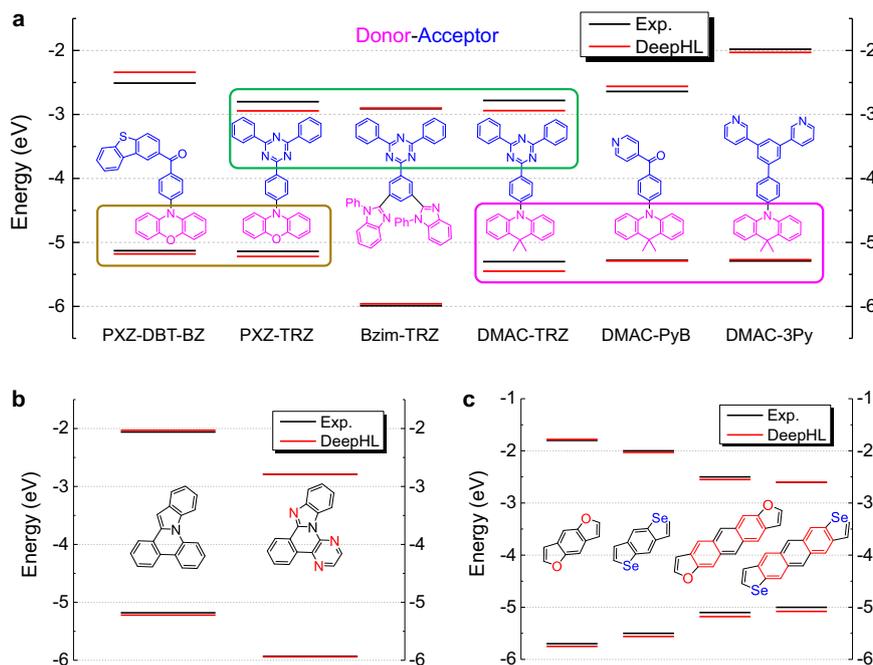


Fig. 4 Experimental and DeepHL-predicted HOMO and LUMO energies. **a** Donor-acceptor-type molecules. Identical moieties are grouped in the colored boxes. **b** The same core structure with a different number of nitrogen atoms. **c** Molecules with oxygen or selenium atoms and different conjugation lengths.

as an acceptor in Fig. 4a were predicted to have similar LUMO energies which agree well with the experimental values within the prediction error. Molecules having the same donor moieties, such as dimethyl acridine (DMAC) and phenoxazine (PXZ), are also predicted to have similar HOMO energies. As shown in Supplementary Fig. 5a, DeepHL predicts the HOMO and LUMO energies of DMAC-TRZ as a combination of the HOMO energy of DMAC and the LUMO energy of TRZ. This result agrees well with the DFT calculations shown in Supplementary Fig. 5, indicating that the HOMO and LUMO in DMAC-TRZ are located at DMAC and TRZ, respectively. Additionally, the HOMO and LUMO energies of D-A-D- and A-D-A-type molecules can be accurately predicted. As shown in Supplementary Fig. 6, the LUMO energies of molecules with A-D-A structures of the same acceptor moieties are predicted to be almost the same as the experimental values, whereas the HOMO energies vary with different donor moieties. In the cases of the D-A-D molecules shown in Supplementary Fig. 7, it is observed that the LUMO energies are predicted to gradually decrease as the number of 1,3,4-thiadiazole acceptor moieties is increased.

DeepHL is also found to accurately represent the effects of functional groups, heteroatoms and heavy atoms, and conjugation on the HOMO and LUMO energies. Functional groups with strong electron-donating ability increase the electron density of the molecules and cause more electron-electron repulsion, leading to increased HOMO and LUMO energies. At the same time, electron-withdrawing groups reduce the HOMO and LUMO energies. The dimethyl amine and nitro groups shown in Supplementary Fig. 8 are electron-donating and electron-withdrawing groups, respectively. The effects of substituents on the HOMO and LUMO energies are precisely reflected in the predicted values. The molecules shown in Fig. 4b have a different number of heteroatoms (nitrogen)⁸⁶. In Fig. 4b, it is clearly shown that the experimentally measured HOMO and LUMO energies decrease as the number of nitrogen atoms is increased due to the strong inductive effect⁸⁷. As demonstrated in Fig. 4b, DeepHL can reproduce the tendency of nitrogen atoms to cause a decrease of HOMO and LUMO energies. Furthermore, DeepHL can capture how the conjugation length can change the

bandgap, similar to a particle-in-a-box model. As can be seen in Fig. 4c, DeepHL predicts that the bandgap will be decreased when the conjugation length is increased⁸⁸ or when oxygen is replaced by selenium (i.e., heavy atom)^{87,89}.

Before this section is closed, it should be highlighted that DeepHL can predict the HOMO and LUMO energies of molecules in different local environments because DeepHL includes the molecule-solvent interactions. As shown in Supplementary Table 4, the HOMO and LUMO energies in solids, dichloromethane, and *N,N*-dimethylformamide are shown to be well predicted within the MAEs.

Deep learning-assisted development of emitter and host molecules

The HOMO and LUMO energies are crucial factors when designing organic molecules for use in high-performance OLED devices. A proper alignment of the electronic energies of materials across the multiple layers of OLED devices can facilitate the transportation of charge carriers (i.e., holes and electrons) to the emitting layer and the efficient energy transfer from host molecules to emitters, achieving a high external quantum efficiency (EQE). DeepHL can be effectively used to quickly and accurately predict the HOMO and LUMO energies of component materials for OLED devices. In the following, we will demonstrate how DeepHL can be used to efficiently prescreen newly designed emitter and host molecules for deep-blue OLEDs.

Let's first consider a typical structure of OLED devices comprising indium-tin-oxide (ITO, anode), poly(3,4-ethylenedioxythiophene):poly(styrene sulfonate) (PEDOT:PSS, hole injection layer), poly(*N*-vinylcarbazole) (PVK, hole transport layer), 1,3,5-tris(2-*N*-phenylbenzimidazolyl) benzene (TPBi, electron transport layer), and lithium fluoride/aluminum (LiF/Al, cathode), as shown in Fig. 5b. For this given OLED device structure with deep-blue emission, newly designed emitters and host molecules must satisfy the following requirements. First, the HOMO (LUMO) energy of host molecules should be between the HOMO (LUMO) energies of the electron transport layer (TPBi) and hole transport

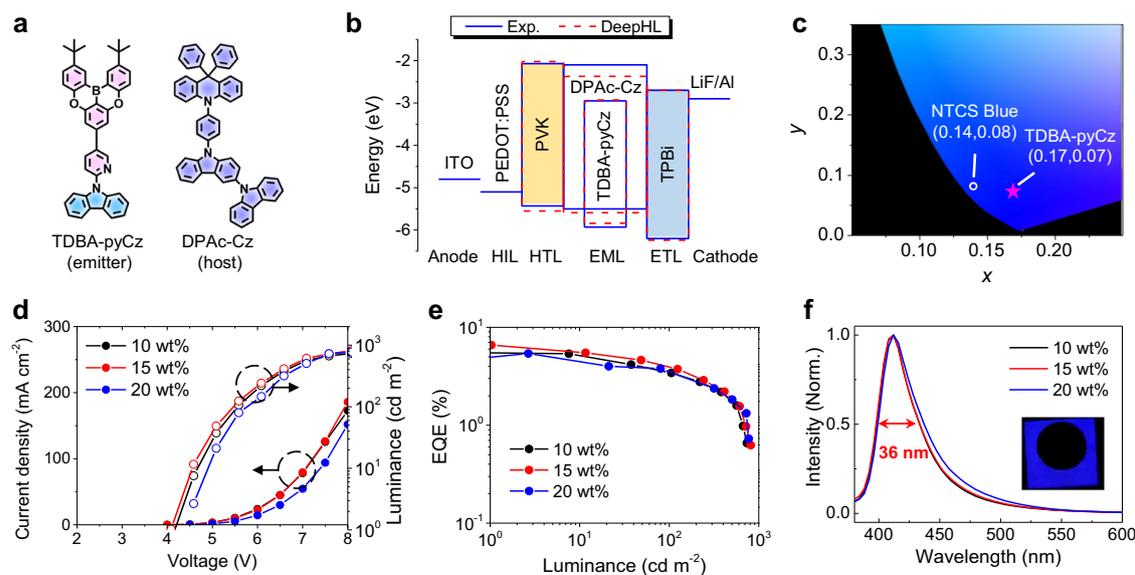


Fig. 5 Deep-blue fluorescent OLED device with TDBA-pyCz and DPAC-Cz developed by DL-assisted virtual screening. **a** Molecular structures of DPAC-Cz (host) and TDBA-pyCz (emitter). **b** Device structure. **c** CIE 1931 chromaticity diagram. **d** Current density (J)-Voltage (V)-Luminance (L) curves. **e** EQE-Luminance (L) curves. **f** EL spectra at different doping concentrations of emitters at a luminance of 500 cd m^{-2} . A photograph of the device emission is shown in the inset.

Table 1. DeepHL-predicted HOMO and LUMO energies and DLOS-predicted optical and photophysical properties of TDBA-pyCz and DPAC-Cz (Experiment / Prediction).

	HOMO (eV)	LUMO (eV)	$\lambda_{\text{abs}}^{\text{max}}$ (nm)	$\lambda_{\text{emi}}^{\text{max}}$ (nm)	Φ_{PL} (%)
TDBA-pyCz in CH_2Cl_2	-5.71^{a} / -5.67	-2.64^{a} / -2.74	387^{b} / 379^{c}	404^{b} / 426^{c}	$-^{\text{e}}$ / 62^{c}
TDBA-pyCz in toluene	$-^{\text{e}}$ / -5.71	$-^{\text{e}}$ / -2.79	387^{b} / 381^{c}	404^{b} / 416^{c}	60^{d} / 72^{c}
TDBA-pyCz film	-5.93^{a} / -5.84	-2.95^{a} / -2.93	389^{b} / 382^{c}	420^{b} / 446^{c}	18^{d} / 36^{c}
TDBA-pyCz-doped DPAC-Cz film	$-^{\text{e}}$ / -5.85	$-^{\text{e}}$ / -2.94	343^{b} / 394^{c}	415^{b} / 441^{c}	39^{d} / 74^{c}
DPAC-Cz in CH_2Cl_2	-5.15^{a} / -5.29	-1.75^{a} / -2.05	342^{b} / 320^{c}	383^{b} / 396^{c}	$-^{\text{e}}$ / 25^{c}
DPAC-Cz in toluene	$-^{\text{e}}$ / -5.36	$-^{\text{e}}$ / -2.14	342^{b} / 317^{c}	372^{b} / 385^{c}	34^{d} / 26^{c}
DPAC-Cz film	-5.50^{a} / -5.59	-2.10^{a} / -2.37	343^{b} / 329^{c}	379^{b} / 404^{c}	21^{d} / 37^{c}

^aCyclic voltammograms in Supplementary Fig. 16.

^bUV-visible absorption and fluorescence spectra in Supplementary Figs. 18 and 19.

^cValues predicted by using DLOS⁵⁵.

^dAbsolute PLQY obtained with neat/doped film of TDBA-pyCz ($\lambda_{\text{exc}} = 350 \text{ nm}$) and DPAC-Cz ($\lambda_{\text{exc}} = 320 \text{ nm}$) using an integrating sphere.

^eNot measured. $\lambda_{\text{abs}}^{\text{max}}$: Maximum absorption wavelength, $\lambda_{\text{emi}}^{\text{max}}$: Maximum emission wavelength, Φ_{PL} : Absolute photoluminescence quantum yield.

layer (PVK). Second, host molecules should have larger bandgaps than emitters to efficiently transfer the excited energy of host molecules to emitters. Third, emitters should have a bandgap larger than $\sim 2.9 \text{ eV}$ for deep-blue emission.

As host and emitter molecules, we have designed 8 and 10 molecules in Supplementary Figs. 9 and 10, respectively and then we used DeepHL to predict their HOMO and LUMO energies. Finally, we selected DPAC-Cz and TDBA-pyCz as promising host and emitter molecules because the HOMO/LUMO energies of TDBA-pyCz (emitter) and DPAC-Cz (host) were DeepHL-predicted to be $-5.84/-2.93$ and $-5.59/-2.37$, respectively, which satisfied the aforementioned requirements. It should be noted that all newly designed molecules are not included in our experimental database, and the top 3 molecules from our experimental database with the highest similarity scores for TDBA-pyCz and DPAC-Cz are presented in Supplementary Fig. 11. In addition, we further used the previously developed DLOS to predict the optical and photophysical properties of TDBA-pyCz and DPAC-Cz

to ensure that the absorption and emission properties of TDBA-pyCz and DPAC-Cz were suitable for deep-blue emission and were also favorable for an efficient energy transfer from DPAC-Cz to TDBA-pyCz.

As the suitability of TDBA-pyCz and DPAC-Cz as emitter and host molecules was confirmed by DL-prediction, TDBA-pyCz and DPAC-Cz were synthesized (See the Supplementary Information for details), and their optical, photophysical, and electrochemical properties were measured as summarized in Table 1 and Supplementary Table 5. DL-predicted and experimentally-measured optical, photophysical, and electrochemical properties of TDBA-pyCz and DPAC-Cz were found to be in an agreement within the prediction error. Notably, DeepHL-predicted HOMO and LUMO energies agreed very well with the experimental values obtained from cyclic voltammetry (Supplementary Fig. 16), as shown in Fig. 5b, and the emission properties of TDBA-pyCz in DPAC-Cz were suitable for deep-blue emission (Table 1). Furthermore, DL-prediction revealed that the efficient energy

Table 2. EL performance of deep-blue OLEDs with TDBA-pyCz and DPAC-Cz.

Doping conc.	V_{on}^a (V)	CE_{max}^b (cd A $^{-1}$)	PE_{max}^c (lm W $^{-1}$)	EQE_{max}^d (%)	Luminance e (cd m $^{-2}$)	EL_{max} (nm)	CIE (x, y) f
10 wt%	4.1	1.81	1.26	5.71	936	412	(0.18, 0.09)
15 wt%	4.0	2.36	1.28	6.58	1052	412	(0.17, 0.07)
20 wt%	3.7	2.28	1.79	5.40	1183	412	(0.17, 0.08)

^aTurn-on voltage at a luminance of 1.0 cd m $^{-2}$.

^bMaximum current efficiency.

^cMaximum power efficiency.

^dMaximum external quantum efficiency

^eMaximum luminance.

^fCIE color coordinates at the luminance of 500 cd m $^{-2}$.

transfer from DPAC-Cz to TDBA-pyCz was possible because the emission spectrum of DPAC-Cz overlapped well with the absorption spectrum of TDBA-pyCz (Supplementary Fig. 17), which was confirmed by UV-visible absorption and fluorescence spectra measured with TDBA-pyCz, DPAC-Cz, and TDBA-pyCz-doped DPAC-Cz films (Supplementary Fig. 18).

Fabrication of deep-blue fluorescent OLED device

Using TDBA-pyCz and DPAC-Cz, OLED devices were fabricated by solution processes, and their performance was fully characterized, as shown in Fig. 5 and Table 2. The OLED devices with TDBA-pyCz and DPAC-Cz were found to exhibit a narrow deep-blue emission at 412 nm (CIE: $x = 0.17$, $y = 0.07$; FWHM = 36 nm), satisfying the National Television System Committee standard, and the maximum EQE (EQE_{max}) was measured to be as high as 6.58%. The time-resolved fluorescence (TRF) signals were measured with TDBA-pyCz-doped DPAC-Cz film (Supplementary Fig. 20), and the fluorescence lifetime was determined to be 2.4 ns. DFT calculations revealed that TDBA-pyCz exhibits a relatively large singlet-triplet state energy gap ($\Delta E_{\text{ST}} = 0.42$ eV) (Supplementary Fig. 21). TRF experimental results and DFT calculations confirmed that TDBA-pyCz is a pure fluorescent emitter. The orientation factor of TDBA-pyCz in DPAC-Cz film was measured to $\Theta = 0.187$ using the angle-dependent fluorescence experiments (Supplementary Figs. 22 and 23), indicating that ~81% of TDBA-pyCz is horizontally aligned along the glass substrate, which is expected to lead to a high out-coupling efficiency. The surface morphology of TDBA-pyCz-doped DPAC-Cz films examined by atomic force microscopy (AFM) shows a small root-mean-square roughness of 0.291 nm in Supplementary Fig. 24, implying that our host and emitter molecules are suitable for the solution process. The current density–voltage curves of the hole-only device and electron-only device presented in Supplementary Fig. 25 reveal that the hole and electron mobilities are reasonably well-balanced in the OLED device.

In short, the deep-blue fluorescent OLED device with TDBA-pyCz and DPAC-Cz was successfully developed by DL-assisted virtual screening, and exhibited the high EQE ($EQE_{\text{max}} = 6.58\%$), high horizontal emitter orientation ($\Theta = 0.187$), and reasonably well-balanced hole and electron mobilities. The performance of our solution-processed deep-blue fluorescent OLED device was found to be superior to those previously reported in terms of EQE, emission bandwidth, and emitter orientation as shown in Supplementary Table 6 and Supplementary Fig. 26.

In this study, we built an experimental database of the HOMO and LUMO energies of 3026 organic molecules in solutions and solid states. We successfully developed a DL model to reliably and quickly predict the HOMO and LUMO energies of molecules in different local environments (DeepHL). The high accuracy of DeepHL was shown to result from an understanding of the

effect of molecular structure on the HOMO and LUMO energies, including the donor–acceptor structures, functional groups, heteroatoms and heavy atoms, and conjugation. Lastly, we demonstrated that DeepHL combined with DLOS were efficiently used to prescreen newly designed emitter and host molecules optimized for a given OLED device structure. The optical, photophysical, and electrochemical properties of DPAC-Cz and TDBA-pyCz that were predicted by DeepHL and DLOS were shown to agree very well with the experimental ones. Solution-processed deep-blue fluorescent OLEDs, successfully developed with the aid of our DL-prediction (DeepHL and DLOS), exhibited an EQE of 6.58% and narrow emission bandwidth. Overall, our DL-assisted virtual screening will be able to revolutionize the development of component materials in optoelectronics.

METHODS

Building the experimental database of HOMO and LUMO energies

Our experimental database was built by collecting the HOMO and LUMO energies of organic compounds from 860 articles. Overall, our database included the HOMO and LUMO energies of 3026 organic molecules in solvents or solids, yielding 3362 molecule/solvent combinations. In our database, the solution and solid states were labeled to reflect the local environment of given molecules. In the literature, the HOMO and LUMO energies have been measured using ultraviolet photoelectron spectroscopy, inverse photoelectron spectroscopy, and cyclic voltammetry (CV). Additionally, ultraviolet-visible (UV-visible) absorption spectroscopy was applied to measure the optical bandgaps that were used to determine either HOMO or LUMO energy, in the case where one of the energies was not able to be directly measured because of the technical difficulty of the measurements. We found that the HOMO and LUMO energies of some molecules were measured with different experimental methods and/or under different experimental conditions (i.e. in solvents and solid states). The experimental errors of the HOMO and LUMO energies of the same molecules with different experimental methods and/or under different experimental conditions were found to be 0.089 and 0.112 eV for the HOMO and LUMO energies, respectively.

Graph representation of molecules

Molecules and their structural features can be represented by using an adjacency matrix (\mathbf{A}_j) and a feature matrix (\mathbf{X}_j). The adjacency matrix describes the connectivity of atoms in a given molecule. The single, aromatic, double, and triple bonds are encoded as 1, 1.5, 2, and 3 in the adjacency matrix, respectively, as shown in Fig. 6b. In addition, the diagonal elements are encoded as 1 to represent the atom itself. In our DL model, the maximum number of atoms is 150, which can be readily extended if necessary, giving the adjacency matrix of 150 × 150 elements. The feature matrix is one-hot-encoded consisting of the identity of the atoms, the number of hydrogen atoms, the number of connected atoms, aromaticity, hybridization state, ring, and formal charge. The total feature matrix size is 150 × 43. The example of the adjacency and feature matrices for 2,6-dinitrotoluene is shown in Fig. 6.

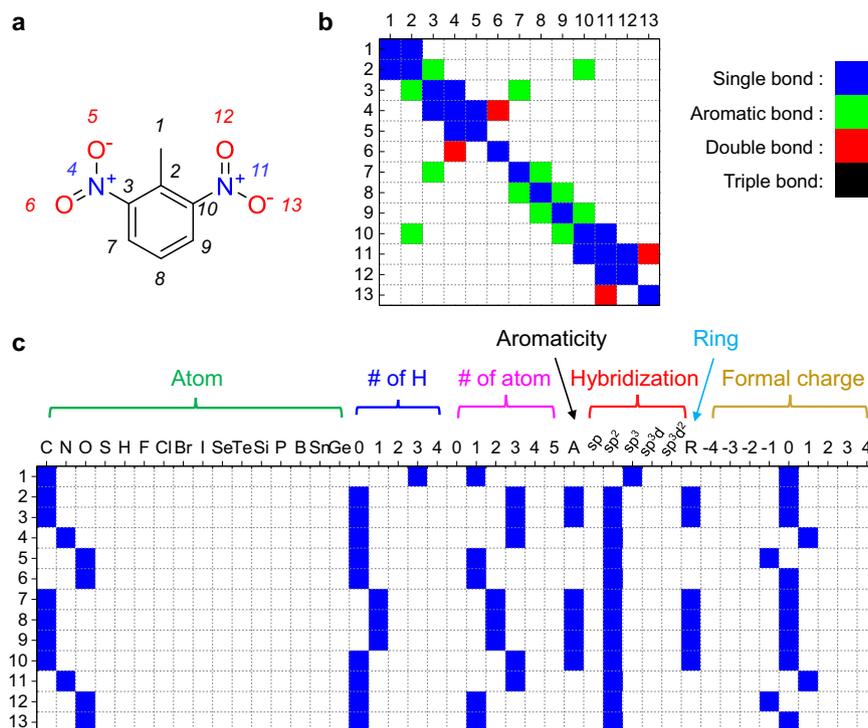


Fig. 6 Graph representation of a molecule. **a** Molecular structure. **b** adjacency matrix (A_j), and **c** feature matrix (X_j) of 2,6-dinitrotoluene.

Algorithm of our deep learning model

The detailed algorithm of our DL model is described in Algorithm 1. A_j and X_j are the adjacency and feature matrices of j , respectively, H_j is the hidden matrix which is an updated feature matrix of j , $reduce_sum$ represents the summation of all row vectors of the hidden matrix, $concat$ represents the concatenation of vectors, and \circ denotes the function composition. In addition, GCN and MLP stand for the graph convolutional network and multi-layer perceptron, respectively.

Algorithm 1. Deep learning model algorithm

Input: A_{mol} , X_{mol} , A_{sol} , X_{sol}	# Molecule and solvent graphs
Output: Properties y	
1 $H_{mol}^{(0)} \leftarrow X_{mol}$	
2 $H_{sol}^{(0)} \leftarrow X_{sol}$	
3 for k in range(Number of GCN layers)	# GCN layers
4 $H_{mol} \leftarrow GCN_{k}^{mol}(A_{mol}, H_{mol})$	
5 $H_{sol} \leftarrow GCN_{k}^{sol}(A_{sol}, H_{sol})$	
6 endfor	
7 $z_{mol}^{(0)} \leftarrow reduce_sum(H_{mol})$	# Chemical space layers
8 $z_{sol}^{(0)} \leftarrow reduce_sum(H_{sol})$	
9 for l in range(Number of MLPs)	
10 $z_{mol} \leftarrow MLP_{mol}^l(z_{mol})$	
11 $z_{sol} \leftarrow MLP_{sol}^l(z_{sol})$	
12 endfor	
13 for m in range(Number of interaction layers)	# Interaction layers
14 $z \leftarrow MLP_m \circ concat(z_{mol}, z_{sol})$	
15 endfor	
16 $y \leftarrow MLP(z)$	# Output
17 return y	

In our DL model (DeepHL), the GCN updates the $l+1$ -th hidden matrix of j (H_j^{l+1}) as follows:

$$H_j^{l+1} = \sigma(A_j \cdot H_j^l \cdot W_j^l + b_j^l + H_j^l) \quad (1)$$

where σ is the rectified linear unit (ReLU), A_j is the adjacency matrix of j , and W_j^l and b_j^l are weight and bias of the l -th layer, respectively.

After passing through the GCN layers, the row vectors ($h_{i,j}$) of H_j (i.e., $H_j = (h_{1,j}, h_{2,j}, \dots, h_{i,j})^T$) are summed to ensure permutation invariance and to produce a chemical space vector of j (z_j) as follows:

$$z_j = \sum_i h_{i,j} \quad (2)$$

In our DL model, we used the multi-layer perceptron (MLP)

$$z_j^{l+1} = \sigma(z_j^l \cdot W_j^l + b_j^l) \quad (3)$$

It should be noted that the last MLP, which is used to calculate the HOMO and LUMO energies, does not have the activation function for the regression task. The architecture of our DL model, which is in-house code based on the RDKit and Keras packages^{90,91}, has been also described in great detail elsewhere⁵⁵.

Training procedure of DL model (DeepHL)

A total of 3362 molecule/solvent combinations were randomly divided into 2688, 336, and 338 for the training, validation, and test datasets, respectively. The distributions of HOMO and LUMO energies and molecular weights in the training, validation, and test datasets were found to be very similar as shown in Supplementary Fig. 4. In addition, we confirmed that there were no duplicate molecule/solvent combinations among the training, validation, and test datasets. The HOMO and LUMO energies were normalized to follow the standard normal distribution. Here, we used two learning methods: learning from scratch (LS) and transfer learning (TL). In the LS method, all parameters (i.e., weights and biases) in the DL model were initialized and optimized over 3000 epochs. The parameters with the lowest validation loss were finally selected. The TL method was performed in two stages. First, the DL model used the optimized parameters of the previously trained DLOS⁵⁵,

and only the parameters in the last hidden layer were optimized. Second, all the parameters in the DL model were fine-tuned by being trained with 10^3 times smaller learning rate. We compared the performances of the two methods and further investigated the dependence of the two methods on the training dataset size. To examine the dependence of the performance of our DL model on the training dataset size, a group of sub-datasets from the training dataset were constructed with different sizes of 50, 100, 200, 400, 800, 1000, 1500, 2000, and 2688. In addition, the larger sub-datasets were set to contain the smaller sub-datasets. Because small datasets can be easily biased to certain composition, total 10 groups of sub-datasets with different compositions were used. The dependence of the two training methods (i.e. LS and TL methods) on the training dataset size was quantified by the MAEs that were calculated using the test dataset, as shown in Fig. 2b, Supplementary Tables 1 and 2. It was found that the MAEs of the TL method were always smaller than those of the LS method. The TL-based DL model (DeepHL) was finally selected to be trained by using both training set and validation set, and the MAEs of the HOMO and LUMO energies of molecules in the test dataset were 0.148 and 0.163 eV, respectively.

Although the HOMO and LUMO energies of a given molecule depend on the molecule's local environment, the MAE of DL predictions can be quite small when combinations of the same molecule and different solvents are divided into the training and test datasets. To avoid such data leakages, the molecules classified as the same scaffold by the MurckoScaffold module in the RDKit were grouped together and split into the training and test datasets regardless of solvents. In this way, the molecules with the same scaffold cannot be included in the training and test datasets at the same time. The prediction MAE of the DL model trained with the scaffold-split datasets was 0.170 eV, which was slightly larger than that of randomly split datasets (0.155 eV). In short, our DL model shows similar performance regardless of data splitting methods.

Materials and synthesis

All reagents were purchased from Sigma-Aldrich, TCI, Acros and Alfa Aesar and used without purification. All reactions were carried out under a nitrogen gas. The chemical structures of the synthesized compounds were analysed through ^1H nuclear magnetic resonance (NMR) and ^{13}C NMR spectra recorded in deuterated chloroform using a Varian Mercury 500 MHz spectrometer (Cambridge Isotope Laboratories). ^1H and ^{13}C NMR spectrum of synthesized compounds are presented in Supplementary Figs. 11–14.

DATA AVAILABILITY

The experimental database of HOMO and LUMO energies and the prediction of HOMO and LUMO energies by our DeepHL are available (<http://deep4chem.korea.ac.kr>). Additional data that are not found in the main text and Supplementary Information are available from the corresponding authors upon reasonable request.

Received: 19 January 2022; Accepted: 19 June 2022;

Published online: 11 July 2022

REFERENCES

- Capelli, R. et al. Organic light-emitting transistors with an efficiency that outperforms the equivalent light-emitting diodes. *Nat. Mater.* **9**, 496–503 (2010).
- Gao, Y. et al. Highly efficient organic tandem solar cell with a SubPc interlayer based on TAPC:C70 bulk heterojunction. *Sci. Rep.* **6**, 23916 (2016).
- Salehi, A. et al. Realization of high-efficiency fluorescent organic light-emitting diodes with low driving voltage. *Nat. Commun.* **10**, 2305 (2019).
- Fukagawa, H. et al. Understanding coordination reaction for producing stable electrode with various low work functions. *Nat. Commun.* **11**, 3700 (2020).
- Hou, B. L. et al. Facile generation of bridged medium-sized polycyclic systems by rhodium-catalysed intramolecular (3+2) dipolar cycloadditions. *Nat. Commun.* **12**, 5239 (2021).
- Wan, Y. et al. Data driven discovery of conjugated polyelectrolytes for optoelectronic and photocatalytic applications. *npj Comput. Mater.* **7**, 1–9 (2021).
- Vasilopoulou, M. et al. High efficiency blue organic light-emitting diodes with below-bandgap electroluminescence. *Nat. Commun.* **12**, 4868 (2021).
- Vebber, M. C., Rice, N. A., Brusso, J. L. & Lessard, B. H. Variance-resistant PTB7 and axially-substituted silicon phthalocyanines as active materials for high-Voc organic photovoltaics. *Sci. Rep.* **11**, 15347 (2021).
- Fukui, K., Yonezawa, T. & Shingu, H. A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Chem. Phys.* **20**, 722–725 (1952).
- Fukui, K., Yonezawa, T., Nagata, C. & Shingu, H. Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *J. Chem. Phys.* **22**, 1433–1442 (1954).
- Shockley, W. & Queisser, H. J. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.* **32**, 510–519 (1961).
- Son, H. J., He, F., Carsten, B. & Yu, L. Are we there yet? Design of better conjugated polymers for polymer solar cells. *J. Mater. Chem.* **21**, 18934–18945 (2011).
- Fukagawa, H., Shimizu, T., Iwasaki, Y. & Yamamoto, T. Operational lifetimes of organic light-emitting diodes dominated by Forster resonance energy transfer. *Sci. Rep.* **7**, 1735 (2017).
- Chang, C.-H. et al. Aligned energy-level design for decreasing operation voltage of tandem white organic light-emitting diodes. *Thin Solid Films* **548**, 389–397 (2013).
- Yadav, R. A. K., Dubey, D. K., Chen, S. Z., Liang, T. W. & Jou, J. H. Role of molecular orbital energy levels in OLED performance. *Sci. Rep.* **10**, 9915 (2020).
- Bauschlicher, C. W. TaFn and TaCln atomization energies for $n = 1-5$. *J. Phys. Chem. A* **104**, 5843–5849 (2000).
- Zhang, Y.-Y. et al. A DFT study on the enthalpies of thermite reactions and enthalpies of formation of metal composite oxide. *Chem. Phys.* **507**, 19–27 (2018).
- Joung, J. F., Kim, S. & Park, S. Cationic effect on the equilibria and kinetics of the excited-state proton transfer reaction of a photoacid in aqueous solutions. *J. Phys. Chem. B* **122**, 5087–5093 (2018).
- Mumit, M. A. et al. DFT studies on vibrational and electronic spectra, HOMO-LUMO, MEP, HOMA, NBO and molecular docking analysis of benzyl-3-N-(2,4,5-trimethoxyphenylmethylene)hydrazinecarbodithioate. *J. Mol. Struct.* **1220**, 128715 (2020).
- Kim, H. J. et al. Ultra-deep-blue aggregation-induced delayed fluorescence emitters: achieving nearly 16% EQE in solution-processed non-doped and doped OLEDs with CIE $y < 0.1$. *Adv. Funct. Mater.* **31**, 2102588 (2021).
- Ha, J. M. et al. Rational molecular design of azaacene-based narrowband green-emitting fluorophores: modulation of spectral bandwidth and vibronic transitions. *ACS Appl. Mater. Interfaces* **13**, 26227–26236 (2021).
- Montavon, G. et al. Machine learning of molecular electronic properties in chemical compound space. *N. J. Phys.* **15**, 095003 (2013).
- Pereira, F. et al. Machine learning methods to predict density functional theory B3LYP energies of HOMO and LUMO orbitals. *J. Chem. Inf. Model.* **57**, 11–21 (2017).
- Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
- Nagasawa, S., Al-Naamani, E. & Saeki, A. Computer-aided screening of conjugated polymers for organic solar cell: classification by Random Forest. *J. Phys. Chem. Lett.* **9**, 2639–2646 (2018).
- Coley, C. W. et al. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2019).
- Jha, D. et al. Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nat. Commun.* **10**, 5316 (2019).
- Mater, A. C. & Coote, M. L. Deep learning in chemistry. *J. Chem. Inf. Model.* **59**, 2545–2559 (2019).
- Lim, J. et al. Predicting drug-target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J. Chem. Inf. Model.* **59**, 3981–3988 (2019).
- Kang, B., Seok, C. & Lee, J. Prediction of molecular electronic transitions using random forests. *J. Chem. Inf. Model.* **60**, 5984–5994 (2020).
- Meftahi, N. et al. Machine learning property prediction for organic photovoltaic devices. *npj Comput. Mater.* **6**, 166 (2020).
- Haghighatari, M. et al. Learning to make chemical predictions: the interplay of feature representation, data, and machine learning methods. *Chem* **6**, 1527–1542 (2020).
- Sandfort, F., Strieth-Kalthoff, F., Kühnemund, M., Beecks, C. & Glorius, F. A structure-based platform for predicting chemical reactivity. *Chem* **6**, 1379–1390 (2020).
- Qiao, B. et al. Quantitative mapping of molecular substituents to macroscopic properties enables predictive design of oligoethylene glycol-based lithium electrolytes. *ACS Cent. Sci.* **6**, 1115–1128 (2020).
- Wu, Y., Guo, J., Sun, R. & Min, J. Machine learning for accelerating the discovery of high-performance donor/acceptor pairs in non-fullerene organic solar cells. *npj Comput. Mater.* **6**, 1–8 (2020).

36. Lee, S. et al. Computational screening of trillions of metal-organic frameworks for high-performance methane storage. *ACS Appl. Mater. Interfaces* **13**, 23647–23654 (2021).
37. Mamede, R., Pereira, F. & Aires-de-Sousa, J. Machine learning prediction of UV-Vis spectra features of organic compounds related to photoreactive potential. *Sci. Rep.* **11**, 23720 (2021).
38. Kang, B., Seok, C. & Lee, J. A benchmark study of machine learning methods for molecular electronic transition: Tree-based ensemble learning versus graph neural network. *Bull. Korean Chem. Soc.* **43**, 328–335 (2022).
39. Ksenofontov, A. A., Lukanov, M. M., Bocharov, P. S., Berezin, M. B. & Tetko, I. V. Deep neural network model for highly accurate prediction of BODIPYs absorption. *Spectrosc. Acta Pt. A-Molec. BioMolec. Spectr.* **267**, 120577 (2022).
40. Schutt, K. T., Saucedo, H. E., Kindermans, P. J., Tkatchenko, A. & Muller, K. R. SchNet - a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
41. Hou, F. et al. Comparison study on the prediction of multiple molecular properties by various neural networks. *J. Phys. Chem. A* **122**, 9128–9134 (2018).
42. Anderson, B., Hy, T. S. & Kondor, R. Cormorant: Covariant Molecular Neural Networks. In *Advances in neural information processing systems*. Vol. 32, 14537–14546 (NIPS, 2019).
43. Lu, C. et al. Molecular Property Prediction: A Multilevel Quantum Interactions Modeling Perspective. In *Proc. AAAI Conference on Artificial Intelligence*. Vol. 33, 1052–1060 (AAAI, 2019).
44. Klicpera, J., Groß, J. & Günnemann, S. Directional Message Passing for Molecular Graphs. In *Proc. 8th International Conference on Learning Representations*. (ICLR, 2020).
45. Ye, S. et al. Asymmetric anthracene derivatives as multifunctional electronic materials for constructing simplified and efficient non-doped homogeneous deep blue fluorescent OLEDs. *Chem. Eng. J.* **393**, 124694 (2020).
46. Rahaman, O. & Gagliardi, A. Deep learning total energies and orbital energies of large organic molecules using hybridization of molecular fingerprints. *J. Chem. Inf. Model.* **60**, 5971–5983 (2020).
47. Yang, G.-X. et al. Rational design of pyridine-containing emissive materials for high performance deep-blue organic light-emitting diodes with CIE_y ~ 0.06. *Dyes Pigment.* **187**, 109088 (2021).
48. Liu, Z. et al. Transferable multilevel attention neural network for accurate prediction of quantum chemistry properties via multitask learning. *J. Chem. Inf. Model.* **61**, 1066–1082 (2021).
49. Kwon, Y., Kang, S., Choi, Y. S. & Kim, I. Evolutionary design of molecules based on deep learning and a genetic algorithm. *Sci. Rep.* **11**, 17304 (2021).
50. Blum, L. C. & Reymond, J. L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **131**, 8732–8733 (2009).
51. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
52. Kim, S. et al. PubChem substance and compound databases. *Nucleic Acids Res.* **44**, D1202–D1213 (2016).
53. Stuke, A. et al. Atomic structures and orbital energies of 61,489 crystal-forming organic molecules. *Sci. Data* **7**, 58 (2020).
54. Zhang, G. & Musgrave, C. B. Comparison of DFT methods for molecular orbital eigenvalue calculations. *J. Phys. Chem. A* **111**, 1554–1561 (2007).
55. Joung, J. F. et al. Deep learning optical spectroscopy based on experimental database: potential applications to molecular design. *JACS Au* **1**, 427–438 (2021).
56. Bucinkas, A. et al. Can attachment of tert-butyl substituents to methoxycarbazole moiety induce efficient TADF in diphenylsulfone-based blue OLED emitters? *Org. Electron.* **86**, 105894 (2020).
57. Sun, K. et al. Novel aggregation-induced emission and thermally activated delayed fluorescence materials based on thianthrene-9,9',10,10'-tetraoxide derivatives. *RSC Adv.* **6**, 22137–22143 (2016).
58. Zhao, W. et al. Molecular optimization enables over 13% efficiency in organic solar cells. *J. Am. Chem. Soc.* **139**, 7148–7151 (2017).
59. Ge, J. et al. Improved efficiency in all-small-molecule organic solar cells with ternary blend of nonfullerene acceptor and chlorinated and nonchlorinated donors. *ACS Appl. Mater. Interfaces* **11**, 44528–44535 (2019).
60. Joung, J. F., Han, M., Jeong, M. & Park, S. Experimental database of optical properties of organic compounds. *Sci. Data* **7**, 295 (2020).
61. Gao, Y. & Cui, Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nat. Commun.* **11**, 5131 (2020).
62. Zhu, R. et al. Phase-to-pattern inverse design paradigm for fast realization of functional metasurfaces via transfer learning. *Nat. Commun.* **12**, 2974 (2021).
63. Lu, T., Han, B., Chen, L., Yu, F. & Xue, C. A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Sci. Rep.* **11**, 15824 (2021).
64. Kim, Y. et al. Deep learning framework for material design space exploration using active transfer learning and data augmentation. *npj Comput. Mater.* **7**, 140 (2021).
65. Zhuang, F. et al. A comprehensive survey on transfer learning. In *Proceedings of the IEEE*. Vol. 109, 43–76 (IEEE, 2021).
66. Wang, Z. et al. Predicting adsorption ability of adsorbents at arbitrary sites for pollutants using deep transfer learning. *npj Comput. Mater.* **7**, 1–9 (2021).
67. Gaussian 16 (Gaussian Inc., Wallingford, CT, 2016).
68. Nakata, M. & Shimazaki, T. PubChemQC Project: a large-scale first-principles electronic structure database for data-driven chemistry. *J. Chem. Inf. Model.* **57**, 1300–1308 (2017).
69. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70, 1263–1272 (PMLR, 2017).
70. Kylberg, W. et al. Synthesis, thin-film morphology, and comparative study of bulk and bilayer heterojunction organic photovoltaic devices using soluble diketopyrrolopyrrole molecules. *Energy Environ. Sci.* **4**, 3617–3624 (2011).
71. Yang, D. et al. Novel high performance asymmetrical squaraines for small molecule organic solar cells with a high open circuit voltage of 1.12 V. *Chem. Commun.* **49**, 10465–10467 (2013).
72. Park, J. B., Ha, J.-W., Jung, I. H. & Hwang, D.-H. High-performance nonfullerene organic photovoltaic cells using a TPD-based wide bandgap donor polymer. *ACS Appl. Energy Mater.* **2**, 5692–5697 (2019).
73. Ma, J., Liu, T. X., Zhang, P., Zhao, X. & Zhang, G. Metal-free-catalyzed three-component [2+2+2] annulation reaction of [60]Fullerene, ketones, and indoles: access to diverse [60]Fullerene-fused 1,2-tetrahydrocarbazoles. *Org. Lett.* **23**, 1775–1781 (2021).
74. Kawamura, Y. et al. 100% phosphorescence quantum efficiency of Ir(III) complexes in organic semiconductor films. *Appl. Phys. Lett.* **86**, 071104 (2005).
75. Jeong, S. H. & Lee, J. Y. Dibenzothiophene derivatives as host materials for high efficiency in deep blue phosphorescent organic light emitting diodes. *J. Mater. Chem.* **21**, 14604–14609 (2011).
76. Zhang, Q. et al. Triplet exciton confinement in green organic light-emitting diodes containing luminescent charge-transfer Cu(I) complexes. *Adv. Funct. Mater.* **22**, 2327–2336 (2012).
77. Wang, H. et al. Novel thermally activated delayed fluorescence materials-thioxanthone derivatives and their applications for highly efficient OLEDs. *Adv. Mater.* **26**, 5198–5204 (2014).
78. Zhang, Q. et al. Efficient blue organic light-emitting diodes employing thermally activated delayed fluorescence. *Nat. Photonics* **8**, 326–332 (2014).
79. Baranoff, E. & Curchod, B. F. Flrpic: archetypal blue phosphorescent emitter for electroluminescence. *Dalton Trans.* **44**, 8318–8329 (2015).
80. Hirai, H. et al. One-Step Borylation of 1,3-Diaryloxybenzenes towards efficient materials for organic light-emitting diodes. *Angew. Chem. Int. Ed.* **54**, 13581–13585 (2015).
81. Cho, Y. J., Chin, B. D., Jeon, S. K. & Lee, J. Y. 20% external quantum efficiency in solution-processed blue thermally activated delayed fluorescent devices. *Adv. Funct. Mater.* **25**, 6786–6792 (2015).
82. Shirota, Y. et al. Starburst molecules based on π -electron systems as materials for organic electroluminescent devices. *J. Lumines* **72–74**, 985–991 (1997).
83. Chen, M.-H. et al. Electronic and chemical properties of cathode structures using 4,7-diphenyl-1,10-phenanthroline doped with rubidium carbonate as electron injection layers. *J. Appl. Phys.* **105**, 113714 (2009).
84. Lee, C. W. & Lee, J. Y. Comparison of tetraphenylmethane and tetraphenylsilane as core structures of high-triplet-energy hole- and electron-transport materials. *Chem. Eur. J.* **18**, 6457–6461 (2012).
85. Wang, J. et al. High efficiency green phosphorescent organic light-emitting diodes with a low roll-off at high brightness. *Org. Electron.* **14**, 2854–2858 (2013).
86. Yan, L. et al. Palladium-catalyzed tandem N-H/C-H arylation: regioselective synthesis of N-heterocycle-fused phenanthridines as versatile blue-emitting luminophores. *Org. Biomol. Chem.* **11**, 7966–7977 (2013).
87. Chen, Y., Shen, L. & Li, X. Effects of heteroatoms of tetracene and pentacene derivatives on their stability and singlet fission. *J. Phys. Chem. A* **118**, 5700–5708 (2014).
88. Nakano, M., Niimi, K., Miyazaki, E., Osaka, I. & Takimiya, K. Isomerically pure anthra[2,3-b:6,7-b']difuran (anti-ADF), -dithiophene (anti-ADT), and -diselenophene (anti-ADS): selective synthesis, electronic structures, and application to organic field-effect transistors. *J. Org. Chem.* **77**, 8099–8111 (2012).
89. Huang, J. et al. Tuning frontier orbital energetics of azaisoindigo-based polymeric semiconductors to enhance the charge-transport properties. *Adv. Electron. Mater.* **3**, 1700078 (2017).
90. Landrum, G. Open-source cheminformatics; <http://www.rdkit.org>.
91. Chollet, F. et al. Keras; <https://keras.io> (2015).

ACKNOWLEDGEMENTS

This study was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean government (No. 2019R1A6A1A11044070 and 2022R1A2C1003627) and LG Display.

AUTHOR CONTRIBUTIONS

M.J., J.F.J., J.H., and M.H. contributed equally to this work. S.P. and D.H.C. conceived and supervised the research. M.J., J.F.J., and M.H. built the experimental database and developed the deep learning algorithm and the web-based application. J.H. synthesized compounds and fabricated OLED devices. M.J., J.F.J., J.H., D.H.C., and S.P. wrote the manuscript. All authors have read and commented the manuscript and Supplementary Information.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-022-00834-3>.

Correspondence and requests for materials should be addressed to Dong Hoon Choi or Sungnam Park.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022