# Supporting Information

# Machine Learning-Assisted High-Throughput Screening of Transparent Organic Light-Emitting Diodes Anode Materials

*Liying Cui[#], Qing Li[#], Yanchang Zhang, Jiao Zhang, Zhe Wang, Jiankang Chen, and Bing Zheng**

L. Y. Cui, Q. Li, Y. C. Zhang, J. Zhang, Z. Wang, J. K. Chen, Prof. B. Zheng
Key Laboratory of Functional Inorganic Materials Chemistry (Ministry of Education), School of Chemistry and Materials Science, Heilongjiang University, Harbin, 150080, P. R. China.
E-mail: zhengbing0106@163.com, zhengbing@hlju.edu.cn (B. Z.)

# Contents of the Supporting Information:

# S1 Calculation details

## S1.1 Calculation details of materials in 2DMatPedia database

The DFT calculations of materials in 2DMatPedia database are performed using the Vienna Ab initio Simulation Package (VASP) with the frozen-core all-electron projector-augmented wave (PAW) method for the electron-ion interaction. All calculations are performed with spin polarization and high initial magnetic moments for magnetic ions. Calibrated Hubbard U values are applied for the transition metals ions in transition metal oxides and fluorides. The cutoff energy for the plane wave expansion of electron wavefunction is set to 520 eV.[1] The interlayer dispersion interaction in layered materials is included via the dispersion-corrected vdW-optB88 exchange-correlation functional.

## S1.2 Calculation details of Pearson correlation coefficient

In the process of feature engineering, we use Pearson correlation coefficient to measure the degree of linear correlation between every pair of features, with values between −1 and 1. A larger absolute value of the correlation coefficient indicates a stronger degree of correlation, whereas a smaller absolute value indicates a weaker degree of correlation. The formula be calculated using Equation (S1).

$$\rho(X,Y)=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2\sum(Y-\overline{Y})^2}} \tag{S1}$$

where $X$ and $Y$ are two features, $\overline{X}$ and $\overline{Y}$ are their average values.[2]

## S1.3 Calculation details of random.shuffle() function

The random.shuffle() is a function in the Python standard library used to randomly shuffle elements in a list or variable sequence. Its algorithm mechanism is based on the Fisher Yates shuffling algorithm (also known as the Knuth shuffling algorithm), which is a classic random permutation algorithm.

The principle of the Fisher Yates shuffling algorithm is as follows:

1. Traverse forward until the first element, starting from the last element in the list.
2. Randomly select a position (including the current position itself).
3. Swap the two elements in the current position and the randomly selected position.
4. Continue traversing forward and repeat steps 2 and 3 until the first element is reached.

# S1.4 Calculation details of t-SNE algorithm

The t-distributed stochastic neighbor embedding (t-SNE) algorithm, is a powerful technique for reducing and visualizing high-dimensional data, which was first proposed by Maaten and Hinton in 2008. The main contribution of the t-SNE algorithm is the ability to preserve the similarities between any two data points, both in high- and low-dimensional spaces, through a probabilistic approach.[2, 3] It uses the t-distribution to efficiently handle outliers in high-dimensional data and generates better clustering in low-dimensional spaces. The conditional probability ($P_{j|i}$) is used to represent the similarity based on the Euclidean distance, which is defined using Equation (S2).

$$P_{j|i} = \frac{\exp(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2)}{\exp(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2)} \tag{S2}$$

where $x_i$, $x_j$, and $x_k$ represent one of the N high-dimensional data points. $\sigma_i$ has different values for different points $x_i$ and is usually the Gaussian mean square error centered at data point $x_i$.

$q_{j|i}$ is the conditional probability of $P_{j|i}$ using the t-distribution, which is defined using Equation (S3).

$$q_{j|i} = \frac{(1+\left\|y_i - y_j\right\|^2)^{-1}}{\sum_{k \neq i}(1+\left\|y_i - y_k\right\|^2)^{-1}} \tag{S3}$$

where $y_i$, $x_j$, and $y_k$ represent one of the N high-dimensional data points.

If the effect of dimensionality reduction is better and local features remain intact, then $q_{j|i} = P_{j|i}$.

## S1.5 Calculation details of work function

Work function is a fundamental and important physical quantity in surface science, defined in solid state physics as the minimum amount of energy required to move an electron from the interior of a solid to the surface. In theoretical calculations, this physical quantity is expressed using Equation (S4).

$$\phi = E_{\text{vac}} - E_{\text{F}} \tag{S4}$$

where $\phi$ represents the work function of the system, where $E_{\text{vac}}$ and $E_{\text{F}}$ denote the vacuum and Fermi level of the system, respectively.[4] The work function depends not only on the electronic properties of the crystal but also on the features of the surface, such as Miller indices, surface functional group modifications,[5] and surface dipole moments.[6-11] In the absence of adsorbed atoms or molecules on the surface, the intrinsic internal polarity of the compound can also give rise to a dipole at the surface, thereby affecting the work function. Materials with both high and low work functions have important applications. Materials with low work functions can be applied to displays, electron guns, and catalysts, whereas materials with high work function can be used in electrodes, photovoltaic devices, and so on.[12, 13]

# S2 Figures

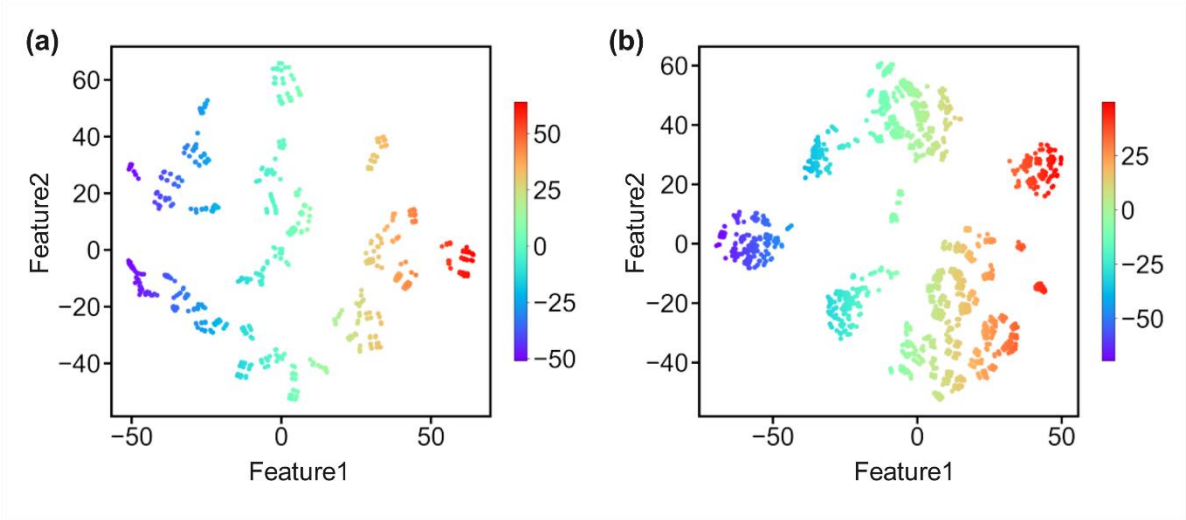


**Figure S1.** The visualization of 2D feature spatial distribution (a) before and (b) after data processing and feature engineering. Feature1 and Feature2 represent the values of the corresponding two features of the dataset after dimensionality reduction. The different material sample point colors represent the Feature1 values.

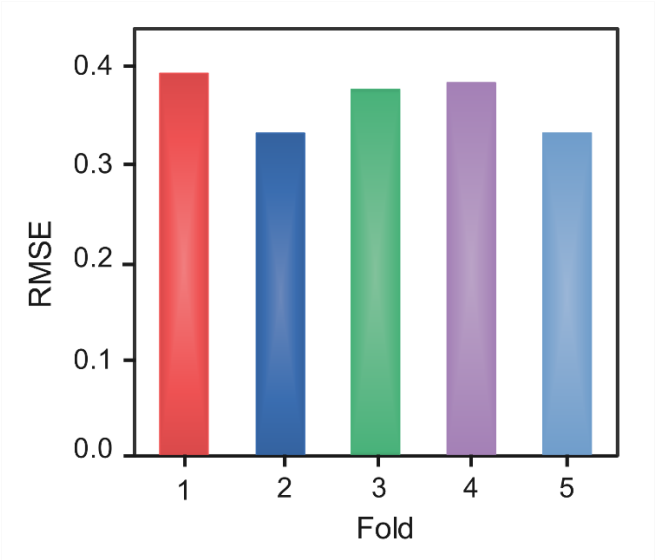


**Figure S2.** The results of 5-fold cross validation of CatBoost model with RMSE as a performance evaluation index.

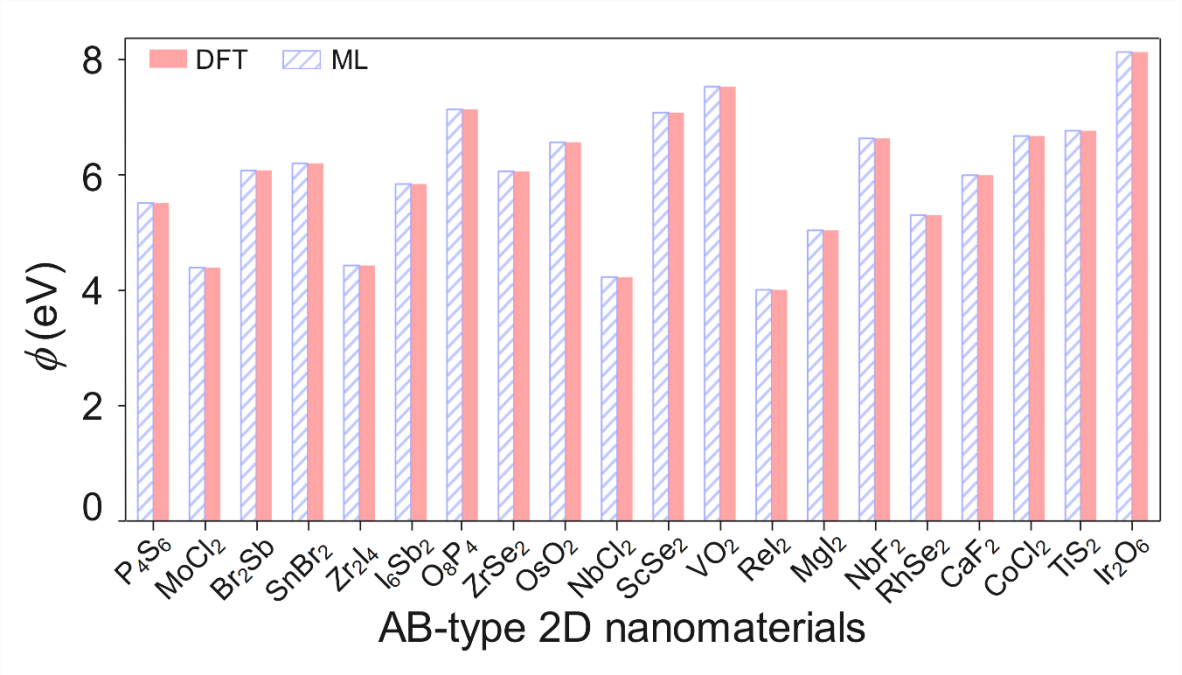


**Figure S3.** The work function ($\phi$) values of AB-type 2D nanomaterials predicted by machine learning (ML) and density functional theory (DFT).

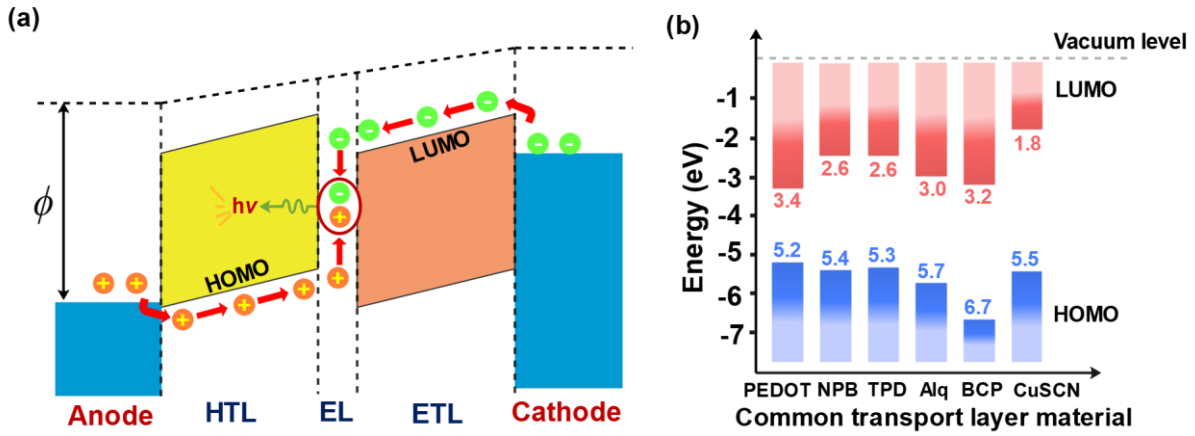


**Figure S4.** (a) The luminous mechanism of OLED devices. Under the action of bias voltage, the holes and the electrons are injected into the HOMO level of the hole transport layer (HTL) from the anode, and the LUMO level of the electron transport layer (ETL) from the cathode, respectively. In the emulsion layer (EL), holes and electrons meet and recombine. Energy is then emitted in the form of photons. $\phi$ represents work function. (b) The energy level of common organic materials.

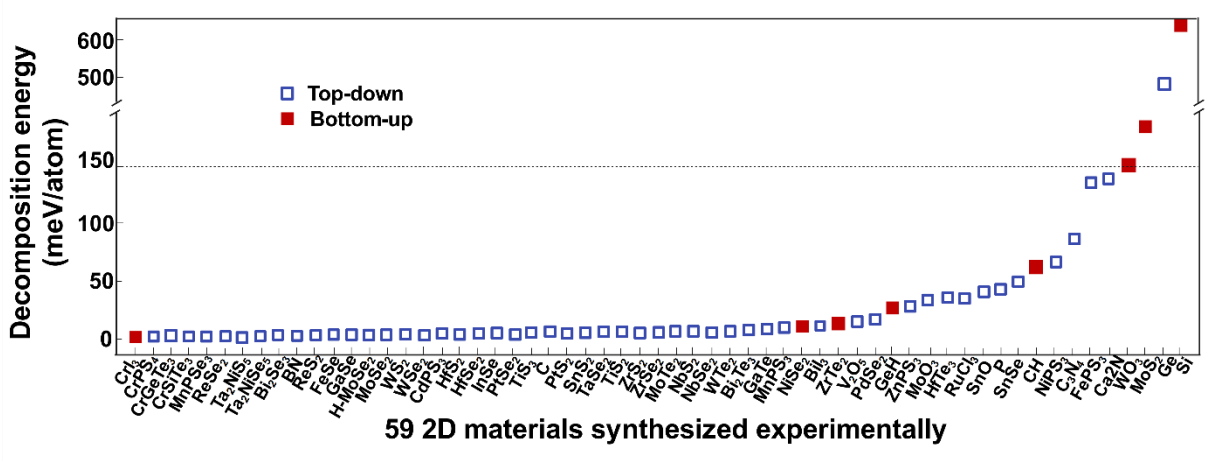


**Figure S5.** The decomposition energy of 59 experimentally obtained 2D materials. The blue open and the red solid blocks are the decomposition energies of 2D materials generated using top-down and bottom-up methods, respectively.

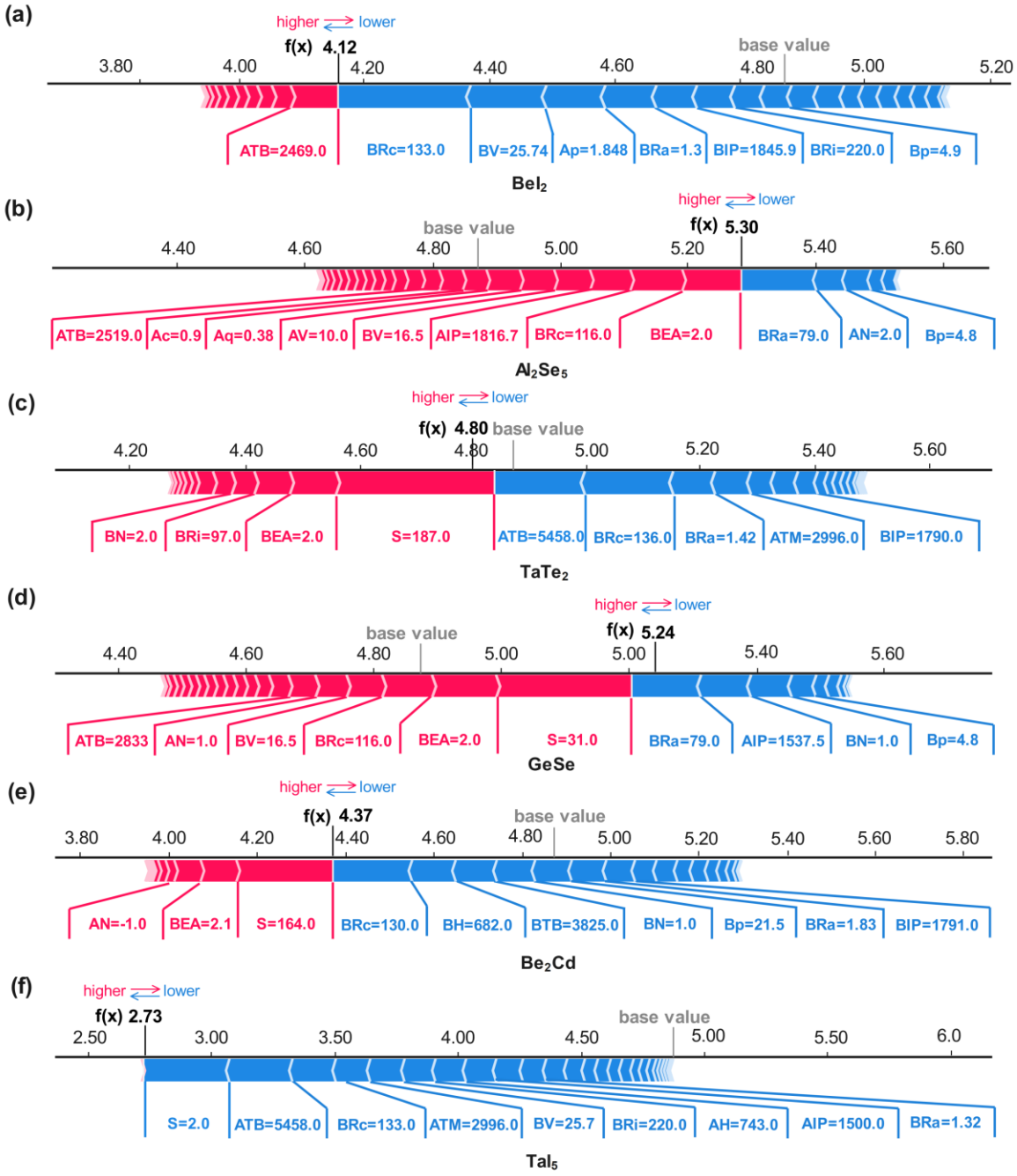


**Figure S6.** Individual SHAP plots of the CatBoost model predictions for (a) $BeI_2$ (*P-3m1*), (b) $Al_2Se_5$ (*C2/m*), (c) $TaTe_2$(*P-6m2*), (d) GeSe (*Pmn2_1*), (e) $Be_2Cd$ (*P4/mmm*) and (f) $TaI_5$ (*P-1*). The SHAP positive (red) and negative (blue) feature weights are given for the six materials. The bar size represents the SHAP value. The base value is the mean value of the work function in 2DMatPedia database during high throughput screening. The output value f(*x*) represents the predicted work function of the three materials.

# S3 Tables

**Table S1.** Properties computed by C2DB single-layer workflow and corresponding methods. A '∗' indicates that spin–orbit coupling (SOC) is included. All calculations are performed with the GPAW code using a plane wave basis except for the Raman calculations, which employ a double-zeta polarised basis of numerical atomic orbitals.[14, 15]

| Property | Method |
|---|---|
| Bader charges | PBE |
| Energy above convex hull | PBE |
| Heat of formation | PBE |
| Orbital projected band structure | PBE |
| Out-of-plane dipole | PBE |
| Phonons (Γ and BZ corners) | PBE |
| Projected density of states | PBE |
| Stiffness tensor | PBE |
| Exchange couplings | PBE |
| Infrared polarisability | PBE |
| Second harmonic generation | PBE |
| Electronic band structure PBE | PBE* |
| Magnetic anisotropies | PBE* |
| Deformation potentials | PBE* |
| Effective masses | PBE* |
| Fermi surface | PBE* |
| Plasma frequency | PBE* |
| Work function | PBE* |
| Optical polarisability | RPA@PBE |
| Electronic band structure | HSE06@PBE* |
| Electronic band structure | $G_0W_0$@PB E* |
| Born charges | PBE, Berry phase |
| Raman spectrum | PBE, LCAO basis set |
| Piezoelectric tensor | PBE, Berry phase |
| Optical absorbance | BSE@$G_0W_0$* |
| Spontaneous polarisation | PBE, Berry phase |
| Topological invariants | PBE*, Berry phase |

**Table S2.** The key variables (name, data type, and a short description.) of materials collection in 2DMatPedia database.

| Key | Datatype | Description |
|---|---|---|
| material_id | string | IDs for entries in the 2Dmatpedia |
| relative_id | string | IDs for where a 2D material is obtained from |
| discovery_process | string | How a 2D materials is generated |
| structure | dictionary | Relaxed crystal structure represented in dictionary |
| formula | string | Chemical formula |
| nelements | string | Number of elements in this material |
| elements | list | List of elements in this material |
| Space group | string | Space group number defined by The International Union of Crystallography |
| point_group | string | Point group in Hermann-Mauguin notation |
| bandgap | float | Energy band gap of this material |
| is_gap_direct | Boolean | Is the material a direct gap |
| is_metal | Boolean | Is the material metallic |
| energy_per_atom | float | Energy per atom in eV without vdW correction |
| energy_vdw_per_atom | float | Energy per atom in eV with vdW correction |
| exfoliation_energy_per_atom | float | Exfoliation energy of the 2D material in eV/atom |
| decomposition_energy-per_atom | float | Decomposition energy of the 2D material in eV/atom |
| total magnetization | float | Total magnetic moment in $\mu_B$ |

**Table S3.** The 24 descriptors and their corresponding symbols after feature engineering and data processing for the AB-type 2D nanomaterials. A and B represent the first and second elements in the formula, respectively.

| Symbol | Feature |
|---|---|
| S | Space group |
| SN | Space group number |
| AN/BN | Element amount of A or B |
| AIP/BIP | Ionization energy of A or B |
| ARi/BRi | Ionic radius of A or B |
| AP/BP | Density of A or B |
| AEA/BEA | Electron affinity energy of A or B |
| ARC/BRC | Covalent radius of A or B |
| AE | Elastic modulus of A |
| Bq | Conductivity of B |
| Ac/Bc | Specific heat capacity of A or B |
| AH/BH | Heat of evaporation of A or B |
| ATB | Boiling point of A |
| ATM/BTM | Melting point of A or B |
| AV | Atomic volume of A |

**Table S4.** The parameter, parameter space, and the best parameter of CatBoost in the process of parameter optimization using random grid search method.

| Parameter | Parameter space | Best parameter of CatBoost |
|---|---|---|
| Iterations | 200–3000 | 1313 |
| Max_depth | 5–12 | 9 |
| Learning_rate | 0.02–0.12 | 0.092 |

**Table S5.** The material_id, chemical formula, space group, band gap ($E_g$, eV), decomposition energy ($E_D$, eV/atom), exfoliation energy ($E_{ex}$, eV/atom), and work function values ($\phi$, eV) of 10 transparent OLED anode candidate materials.

| Material_id | Chemical formula | Space group | $E_g$ | $E_D$ | $E_{ex}$ | $\phi$ |
|---|---|---|---|---|---|---|
| 2dm-4763 | $Bi_1F_4$ | P4/mmm | 0.725 | 0.068 | 0.039 | 6.82 |
| 2dm-748 | $ClO_2$ | P2_1/c | 0.266 | 0.058 | 0.044 | 6.48 |
| 2dm-1472 | BS | P-3m1 | 2.902 | 0.000 | 0.044 | 6.28 |
| 2dm-1127 | $BiCl_3$ | Pmmn | 2.917 | 0.000 | 0.010 | 6.14 |
| 2dm-1126 | BS | P-6m2 | 2.787 | 0.000 | 0.044 | 5.97 |
| 2dm-625 | $Bi_2S_3$ | P-3m1 | 1.104 | 0.005 | 0.059 | 5.95 |
| 2dm-298 | BSe | P-6m2 | 2.577 | 0.000 | 0.047 | 5.53 |
| 2dm-2975 | AlS | P-6m2 | 2.034 | 0.007 | 0.056 | 5.48 |
| 2dm-570 | $BrO_3$ | P-1 | 1.654 | 0.041 | –0.084 | 5.39 |
| 2dm-6204 | PS | P2_1/c | 1.921 | 0.035 | – | 5.11 |

**Table S6.** The deformation potential constant ($E$), elastic constant ($C$), band effective mass ($m$*), and relaxation time ($\tau$) of candidate materials at 300K.

| | Direction | Carrier type | $E$ (eV) | $C$ (J/m$^2$) | $m^*$ | $\tau$ (fs) |
|---|---|---|---|---|---|---|
| $BiCl_3$ | $x$-axis | Electrons | 2.09 | 29.13 | 0.35 | 167.96 |
| | | Holes | 2.09 | 29.13 | 2.88 | 7.09 |
| PS | $x$-axis | Electrons | 1.60 | 43.62 | 0.54 | 250.74 |
| | | Holes | 1.60 | 43.62 | 0.53 | 253.58 |

## References


[1] Zhou J., Shen L., Costa M. D., Persson K. A., Ong S. P., Huck P., Lu Y., Ma X., Chen Y., Tang H., Feng Y. P., 2DMatPedia, an Open Computational Database of Two-Dimensional Materials from Top-Down and Bottom-Up Approaches [J]. *Sci. Data*, 2019, *6*: 86.

[2] Edelmann D., Móri T. F., Székely G. J., On Relationships Between the Pearson and the Distance Correlation Coefficients [J]. *Stat. Probab. Lett.*, 2021, *169*: 6.

[3] Van L., Hinton G., Visualizing Data Using t-SNE [J]. *J. Mach. Learn. Res.*, 2008, *9*: 2579–2605.

[4] Greiner M. T., Chai L., Helander M. G., Tang W. M., Lu Z. H., Transition Metal Oxide Work Functions: The Influence of Cation Oxidation State and Oxygen Vacancies [J]. *Adv. Funct. Mater.*, 2012, *22*: 4557–4568.

[5] Khazaei M., Arai M., Sasaki T., Ranjbar A., Liang Y. Y., Yunoki S., OH-Terminated Two-Dimensional Transition Metal Carbides and Nitrides as Ultralow Work Function Materials [J]. *Phys. Rev. B*, 2015, *92*: 10.

[6] Kolb D., M., Przasnyski M., Gerischer H., Underpotential Deposition of Metals and Work Function Differences [J]. *J. Electroanal. Chem. Interfacial Electrochem.*, 1974, *54*: 25–38.

[7] Zhu J., Zhang H., Zhao L., Xiong W., Huang X., Wang B., Zhang Y. F., Properties of Two-Dimensional Insulators: A DFT Study of Bimetallic Oxide $CrW_2O_9$ Clusters Adsorption on MgO Ultrathin Films [J]. *Appl. Surf. Sci.*, 2016, *379*: 213–222.

[8] Huang J. S., Xu Z., Yang Y., Low-Work-Function Surface Formed by Solution-Processed and Thermally Deposited Nanoscale Layers of Cesium Carbonate [J]. *Adv. Funct. Mater.*, 2007, *17*: 1966–1973.

[9] Prada S., Martinez U., Pacchioni G., Work Function Changes Induced by Deposition of Ultrathin Dielectric Films on Metals: A Theoretical Analysis [J]. *Phys. Rev. B*, 2008, *78*: 8.

[10] Martinez U., Jerratsch J. F., Nilius N., Giordano L., Pacchioni G., Freund H. J., Tailoring the Interaction Strength between Gold Particles and Silica Thin Films Via Work Function Control [J]. *Phys. Rev. Lett.*, 2009, *103*: 4.

[11] Prada S., Giordano L., Pacchioni G., Li, Al, and Ni Substitutional Doping in MgO Ultrathin Films on Metals: Work Function Tuning via Charge Compensation [J]. *J. Phys. Chem. C*, 2012, *116*: 5781–5786.

[12] Stoessel M., Staudigel J., Steuber F., Simmerer J., Winnacker A., Impact of the Cathode Metal Work Function on the Performance of Vacuum-Deposited Organic Light Emitting-Devices [J]. *Appl. Phys. A*, 1999, *68*: 387–390.

[13] Tsiplakides, D., Work Function and Catalytic Activity Measurements of an $IrO_2$ Film Deposited on YSZ Subjected to in Situ Electrochemical Promotion [J]. *J. Electrochem. Soc.*, 1998, *145*: 905.

[14] Haastrup S., Strange M., Pandey M., Deilmann T., Schmidt P. S., Hinsche N. F., Gjerding M. N., Torelli D., Larsen P. M., Riis-Jensen A. C., Gath J., Jacobsen K. W., Mortensen J. J., Olsen T., Thygesen K. S., The Computational 2D Materials Database: High-Throughput Modeling and Discovery of Atomically Thin Crystals [J]. *2D Mater.*, 2018, *5*: 042002.

[15] Gjerding M. N., Taghizadeh A., Rasmussen A., Ali S., Bertoldo F., Deilmann T., Knosgaard N. R., Kruse M., Larsen A. H., Manti S., Pedersen T. G., Petralanda U., Skovhus T., Svendsen M. K., Mortensen J. J., Olsen T., Thygesen K. S., Recent Progress of the Computational 2D Materials Database (C2DB) [J]. *2D Mater.*, 2021, *8*: 044002.