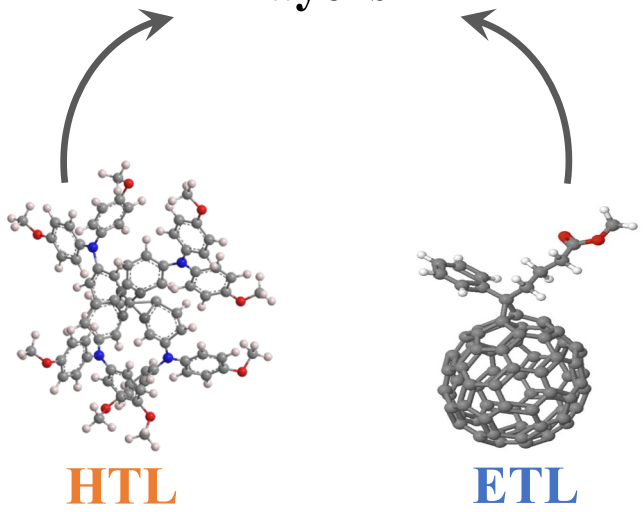
Prediction of novel charge transport layers for efficient perovskite solar cells

Anthonipillai Jenestin Anton, Armadorou Konstantina Kalliopi, Kessler Raluca-Ana

CH-457: AI for Chemistry
2 June 2023

Introduction

Charge transport layers



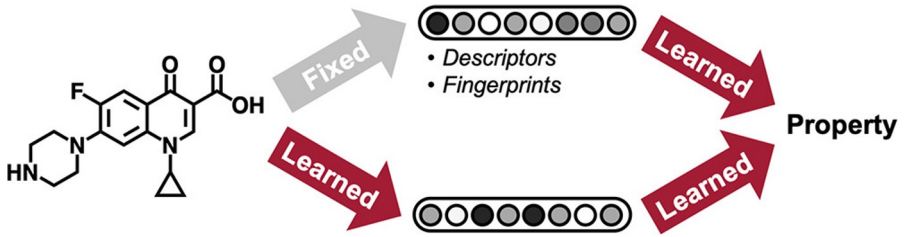
- Various classes of materials
- Non-systematic development of structures

Traditional ML methods

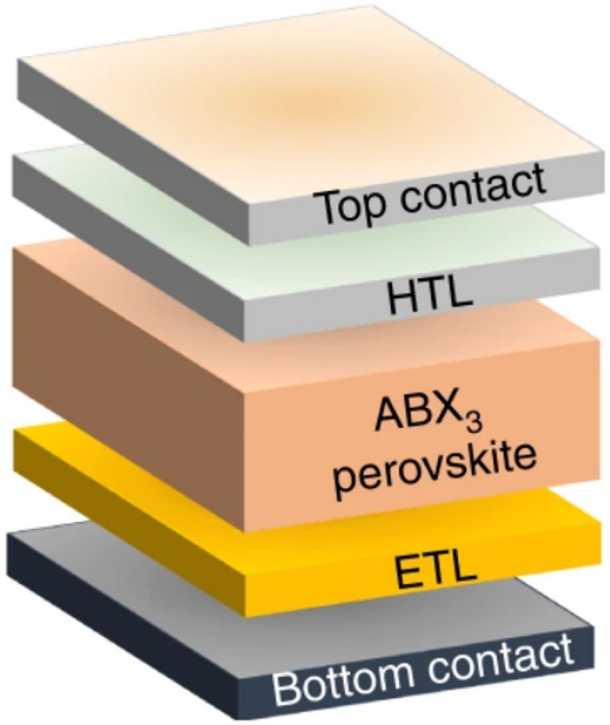
-construct a tree-like model
-able to handle both small and large datasets efficiently

Deep learning models

-based on neural network model
-more efficient but requires large dataset

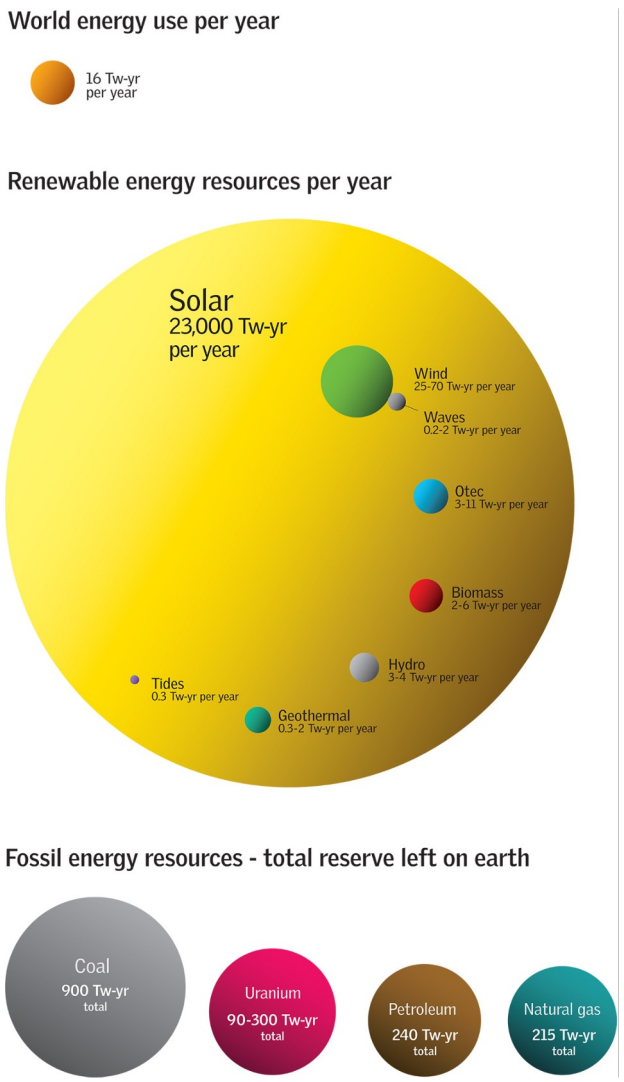


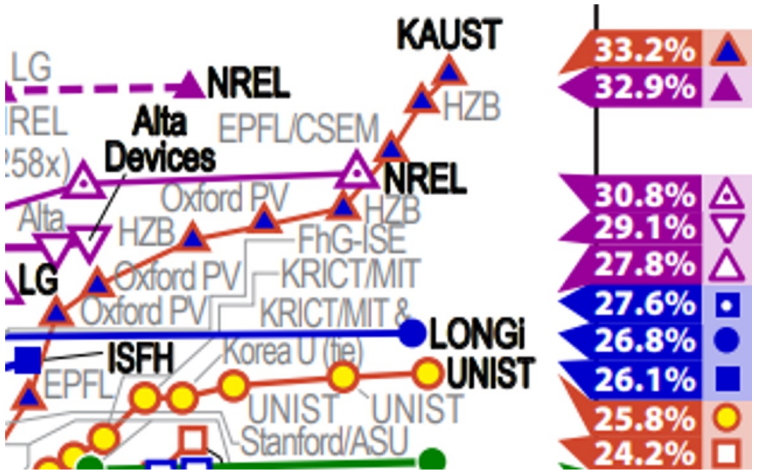
Use in Perovskite solar cells (PSCs)

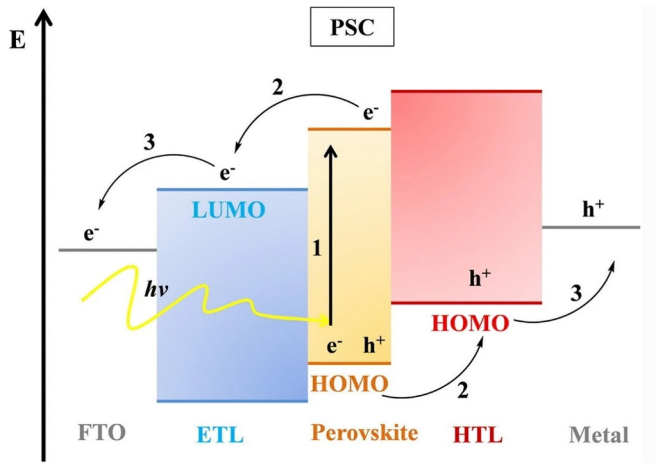


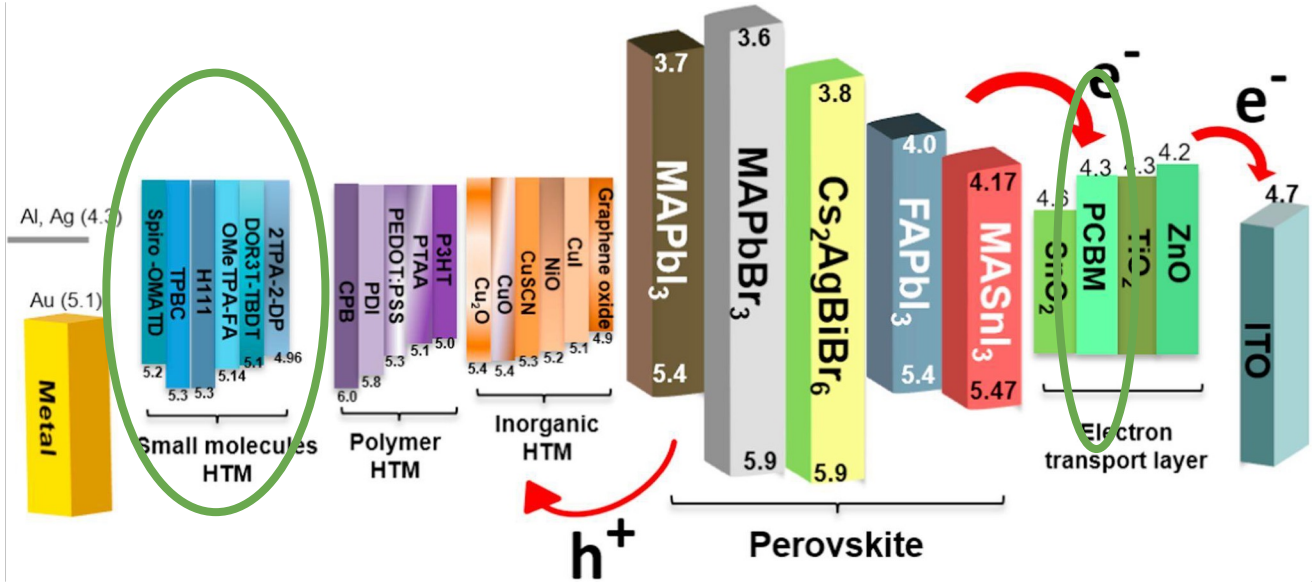
Nat Energy, **2022**, 7, 794–807

Working principle of PSCs







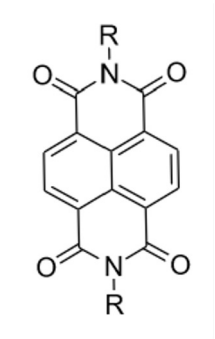


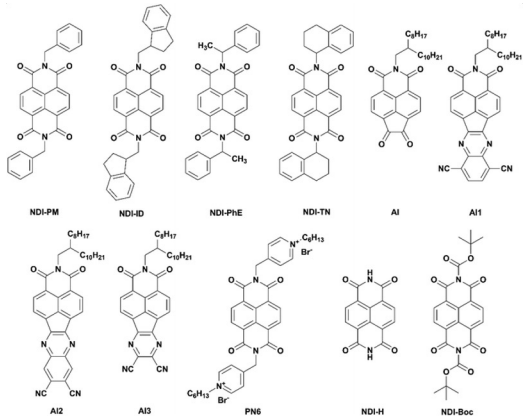
https://www.futureforall.org/energy/solar_energy.htm
https://www.nrel.gov/pv/cell-efficiency.html\
Emergent mater. **2018** 1, 133–154
Nano Energy **2017**, 34, 271-305

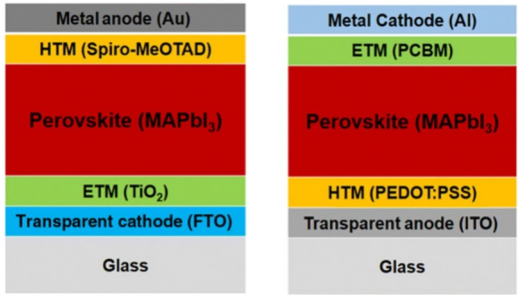
Improvement in solar cell devices can solve the energy issues

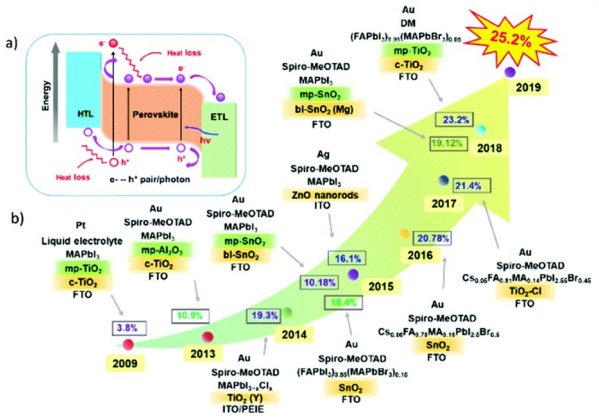
Synthesis scheme of new HTM and ETM

Step 1	Step 2	Step 3	Step 4
Choose a core structure and start the synthesis from there.	Synthesize a large number of molecules with different functional groups	Make multiple batches of perovskite solar cells devices with all the molecules (multiple architectures and combinations can be tried)	Test the performance of the perovskite solar cells and see which one performs better.









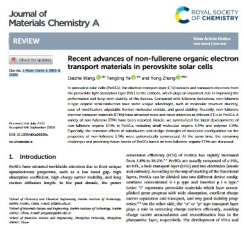
The whole protocol can take between 6 months up to one year

J. Mater. Chem. A, **2020**, 8,20819
Mater. Horiz., **2020**, 7, 2276

Tasks

1. Create the DataSet

- **Literature**
  Extract the compounds and their properties



- **Perovskite database**
  Extract the compounds and their properties

The Perovskite database

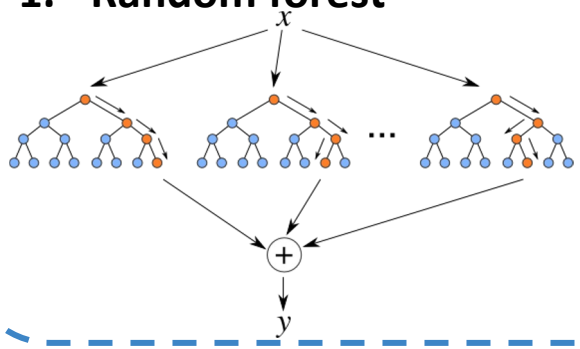
- **Scifinder**
  Extract the majority of SMILES

CAS
A division of the American Chemical Society

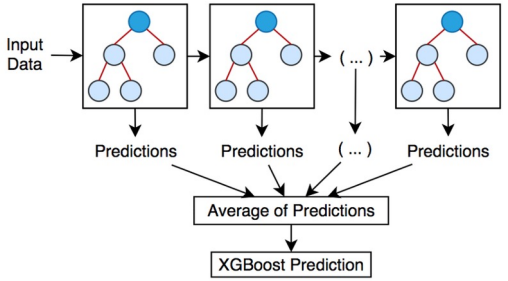
2. Apply ML

- **Traditional ML models**

1. **Random forest**

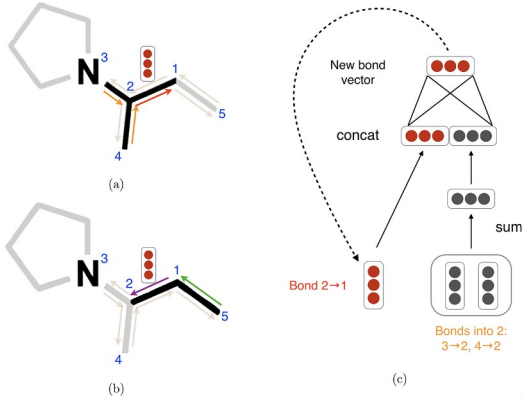


2. **XGBoost**



- **Deep learning models**

chemprop

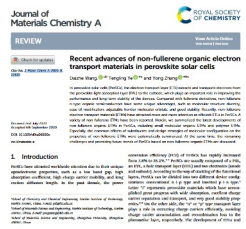


J. Chem. Inf. Model. **2019**, 59, 8, 3370–3388
J. Mater. Chem. A **2020**, 8, 20819
J. Phys. D: Appl. Phys. **2020**, 53 263001

Tasks

1. Create the DataSet

- **Literature**
  Extract the compounds and their properties



- **Perovskite database**
  Extract the compounds and their properties

The Perovskite database

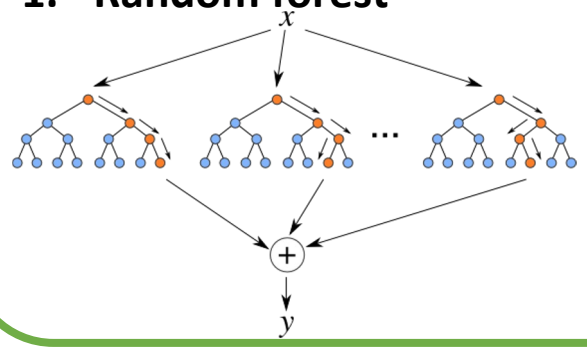
- **Scifinder**
  Extract the majority of SMILES

CAS
A division of the American Chemical Society

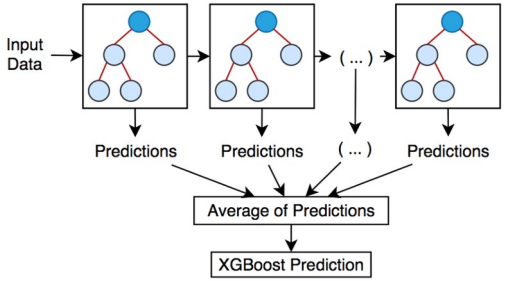
2. Apply ML

- **Traditional ML models**

1. **Random forest**

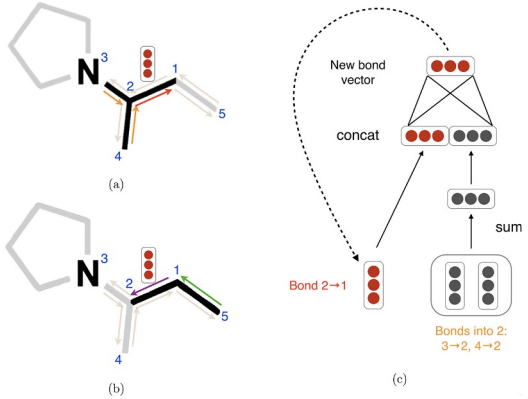


2. **XGBoost**



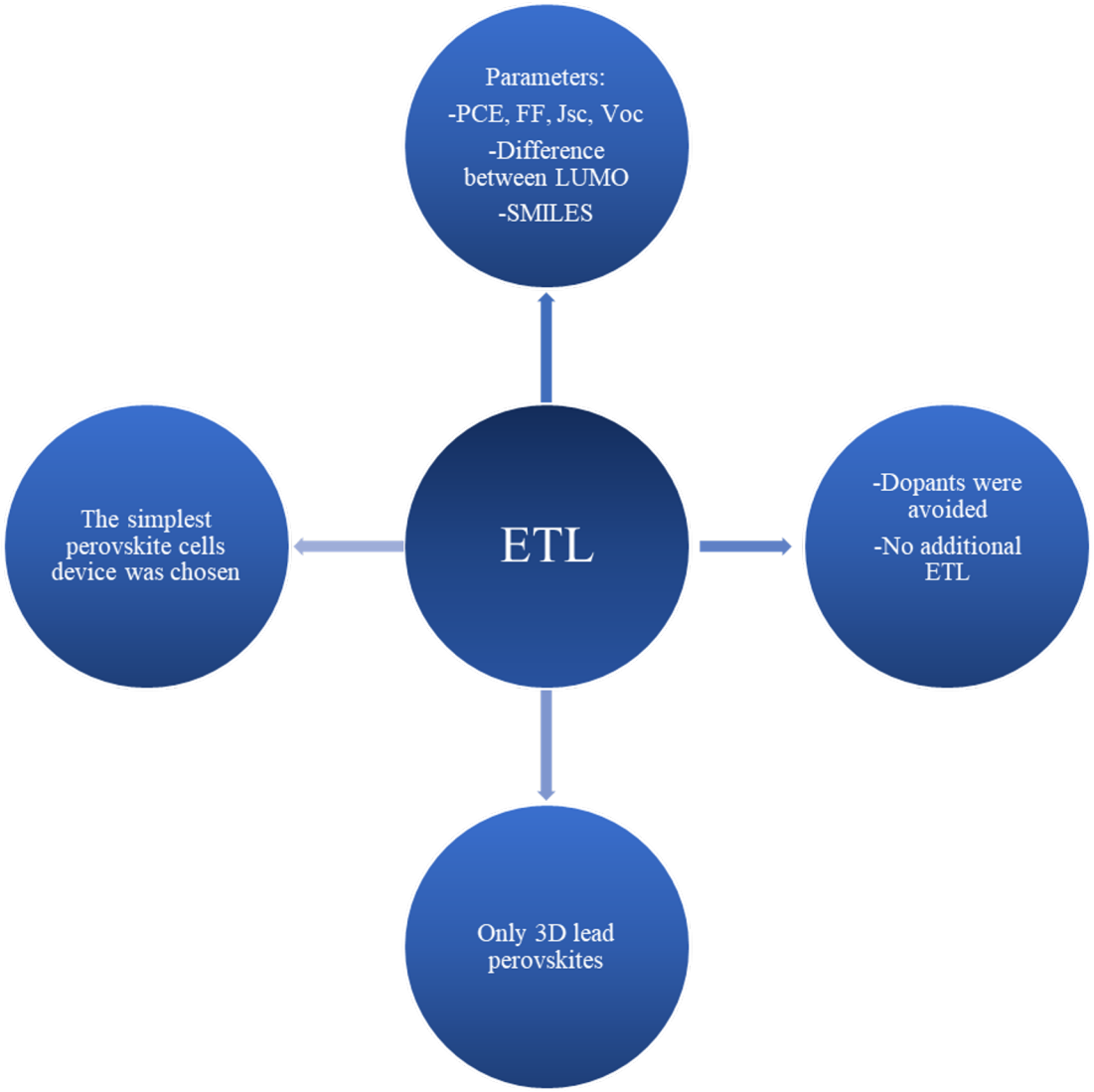
- **Deep learning models**

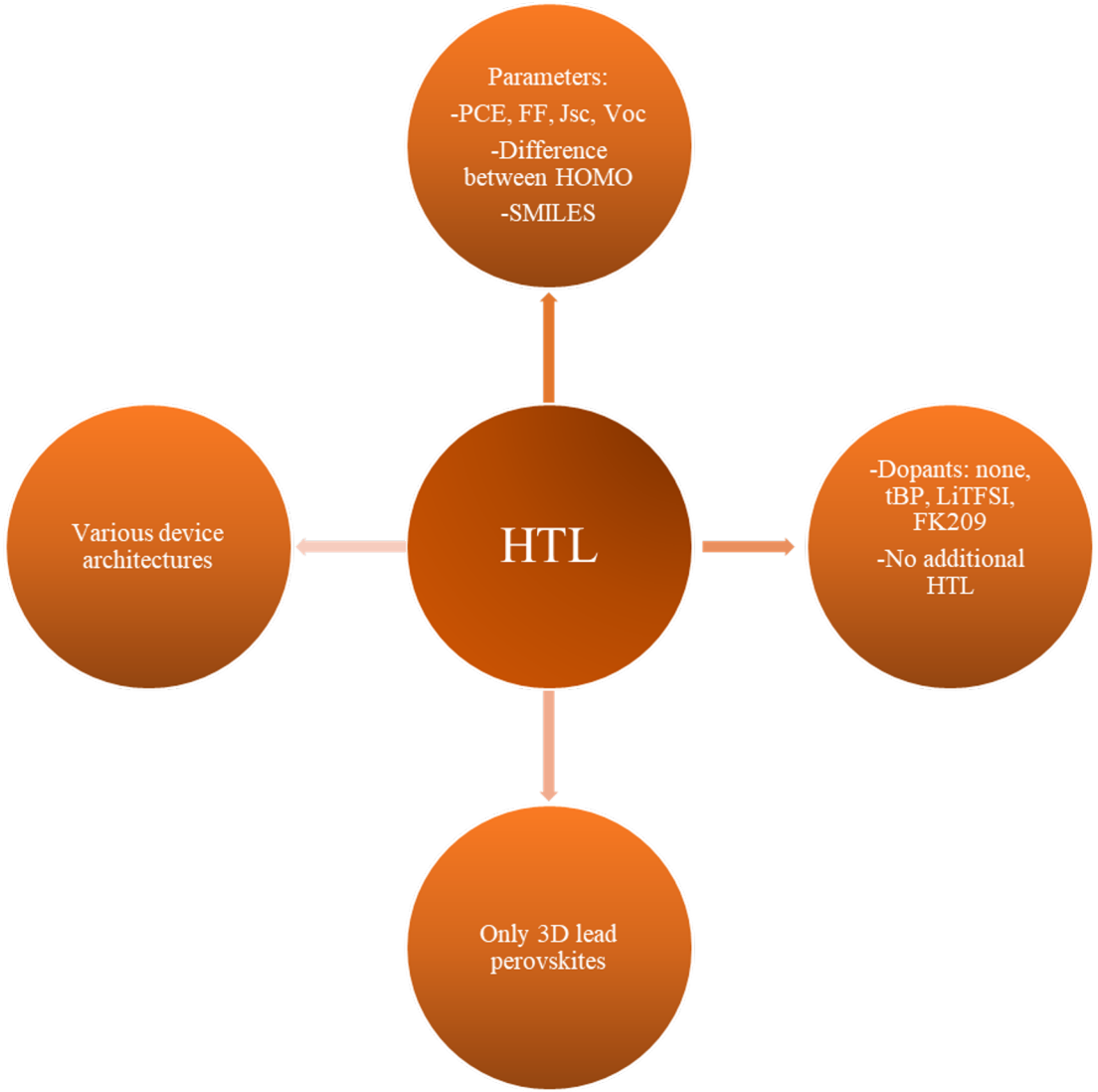
chemprop



J. Chem. Inf. Model. **2019**, 59, 8, 3370–3388
J. Mater. Chem. A **2020**, 8, 20819
J. Phys. D: Appl. Phys. **2020**, 53 263001

Establish Datasets for ETL and HTL





107 datapoints for ETL **220 datapoints for HTL**

Approach

- Data preprocessing:
  - Loading dataset, extracting SMILES strings and corresponding label PCE values
  - Processed SMILES strings → valid and canonical form

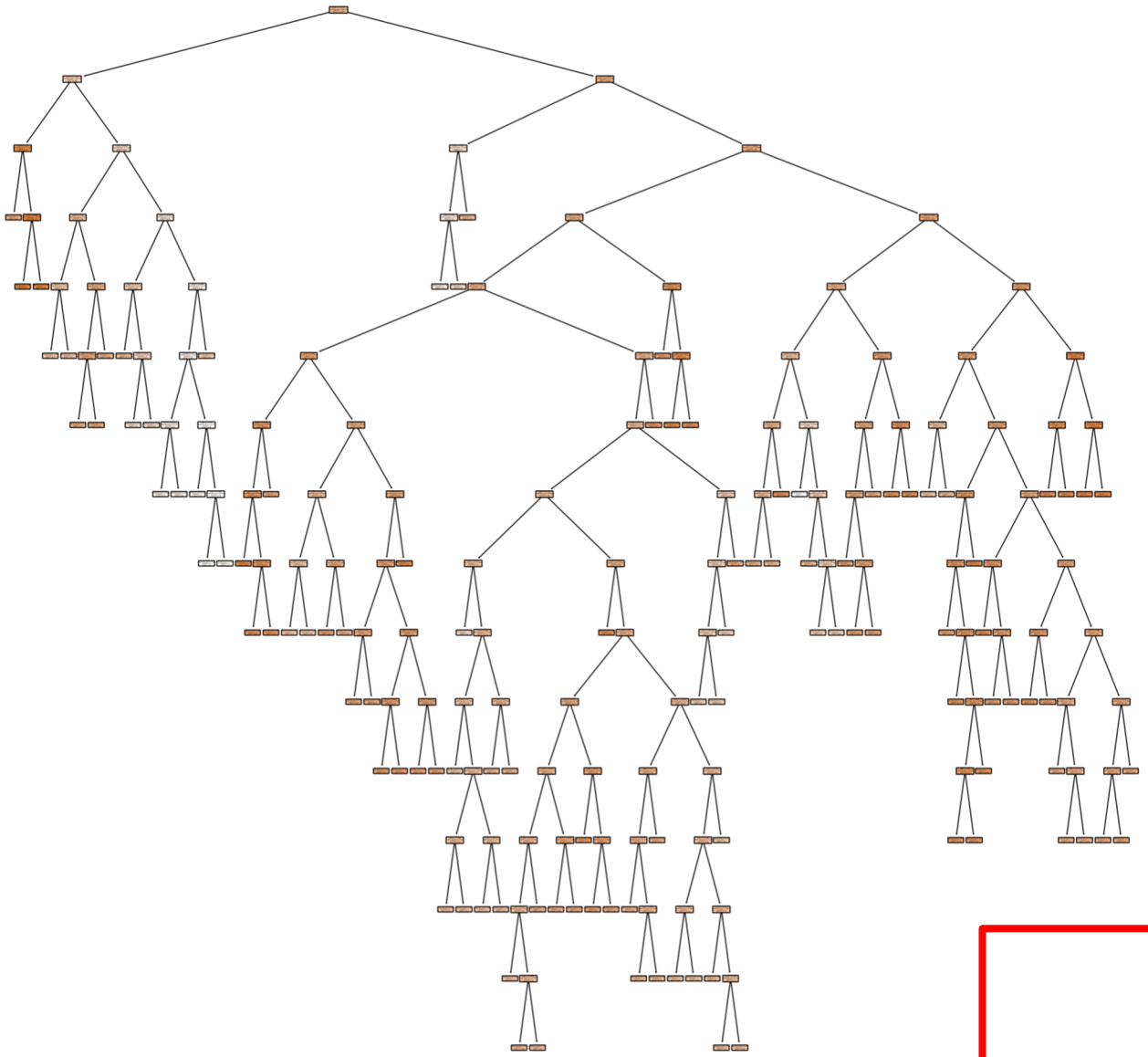
- Feature Extraction:
  - First manually (atomic mass, valence, hybridization, ring structure)
  - Used RDKit toolkit to transform SMILES with descriptors

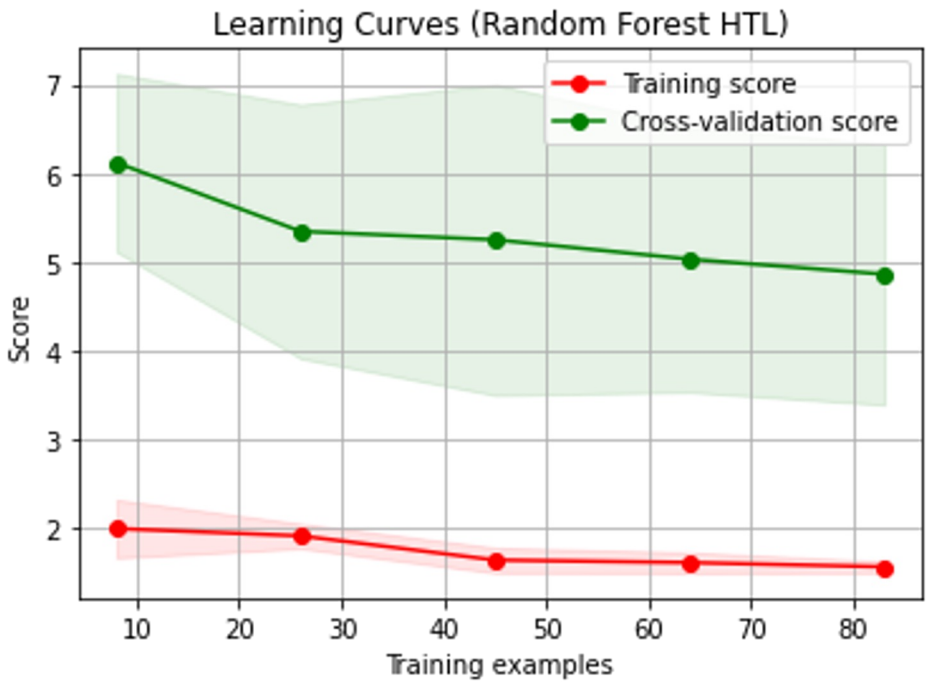
- Data Splitting
  - Split the dataset into training, validation and test sets.
  - 60% training set, 20% for validation set during model tuning and remaining 20% for testing the models final performance.
  - No Normalizing Features but tried

- Model Training and Evaluation:
  - Random Forest Regressor and XGBoost Regressor
  - Trained on training data, used validation set for model tuning (Grid Search)
  - Final model's performance evaluated on test set → unbiased estimation

- Training with initial parameters (Jsc, Voc, FF, PCE, SMILES) overfits → keep only SMILES and PCE as parameters

Results





Random Forest Regression Best

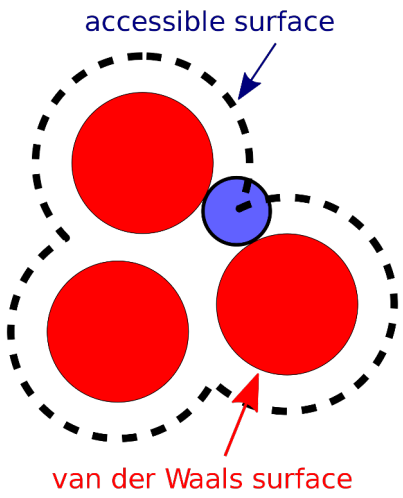
ETL → 63% VS HTL → 21%

Discussion

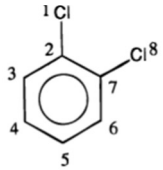
- **Topological molecular features**

Descriptor	Physical meaning	Score for HTLs	Score for ETLs
LabuteASA	Accessible surface area	0.059377	0.003893
EState_VSA3	Intrinsic valence state	0.047809	0.042932
PEOE_VSA7	Orbital electronegativity of atoms	0.038060	0.004212

LabuteASA

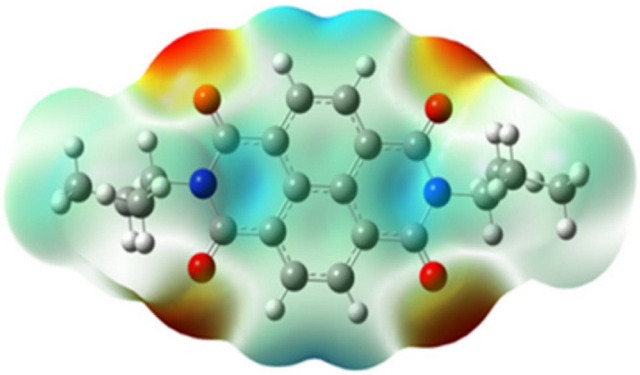


EState_VSA3



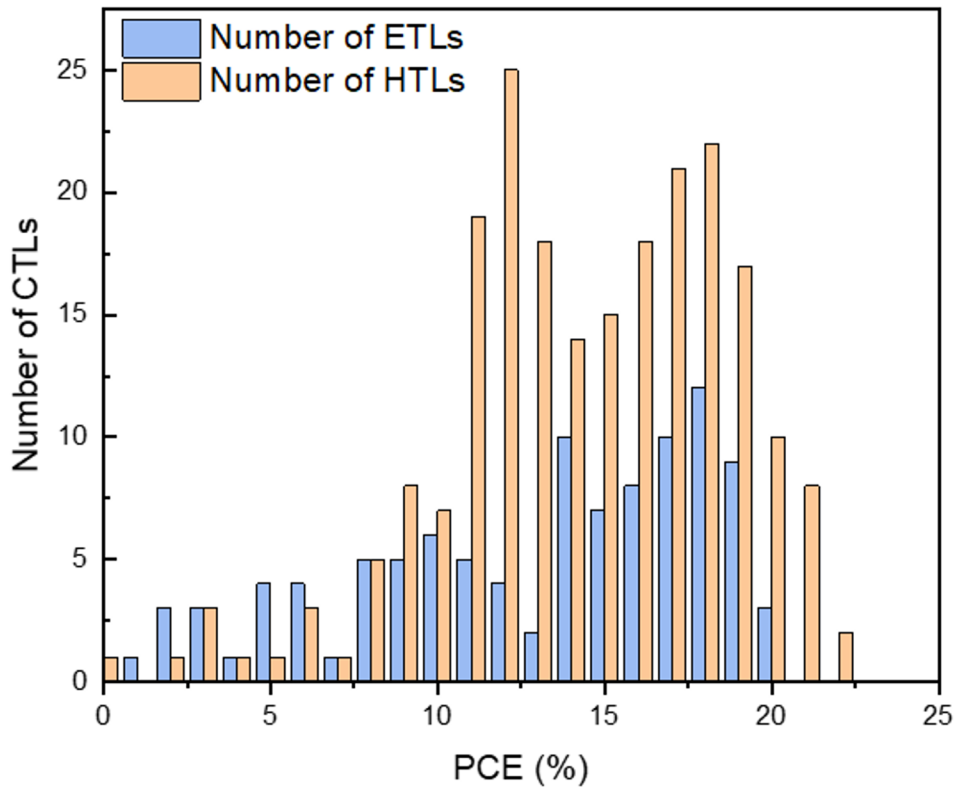
atom	atom id	E-state values
-Cl	1	5.577
-C<	2	0.606
-CH-	3	1.754
-CH-	4	1.841
-CH-	5	1.841
-CH-	6	1.754
-C<	7	0.606
-Cl	8	5.577

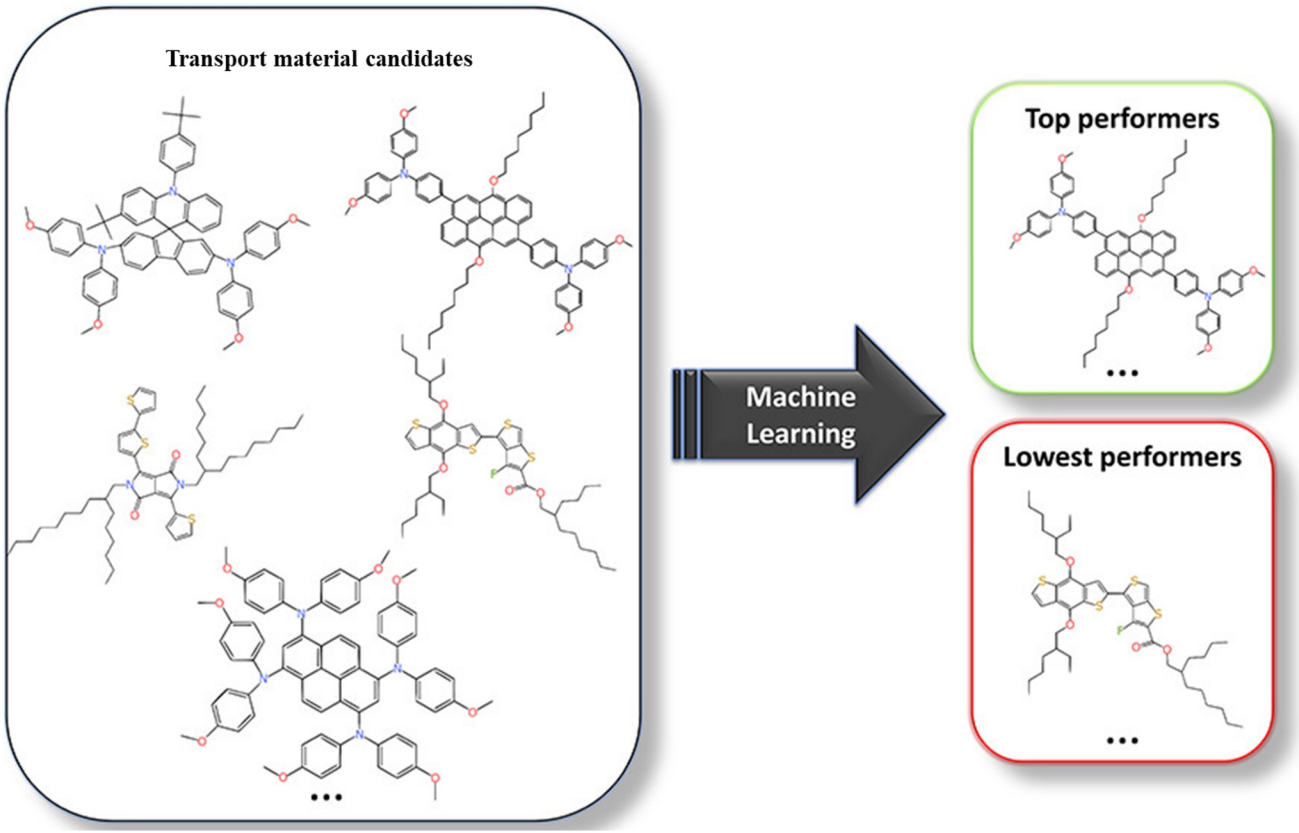
PEOE_VSA7



https://en.wikipedia.org/wiki/Accessible_surface_areai.
Cheminform **2018,** 10, 4
Materials **2021**, 14, 4026.

Limitations, Conclusions and Future Work





Similar average PCE (12.48% vs 14.35%) **BUT** different PCE distribution

Highly efficient inorganic materials not included in dataset
Dopants not explicitly included in the model
Similarities among HTL structures
Problem with missing data in literature (energy levels, mobilities)

We **succeed** to obtain good performance for ETL.
The implementation of this model **can solve many issues** (decrease the amount of work, decrease the amount of material).
Will help **at the evolution** of this new domain.

J. Phys. Chem. C 2022, 126, 31, 13053−13061

Model and Demonstration

Or

http://predskite.sns.network/

Thanks for your attention