RESEARCH PAPER

# Selection of Solar Photovoltaic and Battery Materials with the assistance of AI Based Algorithms

**Nirmal Kumar Agarwal*, Manisha Chaudhary, Vimal Gupta*, Nitima Malsa*, Sai Ramakrishna Buchi Reddy Karri, Ankush Ghosh**
Assistant Professor, Department of Electrical Engineering, JSS University, Noida (UP), India, nirmalnitham@gmail.com*
Assistant Professor, Department of Physics, JSS University, Noida (UP), India, manisha.chaudhary@jssuninoida.edu.in
Assistant Professor, Department of Computer Science Engineering-AIML, JSS University, Noida (UP), India, vimalgupta09@gmail.com
Associate Professor, Department of Computer Science Engineering, JSS University, Noida (UP), India, nitima.malsa@gmail.com
Ford Motor Company, USA, Sairamakrishna.karri@gmail.com
University Center for Research & Development (UCRD), Chandigarh University, India, ankushghosh@gmail.com
***Corresponding Author**
**Email:** nirmalnitham@gmail.com, vimalgupta09@gmail.com, nitima.malsa@gmail.com
**Received:**1stOctober,2026; **Revised:**5thOctober,2026; **Accepted:**8thOctober,2026;
**AvailableOnline:**20thOctober, 2026

**Abstract**
Finding and choosing superior materials for solar photovoltaic (PV) absorbers, battery electrodes, and electrolytes is essential to the shift to a renewable-energy economy. First-principles modeling and traditional trial-and-error testing are too expensive and slow to investigate the vast chemical design space. A unified artificial intelligence (AI) framework for materials selection is presented in this paper. It combines physics-based descriptors and figures of merit for PV and battery materials with machine learning (ML) property predictors, such as regularized regression, ensemble trees, and crystal-graph neural networks; uncertainty-aware Bayesian optimization and active learning; and multi-objective ranking using Pareto analysis, AHP, and TOPSIS.Together with the learning, acquisition, and decision-making equations that link them, all governing equations for photovoltaic efficiency, battery capacity, voltage, diffusion, and degradation are developed. Evaluation metrics, an end-to-end implementation process, and the primary unresolved issues (data scarcity, explainability, synthesisability, and sustainability) are covered. There are no new experimental measures reported in the paper; it is a methodological framework.
**Index Terms—**Graph neural networks, machine learning, materials informatics, multi-objective optimization, battery materials, artificial intelligence, and solar materials.
I. **How to cite this article:** Nirmal Kumar Agarwal* Manisha Chaudhary Vimal Gupta* Nitima Malsa* Sai Ramakrishna Buchi Reddy Karri Ankush Ghosh. Selection of Solar Photovoltaic and Battery Materials with the assistance of AI Based Algorithms. Int J Drug Deliv Technol. 2026;16(89s):2206-2207. DOI: 10.25258/ijddt.16.89s.211.

## INTRODUCTION

The two main components of decarbonised power systems are solar photovoltaic and electrochemical storage. Both are constrained at the device level by materials: a PV absorber's band gap, absorption, and defect tolerance determine its possible efficiency, whereas a battery's energy density, power, and lifespan are determined by the electrode and electrolyte materials' capacity, voltage, ionic transport, and stability. While a single synthesis-and-characterization cycle may take days or weeks, and a density functional theory (DFT) computation may take hours to days, the number of conceivable inorganic compositions, dopants, interfaces, and organic-inorganic hybrids is astronomically huge.
Sometimes referred to as the "fourth paradigm" of research, materials informatics reframes discovery as a data-driven challenge [1], [2]. ML models can learn the mapping from structure to property and filter candidates at a cost several orders of magnitude cheaper than DFT thanks to large open databases like the Materials Project [3] and compositional or structural featurization approaches [4]. Particularly for energy materials, ML has been used for lifetime prediction, electrode and solid-electrolyte discovery, and absorber screening [5]–[7]. The following research gap has been identified.

**1.1. PV absorber computed screening:**The bandgap dependency of single-junction efficiency was shown by Shockley and Queisser's detailed-balance limit [8], with a maximum of approximately 33% close to a bandgap of 1.3–1.4 eV under the AM1.5G spectrum. In order to screen potential absorbers from first principles, Yu and Zunger expanded on this by developing the spectroscopic limited maximum efficiency (SLME), which takes into consideration the actual absorption spectrum, absorber thickness, and non-radiative recombination [9].

**1.2. Battery materials:**Capacity, voltage, rate capability, safety, and cost are all traded off by layered oxides, spinels, olivines, silicon and lithium-metal anodes, and sulfide, oxide, and polymer electrolytes [10]. Voltage, volume change, and migration barriers are provided by first-principles computations, but their breadth is limited by their cost.

**1.3. Materials science under machine learning.** Crystal graph convolutional neural networks (CGCNN) [11], general-purpose composition-based models [4], and reviews of best practices [12]–[14] demonstrate that machine learning (ML) can accurately predict formation energy, bandgap, and elastic characteristics at a negligible cost. Battery cycle life has also been estimated by data-driven models using early-cycle signals [15]. Dedicated solutions have been prompted by the persistent challenge of small datasets [16]. The loop between synthesis and prediction is closed by autonomous experimental platforms [17].

## II. MATHEMATICAL MODELING OF SOLAR PHOTOVOLTAIC & BATTERY

This section defines the target properties that the AI models in Section IV predict or optimise.

### 2.1Materials of the Photovoltaic Absorber

Equation 1 provides a solar cell's power conversion efficiency.

$$\eta = (J_{SC} \cdot V_{OC} \cdot FF) / P_{in} \quad (1)$$

where P_in is the incidence power density (1000 W/m² for AM1.5G), FF is the fill factor, V_OC is the open-circuit voltage, and $J_{SC}$ is the short-circuit current density. Equation 2 displays the photogenerated current for an absorber with thickness d.

$$J_{SC} = q \int \Phi(E) \cdot A(E) \cdot IQE^*(E)\, dE \quad (2)$$

where q is the fundamental charge, Φ(E) is the event's photon flux density, A(E) is the absorption coefficient, and IQE is the inner quantum effectiveness. Regarding a single-pass absorber,$A(E) = 1 - \exp(-^{\alpha(E)\, d})$. An estimate of the absorption coefficient close to a direct-gap semiconductor's band edge is used in Equation 3.

$$\alpha(E) \approx A_0 \sqrt{(E - E_g) / E}, \quad E \geq E_g \quad (3)$$

where $A_0$ is a material-specific constant and E_g is the bandgap. The integral in (2) begins at E_g because in the ideal (Shockley–Queisser) situation, A(E) = 1 for E ≥ E_g and 0 otherwise.The open-circuit voltage follows from the diode equation:

$$V_{OC} = (n\, k_B\, T / q) \cdot \ln( J_{SC} / J_0 + 1 ) \quad (4)$$

when n is the ideality factor, $J_0$ is the reverse saturating current density, T is the cell temperature, and k_B is the Boltzmann's constants. The radiative recombination fraction f_r introduces non-radiative recombination in the SLME formulation [9]:

$$J_0 = J_0^{rad} / f_r, \quad f_r = \exp( -\Delta E / (k_B\, T) ) \quad (5)$$

where $J_0^{rad}$ is the radiation saturation current and ΔE is the difference between the direct permissible gap and an essential gap. Greater ΔE results in lower V_OC as well as higher non-radiative loss.Green's empirical expression can be used to estimate the fill factor:

$$FF_0 = ( v_{oc} - \ln(v_{oc} + 0.72) ) / ( v_{oc} + 1 ), \quad v_{oc} = q\, V_{OC} / (n\, k_B\, T) \quad (6)$$

and Eq. 7 yields the maximum-power point.

$$P_{max} = \max \text{ over } V \text{ of } [\, V \cdot J(V) \,], \quad J(V) = J_{SC} - J_0 ( \exp( qV / (n\, k_B\, T) ) - 1 ) \quad (7)$$

which gives the SLME-type efficiency η = P_max / P_in.

$$E_{hull} = E_f(\text{compound}) - E_f(\text{hull}) \quad (8)$$

Thermodynamics & structural stability are also necessary for practical selection. Equation 8 provides the energy beyond the convex hull. E_f(hull) is the resultant energy of the best stable configuration of contending phases at the same content. When E_hull is tiny, often less than a few tens of meV/atom, stability is expected.Equation 9 provides Goldschmidt tolerance factor for perovskites with the formula $ABX_3$.

$$t = ( r_A + r_X ) / ( \sqrt{2} \cdot ( r_B + r_X ) ) \quad (9)$$

Ionic radii r_A, r_B, and r_X show geometric feasibility. T = 0.8–1.0 is frequently linked to stable perovskite phases. The carrier effective mass, defect generation energies, & carrier mobility μ_c ultimately define the diffusion length $L = \sqrt{(D \tau)}$, using $D = \mu_c k_B T / q$ (Einstein relation) and carrier lifespan τ.

### 2.2 Battery Electrode and Electrolyte Materials

The assumed gravimetric capacity for a material for an electrode is given by Equation 10. where n is the total number of electrons moved per chemical unit, F = 96485 C/mol is Faraday's constant, and M is the molar mass in g/mol. The assumed particular energy of a half-cell is given by equation 11. where V_avg is the mean operational voltage and C_th in Ah/g is appropriately scaled. The optimum cell voltage is determined by the Nernst equation in Equation 12, which uses the gas constant R with reaction quotient Q. From fundamental principles, equation 13 gives the average intercalation voltage that exists between the lithium concentrations $x_1$ and $x_2$.

$$C_{th} = n\, F / (3.6\, M) \ [mAh/g] \quad (10)$$

$$E_{sp} = C_{th} \cdot V_{avg} \ [Wh/kg] \quad (11)$$

$$E = E^0 - ( R\, T / (n\, F) ) \cdot \ln Q \quad (12)$$

$$V_{avg} = - [\, E(Li_{x_2}\ Host) - E(Li_{x_1}\ Host) - (x_2 - x_1)\, E(Li) \,] / ( (x_2 - x_1)\, e ) \quad (13)$$

where E(·) are DFT total energies and e is the electron charge. Volumetric change on cycling, a key indicator of mechanical degradation, is given in Eq.14

$$\Delta V / V_0 = ( V_{lithiated} - V_{delithiated} ) / V_{delithiated} \quad (14)$$

Ionic transportation in electrolyte & electrodes is described by the Arrhenius relationships, which are given in Eq. 15. where E_a is an activation movement barrier, σ_i is the ionic conductivity value, and D is the diffusion - ratio. The nudged elastic bands approach is typically used to

approximate such values. Solid electrolytes are also screened for an electrochemically stable window and a $Li^+$ transfer number close to unity. An empirical representation of capacity fade is given in Equation 16. where N is the cycle number, $Q_0$ is the initial maximum capacity, and a and b are fitted constants. SoH = 0.8 is the conventional definition of the ending a life. The structure of the proposed system is shown in Fig.1

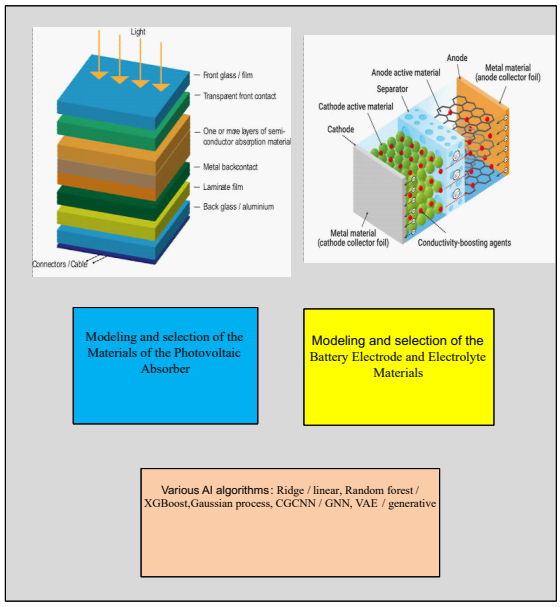



**Fig.1 Structure of the proposed system**

**TABLE I. Representative target properties and AI input descriptors.**

| Application | Target properties (output y) | Typical input descriptors (x) |
|---|---|---|
| PV absorber | $E_g$, $\alpha(E)$, SLME, $E_{hull}$, effective | Crystal structure, tolerance factor, density of states, and elemental statistics (electronegativity, |

## III. AI METHODS FOR MATERIAL PROPERTY PREDICTION AND DISCOVERY

### 3.1 Problem Formulation

Let $\mathcal{D} = \{(x_i, y_i)\}$, i = 1,...,N, be a dataset where $y_i$ is the target attribute and $x_i \in \Psi^d$ is the feature vector of material i. Equation 18 provides the learning task, which is to identify a function $f_\theta$ that minimizes the regularized empirical risk.

$\theta^* = \text{argmin over } \theta \text{ of } [ (1/N) \Sigma_i \ell( y_i, f_\theta(x_i) ) + \lambda \Omega(\theta) ]$ (18)

where $\ell$ is a loss (e.g., squared error), $\Omega$ a regulariser such as $\|\theta\|^2$ and $\lambda \geq 0$ controls model complexity.

### 3.2 Feature Engineering

Equation 19 for composition-based models uses fraction-weighted statistics [4] to aggregate each property p of constituent element k (atomic radius, electronegativity, etc.).

$D = D_0 \exp( -E_a / (k_B T) )$, $\sigma_i = ( \sigma_0 / T )\exp( -E_a / (k_B T) )$ (15)

$Q(N) = Q_0 ( 1 - a N^b )$, $SoH(N) = Q(N) / Q_0$ (16)

A materials-level cost metric combines raw-material pricesis given in Eq.17

$Cost = ( \Sigma_i m_i c_i ) / E_{cell}$ [$/kWh] (17)

where $m_i$ and $c_i$ are the mass and unit price of constituent i per cell, and $E_{cell}$ is the cell energy.

| Application | Target properties (output y) | Typical input descriptors (x) |
|---|---|---|
| | mass, defect tolerance | radius, and valence) |
| Cathode | $V_{avg}$, $C_{th}$, $\Delta V/V$, $E_{hull}$ | Li site topology, lattice parameters, composition, and transition-metal oxidation state |
| Anode | $C_{th}$, $\Delta V/V$, D | host volume, coordination, bond lengths, and composition |
| Solid electrolyte | $\sigma_i$, $E_a$, stability window | Li-site connectivity, bond ionicity, anion framework, and phonon characteristics |
| Cell-level | Cycle life, SoH, cost | Features of early-cycle voltage and capacity, materials cost vector |

$\bar{f}_p = \Sigma_k w_k p_k$, $MAD_p = \Sigma_k w_k | p_k - \bar{f}_p |$, $p_{min}$, $p_{max}$, $range_p$ (19)

where the atomic fraction of element k is denoted by $w_k$. x is created by concatenating these statistics. Highly correlated or low-variance features are eliminated prior to training, and features are standardized as $\bar{x} = (x - \mu_x)/\sigma_x$.

### 3.3 Regression Models

1) Ridge regression provides a linear baselineis given in Eq.20

$\hat{y} = w^T x + b$, $w^* = (X^T X + \lambda I)^{-1} X^T y$ (20)

2) Random forest averages B decorrelated decision trees trained on bootstrap samplesis given in Eq.21

$\hat{y}(x) = (1/B) \Sigma_bT_b(x)$ (21)

where the spread of $\{T_b(x)\}$ also yields a simple uncertainty estimate.

3) Gradient boosting (e.g., XGBoost) builds an additive model stage-wiseis given in Eq.22

F_m(x) = F_(m−1)(x) + ν h_m(x), $\mathcal{L} = \Sigma_i \ell(y_i, \hat{y}_i) + \Sigma_m \Omega(h_m)$, $\Omega(h) = \gamma T_{leaf} + \frac{1}{2} \lambda \|w\|^2$ (22)
where ν is the learning rate, h_m the m-th tree, T_leaf the number of leaves and w the leaf weights [18].
4) Crystal graph convolutional neural network (CGCNN).A graph with atoms as nodes and edges connecting neighbors within a cutoff radius is used to depict a crystal. Equation 23 provides an expansion of interatomic distances using Gaussian basis functions.
$u_t(i,j) = \exp( -( d_{ij} - \mu_t )^2 / \sigma^2 )$ (23)
and node features are updated by the gated convolution [11]
$v_i^{(t+1)} = v_i^{(t)} + \Sigma_{(j \in N(i))} \sigma( z_{ij} W_f^{(t)} + b_f^{(t)} ) \odot g( z_{ij} W_s^{(t)} + b_s^{(t)} )$, $z_{ij} = v_i^{(t)} \oplus v_j^{(t)} \oplus u(i,j)$ (24)
where W, b are learnable weights and biases, g is a softplus activation, Ψ is concatenation, Ψ is element-wise multiplication, and σ is the sigmoid. To predict y, a pooled crystal vector v_c = pool($v_1^{(T)}$,…,$v_n^{(T)}$) is passed through completely connected layers. The loss in (18) is used to train the network using the Adam optimizer [19].

**3.4 Uncertainty Quantification**

Because predictions are used to determine which materials are synthesized, uncertainty is crucial. A closed-form predictive distribution is produced using Gaussian process (GP) regression [20]. Equation 25 provides the radial-basis-function kernel with $f \sim (m(x), k(x, x'))$.
$k(x, x') = \sigma_f^2 \exp( -\|x - x'\|^2 / (2\ell^2) )$ (25)
the posterior mean and variance at a new point x* are given in Eq.26 and Eq.27
$\mu(x^*) = k^{*T} ( K + \sigma_n^2 I )^{-1} y$ (26)
$\sigma^2(x^*) = k(x^*, x^*) - k^T ( K + \sigma_n^2 I )^{-1} k$ (27)
where $\sigma_n^2$ is the noise variance, K is the training covariance matrix, and k* is the covariance vector between x* and the training inputs.For neural networks, an ensemble of M models yields $\hat{\mu} = (1/M) \Sigma f_m(x)$ and $\sigma^2 = (1/M) \Sigma ( f_m(x) - \hat{\mu} )^2$.

**TABLE II. Qualitative comparison of model families for materials selection.**

| Model | Input type | Strengths | Limitations |
|---|---|---|---|
| Ridge / linear | Features of composition | Quicker , interpretable | It is impossible to accurately represent strong non-linearity. |
| Random forest / XGBoost | Composition features | Sturdy on little to medium-sized data, inherent | Absence of structural awareness and limited projection |

**3.5 Bayesian Optimisation and Active Learning**

By optimizing an acquisition function based on (26)–(27), Bayesian optimization (BO) chooses the next candidate to assess [21]. Equation 28 provides the predicted improvement with the best observed value $f^+$.
$EI(x) = ( \mu(x) - f^+ - \xi ) \Phi(Z) + \sigma(x) \varphi(Z)$, $Z = ( \mu(x) - f^+ - \xi ) / \sigma(x)$ (28)
where Φ and φ are the standard normal CDF and PDF and $\xi \geq 0$ controls exploration. The upper confidence bound alternative is given in Eq.29
$UCB(x) = \mu(x) + \kappa \sigma(x)$ (29)
The model is retrained on the larger dataset following the subsequent experiment or DFT calculation, x_(next) = argmax over x $\in \mathcal{X}$_cand of α_acq(x). The acquisition value is multiplied by the probability of feasibility, $P(E_{hull}(x) \leq \varepsilon)$, to accommodate constraints like $E_{hull} \leq \varepsilon$.

**3.6 Generative Design**

A variational autoencoder (VAE) may learn a continuous latent representation z of compositions or structures when the search space is not an enumerated list. It can be trained by maximizing the evidence lower bound, which is provided in Eq. 30.
$ELBO = E_{(q(z|x))} [ \log p(x \mid z) ] - D_{KL}( q(z \mid x) \| p(z) )$ (30)
The aforementioned property predictors are used to direct optimization in z-space, which yields unique candidates that are subsequently decoded and filtered for stability and validity.

**3.7 Explainability**

Shapley values [22], which are provided in Eq. 31, are used to quantify feature contributions in order to make the model scientifically meaningful.
$\varphi_i = \Sigma$ over $S \subseteq F \setminus \{i\}$ of $[ |S|! ( |F| - |S| - 1 )! / |F|! ] \cdot [ f(S \cup \{i\}) - f(S) ]$ (31)
where f(S) is the model prediction using only the features in S and F is the whole feature set. Physically significant design principles are indicated by descriptors with large $|\varphi_i|$, such as electronegativity difference or Li–O bond length.

| Model | Input type | Strengths | Limitations |
|---|---|---|---|
| | | significance | |
| Gaussian process | Features/kernels | Calibrated uncertainty, ideal for BO | Cubic cost |
| CGCNN / GNN | structure of the Crystal | acquires knowledge of the structure-property map | Needs structures and more data |
| VAE / | Latent of the | Proposes | Validity |

| Model | Input type | Strengths | Limitations |
|---|---|---|---|
| generative | vectors | novel materials | and synthesisability checks required |

## IV. FORMULATION OF MULTI-OBJECTIVE MATERIAL SELECTION

High efficiency or capacity, low cost, high stability, and minimal toxicity are some of the conflicting requirements that a viable PV or battery material must meet. Let $F(x) = [f_1(x),\ldots, f_m(x)]$ be the objective vector to be maximized (cost-like objectives are negated).

### 4. 1 Pareto Optimality

Candidate $x_1$ dominates $x_2$ if

$f_k(x_1) \geq f_k(x_2)$ for all k, and $f_j(x_1) > f_j(x_2)$ for at least one j (32)

The group of non-dominated candidates is known as the Pareto front. Non-dominated sorting and a crowding-distance metric are used by evolutionary algorithms like NSGA-II [23] to locate this front while maintaining diversity.

### 4.2 Weighted Normalised Scoring

Objectives are first scaled to a common range,

$\tilde{f}_k = ( f_k - f_k\text{^min} ) / ( f_k\text{^max} - f_k\text{^min} )$ (33)

and combined into a scalar score $J(x) = \Sigma_k w_k \tilde{f}_k(x)$, with $\Sigma_k w_k = 1$ and $w_k \geq 0$.

### 4.3 AHP Weighting

The Analytic Hierarchy Process (AHP) [24] constructs a pairwise comparison matrix **A** from expert judgment and solves the eigenvalue issue, which is provided in Eq. 34, in order to obtain $w_k$ in a clear manner.

$\mathbf{A}\, w = \lambda_max w,\ CI = ( \lambda_max - m ) / ( m - 1 ),\ CR = CI / RI$ (34)

where RI is the random index and CI is the consistency index for a matrix of size m. When CR is less than 0.1, judgments are acceptable.

### 4.4 TOPSIS Ranking

For L candidates and m criteria with decision matrix x_ij, TOPSIS [25] normalises the matrixis given in Eq.35

$r_{ij} = x_{ij} / \sqrt{( \Sigma_i x_{ij}^2 )}$ (35)

Equation 35 shows how TOPSIS [25] normalizes the decision matrix x_ij for L candidates and m criteria. finds the ideal and anti-ideal solutions by applying the weights, $v_{ij} = w_j r_{ij}$. Equation 36 calculates the separation measures using $A^+ = \{v_j^+\}$ and $A^- = \{v_j^-\}$, which represent the greatest and worst values of each criterion.

$S_i^+ = \sqrt{( \Sigma_j ( v_{ij} - v_j^+ )^2 )},\ S_i^- = \sqrt{( \Sigma_j ( v_{ij} - v_j^- )^2 )}$ (36)

The closeness coefficientis given in Eq.37

$C_i = S_i^- / ( S_i^+ + S_i^- ),\ 0 \leq C_i \leq 1$ (37)

Candidates are ranked, and a higher $C_i$ indicates superior overall content. The ranking can be repeated over Monte Carlo draws $y \sim (\mu, \sigma^2)$ to report the chance that each material ranks in the top-k when uncertainty is available from (27). As an additional criterion, a sustainability index could be introduced. For example, in Eq. 38, the supply-risk score ($\rho$), environmental impact ($\varepsilon$), and toxicity ($\tau$) for every component i are normalized to [0, 1] (lower is preferable, therefore the index is ignored before ranking).

$S_sus = \Sigma_i w_i ( \rho_i + \varepsilon_i + \tau_i )$ (38)

## V. IMPLEMENTATION WORKFLOW

There are six stages in the proposed pipeline:

**5.1 Information collection.** Compile structures and characteristics from internal experiments, internet databases (such as the Materials Project [3]), and the literature. Harmonize units, remove duplication, and flag values from discordant DFT functionals.

**5.2 Including features.** Compute compositional statistics using (19) and, if structures exist, graph representations for (23)–(24).

**5.3 Pattern training.** Use categorized splits (e.g., by chemical system) to separate the data into training, validation, test, and standard sets in order to avoid exuberant predictions. For each target in Table I, train a different model and use cross-validation to modify hyperparameters..

**5.4 Assessment and learning by doing.** Predict the characteristics of the candidate library, remove candidates that don't fit strict criteria (such E_hull> ε or poisonous items), and then select the next batch using (28) or (29).

**5.5 Placing orders.** Use (33)–(38) to combine projected qualities with cost as well as sustainability data in order to choose the nominated materials and display Pareto-optimal sets.

**5.6 Checking.** To verify the optimal choices, employ DFT (bandgap, voltage, E_hull, barriers) and, if possible, synthesis and device test. Retrain once the results have been added to the dataset.

## VI. RESULTS AND DISCUSSION

### 6.1 Evaluation Metrics

Predictive accuracy is reported in Eq.39

$MAE = (1/N) \Sigma | y_i - \hat{y}_i |,\ RMSE = \sqrt{( (1/N) \Sigma ( y_i - \hat{y}_i )^2 )},\ R^2 = 1 - \Sigma ( y_i - \hat{y}_i )^2 / \Sigma ( y_i - \bar{y} )^2$ (39)

By comparing the percentage of observations that fall inside the expected confidence interval with its nominal level, uncertainty calibration is verified. The enrichment factor is used to measure discovery performance.

$EF = ( \text{hits in top-k} / k ) / ( \text{hits in library} / N_lib )$ (40)

along with the quantity of costly assessments required in comparison to random search in order to achieve a target property. The total speed-up in screening can be written as in Eq.41

$$S = ( N_{lib} \cdot t_{DFT} ) / ( N_{lib} \cdot t_{ML} + N_{val} \cdot t_{DFT} ) \quad (41)$$

where N_val is the number of candidates that were subsequently verified, and t_ML and t_DFT are the costs of ML inference and DFT per candidate.

**6.2 Worked Reference Point**

To verify the physical models, the well-known Shockley–Queisser finding [8] is recovered by setting n = 1 and A(E) = 1 for E ≥ E_g in (1)–(7) under AM1.5G: a maximum efficiency of around 33–34% at E_g ≈ 1.3–1.4 eV, dropping at both smaller and bigger gaps. Since efficiency is most susceptible to errors close to the optimum, each trained bandgap predictor should be evaluated not only by its error on E_g but also by its error propagated into η through (2)–(7).Values or weights of solar photovoltaic material and battery material are shown in Fig.2 and Fig.3 while the confusion matrix is shown in Fig.4

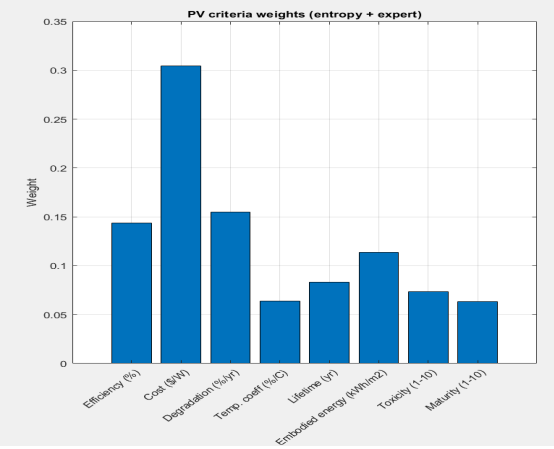



Fig.2 Values or weights of solar photovoltaic material

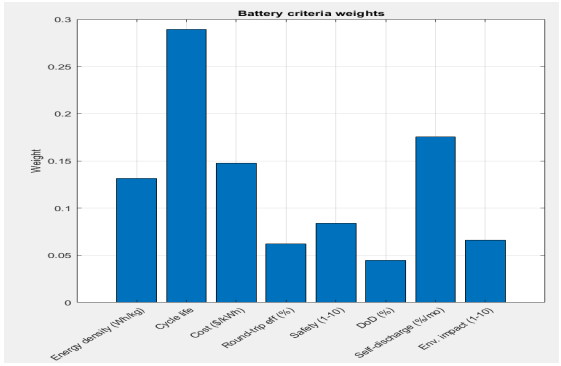



Fig.3 Values or weights of battery material

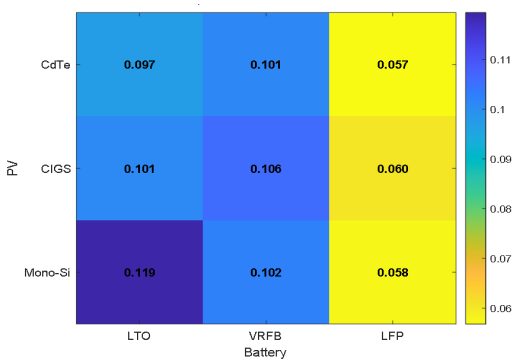



Fig.4Confusion matrix of the battery and SPV

## VII. CONCLUSION

The AI framework for choosing solar PV and battery materials provided in this study combines physics-based figures of merit, machine learning predictors, uncertainty-aware active learning, and multi-criteria decision analysis into a single equation-level pipeline. The framework keeps cost, stability, and sustainability apparent in the decision-making process while reducing the length of time from a very vast candidate space to a small, prioritized list of materials for verification. Future work will involve extending the models to multi-fidelity and interface-aware prediction, combining with autonomous experimentation for closed-loop validation, and implementing the pipeline on particular databases and publishing benchmark results.

## REFERENCES

[1] A. Agrawal and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the 'fourth paradigm' of science in materials science," APL Mater., vol. 4, no. 5, Art. no. 053208, 2016.

[2] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," Nature, vol. 559, pp. 547–555, 2018.

[3] A. Jain et al., "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," APL Mater., vol. 1, no. 1, Art. no. 011002, 2013.

[4] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," npjComput. Mater., vol. 2, Art. no. 16028, 2016.

[5] C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, and S. P. Ong, "A critical review of machine learning of energy materials," Adv. Energy Mater., vol. 10, no. 8, Art. no. 1903242, 2020.

[6] A. Chen, X. Zhang, and Z. Zhou, "Machine learning: Accelerating materials development for energy storage and conversion," InfoMat, vol. 2, no. 3, pp. 553–576, 2020.

[7] Y. Liu, T. Zhao, W. Ju, and S. Shi, "Materials discovery and design using machine learning," J. Materiomics, vol. 3, no. 3, pp. 159–177, 2017.

[8] W. Shockley and H. J. Queisser, “Detailed balance limit of efficiency of p-n junction solar cells,” J. Appl. Phys., vol. 32, no. 3, pp. 510–519, 1961.
[9] L. Yu and A. Zunger, “Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials,” Phys. Rev. Lett., vol. 108, Art. no. 068701, 2012.
[10] J. B. Goodenough and K.-S. Park, “The Li-ion rechargeable battery: A perspective,” J. Amer. Chem. Soc., vol. 135, no. 4, pp. 1167–1176, 2013.
[11] T. Xie and J. C. Grossman, “Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties,” Phys. Rev. Lett., vol. 120, Art. no. 145301, 2018.
[12] J. Mueller, A. G. Kusne, and R. Ramprasad, “Machine learning in materials science: Recent progress and emerging applications,” in Reviews in Computational Chemistry, vol. 29, Wiley, 2016, pp. 186–273.
[13] A. Y.-T. Wang et al., “Machine learning for materials scientists: An introductory guide toward best practices,” Chem. Mater., vol. 32, no. 12, pp. 4954–4965, 2020.
[14] Y. Zhang and C. Ling, “A strategy to apply machine learning to small datasets in materials science,” npjComput. Mater., vol. 4, Art. no. 25, 2018.
[15] K. A. Severson et al., “Data-driven prediction of battery cycle life before capacity degradation,” Nature Energy, vol. 4, pp. 383–391, 2019.
[16] Y. Zhang and C. Ling, op. cit. [14]; see also [13] for transfer-learning and small-data guidance.
[17] B. Burger et al., “A mobile robotic chemist,” Nature, vol. 583, pp. 237–241, 2020.
[18] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785–794.
[19] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv:1412.6980, 2014.
[20] C. E. Rasmussen and C. K. I. Williams, Gaussian Processes for Machine Learning. Cambridge, MA, USA: MIT Press, 2006.
[21] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, “Taking the human out of the loop: A review of Bayesian optimization,” Proc. IEEE, vol. 104, no. 1, pp. 148–175, Jan. 2016.
[22] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 4765–4774.
[23] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, “A fast and elitist multiobjective genetic algorithm: NSGA-II,” IEEE Trans. Evol. Comput., vol. 6, no. 2, pp. 182–197, Apr. 2002.
[24] T. L. Saaty, The Analytic Hierarchy Process. New York, NY, USA: McGraw-Hill, 1980.
[25] C.-L. Hwang and K. Yoon, Multiple Attribute Decision Making: Methods and Applications. Berlin, Germany: Springer-Verlag, 1981.