International Journal of Electrical and Computer Engineering (IJECE)
Vol. 16, No. 3, June 2026, pp. 1508~1517
ISSN: 2088-8708, DOI: 10.11591/ijece.v16i3.pp1508-1517

Tuning feature selection to enhance machine learning predictions of bandgap and efficiency in chalcogenide perovskites

Osphanie Mentari Primadianti[1], Ryan Nur Iman[1], Muhammad Zimamul Adli[1], Agung Muhamad Toha[1], Agung Surya Wibowo[2]
[1]Department of Electrical Engineering, Faculty of Engineering, University Islam Nusantara, Bandung, Indonesia
[2]Department of Electrical Engineering, Faculty of Engineering, Telkom University, Bandung, Indonesia

Article Info

Article history:

Received Oct 23, 2025
Revised Mar 18, 2026
Accepted Mar 26, 2026

Keywords:

Band gap
CatBoost Regressor
Feature fusion
Feature selection
Machine learning

ABSTRACT

Solar cell technology has advanced rapidly in efficiency and material innovation. As a renewable energy source, solar cells help mitigate the global energy crisis. Perovskite-based solar cells have recently achieved efficiencies above 25%, surpassing conventional silicon cells. Among emerging materials, chalcogenide perovskites show great promise due to their superior stability compared to halide perovskites. However, they remain in the exploration stage, making accurate predictions of their electrical properties, especially bandgap, essential for assessing potential in solar cell applications. This study predicts bandgap values using computational methods, emphasizing efficiency and cost reduction compared to experimental approaches. Key features derived from collected data include oxidation state, electronegativity, coordination number, ionic radius, and density. Several machine learning (ML) algorithms: AdaBoost Regressor, gradient boosting regressor, support vector regressor, CatBoost Regressor, and k-neighbor regressor, were implemented using Python. The research process involved data collection, preprocessing (feature scaling, fusion, reduction, and selection), model training and testing with 5-fold cross-validation, and hyperparameter optimization to achieve optimal results. Among the tested models, CatBoost Regressor yielded the best performance, achieving a coefficient of determination (R^2) of 69.34%, a mean absolute error (MAE) of 23.1%, and root-mean-square error (RMSE) of 29.49%, demonstrating its effectiveness in predicting chalcogenide perovskite bandgaps.

Corresponding Author:

Agung Surya Wibowo
Department of Electrical Engineering, Faculty of Engineering, Telkom University
Jl. Telekomunikasi Terusan Buahbatu No.1, Sukapura, Kec. Dayeuhkolot, Kab. Bandung, Jawa Barat, Indonesia
Email: agungsw@telkomuniversity.ac.id

1. INTRODUCTION

The rising global demand for sustainable energy has accelerated the search for advanced materials for next-generation photovoltaic (PV) technology. Solar energy is widely recognized as a clean and environmentally friendly source, driving efforts to efficiently convert sunlight into usable energy [1]. In this context, perovskite materials have emerged as promising candidates due to their structural flexibility, reasonable efficiency, and potential for enhanced stability and environmental compatibility.

The perovskite family, with the general formula ABX_3 [1], has a flexible crystal structure that allows substitutions at multiple sites, leading to diverse electronic properties. While halide perovskites offer high photovoltaic performance, they suffer from instability and toxicity. In contrast, chalcogenide perovskites (CPs) provide better stability, non-toxicity, and tunable bandgaps. Their strong light absorption and structural integrity make them promise for solar cells and other applications, supporting the development of next-generation sustainable PV technologies.

Perovskite structures are widely used beyond solar cells, including in optoelectronics, magnetoresistance, superconductivity, dielectric, and piezoelectric devices, due to their flexible structure and physical properties. Chalcogenide perovskites are stable, lead-free materials with strong optoelectronic performance [2]. Additionally, studies on $LaSrMnO_3$ and $LaBaMnO_3$ show their effectiveness in magnetoresistance applications under varying temperatures and magnetic fields [3].

Studies show that perovskite structures exhibit diverse advanced properties. Metallic compounds such as Ba-La-Cu-O demonstrate superconductivity behavior [4]. Research on $BaTiO_3$ highlights its application as a dielectric material, improving energy storage efficiency through nanostructure engineering [5]. Additionally, perovskites like $BaTiO_3$ and $BiFeO_3$ are widely used in piezoelectric devices, enabling efficient conversion between mechanical and electrical energy [6].

Chalcogenide perovskites have emerged as promising alternatives due to their superior thermal stability and tunable electrical properties, enhancing energy conversion efficiency. They are considered stable and effective materials for solar cell applications, though their development remains limited [7]. However, studies on their electrical and optical properties, particularly bandgap, are still scarce, and both experimental methods and density functional theory (DFT) approaches face challenges in cost, time, and accuracy [8].

The discovery of new chalcogenide perovskites is costly and time-consuming, making machine learning (ML) a valuable alternative. Previous studies applied ML for bandgap prediction, such as using random forest for $BaZrS_3$ without comparing multiple models [9], and combining DFT-based descriptors with ML methods without feature selection [10]. Other work used graph neural networks but also lacked feature selection despite achieving good accuracy [11]. Additionally, studies on different perovskite materials highlight the importance of feature engineering, such as oxidation state, electronegativity, coordination number, and ionic radii, to improve prediction performance [12].

This study focuses on predicting bandgaps and efficiency of chalcogenide perovskites using ML with key features such as ionic radius, electronegativity, oxidation state, and coordination number. By applying feature fusion and selection, ML enables fast and accurate prediction of bandgap energy, a critical factor in solar cell performance. Compared to conventional methods, this approach is more efficient and cost-effective, supporting the design of sustainable, high-efficiency solar cells.

2. METHOD

This section describes the materials and methods used in the study. The dataset is a balanced collection compiled from published peer-reviewed articles [13]–[20] and verified databases. The methodology focuses on predicting bandgap and efficiency using feature fusion and feature selection. Key steps include data preprocessing to reduce dimensionality and identify important features, dataset augmentation to improve model performance, and hyperparameter optimization using OPTUNA. This automated optimization enhances model accuracy and generalization.

2.1. Dataset and features

We compiled a dataset totaling 118 compounds from various sources, including AFlowlib [21], Materials Project [22], PubChem [23], WebElements, Chemglobe, and published journal articles [13]–[20]. We used five key features: a_ions, b_ions, and x_ions, as shown in Table 1. These descriptors (electronegativity, ionic radius, oxidation state, density, and coordination number) link atomic properties to electronic behavior. Their integration in the ML model captures chemical and structural effects, enabling accurate bandgap prediction for optimizing chalcogenide perovskite solar cells.

Table 1. Summary of features and references

Features	Number of features	Reference
Electronegativity	3	WebelEments
Ionic radii	3	Chemglobe.org
Oxidation state	3	[23]
Density	3	Chemglobe.org
Coordination number	3	Chemglobe.org

All selected features influence bandgap and efficiency. Electronegativity differences affect bonding (ionic vs. covalent), ionic radii determine lattice structure, and oxidation states define electronic configurations. Density reflects atomic packing, while coordination number describes the local bonding environment. Together, these features capture key chemical and structural factors for accurate prediction.

2.2. ML processing

Figure 1 presents the ML workflow, starting with data collection and preprocessing. Model performance depends on data quality and quantity, using data from the Materials Project [22] and prior studies [13]–[20]. Data processing includes cleaning to ensure accuracy. Feature fusion integrates multiple features [24], while filter-based feature selection removes irrelevant data, improving efficiency, accuracy, and interpretability [25].

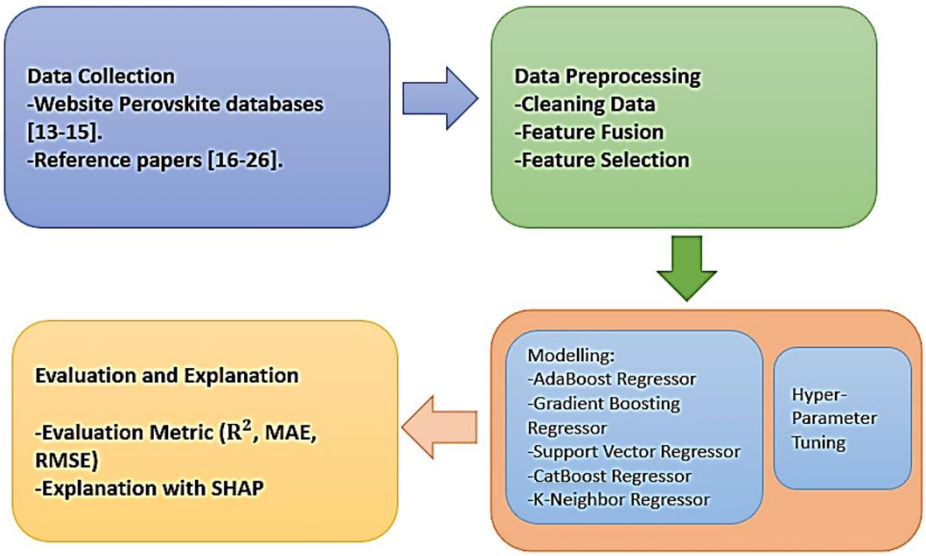


Figure 1. Research methodology workflow

In the modeling stage, multiple ML algorithms: AdaBoost, Gradient Boosting, support vector regression (SVR), CatBoost, and k-nearest neighbor (KNN) are implemented in Python [12]. Hyperparameter tuning is applied to improve performance, using OPTUNA with Bayesian optimization (TPE) [26], [27]. The final stage includes evaluation and interpretation using root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) metrics [28]–[30], while Shapley additive explanation (SHAP) explains the contribution of each feature to the model predictions [31].

In Figure 2 we identified five characteristics: oxidation state (OS), electronegativity (E), coordination number (CN), ionic radius (IR), and density (D), each exhibiting several numerical ranges, including:

a. Oxidation state (OS) ranges from -2 to +6.
b. Electronegativity (E) ranges from 0.79 to 3.44.
c. Coordination number (CN) ranges from 2 to 12
d. Ionic radius (IR) ranges from 1.7E-11 to 2.2E-10.
e. Density (D) ranges from 0.001429 to 18.9.

Each feature group comprises three values or representations, followed by the use of normalization. Prior to amalgamation, each feature underwent normalization via a min/max scaler, which adjusts values to a range between 0 and 1 to eliminate discrepancies in scale. Min-Max Scaler is a normalization technique that scales all signal values to a range between 0 and 1. As in (1) and (2) delineate the Min-Max Scaler normalizing technique [32].

$$X_{std} = \frac{(x - x.min)}{(x.max - xmin)} \quad (1)$$

$$X_{scaled} = X_{xtd} * (X_{max} - X_{min}) + X_{min} \quad (2)$$

The data were normalized using the min-max scaling method presented in (1)–(2). Feature fusion produced 15 features, after which monotonic data were removed. Feature reduction using a variance threshold reduced the set to 14 features, followed by filter-based selection to retain the most relevant ones. Finally, 8 features were used for training and testing. This study applies multiple ML regression models to predict bandgap and efficiency, as described in the following sections.

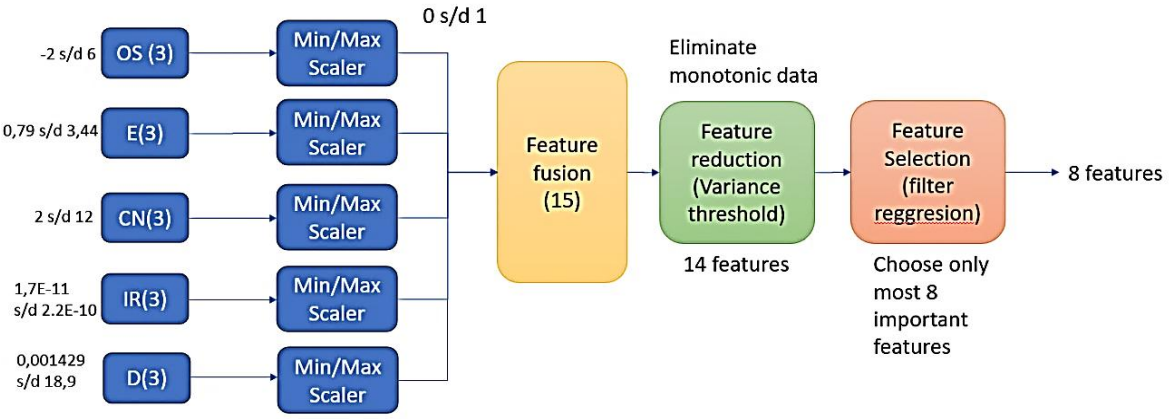


Figure 2. The data processing employed for this research

2.3. CatBoost Regressor

The categorical boosting (CatBoost) regressor represents an ensemble technique derived from gradient boosting [33]. CatBoost is an ensemble machine learning method based on gradient boosting decision trees (GBDT), particularly well-suited for handling heterogeneous and categorical features [34]–[36]. The CatBoost algorithm naturally integrates a method for effectively transforming non-numerical data into numerical formats without requiring parametric adjustments, producing favorable outcomes in a single run [37]. Similar to other gradient boosting methods, CatBoost constructs new trees while addressing the overfitting challenges typical of conventional algorithms. It applies a random permutation strategy to organize the data and then encodes each categorical feature with numerical values [38]. By applying priority factors and weight coefficients, the impact of low-frequency and noisy data is minimized. Equations (3) and (4) were employed for training the dataset.

$$D = (X_j, Y_j) \tag{3}$$

Let j = the number of samples (1, 2, … n), $X_j = j^{th}$ goal value of X $\left(x_j^1, x_j^2, \ldots x_j^i\right)$, and $Y_j = j^{th}$ target value of Y.

$$x_j^i = \frac{\sum_{k=1}^{n} \varphi\,(x_k^i = x_j^i) Y_j + \alpha p}{\sum_{k=1}^{n} \varphi\,(x_k^i = x_j^i) + \alpha} \tag{4}$$

where φ represents the indicator function, α denotes the initial weight, and ρ signifies the starting value.

2.4. AdaBoostRegressor

AdaBoostRegressor [39] is a commonly utilized ML regression technique recognized for its proficiency in reliably predicting target variables. It is part of the boosting algorithm family, which incrementally refines weak models to the data prior to amalgamating them into a more robust model. The boosting technique effectively reduces bias and improves the model's predictive capacity.

2.5. GradientBoostingRegressor

The GradientBoostingRegressor [40] forms an ensemble model through the stepwise integration of predictors, where each one enhances the preceding performance. In contrast to AdaBoost, it applies gradient descent to target residual errors of earlier learners, thereby creating a robust framework for minimizing prediction errors over successive iterations.

2.6. KNeighborsRegressor

KNeighborsRegressor [41] employs similarity-based prediction by locating the k-nearest neighbors of a given data point using a distance measure like Euclidean distance. The predicted value is then derived from the average target values of those neighbors. In our work, we refined the hyperparameters—number of neighbors, leaf size, ρ, and number of tasks—while keeping the others at default, as they contributed little to performance improvement.

2.7. Support vector regression

Support vector regression (SVR) [42] is a ML approach that utilizes a linear function to represent data in a vector space. The SVR model aims to minimize the aggregate of the distances between the actual placements of all samples and this linear function, referred to as the loss function. The SVR algorithm determines the optimal parameters of a linear function by minimizing the associated loss function

3. RESULTS AND DISCUSSION

This section presents the results and discussion of ML models for bandgap prediction in chalcogenide perovskites. Results are shown through figures, graphs, and tables for clear comparison [12]. The discussion highlights the effects of feature selection, feature fusion, and hyperparameter optimization on model performance, including evaluation metrics, feature importance, and algorithm comparisons

3.1. Evaluation metric

Tables 2 and 3 display the results of our simulation concerning bandgap and efficiency. Table 2 illustrates the results of the bandgap simulation, which compares six regression models using 5-fold cross-validation. The best model from the bandgap simulation was the CatBoost Regressor, which achieved the lowest MAE of 0.2310 eV and RMSE of 0.2949 eV, along with the highest R^2 of 0.6934 under 5-fold cross-validation. This indicates CatBoost delivered the most accurate and dependable forecasts among all models. On the other hand, AdaBoost Regressor also did relatively well MAE 0.2708, R^2 0.616, but not as strong as CatBoost. KNeighbors and RandomForest exhibited inferior results with greater mistakes and substantially lower R^2 values, suggesting poor generalization. Support vector regressor with Gradient Boosting the regressor showed reasonable performance, better than RandomForest/KNeighbors but below CatBoost.

Table 2. Results of bandgap simulation

Methods	MAE (eV)	RMSE (eV)	R2
CatboostRegressor	0.231	0.2949	0.6934
AdaboostRegressor	0.2708	0.2949	0.616
KneighborsRegressor	0.3556	0.4119	0.402
RandomForestRegressor	0.3537	0.4428	0.2626
SupportVectorRegressor	0.2867	0.3679	0.5228
GradientBoostingRegressor	0.3059	0.36	0.5432

Table 3 shows the results of the efficiency simulation for forecasting solar cells. The best model from the efficiency simulation was the CatBoost Regressor as the best performer, achieving a minimum MAE of 0.2290 eV, a minimum RMSE of 0.2959 eV, and a maximum R^2 of 0.6914. This shows good prediction accuracy and constant reliability. Other observations were that the AdaBoost Regressor performed second-best with MAE 0.2383 and R^2 0.6624, close to CatBoost. Gradient Boosting Regressor again demonstrated modest performance with an R^2 of 0.5450. KNeighbors, RandomForest, and SVR demonstrated worse accuracy with larger mistakes and low R^2, suggesting they are less appropriate for this task.

Table 3. Outcomes of efficiency simulation

Methods	MAE (eV)	RMSE (eV)	R^2
CatboostRegressor	0.2290	0.2959	0.6914
AdaboostRegressor	0.2383	0.3301	0.6624
KneighborsRegressor	0.3463	0.4034	0.4638
RandomForestRegressor	0.3460	0.4372	0.3597
SupportVectorRegressor	0.3132	0.6329	0.3929
GradientBoostingRegressor	0.3051	0.3600	0.5450

Table 4 compares the methods and findings of Khan *et al.* [12] and our current investigation while employing CatBoost Regressor for bandgap simulation Khan *et al.* [12]. Method Khan *et al.* [12] applied strong predictive accuracy. However, in our investigation, we simulated CatBoost Regressor especially for chalcogenide perovskites, which are the future perovskites, using the Njema *et al*. publication [1]. Then, unlike Khan *et al*. [12], we used feature fusion and feature selection to focus solely on the most important descriptors (electronegativity, ionic radii, oxidation state, density, and coordination number). The findings yielded $R^2 = 0.6934$. Although lower than Khan *et al*. [12], this reflects the increased difficulty of predicting chalcogenide perovskites, which are less researched and may have more complex electronic interactions.

Table 4. Comparison of our research and Khan *et al* [12]

Methods	Differences	Ref.
CatboostRegressor for Bandgap simulation	Perovskites	[12]
CatboostRegressor for Bandgap simulation [our work]	Feature Fusion	This study
	Feature Selection	This study
	Chalcogenide	This study

Figure 3 explains that CatBoost Regressor exhibits the strongest predictive capability—its predictions are closest to the diagonal, agreeing with previous measures (lowest MAE and highest R^2). AdaBoost Regressor is the second-best performance, with decent alignment but somewhat less accuracy. KNeighbors and random forest regressors perform badly, demonstrating significant prediction errors and poor alignment. Support vector and gradient boosting regressors offer moderate predictive power but are less dependable than boosting-based approaches. The red line gives a benchmark: the closer the scatter points lie to this line, the better the model. CatBoost definitely outperforms others.

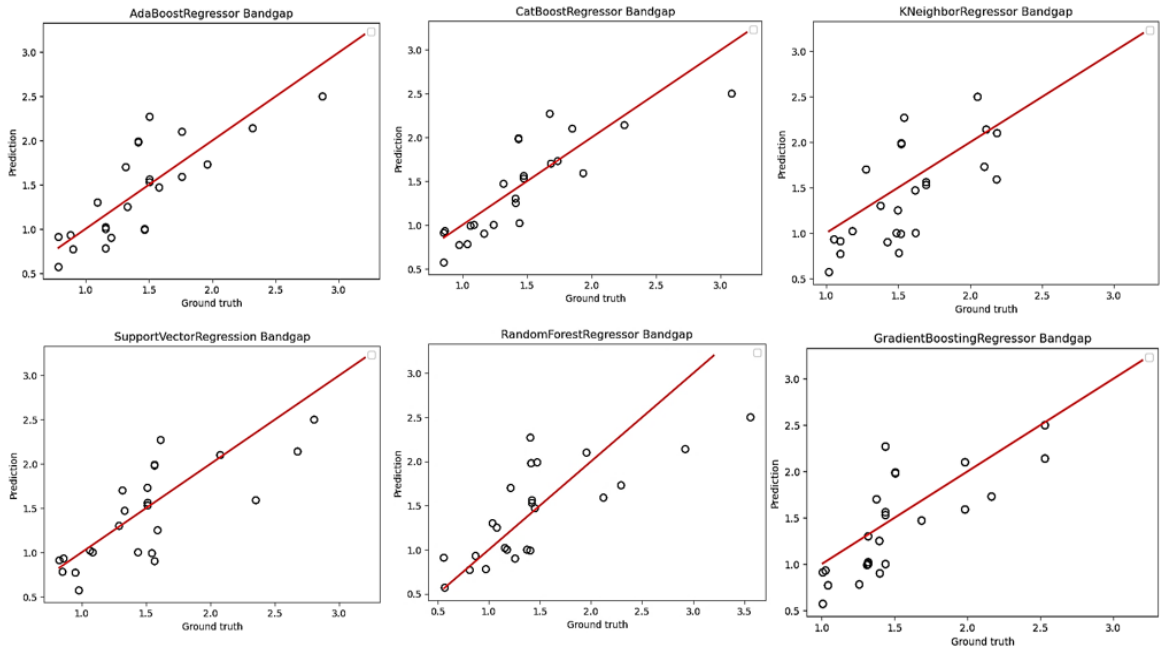
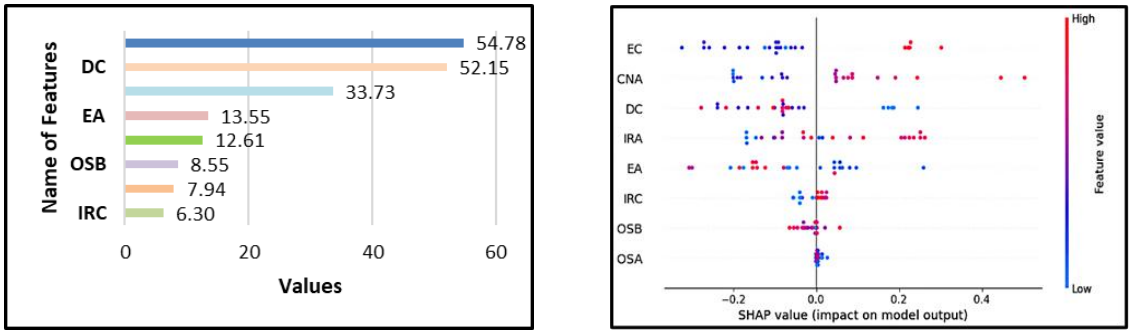


Figure 3. Analysis of the experimental versus predicted bandgap values utilizing our ML models

3.2. Feature importance with filter regression and Shapley additive explanation

This work used feature importance with filter regression, with which we rated our features as shown in Figure 4. Figure 4(a) highlights feature importance. Electronegativity (EC) is the most influential, directly affecting bonding and bandgap. Density (DC) and coordination number (CNA) also play major roles by influencing atomic packing and electronic structure. Moderate features include EA and OSA, which impact local bonding. Lower-impact features—OSB, IRA, and IRC—have smaller contributions, mainly refining the predictions.

Figure 4(b) shows that electronegativity (EC) has the strongest impact on bandgap, where higher values increase predictions. Coordination number (CNA) also significantly raises bandgap values. Density (DC) and ionic radius (IRA) have moderate effects, while oxidation states (OSB, OSA), ionic radius C (IRC), and electronegativity A (EA) contribute less. Overall, electronegativity and coordination number are the most influential factors.



(a) (b)

Figure 4. Visual summary of the feature influence, with (a) utilizing SHAP values for ranking feature importance in our ML and (b) CatBoost Regressor bandgap

3.4. Hyperparameter tuning with OPTUNA

In this work, we used OPTUNAto tune the hyperparameters of various ML models, as shown in Table 5. Table 5 summarizes OPTUNA-based hyperparameter optimization for each model, balancing complexity and generalization. CatBoost and Gradient Boosting use low learning rates with many iterations for higher accuracy. AdaBoost uses fewer estimators to reduce overfitting, KNN uses a large k (31) for smoother predictions, Random Forest balances tree depth and size, and SVR employs a non-linear kernel to capture complex relationships.

Table 5. Hyperparameter tuning settings with OPTUNA

Methods	Hyperparameters
CatboostRegressor	'iterations': 507, 'depth': 6, 'learning_rate': 0.0035589304597655382, 'l2_leaf_reg': 0.09153952172281106, 'random_strength': 0.013416159875383185, 'bagging_temperature': 0.13700779150148432, 'border_count': 51
AdaboostRegressor	'n_estimators': 49, 'learning_rate': 0.09140896209090747, 'random_state': 40
KneighborsRegressor	'n_neighbors': 31, 'leaf_size': 71, 'p': 1, 'n_jobs': 5
RandomForestRegressor	'n_estimators': 163, 'max_depth': 16, 'min_samples_split': 2, 'min_samples_leaf': 2
SupportVectorRegressor	'coef0': 8.355814663808525, 'tol': 0.8122822323306137, 'epsilon': 0.3884557915550403, 'C': 2.427599716333231, 'degree': 8, 'max_iter': 80, 'cache_size': 359
GradientBoostingRegressor	'learning_rate': 0.0026024933851524555, 'alpha': 0.6888940259094042, 'n_estimators': 444, 'min_samples_leaf': 0.03288313515702686, 'min_samples_split': 0.3685733075339944, 'min_weight_fraction_leaf': 0.06567934584629062, 'max_depth': 40, 'min_impurity_decrease': 0.022532313181025488

4. CONCLUSION

This study involved predicting the bandgap and efficiency of chalcogenide perovskite solar cells through the application of various ML models, such as AdaBoost Regressor, CatBoost Regressor, Gradient Boosting Regressor, KNeighbors Regressor, and SVR. To test the success of the ML models, we applied three performance metrics: RMSE, MAE, and R^2. Among all the models, CatBoostRegressor displayed the best performance on bandgap and efficiency forecasts. The CatBoost Regressor achieved the best performance in the bandgap simulation, obtaining the lowest MAE of 0.2310 eV and RMSE of 0.2949 eV, along with the highest R^2 value of 0.6934. For the efficiency simulation, CatBoost gave the most accurate and dependable forecasts among all models. lowest RMSE of 0.2959 eV, and the highest R^2 of 0.6914. Good forecast accuracy and continuous reliability. Both bandgap and efficiency forecasts highlight electronegativity (EC) and coordination number (CNA/CAN) as the most significant aspects. Then, density and ionic radii serve as secondary yet crucial structural descriptors. After that, oxidation status influences predictions moderately, consistent with its involvement in orbital energies. The SHAP values for bandgap

and efficiency give interpretability, indicating that the model's decisions fit with established physical and chemical principles of perovskites.

ACKNOWLEDGMENTS

The authors would like to acknowledge the Department of Electrical Engineering, Islamic Nusantara University, Bandung, and the Indonesian Government under the Ministry of Education and Science and Technology (Kemdiktisaintek) for their support of this work.

FUNDING INFORMATION

This work was funded by the Indonesian Government under the Ministry of Education and Science and Technology (Kemdiktisaintek).

AUTHOR CONTRIBUTIONS STATEMENT

The corresponding author assumes full responsibility for all correspondence associated with this publication.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Osphanie Mentari Primadianti	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Ryan Nur Iman	✓					✓		✓						
Muhammad Zimamul Adli							✓					✓	✓	
Agung Muhamad Toha							✓						✓	
Agung Surya Wibowo		✓	✓	✓		✓		✓		✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data are available from the corresponding author upon reasonable request.

REFERENCES

[1] G. G. Njema and J. K. Kibet, "A review of chalcogenide-based perovskites as the next novel materials: Solar cell and optoelectronic applications, catalysis and future perspectives," *Next Nanotechnology*, vol. 7, p. 100102, 2025, doi: 10.1016/j.nxnano.2024.100102.

[2] M. Kumar, A. Singh, D. Gill, and S. Bhattacharya, "Optoelectronic properties of chalcogenide perovskites by many-body perturbation theory," *Journal of Physical Chemistry Letters*, vol. 12, no. 22, pp. 5301–5307, 2021, doi: 10.1021/acs.jpclett.1c01034.

[3] Z. J. Razi, S. A. Sebt, and A. Khajehnezhad, "Magnetoresistance temperature dependence of LSMO and LBMO perovskite manganites," *Journal of Theoretical and Applied Physics*, vol. 12, no. 4, pp. 243–248, 2018, doi: 10.1007/s40094-018-0310-3.

[4] J. G. Bednorz and K. A. Muller, "Possible high Tc superconductivity in the Ba-La-Cu-O system," *Zeitschrift for Physik B Condensed Matter*, vol. 64, no. 2, pp. 189–193, Jun. 1986, doi: 10.1007/BF01303701.

[5] X. Jiang *et al.*, "Superior energy storage BaTiO3-based amorphous dielectric film with polymorphic hexagonal and cubic nanostructures," *Chemical Engineering Journal*, vol. 431, p. 133447, 2022, doi: 10.1016/j.cej.2021.133447.

[6] J. Wu, "Perovskite lead-free piezoelectric ceramics," *Journal of Applied Physics*, vol. 127, no. 19, May 2020, doi: 10.1063/5.0006261.

[7] M. Suhail, H. Abbas, M. B. Khan, and Z. H. Khan, "Chalcogenide perovskites for photovoltaic applications: a review," *Journal of Nanoparticle Research*, vol. 24, no. 7, p. 142, 2022, doi: 10.1007/s11051-022-05525-0.

[8] B. Kaduk, T. Kowalczyk, and T. Van Voorhis, "Constrained density functional theory," *Chemical Reviews*, vol. 112, no. 1, pp. 321–370, 2012, doi: 10.1021/cr200148b.

[9] S. Sharma *et al.*, "Machine learning-aided band gap engineering of BaZrS3 Chalcogenide Perovskite," *ACS Applied Materials and Interfaces*, vol. 15, no. 15, pp. 18962–18972, 2023, doi: 10.1021/acsami.3c00618.

[10] J. Yang and A. Mannodi-Kanakkithodi, "High-throughput computations and machine learning for halide perovskite discovery,"

MRS Bulletin, vol. 47, no. 9, pp. 940–948, 2022, doi: 10.1557/s43577-022-00414-2.
[11] P. Omprakash, B. Manikandan, A. Sandeep, R. Shrivastava, P. Viswesh, and D. B. Panemangalore, "Graph representational learning for bandgap prediction in varied perovskite crystals," *Computational Materials Science*, vol. 196, p. 110530, 2021, doi: 10.1016/j.commatsci.2021.110530.
[12] A. Khan, J. Kandel, H. Tayara, and K. T. Chong, "Predicting the bandgap and efficiency of perovskite solar cells using machine learning methods," *Molecular Informatics*, vol. 43, no. 2, p. e202300217, 2024, doi: 10.1002/minf.202300217.
[13] S. O'Brien, L. Brus, and C. B. Murray, "Synthesis of monodisperse nanoparticles of barium titanate: Toward a generalized strategy of oxide nanoparticle synthesis," *Journal of the American Chemical Society*, vol. 123, no. 48, pp. 12085–12086, 2001, doi: 10.1021/ja011414a.
[14] M. Buffiere, D. S. Dhawale, and F. El-Mellouhi, "Chalcogenide materials and derivatives for photovoltaic applications," *Energy Technology*, vol. 7, no. 11, p. 1900819, 2019, doi: 10.1002/ente.201900819.
[15] N. Sata, M. Ishigame, and S. Shin, "Optical absorption spectra of acceptor-doped SrZrO3 and SrTiO3 perovskite-type proton conductors," *Solid State Ionics*, vol. 86–88, no. PART 1, pp. 629–632, 1996, doi: 10.1016/0167-2738(96)00226-3.
[16] Y. Gan, N. Miao, P. Lan, J. Zhou, S. R. Elliott, and Z. Sun, "Robust design of high-performance optoelectronic chalcogenide crystals from high-throughput computation," *Journal of the American Chemical Society*, vol. 144, no. 13, pp. 5878–5886, 2022, doi: 10.1021/jacs.1c12620.
[17] Y. Zhang, T. Shimada, T. Kitamura, and J. Wang, "Ferroelectricity in Ruddlesden–Popper chalcogenide perovskites for photovoltaic application: the role of tolerance factor," *The Journal of Physical Chemistry Letters*, vol. 8, no. 23, pp. 5834–5839, Dec. 2017, doi: 10.1021/acs.jpclett.7b02591.
[18] N. A. Moroz *et al.*, "Insights on the synthesis, crystal and electronic structures, and optical and thermoelectric properties of $Sr_{1-x}Sb_xHfSe_3$ orthorhombic perovskite," *Inorganic Chemistry*, vol. 57, no. 12, pp. 7402–7411, Jun. 2018, doi: 10.1021/acs.inorgchem.8b01038.
[19] K. Kuhar *et al.*, "Sulfide perovskites for solar energy conversion applications: computational screening and synthesis of the selected compound LaYS 3," *Energy & Environmental Science*, vol. 10, no. 12, pp. 2579–2593, 2017, doi: 10.1039/C7EE02702H.
[20] Y. Nishigaki *et al.*, "Extraordinary strong band-edge absorption in distorted chalcogenide perovskites," *Solar RRL*, vol. 4, no. 5, p. 1900555, 2020, doi: 10.1002/solr.201900555.
[21] S. Curtarolo *et al.*, "AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations," *Computational Materials Science*, vol. 58, pp. 227–235, 2012, doi: 10.1016/j.commatsci.2012.02.002.
[22] A. Jain *et al.*, "Commentary: The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, 2013, doi: 10.1063/1.4812323.
[23] S. Kim *et al.*, "PubChem 2019 update: Improved access to chemical data," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1102--D1109, 2019, doi: 10.1093/nar/gky1033.
[24] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, 2021, pp. 3559–3568, doi: 10.1109/WACV48630.2021.00360.
[25] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018, doi: 10.1016/j.neucom.2017.11.077.
[26] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," *Ecological Modelling*, vol. 406, pp. 109–120, Aug. 2019, doi: 10.1016/j.ecolmodel.2019.06.002.
[27] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Jul. 2019, pp. 2623–2631, doi: 10.1145/3292500.3330701.
[28] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics (Switzerland)*, vol. 10, no. 5, pp. 1–19, 2021, doi: 10.3390/electronics10050593.
[29] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, pp. 79–82, 2005.
[30] A. Clarke, "Do you know your obligations?," *Journal of the Irish Dental Association*, vol. 53, no. 1, pp. 48–49, 2007.
[31] Y. Nohara, K. Matsumoto, H. Soejima, and N. Nakashima, "Explanation of machine learning models using Shapley additive explanation and application for real data in hospital," *Computer Methods and Programs in Biomedicine*, vol. 214, p. 106584, Feb. 2022, doi: 10.1016/j.cmpb.2021.106584.
[32] S. T. Jaafar and M. Mohammadi, "Epileptic seizure detection using deep learning approach," *UHD Journal of Science and Technology*, vol. 3, no. 2, pp. 41–50, 2019, doi: 10.21928/uhdjst.v3n2y2019.pp41-50.
[33] J. M. Ahn, J. Kim, and J. Kim, "Ensemble machine learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and attention-based CNN-LSTM for harmful algal blooms forecasting," *Toxins*, vol. 15, no. 10, p. 608, 2023.
[34] J. T. Hancock and T. M. Khoshgoftaar, "CatBoost for big data: an interdisciplinary review," *Journal of Big Data*, vol. 7, no. 1, p. 94, 2020, doi: 10.1186/s40537-020-00369-8.
[35] R. S. Ajin, S. Segoni, and R. Fanti, "Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility," *Scientific Reports*, vol. 14, no. 1, p. 24851, Oct. 2024, doi: 10.1038/s41598-024-72663-x.
[36] Y. Zhang, Z. Zhao, and J. Zheng, "CatBoost: A new approach for estimating daily reference crop evapotranspiration in arid and semi-arid regions of Northern China," *Journal of Hydrology*, vol. 588, p. 125087, Sep. 2020, doi: 10.1016/j.jhydrol.2020.125087.
[37] N. S. Bhati and M. Khari, "A new intrusion detection scheme using CatBoost classifier," in *Forthcoming Networks and Sustainability in the IoT Era, First EAI International Conference, FoNeS – IoT 2020, Virtual Event*, 2020, pp. 169–176, doi: 10.1007/978-3-030-69431-9_13.
[38] N. Prasanna Venkatesh, R. Pradeep Kumar, B. Chakravarthy Neelapu, K. Pal, and J. Sivaraman, "CatBoost-based improved detection of P-wave changes in sinus rhythm and tachycardia conditions: a lead selection study," *Physical and Engineering Sciences in Medicine*, vol. 46, no. 2, pp. 925–944, Jun. 2023, doi: 10.1007/s13246-023-01274-z.
[39] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: a boosting algorithm for regression problems," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, 2004, vol. 2, pp. 1163–1168, doi: 10.1109/IJCNN.2004.1380102.
[40] T. Duan *et al.*, "NGBoost: natural gradient boosting for probabilistic prediction," in *Proceedings of the 37th International Conference on Machine Learning, PMLR*, 2020, pp. 2690–2700.
[41] O. Kramer, "K-nearest neighbors," in *Dimensionality Reduction with Unsupervised Nearest Neighbors*, Berlin, Germany: Springer, 2013, pp. 13–23.
[42] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004, doi: 10.1023/B:STCO.0000035301.49549.88.

BIOGRAPHIES OF AUTHORS

Osphanie Mentari Primadianti received the B.Eng. degree in electrical engineering from Telkom University, Indonesia in 2012 and the master's degrees in electric engineering from University of Indonesia, Indonesia, in 2017. Currently, she is a lecturer at the Department of Electrical Engineering, University Islam Nusantara, Bandung University. Her research interests include renewable energy, artificial intelligence, bioinformatic especially molecular and toxicity. She will continue her Ph.D. soon in electrical engineering in Indonesia. She likes going to nature places, reads some books, listen to music in leisure time. She can be contacted at email: osphanie88@mail.com.

Ryan Nur Iman is currently a Ph.D. student in materials chemistry at Kanazawa University, Japan. His research interests include semiconducting devices, magnetic materials, and carbon-based materials. He is a lecturer in the Department of Electrical Engineering at Universitas Islam Nusantara, Indonesia.

Muhammad Zimamul Adli has a strong research interest in sensors, biosensors, nanomaterials, and nanotechnology. He is currently a Ph.D. student at the Bandung Institute of Technology (ITB), Bandung, Indonesia. In addition to his academic pursuits, he serves as the head of the Electrical Engineering Study Program at UNINUS University, Indonesia. He is also actively involved in organizational activities, including Nahdlatul Ulama in Bandung. He can be contacted at email: zimamuladli@uninus.ac.id.

Agung Muhamad Toha received his master's degree in electrical engineering from the Bandung Institute of Technology (ITB), Indonesia. He is currently a lecturer in the Department of Electrical Engineering at Universitas Islam Nusantara, Indonesia. Outside his academic activities, he enjoys nature, particularly mountain environments. He is a devoted family man, with one son, and enjoys spending quality time with his family. He can be contacted at email: agungmuhamadtoha@gmail.com.

Agung Surya Wibowo received his Ph.D. degree from Jeonbuk National University, South Korea. He is currently a lecturer in the Department of Electrical Engineering at Telkom University, Indonesia. His areas of expertise include control engineering, artificial intelligence (machine learning and deep learning), and bioinformatics. He can be contacted at email: agungsw@telkomuniversity.ac.id.