



OPEN

High-accuracy machine learning approach for predicting J–V characteristics of perovskite solar cells under variable irradiance

Ayşegül Toprak

Perovskite solar cells (PSCs) have attracted significant attention in recent years due to their exceptional power conversion efficiencies and low-cost fabrication potential. However, accurately modeling their J–V characteristics under varying irradiance conditions remains challenging, as conventional experimental methods require considerable time, cost, and experimental effort. In this work, a machine learning-based approach was developed to overcome these limitations by training a Multi-Layer Perceptron (MLP) artificial neural network capable of predicting PSC performance with high precision. The model was trained on a large-scale, simulation-generated dataset covering diverse irradiance levels, using irradiance intensity and voltage as inputs and current as the output. The Levenberg-Marquardt algorithm enabled fast convergence and low prediction error. The proposed Artificial Neural Network (ANN) achieved correlation coefficients above 0.9996 and very low Mean Squared Error (MSE) values across training, validation, and testing. Comparative analysis showed a close match between the predicted J–V curves and the simulation data, confirming the model's reliability. These results indicate that the proposed ANN offers a cost-effective and scalable solution for PSC performance modeling, potentially accelerating the optimization and deployment of next-generation photovoltaic technologies.

Keywords Perovskite solar cells (PSCs), Machine learning, Artificial neural networks (ANN), Prediction

Perovskite solar cells (PSCs) have emerged as a promising photovoltaic technology due to their exceptional power conversion efficiency (PCE), tunable optoelectronic properties, and low fabrication costs¹. Since their initial demonstration with a modest efficiency of 3.8% in 2009, PSCs have experienced a rapid rise in performance, with certified efficiencies exceeding 26% in recent years². This progress has positioned PSCs as a competitive alternative to conventional silicon-based solar technologies. Despite their impressive laboratory-scale performance, the practical optimization and deployment of PSCs remain challenging. One of the main difficulties is that modeling the current-voltage (J–V) characteristics of PSCs under different irradiation conditions is done by experimental methods, which is both costly and time-consuming. Moreover, the performance of PSCs is significantly influenced by environmental factors, particularly irradiance levels that vary throughout the day and across different geographic regions³. This sensitivity to fluctuating irradiance poses a major barrier to real-world deployment, as performance instability can undermine reliability under practical operating conditions⁴. Conventional experimental characterization under such dynamic conditions often exacerbates these limitations, restricting the scalability of PSC evaluation for comprehensive and realistic scenario analysis. Therefore, incorporating machine learning (ML) based methods—such as artificial neural networks that can simulate or predict PSC behavior under varying irradiance—offers a scalable and cost-effective alternative to traditional experimental procedures. In recent years, data-driven methodologies have gained significant momentum, especially in the fields of machine learning (ML), materials science and photovoltaics, enabling analysing, predicting and optimizing device performance.

In the context of perovskite solar cells (PSCs), ML approaches offer the capability to process vast and complex datasets—encompassing compositional variables, fabrication parameters, and operational conditions—to uncover patterns and correlations that are often hidden from conventional analytical models⁵. By leveraging advanced algorithms, such as artificial neural networks (ANNs), support vector machines (SVMs), and ensemble learning methods, researchers can construct predictive models that capture the nonlinear dependencies between material

Department of Electronics and Automation, Selçuk University, Konya, Türkiye, Turkey. email: aytoprak@selcuk.edu.tr

properties and device performance metrics^{6,7}. Recent studies further illustrate these trends, including additive/interface engineering for carrier dynamics improvement⁸, stability analysis in lead-free perovskite devices⁹, unified simulation-ML frameworks for tin-based device optimization¹⁰, irradiance-dependent performance enhancement in tandem structures¹¹, and ML-based optimization strategies under varied operating conditions¹². Supported by comprehensive experimental datasets, large-scale simulations, and high-throughput screening studies, ML algorithms can not only enhance the prediction accuracy of PSC performance but also assist in identifying underlying physical mechanisms and degradation pathways that traditional models may overlook¹³. Furthermore, ML-driven approaches enable accelerated materials discovery, guiding the selection of optimal fabrication conditions, compositional tuning, and device architecture design¹⁴. These capabilities significantly reduce experimental effort, lower research costs, and shorten development timelines, thus providing a pathway toward rapid prototyping and commercialization of next-generation high-efficiency PSCs. Various machine learning (ML) techniques, including artificial neural networks (ANNs), support vector machines (SVMs), random forests (RF), gradient boosting (GB), and other advanced models, have been applied in photovoltaic research^{15–22}. In particular, they have been successfully employed to predict critical photovoltaic parameters such as power conversion efficiency (PCE), open-circuit voltage (Voc), short-circuit current density (Jsc), and fill factor (FF)^{7,23,24}. For example, a multi-layer perceptron (MLP) based ANN model has achieved low root mean square error (RMSE) values in predicting these key indicators. Recent experimental analyses have shown that including irradiance intensity as a direct input to neural network models significantly improves prediction of Jsc and FF across varying sunlight levels, achieving prediction R^2 values above 0.85 under both low and high irradiance regimes²⁵. In another study, ensemble methods such as Random Forest and gradient boosting combined with incident irradiance data were applied to over 6,000 PSC samples, successfully modeling the dependence of PCE and Voc on light intensity, with model accuracy exceeding 90%²⁶. In further research, parts of the optoelectronic model were trained using irradiance-dependent J–V data for planar p–i–n structured PSCs, providing rapid and accurate performance prediction²⁷. Building on this literature, the present study aims to model the J–V characteristics of PSCs under varying irradiance conditions using an ANN approach, where the model predicts current values as a function of irradiance and compares them with simulation-derived J–V curves, followed by a comprehensive evaluation of performance on training, validation, testing, and overall datasets. Predicting the J–V characteristics under varying irradiance conditions provides an important advantage for evaluating the real performance of solar cells. Since solar irradiance changes continuously throughout the day depending on weather conditions, measurements performed under constant illumination cannot fully represent real operating situations. The developed model overcomes this limitation by making accurate predictions at different light levels, allowing a better understanding of the device behavior under realistic conditions. This approach is expected to provide a faster, scalable, and cost-effective alternative to conventional experimental characterization techniques in PSC research.

Materials and methods

This study combines computational simulations with machine learning techniques to develop a predictive model for perovskite solar cell performance. The flowchart for the dataset characteristics, ANN architecture, training process, and validation approach is shown in Fig. 1.

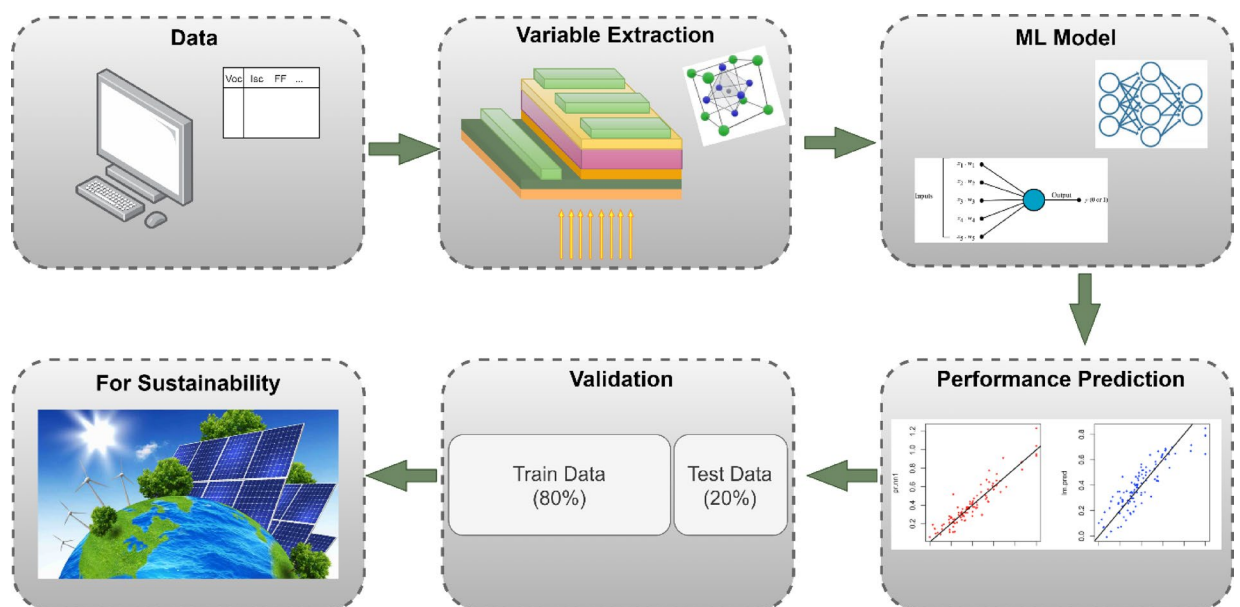


Fig. 1. Flowchart of performance prediction of Perovskite Solar Cells.

Overview of the dataset and collection process

The dataset used in this study was generated through large-scale drift-diffusion (DD) simulations to analyze the dominant recombination processes in perovskite solar cells. A comprehensive set of simulations was carried out, encompassing a wide range of material properties and device parameters, including layer thicknesses, charge carrier mobilities, trap densities, and doping levels. These simulations were designed to replicate realistic PSC behaviors under varying light intensities, providing detailed performance metrics like open-circuit voltage (Voc), short-circuit current (Jsc), fill factor (FF), and ideality factor (n). The dataset was carefully balanced to include scenarios where recombination was dominated by band-to-band, bulk (grain boundary), or interfacial traps. Parasitic effects such as series and shunt resistances were incorporated to more accurately reproduce experimental device performance. The full dataset consists of more than two million simulated data and provides a robust representation of perovskite solar cell behavior. The simulations were conducted using the open-source tool SIMSalabim, and the resulting data was used to train machine learning algorithms, enabling rapid and accurate classification of recombination mechanisms in real-world PSC devices²⁸. Prior to model training, all input and output variables were normalized and downsampled to ensure computational efficiency and stable convergence of the learning process. To achieve reliable performance, the dataset was divided into 80% for training (8000 samples) and 20% for testing (2000 samples). Schematic diagram of perovskite solar cell is illustrated in Fig. 2. This figure shows the layer structure of a perovskite solar cell. The transparent conductive oxide (TCO) permits light entry and ensures conductivity. The electron transport material (ETM) directs electrons to the electrode while blocking holes. The perovskite layer absorbs light and generates charge carriers. The hole transport material (HTM) moves holes to the top electrode while preventing electron leakage. The electrodes collect and deliver charges to the external circuit^{29,30}.

The dataset used in this study is publicly available on Kaggle³¹ and has not been previously applied for current prediction in perovskite solar cells under different irradiance levels using machine learning. In this work, this dataset is employed to develop an Artificial Neural Network (ANN) model capable of predicting the current of PSCs under varying illumination intensities. By leveraging this comprehensive simulation-based dataset, the proposed ANN model provides a fast, reliable, and cost-effective computational tool for optimizing PSC performance without the need for extensive experimental testing.

ANN-Based prediction model

Artificial neural networks (ANNs) have proven to be powerful tools for modeling complex, nonlinear, and multivariate systems, making them particularly suitable for photovoltaic performance prediction³². Unlike traditional empirical or analytical models, ANNs can learn intricate relationships between inputs and outputs directly from data, without the need for explicit physical equations. They have the capacity to adapt to changing environmental conditions and provide accurate predictions at low computational cost. These advantages have led to their increasing adoption in perovskite solar cell (PSC) research, where device behavior is strongly influenced by parameters such as irradiance, temperature, and operating voltage³³. In this study, the ANN model was designed using a Feed-Forward Back Propagation (FF-BP) architecture known as the Multi-Layer Perceptron (MLP). An overview of the model development workflow is presented in Fig. 3 and the operational flow of the MLP model is illustrated in Fig. 4.

The MLP model consists of an input layer, at least one hidden layer, and an output layer. The neurons, which are the basic processing elements within each layer, process input data using weights (w), bias (b), and an activation function (f). The output of a neuron is calculated by the following general mathematical expression³⁴:

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i + b\right) \quad (1)$$

Here, x_i represent the input data, w_i are the weights, and b is the bias term. The tangent sigmoid (tansig) activation function was used in the hidden layer, while a linear (purelin) activation function was applied in the output layer. These functions are defined as hidden layer activation (Eq. (2)) and output layer activation (Eq. (3)):

$$f_{tansig}(z) = \frac{2}{1 + e^{-2z}} - 1 \quad (2)$$

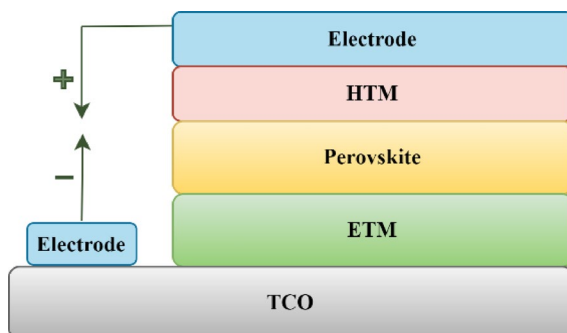


Fig. 2. Schematic diagram of perovskite solar cell.

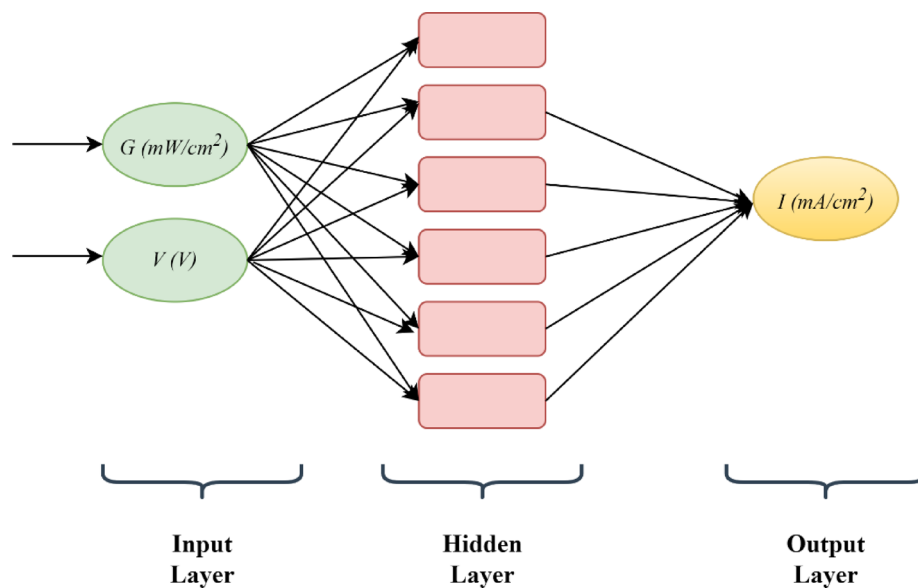


Fig. 3. Fundamental architecture of artificial network model.

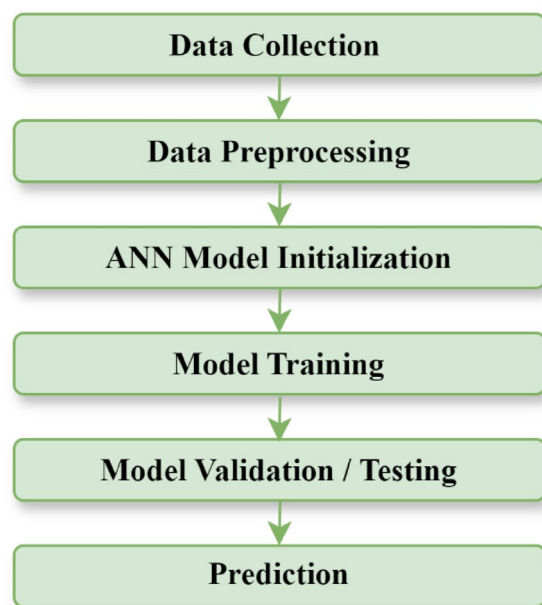


Fig. 4. The process diagram of the MLP model.

$$f_{\text{purelin}}(z) = z \quad (3)$$

For training the network, the Levenberg–Marquardt (LM) algorithm was employed due to its widespread use in the literature and its high performance with small to medium-sized datasets. The LM algorithm offers faster convergence by approximating second-order derivatives of the error function, compared to traditional gradient descent methods³⁵.

To evaluate the prediction performance of the model, three statistical criteria were used: Mean Squared Error (MSE) (Eq. (4)), the correlation coefficient (R) (Eq. (5)), and the Margin of Deviation (MoD) (Eq. (6))^{36,37}:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i^{\text{exp}} - y_i^{\text{ANN}})^2 \quad (4)$$

$$R = \frac{\sum_{i=1}^N (y_i^{exp} - y^{-exp}) (y_i^{ANN} - y^{-ANN})}{\sqrt{\sum_{i=1}^N (y_i^{exp} - y^{-exp})^2} \times \sqrt{\sum_{i=1}^N (y_i^{ANN} - y^{-ANN})^2}} \quad (5)$$

$$MoD = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i^{exp} - y_i^{ANN}}{y_i^{exp}} \right| \times 100 \quad (6)$$

Where y_i^{exp} denotes the actual values, y_i^{ANN} are the ANN predictions, and N is the number of data points. Lower MSE and MoD values, and R values closer to 1 indicate higher model accuracy³⁸.

In this study, an ANN model was developed to predict the J–V characteristics of perovskite solar cells (PSCs) under varying irradiance levels. The input layer of the network receives two variables: irradiance intensity (mW/cm²) and voltage (V), while the output layer yields the predicted current (mA/cm²).

The J–V characteristics obtained under different irradiance conditions were used as the training dataset. To improve convergence speed and numerical stability during training, all input and output variables were normalized. To ensure robust performance, the dataset was split into training data (80%) and test data (20%).

Previous studies have shown that ANN models trained under varying environmental conditions outperform traditional regression-based approaches in terms of accuracy^{33,39,40}. Accordingly, the developed model aims to provide reliable and precise predictions of PSC performance.

Results and discussion

The J–V characteristics obtained from the simulation dataset are presented in Fig. 5. It can be observed that, as the irradiance intensity increases, the solar cell generates a higher current, while the voltage remains largely constant. Despite being derived from simulated data, these characteristics accurately reproduce the physical trends observed in actual J–V curves, demonstrating the strong reliability and realism of the drift–diffusion–based simulations used in this study. The results are consistent with those reported in the literature, confirming that the simulation data effectively capture the real operational behavior of perovskite solar cells under varying irradiance conditions.

The variation of MSE during the training process is shown in Fig. 6. The graph presents the error curves for the training (blue), validation (green), and testing (red) datasets. The training process was conducted over 601 epochs, with the best validation performance achieved at epoch 595, corresponding to an MSE value of 0.00043069. Although the initial MSE values for all datasets were relatively high, they decreased rapidly as the number of epochs increased. A significant reduction in both training and validation errors was observed within the first 100 epochs, and after approximately 200 epochs, the error values stabilized and reached their minimum levels. The error of the test dataset closely matched that of the validation dataset, indicating good generalization capability and the absence of overfitting. These results demonstrate that the model's learning process progressed in a stable manner and produced reliable outcomes.

Figure 7 illustrates the training process of the artificial neural network model developed for solar cell data. At the 601st epoch, the gradient value reached 0.0011745, and the learning rate (μ) decreased to 1×10^{-7} . Due to the absence of improvement over six consecutive epochs, training was terminated by early stopping. The consistent decrease in the gradient demonstrates the stability of the optimization process. The very low value of μ indicates model convergence, while the validation patience confirms the model's ability to generalize without overfitting. These results reveal that the model can analyze solar cell data with high accuracy. The low gradient values and early stopping mechanism further support the model's capability to make reliable predictions. This study highlights the effectiveness of artificial neural networks as a tool for modeling photovoltaic energy systems.

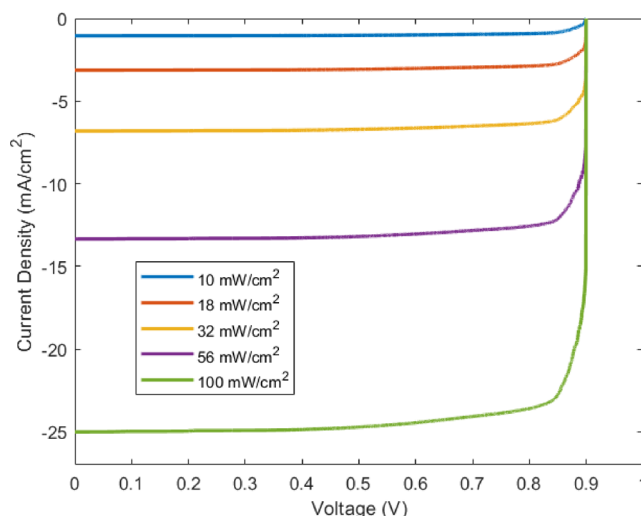


Fig. 5. J–V characteristics of perovskite solar cells.

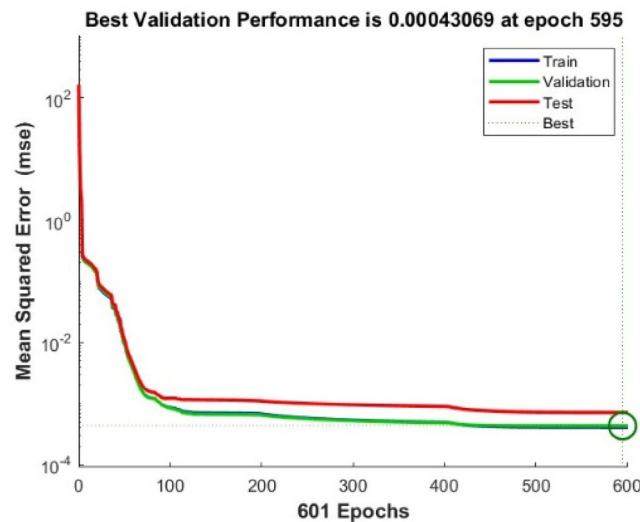


Fig. 6. Training performance of ANN according to epoch.

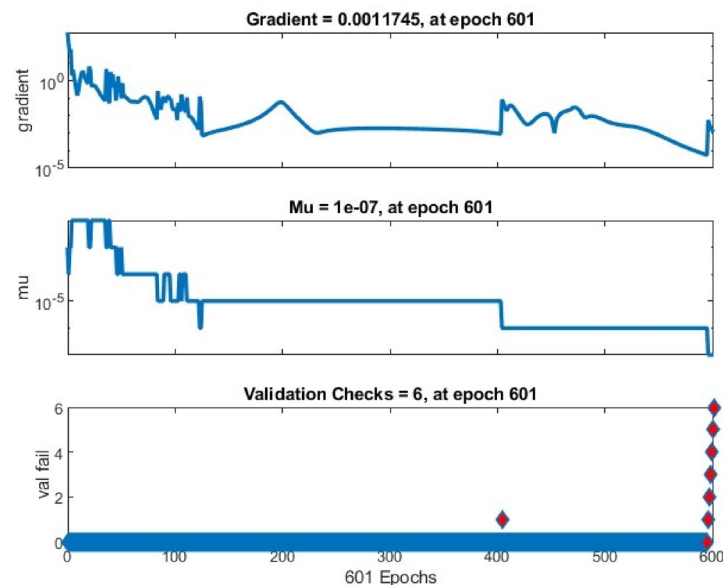


Fig. 7. Training state of the ANN.

Figure 8 presents the error histogram that visualizes the distribution of the differences between the predicted outputs and the actual target values of ANN model across twenty error bins. The horizontal axis represents the prediction errors (Targets-Outputs), while the vertical axis indicates the number of instances within each error range. The majority of the errors are concentrated within the -0.4 to 0.4 range, particularly between -0.00697 and 0.4251 , which indicates that the model performs highly accurate predictions. The clustering of errors around the zero-error reference line (orange) further supports the model's strong predictive accuracy.

In addition, the error distributions for the training (blue), validation (green), and test (red) datasets are similarly centered, indicating that the model maintains consistent and generalizable performance across all data partitions. The very small number of instances with large errors, observed in distant bins (e.g., around -2.0 or above 5.8), demonstrates that the model avoids extreme prediction deviations. Overall, Fig. 8 confirms that the ANN model provides accurate and stable predictive performance.

Figure 9 presents the regression analysis results of the developed ANN model for the training (a), validation (b), testing (c), and overall dataset (d) phases. In Fig. 9 (a), the training phase demonstrates an exceptionally high correlation coefficient ($R = 0.99979$) between the ANN outputs and the actual values. The data points are densely clustered along the ideal equality line ($Y = T$), indicating that the model successfully learned the complex relationship between the input parameters and the output during the training process. Figure 9 (b) illustrates the validation phase, where unseen data during training were used to assess the model's generalization capability.

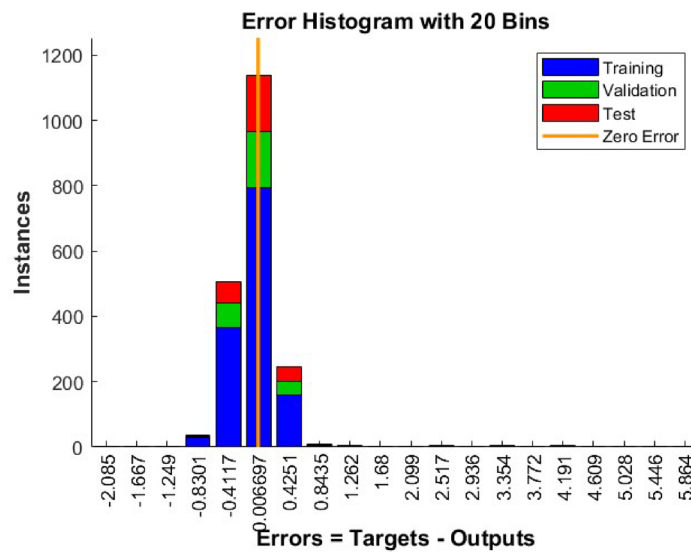


Fig. 8. Error histogram of the ANN model.

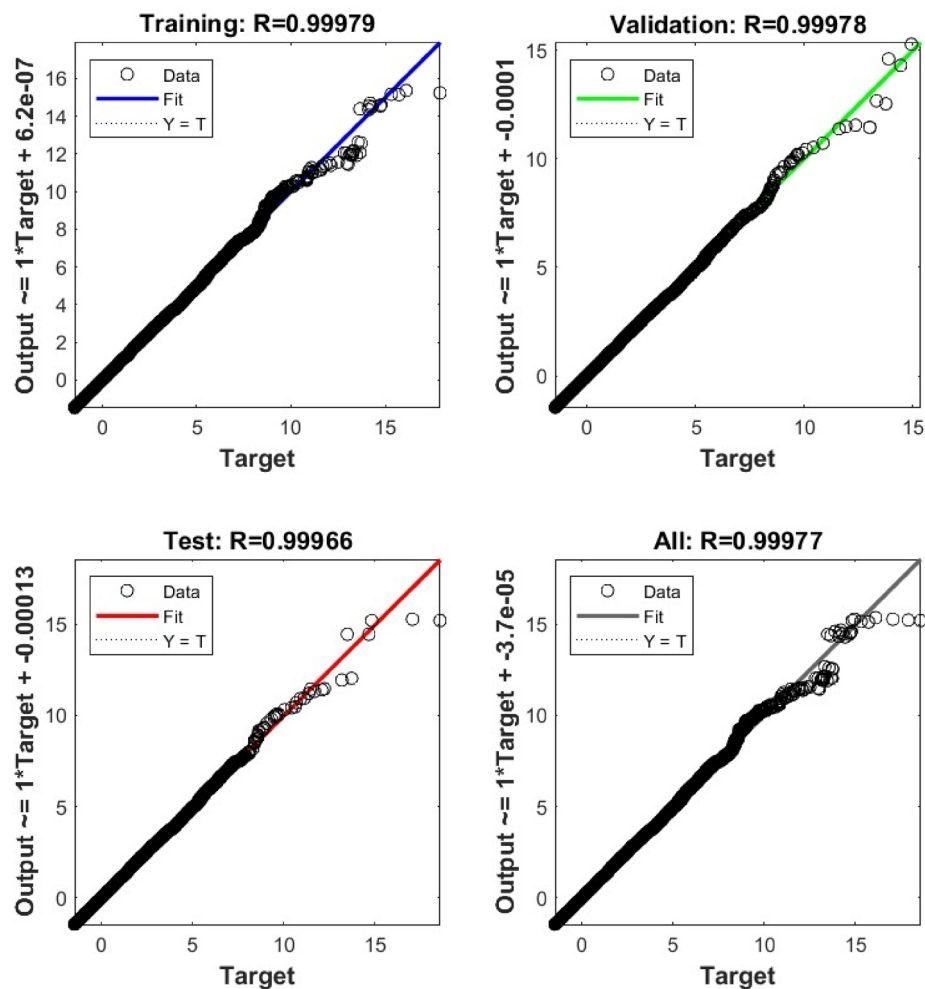


Fig. 9. Performance analysis of the ANN model. (a) Training phase (b) Validation phase (c) Test phase (d) Overall data

Data Set	MSE	MoD (%)	R
Train	1.833E-03	0.909	0.99979
Validation	1.903E-03	-0.429	0.99978
Test	1.861E-03	0.628	0.99966
All Data	1.847E-03	0.678	0.99977

Table 1. The numerical values of the performance parameters of the ANN

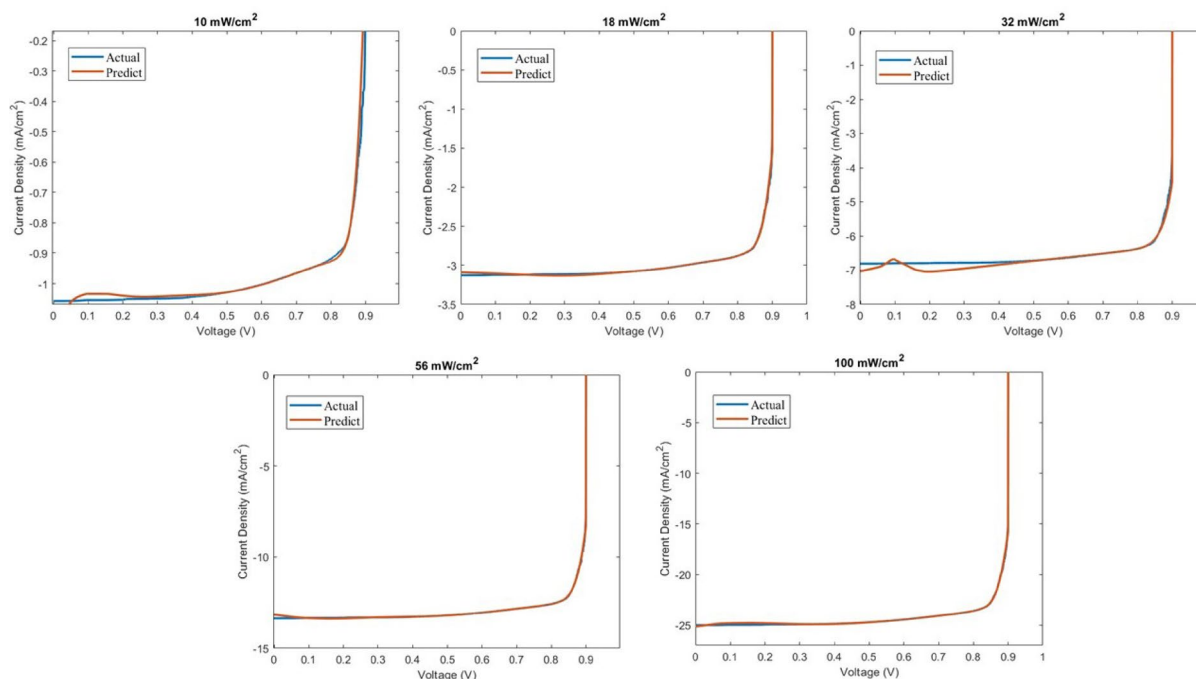


Fig. 10. J–V characteristics at different radiation intensities.

The obtained correlation coefficient ($R = 0.99978$) confirms that the ANN maintains high predictive accuracy beyond the training set, and the data points remain closely aligned with the equality line, indicating minimal overfitting. In Fig. 9 (c), the testing phase results are presented, showing a correlation coefficient of $R = 0.99966$. Similar to the training and validation results, the alignment of the data points with the equality line reflects the model's consistency and reliability when predicting current values for entirely new input conditions.

Finally, Fig. 9 (d) combines all datasets, yielding an overall correlation coefficient of $R = 0.99977$. The tight clustering of all data points along the equality line further validates the ANN model's ability to accurately predict the current values of the perovskite solar cells under varying irradiance conditions across the entire dataset. Collectively, these results confirm that the developed ANN is a highly effective for modeling the current–voltage characteristics of photovoltaic devices. The performance metrics of the trained neural network are presented in Table 1. The MSE values obtained during the training, validation, and testing phases are at notably low levels. This indicates that the model is capable of making highly accurate predictions, not only on the data it was trained on but also on unseen data. The Margin of Deviation (MoD) values further confirm the model's stability. Overall, the MoD values across all datasets remain within $\pm 1\%$, indicating negligible systematic bias and confirming the ANN model's high reliability for practical photovoltaic performance prediction⁴¹.

To further validate the robustness of the proposed ANN model, a 5-fold cross-validation procedure was performed. In each iteration, four groups were used for training and one group for testing. The model achieved low mean squared error (RMSE = 0.0757), low mean absolute error (MAE = 0.0245), and high correlation coefficients ($R^2 = 0.9939$) across all folds, indicating strong generalization capability and stable performance. These results confirm that the network does not rely on a specific data partition and can accurately predict current–voltage characteristics across the full range of irradiance conditions.

Figure 10 shows the comparison between the actual and predicted current–voltage (J–V) characteristics of the perovskite solar cell under different irradiance levels of 10, 18, 32, 56, and 100 mW cm^{-2} . It is clearly observed that the predicted current density values obtained from the machine-learning model are in excellent agreement with the actual data across the entire voltage range. For all illumination intensities, both curves overlap almost completely, indicating that the proposed model can successfully capture the nonlinear J–V behavior of the device.

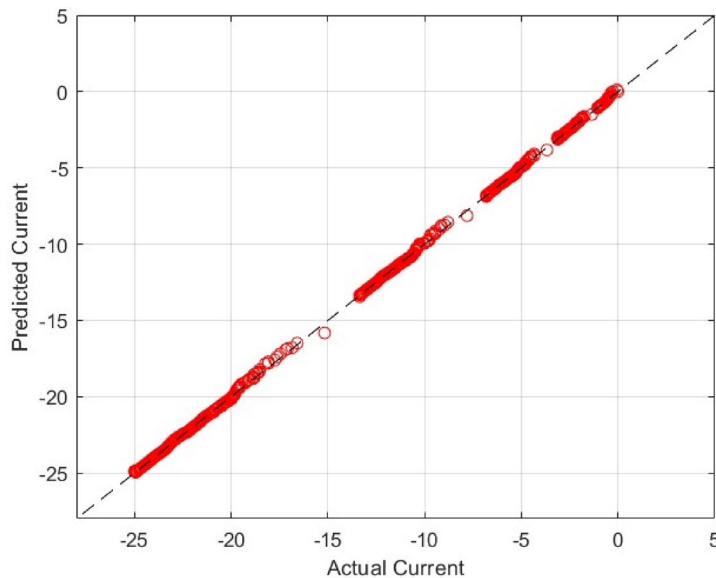


Fig. 11. ANN predicted vs. actual values.

As the irradiance increases, the short-circuit current density (J_{sc}) rises significantly, while the slope near the open-circuit region becomes steeper, which is consistent with the expected photogeneration and recombination dynamics of perovskite devices. The close matching of the model with actual trends at both low and high irradiance conditions demonstrates the model's strong generalization capability and its ability to reproduce the irradiance-dependent charge-transport behavior. This high degree of correlation confirms that the developed model is capable of accurately predicting current values of perovskite solar cells under varying illumination levels, validating its robustness and predictive reliability for photovoltaic performance estimation.

Figure 11 compares the predicted current values obtained from ANN model with the actual values. The majority of the predictions close to the ideal line ($y=x$), indicating that the model performs with high overall accuracy. This strong correlation confirms the high prediction accuracy of the model.

Conclusion

In this study, an Artificial Neural Network (ANN) model based on a Multi-Layer Perceptron (MLP) architecture was developed and validated for the accurate prediction of the J–V characteristics of perovskite solar cells (PSCs) under varying irradiance conditions. Using a comprehensive obtained data dataset, the model was trained with irradiance intensity and voltage as inputs, and current as the output. The Levenberg-Marquardt (LM) training algorithm ensured fast convergence and high prediction accuracy, achieving correlation coefficients (R) exceeding 0.9996 and extremely low Mean Squared Error (MSE) values across training, validation, and testing phases. The low Margin of Deviation (MoD) values, particularly in the validation dataset, indicate minimal prediction bias and confirm the strong generalization capability of the proposed model across unseen data. The ANN predictions exhibited a near-perfect match with actual J–V curves at multiple irradiance levels, demonstrating the model's ability to generalize complex nonlinear relationships and maintain robust performance across a wide range of operating conditions. The proposed ANN model effectively captures the behavior of perovskite solar cells under continuously changing irradiance conditions, which vary throughout the day depending on sunlight intensity and weather. This capability provides a more realistic assessment of device performance and reduces the need for extensive experimental testing, making the analysis process faster, more economical, and reliable.

The proposed ANN framework addresses a critical challenge in PSC research by providing a scalable, computationally efficient, and cost-effective alternative to experimentally demanding and time-consuming characterization. By eliminating the need for repetitive measurements under diverse illumination scenarios, the model accelerates performance assessment and enables rapid optimization of device parameters. This approach not only aligns with current trends in data-driven materials science but also integrates seamlessly with predictive modeling pipelines for renewable energy systems.

From a broader perspective, the successful application of ANN modeling in this work confirms the transformative potential of machine learning in photovoltaics. The methodology presented here can be readily extended to other photovoltaic technologies and adapted for real-time monitoring, fault detection, and adaptive control in large-scale solar energy systems. Furthermore, expanding the model to predict additional parameters such as temperature coefficients, degradation rates, and spectral response could broaden its applicability in PV system design. In the future, these results could play a pivotal role in guiding the design of next-generation high-efficiency PSCs, optimizing operational strategies in dynamic environments, and accelerating the transition towards a more sustainable energy infrastructure.

Data availability

The datasets analysed during the current study are available in the Kaggle repository, <https://www.kaggle.com/datasets/shashwatwork/dataset-of-drp-for-perovskite-solar-cells>.

Received: 10 August 2025; Accepted: 17 October 2025

Published online: 21 November 2025

References

- Roy, P., Ghosh, A., Barclay, F., Khare, A. & Cuce, E. Perovskite solar cells: A review of the recent advances, *Coatings*, vol. 12, no. 8, p. 1089, (2022).
- Khatoun, S. et al. Perovskite solar cell's efficiency, stability and scalability: A review. *Mater. Sci. Energy Technol.* **6**, 437–459 (2023).
- Kumar, M., Namrata, K., Kumar, N. & Saini, G. Solar irradiance prediction using an optimized data driven machine learning models. *J. Grid Comput.* **21** (2), 28 (2023).
- Mao, L. & Xiang, C. A comprehensive review of machine learning applications in perovskite solar cells: materials Discovery, device Performance, process optimization and systems integration. *Mater. Today Energy* **47**, 101742 (2024).
- Lemm, D., von Rudorff, G. F. & Von Lilienfeld, O. A. Improved decision making with similarity based machine learning: applications in chemistry. *Mach. Learning: Sci. Technol.* **4** (4), 045043 (2023).
- Rumman, A. H. et al. Data-driven design for enhanced efficiency of Sn-based perovskite solar cells using machine learning. *APL Mach. Learn.* **1**, 4 (2023).
- Tao, Q., Xu, P., Li, M. & Lu, W. Machine learning for perovskite materials design and discovery. *Npj Comput. Mater.* **7** (1), 23 (2021).
- Kumar, U., Subudhi, P. & Punetha, D. Synergistic enhancement of carrier dynamics in Eco-Friendly perovskite solar cells through fluorinated iodide Additive-Induced crystallographic and interface modifications. *Adv. Theory Simulations.* **8** (5), 2401268 (2025).
- Jain, A., Subudhi, P., Khanal, G. M. & Punetha, D. Simulation-Based Analysis of Environmentally Friendly Perovskite Solar Cells: Projecting 21.61% Efficiency with Lead-Free CsSn0.5Ge0.513, *physica status solidi (a)*, vol. 222, no. 6, p. 2400727, (2025).
- Subudhi, P., Sivapatham, S., Narasimhan, R., Kumar, B. & Punetha, D. Enhancing photovoltaic performance in tin-based perovskite solar cells: A unified approach utilizing numerical simulation and machine learning techniques. *J. Power Sources.* **639**, 236639 (2025).
- Yadav, P., Subudhi, P., Dixit, H. & Punetha, D. Numerical optimization of cesium bismuth iodide-based CIGS/Perovskite tandem solar cells for enhanced photovoltaic performance. *Opt. Laser Technol.* **182**, 112072 (2025).
- Kiruthika, V. & Shoba, S. Hybrid ensemble-deep transfer model for early cassava leaf disease classification, *Heliyon*, vol. 10, no. 16, (2024).
- Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. *Mach. Learn. Mol. Mater. Sci. Nature*, **559**, 7715, 547–555, (2018).
- Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5** (1), 83 (2019).
- Hui, Z. et al. Predicting photovoltaic parameters of perovskite solar cells using machine learning. *J. Phys.: Condens. Matter.* **36** (35), 355901 (2024).
- Liu, Y. et al. How machine learning predicts and explains the performance of perovskite solar cells. *Solar RRL.* **6** (6), 2101100 (2022).
- Abbas, S. R., Mir, B. A., Ryu, J. & Lee, S. W. Toward Sustainable Solar Energy: Predicting Recombination Losses in Perovskite Solar Cells with Deep Learning, *Sustainability*, vol. 17, no. 12, p. 5287, (2025).
- Salah, M. M., Ismail, Z. & Abdellatif, S. Selecting an appropriate machine-learning model for perovskite solar cell datasets. *Mater. Renew. Sustainable Energy.* **12** (3), 187–198 (2023).
- Samantaray, N., Singh, A. & Tonk, A. Identifying the best ML model for predicting the bandgap in a perovskite solar cell. *RSC Sustain.* **2** (11), 3520–3524 (2024).
- Graniero, P. et al. The challenge of studying perovskite solar cells' stability with machine learning. *Front. Energy Res.* **11**, 1118654 (2023).
- Liu, Z. et al. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing, *Joule*, vol. 6, no. 4, pp. 834–849, (2022).
- Ünver, E., Toprak, A., Hussaini, A. A., Kuş, M. & Yıldırım, M. Investigation and calculation of electrical performance of lead-free AgBiI4 perovskite based Schottky photodiode using machine learning. *J. Mater. Sci.: Mater. Electron.* **36** (11), 671 (2025).
- Afre, R. A. & Pugliese, D. Perovskite solar cells: a review of the latest advances in materials, fabrication techniques, and stability enhancement strategies, *Micromachines*, vol. 15, no. 2, p. 192, (2024).
- Paul, A. et al. From 3.8% to over 23.8% power conversion efficiency: commercial perovskite solar cells, significant manufacturing techniques, and future prospects. *Encyclopedia Materials: Electron.* **1**, V1–418 (2023).
- Sadhu, D., Dattatreya, D., Deo, A., Tarafder, K. & De, D. Performance prediction and analysis of perovskite solar cells using machine learning. *J. Alloys Compd. Commun.* **3**, 100022 (2024).
- Hussain, W., Sawar, S. & Sultan, M. Leveraging machine learning to consolidate the diversity in experimental results of perovskite solar cells. *RSC Adv.* **13** (32), 22529–22537 (2023).
- Zhao, X. et al. Accelerating device characterization in perovskite solar cells via neural network approach. *Appl. Energy.* **392**, 125922 (2025).
- Le Corre, V. M., Sherkar, T. S., Koopmans, M. & Koster, L. J. A. Identification of the dominant recombination process for perovskite solar cells based on machine learning. *Cell Rep. Phys. Sci.* **2**, 2 (2021).
- Green, M. A., Ho-Baillie, A. & Snaith, H. J. The emergence of perovskite solar cells. *Nat. Photonics.* **8** (7), 506–514 (2014).
- Mohammad, A. & Mahjabeen, F. Promises and challenges of perovskite solar cells: a comprehensive review. *BULLET: Jurnal Multidisiplin Ilmu.* **2** (5), 1147–1157 (2023).
- <https://www.kaggle.com/datasets/shashwatwork/dataset-of-drp-for-perovskite-solar-cells>
- Wang, S., Zhang, Y., Zhang, C. & Yang, M. Improved artificial neural network method for predicting photovoltaic output performance. *Global Energy Interconnect.* **3** (6), 553–561 (2020).
- Lee, J. & Kim, Y. Comparative estimation of electrical characteristics of a photovoltaic module using regression and artificial neural network models, *Electronics*, vol. 11, no. 24, p. 4228, (2022).
- Karabacak, K. & Cetin, N. Artificial neural networks for controlling wind-PV power systems: A review. *Renew. Sustain. Energy Rev.* **29**, 804–827 (2014).
- Hagan, M. T. & Menhaj, M. B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Networks.* **5** (6), 989–993 (1994).
- Chai, T. & Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **7** (3), 1247–1250 (2014).
- Güzel, T. & Çolak, A. B. Artificial intelligence approach on predicting current values of polymer interface Schottky diode based on temperature and voltage: an experimental study. *Superlattices Microstruct.* **153**, 106864 (2021).

38. Willmott, C. J. & Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Res.* **30** (1), 79–82 (2005).
39. Mellit, A. & Pavan, A. M. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy, *Solar energy*, vol. 84, no. 5, pp. 807–821, (2010).
40. Voyant, C. et al. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy*. **105**, 569–582 (2017).
41. Mellit, A. & Kalogirou, S. A. Artificial intelligence techniques for photovoltaic applications: A review. *Prog. Energy Combust. Sci.* **34** (5), 574–632 (2008).

Author contributions

A.T. wrote the main manuscript text and prepared all figures.

Funding information

The author declares that there is no funding to be acknowledged.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.T.

Reprints and permissions information is available at www.nature.com/reprints.

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

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

© The Author(s) 2025