npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01383-7

Machine learning driven performance for hole transport layer free carbon-based perovskite solar cells

Sreeram Valsalakumar [1] ✉, Shubhranshu Bhandari [1] ✉, Anurag Roy[1], Tapas K. Mallick[1,2], Justin Hinshelwood[3] & Senthilarasu Sundaram[4]

The rapid advancement of machine learning (ML) technology across diverse domains has provided a framework for discovering and rationalising materials and photovoltaic devices. This study introduces a five-step methodology for implementing ML models in fabricating hole transport layer (HTL) free carbon-based PSCs (C-PSC). Our approach leverages various prevalent ML models, and we curated a comprehensive dataset of 700 data points using SCAPS-1D simulation, encompassing variations in the thickness of the electron transport layer (ETL) and perovskite layers, along with bandgap characteristics. Our results indicate that the ANN-based ML model exhibits superior predictive accuracy for C-PSC device parameters, achieving a low root mean square error (RMSE) of 0.028 and a high R-squared value of 0.954. The novelty of this work lies in its systematic use of ML to streamline the optimisation process, reducing the reliance on traditional trial-and-error methods and providing a deeper understanding of the interdependence of key device parameters.

Over the past decade, the perovskite solar cell (PSC) become a promising candidate for the next generation photovoltaics technology[1]. The exponential growth of PSC's power conversion efficiency (PCE) from 3.8% in 2009 to 26.1% in 2024 and the unique characteristics like flexibility, bandgap tenability, long career diffusion lengths and low-cost fabrications attracted solar manufacturers across the world[2,3]. For the commercial silicon solar cells, manufacturing process requires an annealing temperature facility over 1000 °C for the silicon wafer purification. The complex texturing and other coatings need special facilities, which adds to the fabrication cost of silicon cells[4]. Meanwhile, in the PSC fabrication, the solution-processed perovskite materials enable the usage of existing low-maintenance commercial coating technologies[5,6]. Even though the PSC's highest performance was recorded by using lab-scale manufacturing techniques like spin coating, screen printing or blade coating, scalable approaches like slot-die coating or spray coating illustrated a consistent and better performance for the PSCs[6–8].

Although the PSC generally has two architectures, planar (n-i-p) and inverted (p-i-n) structures, both follow comparable energy conversion[9]. In the perovskite device, photon absorption, exciton generation and dissociation occur due to the perovskite material, whereas ETL and HTL layers account for the electron and hole transfer. Finally, the metal electrodes are responsible for collecting charges[10,11]. The PSC performance varies with the optical and electrical properties of the layers regulated by bandgap, carrier mobility, grain size and crystallinity of the materials[12,13].This process is governed by changes in the material's bandgap, which are influenced by factors such as layer thickness, morphology, and porosity. These variables significantly impact the device's swift performance, enhancing both efficiency and stability[14]. Subsequently, it affects the PSC performance, and it marks the importance of the optimisation of PSC layers according to the fabrication techniques[14–16].

However, while aiming for the large-scale developments of PSC, it must overcome various barriers like stability and automated consistent pathways for scalable technologies[17]. Although scalable technologies have been used widely in industries like automotive, batteries and building sectors, PSC manufacturing needs careful control in printing cell layers[17,18]. The coating quality of the deposited films affects the PSC performance due to the structural defects, pinhole formation, and interface conditions[19,20]. While the homogeneity of the absorber or the perovskite layer deteriorates, the defects significantly increase, affecting the electron transport layer (ETL) and hole transport layer (HTL) interaction with the perovskite layer[21,22]. The physics underlying each coating technique is intricate and distinct; typically, the optimisation process relies on a trial-and-error methodology. However, due to its

[1]Solar Energy Research Group, Environment and Sustainability Institute, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, UK. [2]Mechanical and Energy Engineering, College of Engineering, Imam Abdulrahman Bin Faisal University, Dammam, 34212, Saudi Arabia. [3]Faculty of Environment Science and Economy, University of Exeter Penryn Campus, Cornwall, TR10 9FE, UK. [4]School of Computing, Engineering and Design Technologies, Teeside University TeesValley, Middlesbrough, TS1 3BX, UK. ✉e-mail: sv353@exeter.ac.uk; shubhranshu0094@gmail.com

time-intensive nature and the resultant increase in material wastage, integrating new scalable technology into PSC fabrication generally requires approximately a year or more to finalise optimisation and attain reproducibility[23,24]. This is mainly due to the different ejecting and curing mechanisms, precursor solution rheology and other physical factors like speed, temperature and humidity[25–27].

Alternately, a data-driven approach through the Machine Learning (ML) approach can significantly reduce the optimisation time and the wastage[28,29]. The ML-driven approach scans the data and predicts the relationship with less human intervention[28,30]. As the prerequisite for ML applications is a consistent and coherent dataset, higher quality or reliable prediction results need a large dataset[31,32]. Concerning the ML approach and the dataset parameters, various ML techniques like linear regression (LR), artificial neural network (ANN) and random forest (RF) can be used[33,34]. Recently, there has been a noticeable increase in the publications related to ML integration for PSC advancements, and this field starts from the new material identification to the PCE predictions[31,35]. Considering the advantages of the ML approach, a trained model only takes a few seconds for the prediction, and a massive dataset provides higher accuracy results than the DFT analysis[36,37]. In 2023, Mahmood et al. exhibited the ML model analysis of the molecular descriptors and fingerprint for selecting suitable green solvents for organic solar cells[38]. Also, the work demonstrated the potential of fast and easy models in Ml as compared to the DFT/thermodynamic approach[38]. Moreover, the ML analysis can summarise the relationship with various factors difficult to interpret for humans[39]. Liu et al.[40] demonstrated a rapid optimisation technique for the commercialisation of the PSC and the proposed ML framework guided a route for achieving power conversion efficiency (PCE) of 18.5%[40]. Recently Salah et al.[41] provided a comprehensive analysis of various ML model integrations towards the fabrication of PSC and examined ML model's behaviour with the dataset[41]. Perovskite solar cells represent intricate physicochemical systems encompassing perovskite materials, transport layer materials, and electrodes. The predictive modelling of physicochemical properties and the systematic screening of constituent materials pertinent to perovskite solar cells constitute a notable strength of ML. While the ML modelling offers numerous benefits, replicating the complexities of the layer structure and material behaviour in the ambient atmosphere within the ML model is challenging[42].

Considering the commercialisation aspect of the PSC, lowering the production cost is as important as the faster optimisation techniques[43]. While the HTL-free PSC is gaining attention due to its comparatively low production cost, in this work, we focussed on the HTL-free PSC design[43,44]. Since the counter electrode plays a critical role on determining the PSCs performance, we used carbon as the counter electrode for this study. The characteristics such as cost-effectiveness, environmental superiority and abundance make carbon a perfect choice for the PSC[45]. Moreover, the flexibility of adding the HTL characteristics into the carbon layer enhances the PCE and stability of the C-PSC[46]. In 2021, Ye et al. demonstrated a modified carbon-based HTL-free PSC with a PCE of 18.90% and they introduced perfluorotetradecanoic acid with a carbonyl bond to suppress the ion migration and reduce the crystal defects in perovskite[47].

This work introduces a novel approach by developing an ML model that predicts the performance parameters of HTL-free C-PSCs' (carbon-based Perovskite solar cells) performance parameters with a planar (n-i-p) structure. Our proposed ML model predicts the performance parameters of the C-PSC based on the thickness and bandgap for the ETL and perovskite layers, thus easing the optimisation process of these layers. The model also delivers the optimal parameters for the higher-performing HTL-free PSCs. We constructed a database of 700 data points from the SCAPS-1D (Solar Cell Capacitance Simulator) software simulation results[48]. We also analysed the dataset with various ML models to check the trends and performance variation. This work emphasises the effect of the composition of ETL and perovskite layers fabrication parameters on the device performance, enabling an efficient reverse experimental design towards a highly performing C-PSC. This innovative methodology represents a significant advancement in the rational design and fabrication of high-performance PSCs, potentially accelerating their commercialisation.

Methods

In this work, we followed a five-step approach to creating the ANN-based ML model. As shown in the Fig. 1, at first, we identified the C-PSC parameters such as ETL thickness, ETL bandgap, perovskite thickness and perovskite bandgap. Then we set out the range of each parameter needed for the dataset creation. In step 2, we focussed on preparing the dataset through the SCAPS-1D simulation, which has around 790 data points with the

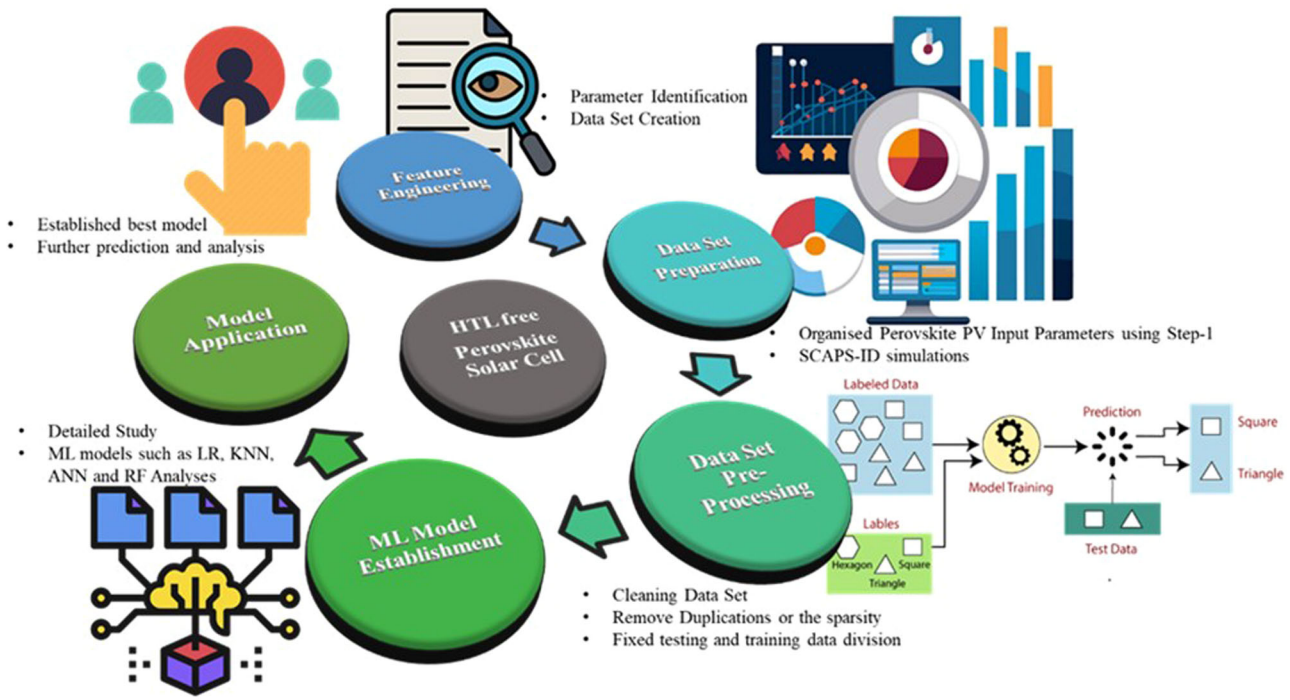


Fig. 1 | Five-step approach for creating an ML model which predicts C-PSC performance.

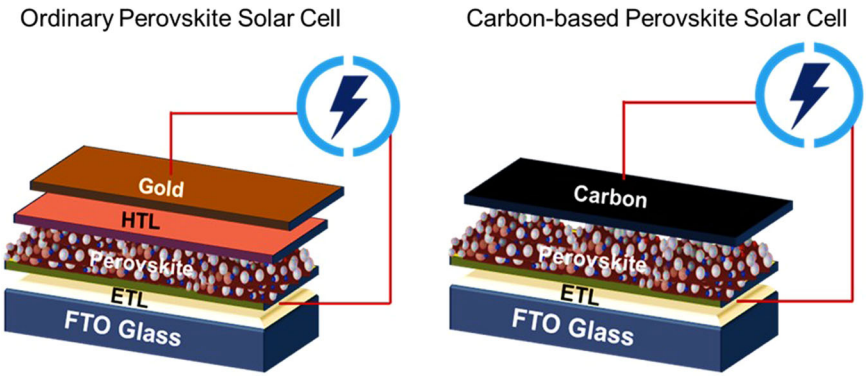


Fig. 2 | Ordinary PSC and HTL Free C-PSC structure, carbon-based structure used for the SCAPS simulation and the dataset generation.

Table 1 | Prior research on machine learning methods in perovskite solar cell studies

No	Dataset acquisition	ML method	Prediction parameter	Ref
1	Previous published papers—333 dataset	ANN	• Perovskite material bandgap • Output performance of the cell	54
2	Previous published papers—regular PSC structure—289 Inverted PSC structure—117	Decision tree model	• Partial stability analysis	65
3	Previous published papers—bandgap database—696 performance database—613	Random forest	• Perovskite materials bandgap • PCE prediction	66
4	Previous published papers—hysteresis database—387 reproducibility database—several small sets	Association rule learning	• Hysteresis analysis for inverted cells • PSC reproducibility analysis	67
5	From experiments: UV−vis spectra database—80 J-V database—90	Random forest	• Bandgap prediction of halide perovskite • Optical properties prediction • Performance evaluation	63
6	From SCAPS simulation: HTL free C-PSC performance dataset—700	ANN	• Performance prediction—Voc, Jsc, PCE and FF • Co-relation between the C-PSC fabrication parameters and performance characterestics.	Current work

device performance characteristics. Subsequently, for step 3, dataset cleaning was performed with data pre-processing techniques and finalised with 700 data points. Finally, in the last steps, the dataset was tested with different ML algorithms and further analysis was carried out using the better-performing ML algorithm.

Device modelling

In this study, we adopted a planar heterojunction HTL free C-PSC structure and used the configuration for the SCAPS-1D simulation. As Fig. 2 depicts, the device consists of a layered configuration of FTO (Fluorine-doped Tin Oxide)/ ETL/ Interface Defect Layer (IDL)/ absorber layer (perovskite – $CH_3NH_3PbI_3$)/ Carbon as an electrode. The model optimisation in the SCAPS-1D software is based on applying a continuity and Poisson's equation for each charge carrier[49]. The physical parameters used for each layer in the SCAPS -1D simulation are defined in Table 1, and all these simulations were under the condition of 300 K room temperature and AM 1.5 G solar spectrum. In this structure, the transparent conductive layer FTO acts as the front contact electrode and the carbon as the back contact. In the simulation analysis, the work function of 5.0 eV is used for the carbon layer[50]. Also, an IDL layer is added to the ETL/Perovskite interface to compensate for the charge recombination due to the variation of the thickness and bandgap[50,51].

Dataset generation

To generate the datasets for the ML modelling, assorted configurations of thickness and bandgap ranges for the perovskite and ETL have been used. The range is taken from the previously reported performance behaviour of the HTL free C-PSC[52,53]. The perovskite thickness and the bandgap range used in this study range from 350 nm to 500 nm and 1.5 eV to 1.7 eV, respectively. Similarly, the ETL's thickness and bandgap are from 155 nm to 165 nm and 3.2 eV to 3.6 eV, respectively. Various arrangements of these range combinations were used for developing PSC samples in the SCAPS software version 3.3.2021. Considering the ML method's accuracy based on the data's quantity and quality, the dataset has to be free from the clones and the sparsity. The dataset includes eight cell characteristics, in which four characters act as input variables, and the other four are the prediction parameters. The input characteristics will include perovskite thickness, perovskite bandgap, ETL thickness, and the ETL bandgap. The output parameters include efficiency, fill factor (FF), short circuit current density (J_{SC}) and open circuit voltage (V_{OC}).

Machine learning methods

We analysed our dataset with various techniques to find the best ML method for the model. The most recognised techniques, such as Linear regression (LR), Random Forest (RF), K-nearest neighbour (KNN) and Artificial Neural Networks (ANN), have been used for the initial analysis. Then, the accuracy parameters like RMSE (root mean square error) and R2 (R-squared) values were generated to compare the techniques. The RMSE value directly indicates the error between the predicted and labelled values, and the smaller RMSE value denotes better prediction results[51]. In contrast, a more considerable R-squared value represents better prediction results[44]. The R-squared value denotes the proportion of variability in the dependent variable to the independent variables in the model[54,55].

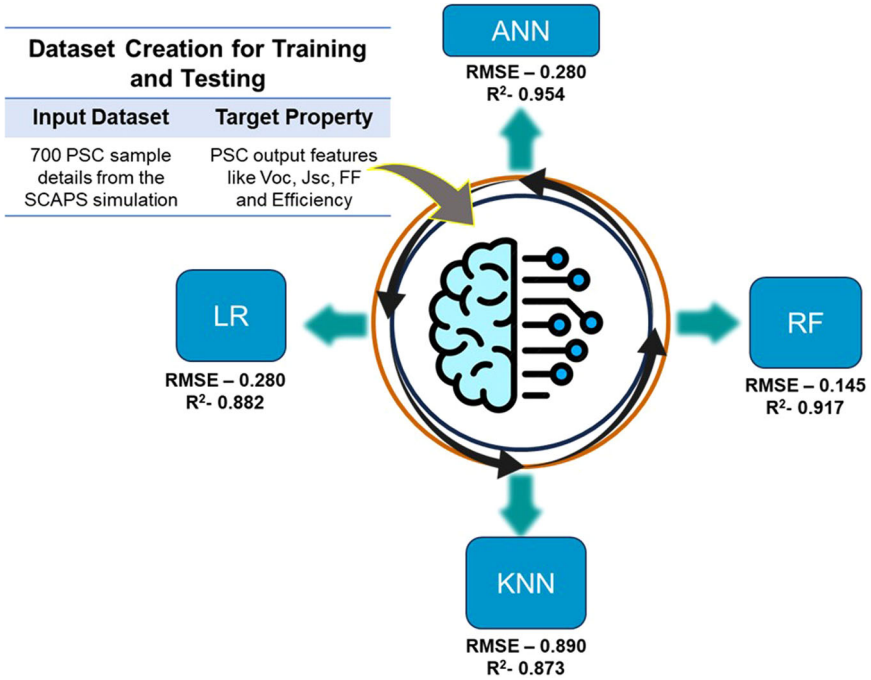


Fig. 3 | Compares the parameterisation of processed data across various ML models. The RMSE and R^2 values represent the ML model's testing accuracy measurements.

Results and discussion

Data pre-processing

Initially, the dataset had around 790 data points with some cloned PSC sample results. Largely, the homogeneity of the results occurred from the minor changes in the ETL thickness and bandgap combination. Considering the homogeneity or the sparsity in the dataset causes a curse of dimensionality, which enhances the complexity of the modelling process[56,57] The cloned samples and the null data were removed from the dataset to boost the accuracy of the ML models. Finally, the dataset remained with 700 data points and was used for further modelling. The ML models' dataset is divided into training and testing data. Figure 3 indicates the RMSE and R-squared values for different ML models after the data processing. As indicated in the Python script in the Supplementary Reference, 560 data points are used for training the ML model and 140 data points are for testing.

ML analysis

The LR is usually considered the simplest model in the ML, and it uses linear equations to fit the data using the relationship between the independent and dependent variables of the observed data[55,58]. Primarily, LR focuses on finding the best fitting line to reduce the gap between the actual and predicted values[58]. The fitting of this model into the dataset provides a R-squared value of 0.882 and RMSE of 0.28. RF is classified as a tree-based model and is well known for avoiding over-fitting efficiently by enhancing the accuracy of the individual decision trees[59]. Even though the RF is flexible and able to manage a wide variety of data, it lacks the capability to predict data outside the training range[55]. By evaluating this algorithm, the RMSE and R-squared values are 0.145 and 0.917, respectively. KNN is catalogued as an instance-based learning algorithm, and the predictions depend on the similarity between the existing data points and the new data points in the training dataset[60]. This model has demonstrated an RMSE and R-squared value of 0.890 and 0.873, respectively. Considering the other three techniques, the ANN model's popularity has increased exponentially in the last few years due to the wide acceptance and usage of deep learning. While applying the ANN algorithm, the model shows the R-squared and RMSE values as 0.954 and 0.028, respectively. While comparing these values for the other algorithms, the ANN demonstrates better results. Subsequently, further device prediction performance analysis was done through the ANN-based ML model.

The ML model was designed to demonstrate the optimisation capability with four key physical properties of perovskite and ETL layers as input and predicting five PSC output parameters. The input parameters are the ETL and perovskite layer's bandgaps and thicknesses. The output parameters are open circuit voltage (Voc), short circuit current density (Jsc), fill factor (FF) and PCE. Different combinations of bandgap and thickness data were used for the SCAPS and PSC performance obtained from the simulation results.

The thickness and bandgap of the electron transport layer (ETL) and perovskite layer significantly impact the efficiency of charge extraction, transport, light absorption, and charge carrier generation within the solar cell[14,61,62]. Thicker ETL layers can reduce charge recombination, while the ETL bandgap determines the energy levels for electron transport[51]. Similarly, a thicker perovskite layer enhances light absorption and charge generation, and its bandgap influences the Voc and energy conversion efficiency[51,62]. Performance metrics such as Voc are influenced by the perovskite bandgap and interface quality, with a higher bandgap generally leading to a higher Voc[63]. The Jsc depends on both layers' optimal thickness and bandgap alignment, ensuring maximum light absorption and efficient charge transfer. The FF is affected by charge transport properties and recombination rates, which are influenced by the ETL and perovskite properties[21,35,62]. PCE is a comprehensive performance measure determined by the interplay between Voc, Jsc, and FF, with optimised ETL and perovskite layers leading to higher PCE.

A correlation matrix is used as a statistical analysis tool for quantifying the degree of relationship between the variables[60,64]. Figure 4 shows the generated correlation matrix from the ANN-based ML model and depicts the correlation insights among the input and output features of the dataset. It provides the features' biases and can offer an understanding of the features more influential on the device's performance. The value represented in this correlation matrix is the Pearson coefficient, and the signs (±) depict the positive (+) and negative (-) correlation[54,64]. The FF and Voc indicate the strongest negative correlation, the same as the Jsc and perovskite bandgap. A similar influence is evident with the efficiency of the cell and perovskite parameters.

In contrast, the ETL thickness and ETL bandgap show less correlation to the device performance features. Even though the Pearson correlation analyses the linear relationship between the variables, ANN considers the

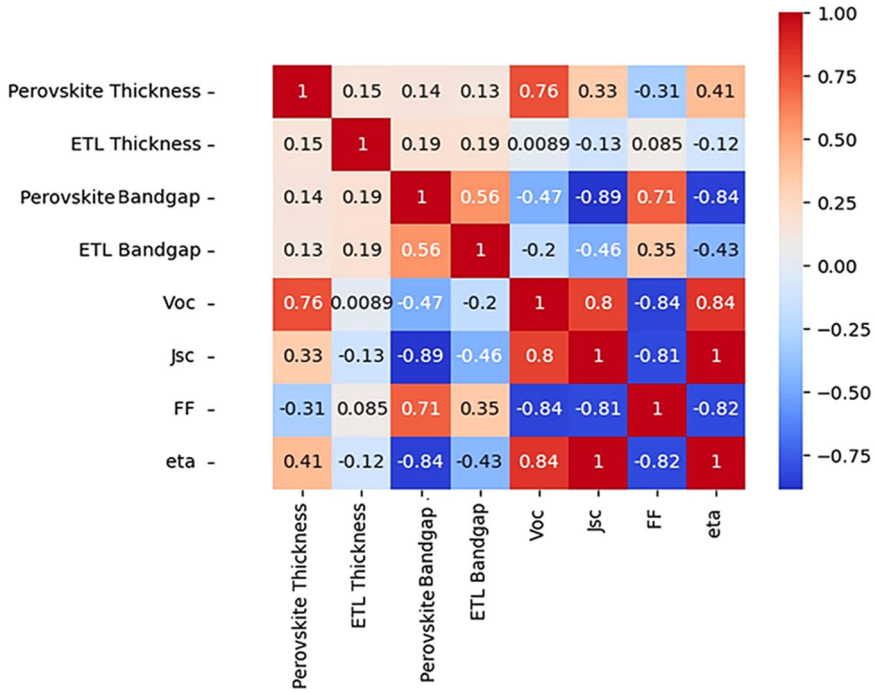


Fig. 4 | Pearson correlation of all the variables in the ANN-based ML model C-PSC performance prediction.

complex non-linear relationships[60,64]. In the context of neural networks in the ANN ML model, the model's performance is not solely based on the linear relationships between the variables. However, the Pearson correlation analysis highlights the need for an effective strategy and guidance of the performance characteristics for PSC fabrication.

The decreased RMSE and increased R^2 value within the model signify heightened accuracy and reliability. RMSE and R^2 values for individual input parameters are delineated in Table 1. Figure 5 portrays a scatter plot elucidating minimal deviations and error margins between predicted values and actual values within the test dataset. In Fig. 5, vertical coordinates represent predicted values, horizontal coordinates depict actual values, and a red line delineates the ideal curve. While the majority of blue data points disperse proximate to the ideal curve, the model showcases diminished error alongside augmented predictive efficacy. However, concerning the parameter Voc, the dispersion percentage notably exceeds that of other predictive parameters. As per Table S2, RMSE and R^2 values stand at 0.0047 and 0.7037, respectively, markedly lower than those of other parameters.

In the ANN based ML model, we investigated the dependency of each characteristic towards the C-PSC performance rather than predicting the performance parameters of the C-PSC. Figure 6 represents the 4D scatter plot of the predicted performance concerning the C-PSC input parameters. As shown in Fig. 6a, the higher Voc prediction values are implied around the combination of lower ETL thickness and higher perovskite thickness with the bandgap of 1.56 eV. This trend is similar to the Jsc and PCE predictions with the thickness and bandgap parameters, as demonstrated in the Fig. 6b and d. In contrast to that, a better FF is shown in the range of higher values of perovskite thickness, bandgap and ETL thickness; Fig. 6c indicates the pattern of the relationship between the parameters and the performance. For higher bandgap perovskite, electron diffusion length is expected to be higher and reduces the possibility of recombination; in turn, this property enhances the fill factor of the C-PSC[49]. With these comparisons, it is evident that several features modify the device performance, and a single-dimensional approach is nearly impossible to control the device optimisation. In this setting, device performance prediction through ML models enumerates all the feasible values for the parameters and gives a multi-dimensional approach to optimisation. However, the ML model optimisations are based on a mathematical model rather than an experimental model, so the rationality of explaining ML model application into the device parameter optimisation can be done by investigating the underlying reliance of each parameter.

Numerous antecedent investigations have relied upon experimental data from diverse published literature or exhibited a paucity of data points. Table 1 delineates the data acquisition strategy, machine learning (ML) methodology, and the predicted parameters inherent to antecedent inquiries. The absence of a standardized testing protocol in fabricating perovskite devices engenders a performance divergence across disparate research endeavours. Furthermore, the efficacy of each device is intricately linked to fabrication methodologies, external variables, and assorted influencing factors. Consequently, assimilating data from prior investigations exacerbates disparities in predictive outcomes, introducing intricacy into the applicability of the ML model for optimization methodologies. Given the present study's focus on a theoretical dataset characterized by a significantly augmented data volume, this model is a more formidable augmentation to optimization techniques for fabricating high-performance HTL-free carbon-based perovskite solar cell (C-PSC) devices. In pursuit of a nuanced understanding of the variations in the properties of ETL and Perovskite layers, it is imperative to delve deeper into additional influencing factors to facilitate a comprehensive performance analysis. Acknowledging the foundational role of the ML prediction dataset, rooted in theoretical calculations, it becomes evident that external variables or fabrication intricacies exert notable influence on experimental outcomes. Thus, a holistic investigation integrating practical observations and experimental data is essential to refine the predictive accuracy of the proposed ML model. By synthesising both empirical and theoretical insights, this model aims to streamline the optimisation processes inherent in PSCs, thereby expediting the transition from laboratory prototypes to commercial viability.

Complexity analysis

The algorithm complexity depends on the number of input parameters required for the algorithm to perform the prediction[41]. In this work, the algorithm designed for four input features and subsequently four output features. As the ANN algorithm consist of different layers, we modified the layers selection to achieve better prediction results. The Figure S1 depicts the layer structure used for the ANN analysis, it consists of three hidden layers with first, second and third layer consists of 128, 64 and 32 neurons, respectively. Both the input and output layers, which is the entry and last

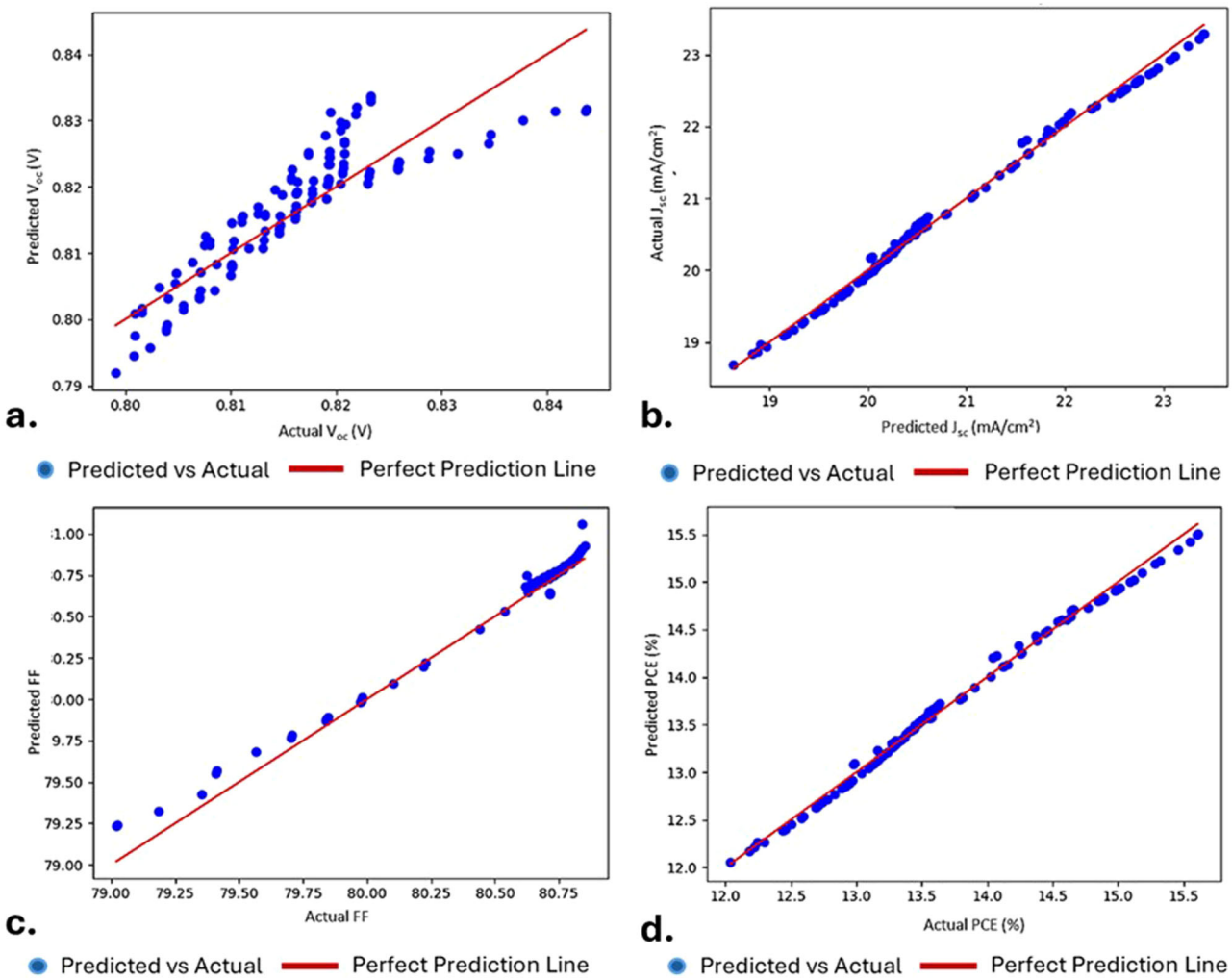


Fig. 5 | Comparison of ANN prediction results with true value and predicted value for each variable. a Represents the open circuit voltage (Voc) results. **b** Current density (Jsc) values (**c**) Fill Factor (FF) values. **d** Efficiency values.

later consist of four neurons each. These layers arrangement provides a flexibility and adaptability in enhancing the prediction accuracy. The progressive reduction of neurons in each layer enables the algorithm to learn the hierarchical representation of the input data. Each parameter's influence on the output prediction were analysed, and Fig. 7 demonstrates the dependency of each parameter in the output prediction. Form the analysis, the perovskite bandgap has the highest influence, followed by the perovskite thickness and ETL layer characteristics. This insight provides a better understanding of the perovskite layer optimisation process in the fabrication of the C-PSC.

In conclusion, an artificial neural network (ANN) machine learning (ML) model was constructed utilizing 700 data points derived from SCAPS-1D simulations to prognosticate parameters of the hole transport layer (HTL)-free carbon-based perovskite solar cells (C-PSCs). These predictions were based on the bandgap and thickness variations of the electron transport layer (ETL) and perovskite layers. The ANN-based ML model exhibited notable predictive prowess for C-PSC device parameters, yielding a low root mean square error (RMSE) value of 0.028 and a high R-squared value of 0.954.

Alternative ML models, including random forest (RF), linear regression (LR), and K-nearest neighbours (KNN), were considered for device performance prediction. Given the superior performance of the ANN model, subsequent analyses were exclusively conducted employing the ANN algorithm. Despite utilising simulation-derived data for device performance prediction, this model is a practical guide for the optimisation process before experimental execution.

A comprehensive examination of each parameter's influence and its correlation to device performance was undertaken. Furthermore, the ML model's predictions on device performance unveiled intricate insights challenging for human validation from randomly generated data. The model delineated perovskite and ETL layer combinations conducive to heightened device performance. Effects of these combinations on device performance parameters, such as Voc, FF, Jsc, and PCE, were scrutinised, streamlining the device optimisation process, and mitigating the time and cost associated with traditional trial-and-error research methodologies.

The interpretable nature of this ML model elucidated the impact of slight variations in thickness and bandgap values on C-PSC performance. Although the ETL bandgap exhibited minimal influence on the C-PSC's power conversion efficiency (PCE), a perovskite bandgap around 1.56 eV yielded superior results compared to higher bandgap perovskite devices. Conversely, higher perovskite bandgap values correlated with diminished device performance, attributable to the sluggish transport of photogenerated carriers.

In contrast to commonly employed device optimisation techniques involving interface layers and additive engineering, which are often benchmarked against standard devices and exhibit variability across research groups or batches, this ML model's swift and reliable predictions facilitate its application across diverse C-PSC fabrication methods. This expedites the device fabrication process and streamlines the commercialisation trajectory of HTL-free C-PSCs.

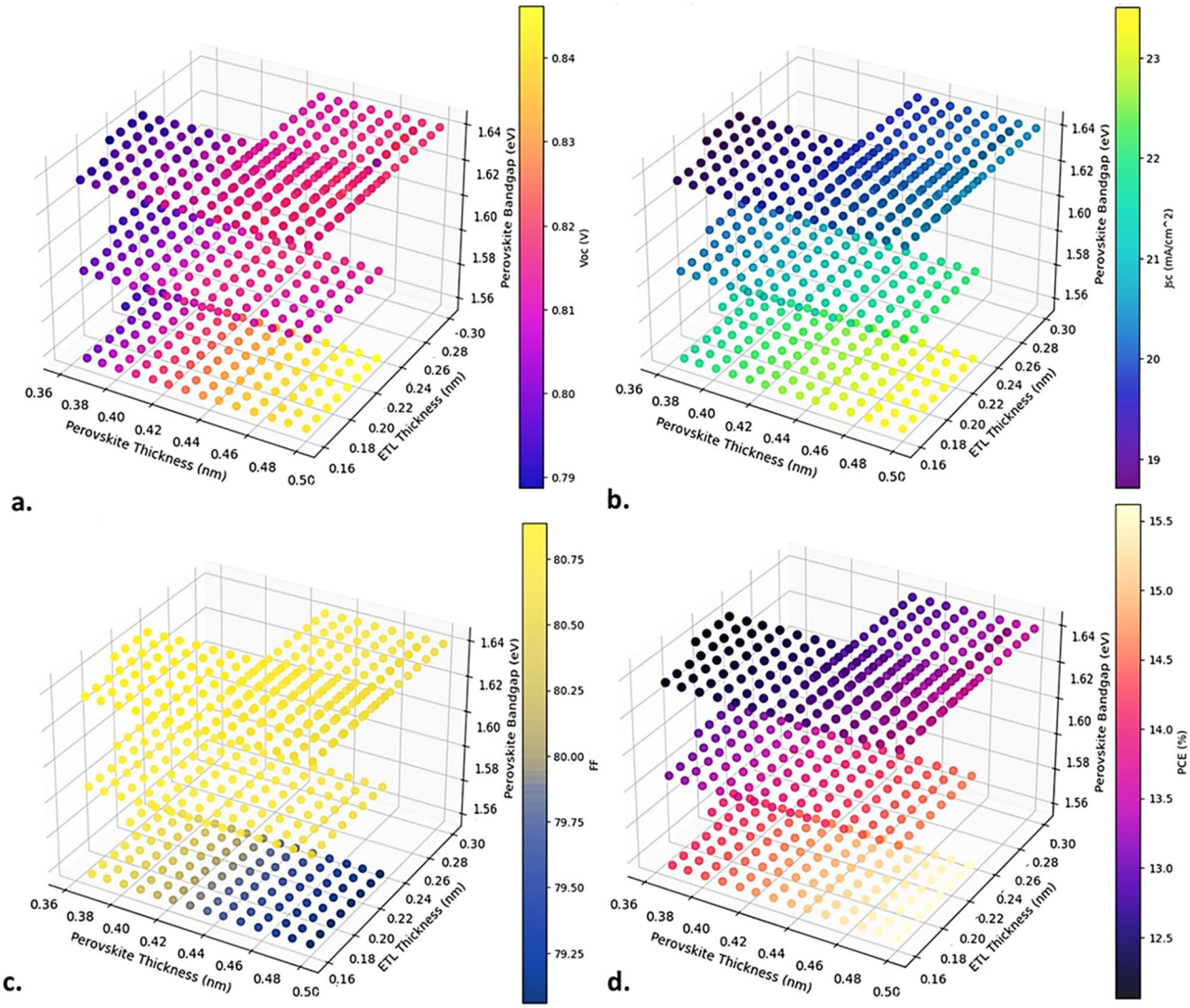
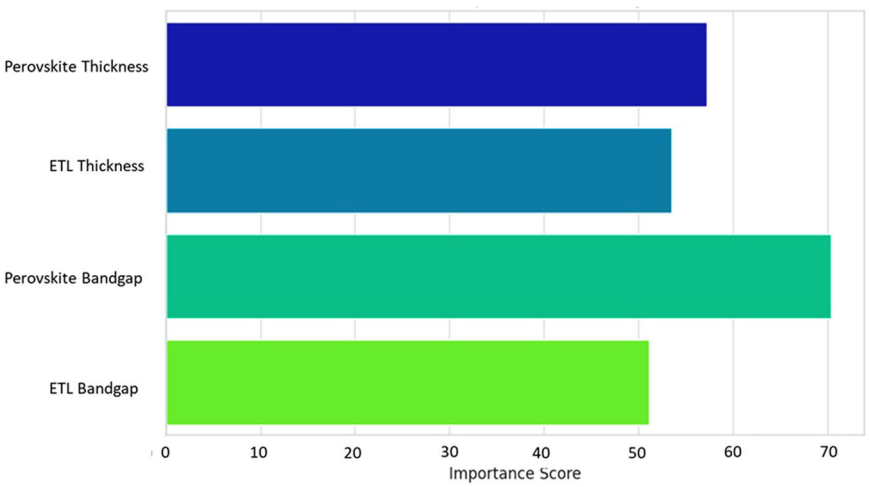


Fig. 6 | ANN output parameters prediction data interpretation through the 4D-plots. a Voc concerning the perovskite layer thickness, ETL thickness and perovskite layer bandgap. **b** Jsc concerning the perovskite bandgap, ETL and perovskite layer thickness. **c** FF concerning the perovskite bandgap, ETL and perovskite layer thickness. **d** PCE concerning the perovskite bandgap, ETL and perovskite layer thickness.

Fig. 7 | Input parameters contribution and importance on the C-PSC performance – according to the ANN model.



Data availability

The dataset is available on request.

Received: 18 March 2024; Accepted: 11 August 2024;
Published online: 10 September 2024

References

1. Zhang, H., Pfeifer, L., Zakeeruddin, S. M., Chu, J. & Grätzel, M. nature reviews chemistry tailoring passivators for highly efficient and stable perovskite solar cells. *Nat. Rev. Chem.* **7**, 632–652 (2023).
2. Photovoltaic Research NREL. *Best Research-Cell Efficiency Chart.* https://www.nrel.gov/pv/cell-efficiency.html (1976).
3. Park, N. G. Research direction toward scalable, stable, and high efficiency perovskite solar cells. *Adv. Energy Mater.* **10**, 1903106 (2020).
4. Zafoschnig, L. A., Nold, S. & Goldschmidt, J. C. The race for lowest costs of electricity production: techno-economic analysis of silicon, perovskite and tandem solar cells. *IEEE J. Photovolt.* **10**, 1632–1641 (2020).
5. Roy, P., Kumar Sinha, N., Tiwari, S. & Khare, A. A review on perovskite solar cells: evolution of architecture, fabrication techniques, commercialization issues and status. *Sol. Energy* **198**, 665–688 (2020).
6. Valsalakumar, S., Roy, A., Mallick, T. K., Hinshelwood, J. & Sundaram, S. An overview of current printing technologies for large-scale perovskite solar cell development. *Energies 2023* **16**, 190 (2022).
7. Parida, B. et al. Recent developments in upscalable printing techniques for perovskite solar cells. *Adv. Sci.* **9**, e2200308 (2022).
8. Huang, B.-J., Guan, C.-K., Huang, S.-H. & Su, W.-F. Development of once-through manufacturing machine for large-area Perovskite solar cell production. *Solar Energy* **205**, 192–201 (2020).
9. Chung, J. et al. Record-efficiency flexible perovskite solar cell and module enabled by a porous-planar structure as an electron transport layer. *J. Energy Environ. Sci.* **13**, 4854 (2020).
10. Huang, Y. et al. Progress report towards simplifying the device structure of high-performance perovskite *Solar Cells*. **30**, 2000863 (2020).
11. Zhou, Y., Herz, L. M., Jen, A. K.-Y. & Saliba, M. Advances and challenges in understanding the microscopic structure-property-performance relationship in perovskite. *Solar Cells.* **7**, 794–807 (2022).
12. Lakhdar, N. & Hima, A. Electron transport material effect on performance of perovskite solar cells based on CH 3 NH 3 GeI 3. *Opt. Mater. (Amst.)* **99**, 109517 (2020).
13. Kim, G.-W. et al. Hole transport materials in conventional structural (n-i-p) perovskite solar cells:from past to the future. *Adv. Energy Mater*. **10**, 1903403 (2020).
14. Valsalakumar, S., Roy, A., Mallick, T. K., Hinshelwood, J. & Sundaram, S. Mesoporous TiO 2-layer's rheological impact on the perovskite solar cell performance. *Mater. Lett.* **337**, 133960 (2023).
15. Bag, A., Radhakrishnan, R., Nekovei, R. & Jeyakumar, R. Effect of absorber layer, hole transport layer thicknesses, and its doping density on the performance of perovskite solar cells by device simulation. *Solar Energy* **196**, 177–182 (2019).
16. Bhandari, S., Valsalakumar, S., Chanchangi, Y., Selvaraj, P. & Mallick, T. K. Effect of novel graphitic carbon/NiO hole transporting electrode on the photovoltaic and optical performance of semi-transparent perovskite solar cells. *RSC Adv.* **13**, 7380–7384 (2023).
17. Yan, J., Savenije, T. J., Mazzarella, L. & Isabella, O. Progress and challenges on scaling up of perovskite solar cell technology. *CrystEngComm* **26**, 312 (2022).
18. Berger, E. et al. Recent developments in perovskite-based precursor inks for scalable architectures of perovskite solar cell technology. *Sustain. Energy Fuels* **6**, 2879–2900 (2022).
19. Zhang, J. et al. Two-step sequential blade-coating of high quality perovskite layers for efficient solar cells and modules. *J. Mater. Chem. A Mater.* **8**, 8447–8454 (2020).
20. Tailor, N. K. et al. Recent progress in morphology optimization in perovskite solar cell. *J. Mater. Chem. A Mater.* **8**, 21356–21386 (2020).
21. Montoya De Los Santos, I. et al. Optimization of CH3NH3PbI3 perovskite solar cells: a theoretical and experimental study. *Sol. Energy* **199**, 198–205 (2020).
22. Tariq Jan, S. & Noman M. Influence of layer thickness, defect density, doping concentration, interface defects, work function, working temperature and reflecting coating on lead-free perovskite solar cell. *Sol. Energy* **237**, 29–43 (2022).
23. Teixeira, C., Fuentes-Pineda, R., Andrade, L., Mendes, A. & Forgács, D. Fabrication of low-cost and flexible perovskite solar cells by slot-die coating for indoor applications. *Mater. Adv.* **4**, 3863–3873 (2023).
24. Glowienka, D., Huang, S. H., Lee, P. H., Tsai, F. Y. & Su, W. F. Understanding the dominant physics mechanisms on the p-i-n perovskite solar cells fabricated by scalable slot-die coating process in ambient air. *Sol. RRL* **8**, 2300791 (2023).
25. Ismail, M., Noman, M., Tariq Jan, S. & Imran, M. Boosting efficiency of eco-friendly perovskite solar cell through optimization of novel charge transport layers. *R. Soc. Open Sci.* **10**, 230331 (2023).
26. Li, F., Lin, F. R. & Jen, A. K. Y. Current state and future perspectives of printable organic and perovskite solar cells. *Adv. Mater.* **36**, 2307161 (2023).
27. Jiang, Q. et al. Towards linking lab and field lifetimes of perovskite solar cells. *Nature* **623**, 313–318 (2023). *2023 623:7986*.
28. Mammeri, M., Dehimi, L., Bencherif, H. & Pezzimenti, F. Paths towards high perovskite solar cells stability using machine learning techniques. *Solar Energy* **249**, 651–660 (2022).
29. Zhang, J. et al. Optimizing perovskite thin-film parameter spaces with machine learning-guided robotic platform for high-performance perovskite solar cells. *Adv. Energy Mater.* **13**, 2302594 (2023).
30. Ismail, Z. S., Sawires, E. F., Amer, F. Z. & Abdellatif, S. O. Perovskites informatics: studying the impact of thicknesses, doping, and defects on the perovskite solar cell efficiency using a machine learning algorithm. *Int. J. Numer. Modelling Electron. Netw. Devices Fields* **37**, e3164 (2023).
31. Kumar et al. Machine learning in perovskite solar cells: recent developments and future perspectives. *Energy Technol*. **11**, 2300735 (2023).
32. Liu, Y. et al. Machine learning for perovskite solar cells and component materials: key technologies and prospects. *Adv. Funct. Mater.* **33**, 2214271 (2023).
33. Hui, Z., Wang, M., Yin, X., Wang, Y. & Yue, Y. Machine learning for perovskite solar cell design. *Comput. Mater. Sci.* **226**, 112215 (2023).
34. Kumar, A., Singh, S., Mohammed, M. K. A. & Sharma, D. K. Accelerated innovation in developing high-performance metal halide perovskite solar cell using machine learning. *Int. J. Modern Phys. B* **37**, 2350067 (2022).
35. Lu, Y. et al. Predicting the device performance of the perovskite solar cells from the experimental parameters through machine learning of existing experimental results. *J. Energy Chem.* **77**, 200–208 (2023).
36. Liu, W. et al. Machine learning enables intelligent screening of interface materials towards minimizing voltage losses for p-i-n type perovskite solar cells. *J. Energy Chem.* **83**, 128–137 (2023).
37. Rauf, A. et al. Effect of bromide incorporation on the electronic & photovoltaic properties of Sn-based perovskite devices: a multiscale investigation utilizing first principles approach and numerical simulation, aided by machine learning models. *Sol. Energy* **253**, 375–388 (2023).
38. Mahmood, A., Sandali, Y. & Wang, J.-L. Easy and fast prediction of green solvents for small molecule donor-based organic solar cells through machine learning. *Phys. Chem. Chem. Phys.* **25**, 10417 (2023).
39. Jaafreh, R. et al. A machine learning-assisted approach to a rapid and reliable screening for mechanically stable perovskite-based materials. *Adv. Funct. Mater.* **33**, 2210374 (2023).
40. Liu, Z. et al. Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. *Joule* **6**, 834–849 (2022).

41. Salah, M. M., Ismail, Z. & Abdellatif, S. Selecting an appropriate machine-learning model for perovskite solar cell datasets. *Mater. Renew. Sustain Energy* **12**, 187–198 (2023).
42. Mahmood, A. & Wang, J.-L. Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy Environ. Sci.* **14**, 90 (2021).
43. Zhou, Z. & Pang, S. Highly efficient inverted hole-transport-layer-free perovskite solar cells. *J. Mater. Chem. A* **9**, 18148 (2020).
44. Iqbal, S. et al. Fully printed HTL-Free MAPbI3 perovskite solar cells with carbon electrodes. *Coatings* **13**, 1338 (2023).
45. Bhandari, S., Roy, A., Ghosh, A., Mallick, T. K. & Sundaram, S. Perceiving the temperature coefficients of carbon-based perovskite solar cells †. *Sustain. Energy Fuels* **4**, 6283–6298 (2020).
46. Bhandari, S., Roy, A., Ghosh, A., Mallick, T. K. & Sundaram, S. Performance of WO 3-incorporated carbon electrodes for ambient mesoscopic perovskite solar cells. *ASC Omega* **30**, 422–429 (2019).
47. Ye, T. et al. Cost-effective high-performance charge-carrier-transport-layer-free perovskite solar cells achieved by suppressing ion migration. *Cite This: ACS Energy Lett.* **6**, 3052 (2021).
48. Burgelman, M., Nollet, P. & Degrave, S. Modelling polycrystalline semiconductor solar cells. *Thin Solid Films* **361–362**, 527–532 (2000).
49. Aliaghayee, M. Optimization of the perovskite solar cell design with layer thickness engineering for improving the photovoltaic response using SCAPS-1D. *J. Electron. Mater.* **52**, 2475–2491 (2023).
50. Ijaz, S. et al. Numerical simulation to optimize the efficiency of HTM-free perovskite solar cells by ETM engineering. *Sol. Energy* **250**, 108–118 (2023).
51. Sinha, N. K., Ghosh, D. S. & Khare, A. Role of built-in potential over ETL/perovskite interface on the performance of HTL-free perovskite solar cells. *Opt. Mater. (Amst.)* **129**, 112517 (2022).
52. Chen, R. et al. Carbon-based HTL-free modular perovskite solar cells with improved contact at perovskite/carbon interfaces. *J. Mater. Chem. C. Mater.* **8**, 9262–9270 (2020).
53. Huang, L., Ge, Z. & Simple Robust, and going more efficient: recent advance on electron transport layer-free perovskite solar cells. *Adv. Energy Mater.* **9**, 1900248 (2019).
54. Li, J., Pradhan, B., Gaur, S. & Thomas, J. Predictions and strategies learned from machine learning to develop high-performing perovskite solar cells. *Adv. Energy Mater.* **9**, 1901891 (2019).
55. Li, W. et al. Performance prediction and optimization of perovskite solar cells based on the Bayesian approach. *Sol. Energy* **262**, 111853 (2023).
56. Parikh, N. et al. Is machine learning redefining the perovskite solar cells? *J. Energy Chem.* **66**, 74–90 (2022).
57. Klein, L. et al. Discovering pocess dynamics for scalable perovskite solar cell manufacturing with explainable AI. *Adv. Mater.* **36**, 2307160 (2023).
58. Hartono, N. T. P. et al. Stability follows efficiency based on the analysis of a large perovskite solar cells ageing dataset. *Nat. Commun.* **14**, 1–7 (2023).
59. Odabaşı Özer, Ç. & Yıldırım, R. Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy* **56**, 770–791 (2019).
60. Salah, M. M., Ismail, Z. & Abdellatif, S. Selecting an appropriate machine-learning model for perovskite solar cell datasets. *Mater. Renew. Sustain. Energy* **12**, 187–198 (2023).
61. Al-Mousoi, A. K., Mustafa, •, Mohammed, K. A. & Mohammed, M. K. A. Engineered surface properties of MAPI using different antisolvents for hole transport layer-free perovskite solar cell (HTL-free PSC). *J. Solgel Sci. Technol.* **96**, 659–668 (2020).
62. Valsalakumar, S., Bhandari, S., Mallick, T. K., Hinshelwood, J. & Sundaram, S. Experimental validation of optimized solar cell capacitance simulation for rheology-modulated carbon-based hole transport layer-free perovskite solar cell. *Adv. Sustain. Energy* **5**, 2300244 (2024).
63. Gok, E. C. et al. Predicting perovskite bandgap and solar cell performance with machine learning. *Sol. RRL* **6**, 2100927 (2022).
64. Hu, Y. et al. Machine-learning modeling for ultra-stable high-efficiency perovskite solar cells. *Adv. Energy Mater.* **12**, 2201463 (2022).
65. Gla, Ç., Odabaş¸ı, O. & Yıldırım, R. Machine learning analysis on stability of perovskite solar cells. *Solar Enery Mater. Solar Cell* **205**, 110284 (2019).
66. Hussain, W., Sawar, S. & Sultan, M. Leveraging machine learning to consolidate the diversity in experimental results of perovskite solar cells. *RSC Adv.* **13**, 22529–22537 (2023).
67. Gla, Ç., Odabaş¸ı, O. & Yıldırım, R. Assessment of reproducibility, hysteresis, and stability relations in perovskite solar cells using machine learning. *Energy Technol.* 8 1901449 (2020).

Acknowledgements

SV acknowledges the Engineering and Physical Sciences Research Council for the PhD fellowship. This research was funded by the Engineering and Physical Sciences Research Council, UK: PhD Fellowship and Engineering and Physical Sciences Research Council, UK: EP/T025875/1.

Author contributions

S.V.: Conceptualization, investigation, S.V., S.B., A.R.: methodology, data curation, analysis, writing original draft, writing review and editing, visualisation. T.K.M., J.H., S.S.: Writing—review and editing, supervision, project administration, funding acquisition.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01383-7.

Correspondence and requests for materials should be addressed to Sreeram Valsalakumar or Shubhranshu Bhandari.