Machine Learning Assisted Analysis, Prediction, and Fabrication of High-Efficiency CZTSSe Thin Film Solar Cells

Vijay C. Karade, Santosh S. Sutar, Seung Wook Shin, Mahesh P. Suryawanshi,*
Jun Sung Jang, Kuldeep Singh Gour, Rajanish K. Kamat, Jae Ho Yun,
Tukaram D. Dongale,* and Jin Hyeok Kim*

The Earth-abundant element-based $Cu_2ZnSn(S,Se)_4$ (CZTSSe) absorber is considered as a promising material for thin-film solar cells (TFSCs). The current record power conversion efficiency (PCE) of CZTSSe TFSCs is ≈13%, and it's still lower than CdTe and CIGS-based TFSCs. A further breakthrough in its PCE mainly relies on deep insights into the various device fabrication conditions; accordingly, the experimental–oriented machine learning (ML) approach can be an effective way to discover key governing factors in improving PCE. The present work aims to identify the key governing factors throughout the device fabrication processes and apply them to break the saturated PCE for CZTSSe TFSCs. For realization, over 25,000 data points were broadly collected by fabricating more than 1300 CZTSSe TFSC devices and analyzed them using various ML techniques. Through extensive ML analysis, the i-ZnO thickness is found to be the first, while Zn/Sn compositional ratio and sulfo-selenization temperature are other key governing factors under thin or thick i-ZnO thickness to achieve over 11% PCE. Based on these key governing factors, the applied random forest ML prediction model for PCE showed Adj. $R^2 = >0.96$. Finally, the best-predicted ML conditions considered for experimental validation showed well-matched experimental outcomes with different ML models.

1. Introduction

In recent years, Earth-abundant kesterite $Cu_2ZnSn(S,Se)_4$ (CZTSSe) absorber material has been widely investigated for thin-film solar cells (TFSCs) application, owing to its low-cost, non-toxic elements, and suitable optoelectronic properties.[1,2] The current record power conversion efficiency (PCE) of the CZTSSe TFSCs (≈13.0%) is far below the Shockley-Queisser limit (31%) as well as $Cu(In,Ga)Se_2$ (CIGS, 23.4%) and CdTe (22.1%) counterparts.[3,4] To improve the PCE of kesterite, numerous strategies have been designed and attempts have been made.[5–8] These studies have shown that the design and fabrication of the individual layer mainly determine the device's performance. For instance, a slight change in the compositional ratio of a CZTSSe absorber layer can result in poor device performance due to the easy formation of high densities of anti-site defects and secondary phases.[7,9] Similarly, the change in

V. C. Karade, K. S. Gour, J. H. Yun
Department of Energy Engineering
Korea Institute of Energy Technology (KENTECH)
Naju 522132, Republic of Korea
V. C. Karade, J. S. Jang, K. S. Gour, J. H. Kim
Optoelectronics Convergence Research Center and Department of
Materials Science and Engineering
Chonnam National University
Gwangju 61186, Republic of Korea
E-mail: jinhyeok@chonnam.ac.kr

S. S. Sutar
Yashwantrao Chavan School of Rural Development
Shivaji University
Kolhapur 416004, India
S. W. Shin
Future Agricultural Research Division
Rural Research Institute
Korea Rural Community Corporation
Ansan-si 15634, Republic of Korea
M. P. Suryawanshi
School of Photovoltaic and Renewable Energy Engineering
University of New South Wales
Sydney, New South Wales 2052, Australia
E-mail: m.suryawanshi@unsw.edu.au
R. K. Kamat
Department of Electronics
Shivaji University
Kolhapur 416004, India
R. K. Kamat
Dr. Homi Bhabha State University
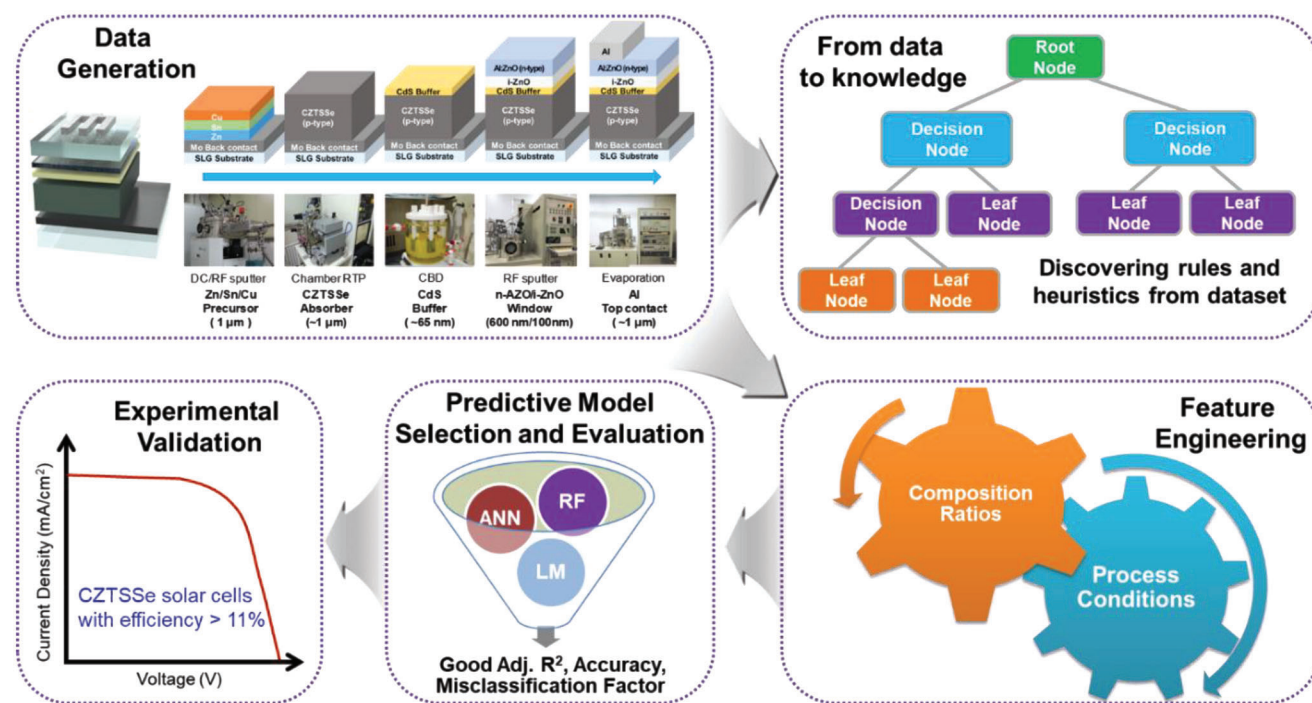15 Madam Cama Road, Mumbai 400032, India

Figure 1. Illustration of ML-guided fabrication of CZTSSe TFSCs. The first method includes the fabrication of devices with a different process condition (data generation) and uses the existing data to discover the key stages/rule in device fabrication. The second method includes the use of generated data to train the model and predict the device parameters with random device fabrication conditions.

transparent conducting oxide (TCO) deposition conditions also causes variations in their optoelectronic properties affecting the device PCE.[10,11] It suggests that exhaustive experiments must be performed to achieve a further breakthrough in PCE. However, in multi-structured TFSCs, a continuous optimization process of each layer causes the wastage of energy and available material resources, ultimately increasing the cost of experimentation and human efforts. Therefore, to shorten the time and overall cost of the experimentations, it is advisable to use high throughput machine learning (ML)-based analysis techniques.[12,13]

The fundamental idea behind the ML approach is to analyze and map the relationships between the non-linear physical, chemical, and optoelectronic properties of materials with their associated output data.[14–16] This early understanding of the optoelectronic device through the ML approach helps the material science community to clearly understand, discover, and optimize the fabrication process to develop highly efficient TFSCs.[17–22] It also provides key steps in the device fabrication process and optimal layer fabrication conditions omitting excessive experimental stages (**Figure 1**).[22–25] Accordingly, high throughput ML techniques are extensively employed for different photovoltaic (PV) materials. For the perovskite TFSCs, the different recom-

bination mechanisms, suitable device architecture, the compositional ratio for materials, and other process conditions have been extensively explored with different ML techniques.[26–28] Lately, these ML techniques have been used for chalcogenide-based TFSCs such as CIGS and CZTSSe.[22,23,29] Zhu et al.[22] established the correlations between the device performances of CIGS TFSCs and different preparative parameters through an artificial neural network (ANN), random forest (RF), and linear regression ML algorithms. Li et al.[23] employed the Gaussian process regression (GPR) ML algorithm to rapidly predict the optimal Mn doping concentration in the CZTSSe absorber layer. Luque et al.[29] successfully employed the linear discriminant analysis (LDA)-based ML techniques to mainly establish the relationship between point defects and PCE in the $Cu_2ZnGeSe_4$ TFSCs.

The previous ML studies on the kesterite mainly fall back in several aspects, i) among a diverse list of ML algorithms, many of the researchers only use either one or two algorithms, ii) the database (number of observations) used to train the model is considerably small ($\approx$<100 data points), iii) considering different device parameters only limited target properties such as only open circuit voltage (V_{OC}) and/or PCE were considered during the analysis, iv) similarly, among different stages of device fabrication process either one or two steps were used for the studies. All these factors can result in incompetent and less reproducible ML models. Therefore, to well establish the ML-guided kesterite device fabrication model, it is necessary to consider a large number of data sets. It should contain all the key stages of the device fabrication process as well as device parameters. As an outcome, establishing the correlation between

T. D. Dongale
Computational Electronics and Nanoscience Research Laboratory
School of Nanoscience and Biotechnology
Shivaji University
Kolhapur 416004, India
E-mail: tdd.snst@unishivaji.ac.in

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

device parameters and process conditions via different sets of ML algorithms delivers a more practical and liable ML model.

Inspired by the above-mentioned insights and previous ML studies, the present work, for the first time, demonstrates the experimental-oriented ML approach to a complete device fabrication process to break saturated PCE of CZTSSe TFSCs. It consists of over 25000 data points collected by fabricating over 1300 devices with different process conditions to build universal and efficient prediction models. This work has the strong motivation of employing different ML algorithms and finding a suitable ML model, which gives the key governing factors (device fabrication step). In addition, it includes the validation of the proposed ML model by fabricating the CZTSSe TFSCs-based on the decision tree (DT) and classification and regression tree (CART) rules.

2. Results and Discussion

2.1. ML-Assisted Prediction of Optimal Device Fabrication Conditions

The first step in the ML-assisted design of TFSCs consists of identifying possible rules- and heuristics based on synthesis conditions such as annealing time, temperature, pressure, compositional ratio, etc. These insights help to understand how synthesis conditions affect the device performances and output parameters. In this regard, the major device fabrication process, annealing conditions, and compositional ratios (precursor composition) are used to build the ML models of the CZTSSe TFSCs. The band gap energy (E_g) engineering and effect of the S/Se ratio on CZTSSe device performance could be another part of the study and thus considered as out of the scope of the present ML study. Therefore, it has been excluded from the present work. For reference, the compositional ratio of the CZTSSe absorber layer after the standard annealing process is given in Table S2 (Supporting Information).

Initially, we employed DT and CART algorithms to find the best suitable algorithm that provides the optimal rules. As DT is a supervised ML algorithm, it is necessary to provide the information of different sets of classes, such as class-1 (high), class-2 (medium), or class-3 (low), to achieve an output of target properties.[30] Accordingly, to build the DT model with PCE as the target property, the dataset was divided into four classes viz. low (PCE from 5.01 to 6.53%), medium (PCE from 6.53 to 7.34%), high (PCE from 7.34 to 8.16%), and very high (PCE from 8.16 to 10.24%) (See Table S1, Supporting Information) and resultant DT model is depicted in **Figure 2**a. It should be noted that the representative DT model for PCE as the target property is only given here, while the similar DT models for V_{OC}, short circuit current (J_{SC}), fill factor (FF), series resistance (R_S), and shunt resistance (R_{Sh}) can be found in Figures S1 to S5 (Supporting Information). After execution of the DT algorithm, each generated node consists of the following details: i) the class of the node, ii) the probabilities of classes, iii) the number of observations used, and iv) the decision rule, the similar information, and node details (Figure 2a upper left corner). As explained earlier, the present DT model was obtained with several device fabrication process conditions including individual optimization of each layer; it shows the probability of the key process conditions that need to be controlled during the fabrication of high PCE TF-

SCs. Hence, the set of DT rules can be defined and/or read as follows; the DT model of the PCE split from the root node and gave the first decision rule about TCO conditions, i.e., i) if the i-ZnO thickness is ≥ 29 nm, then low PCE can be obtained; otherwise, very high PCE can be obtained. The probability of getting a low PCE is 87%, while the probability of getting a very high PCE is 13%. This first rule demonstrates that controlling the i-ZnO thickness can be a key governing factor in fabricating very high PCE CZTSSe TFSCs. In the next step, the DT establishes the second decision rule, i.e., if the compositional ratio of Zn/Sn is ≥1, the probabilities of low and high PCE are 65% and 21%, respectively. Similarly, the other rules can be established in the DT model. The DT sub-rules, which use less than 10% of total observation, make the model irrational. In other words, DT sub-rules, which use minimum observations out of the total to construct the rule/model, result in building less efficient or reproducible conditions. Therefore, one should be conscious while following the rule.

In the next work stage, we build a CART-based decision model of the CZTSSe TFSCs. The CART is an unsupervised ML algorithm, indicating that the data labeling is unnecessary and the models themselves learn from the dataset.[30] In some cases, the CART can also be used for classification problems as a supervised ML algorithm. The CART ML algorithm lowers the manual work and easily generates patterns from the dataset, resulting in a lower computational cost.[31] The CART ML model can provide different clusters of target properties. Figure 2b shows the CART model of the CZTSSe TFSCs with PCE. In each decision node, the percentage and fraction numbers represent the total number of observations used and the average PCE values of that cluster, respectively, while the decision rule is placed at the bottom of each node (Figure 2b upper left corner). After executing the CART model, we observed that the root node is divided into two parts. The decision rule can be read as follows: if the i-ZnO thickness is ≥ 29 nm, then an average PCE of ≈7.2% can be obtained; otherwise, ≈8.6% can be obtained. Further, the CART model checks the second rule: if the composition ratio of Cu/Sn ≥ 1.3, then an average PCE of 6.7% can be obtained otherwise 7.4% can be obtained; similarly, one can read further rules. As shown in Figure 2b, the CART model provides more detailed and simplified decisions than the DT model. This CART model suggests optimal conditions to fabricate the highly efficient CZTSSe TFSCs-based on the present data. Further, the CART models of the other device parameters are depicted in Figures S6 to S10 (Supporting Information).

Briefly, DT and CART models clearly demonstrate that the influence of i-ZnO thickness, compositional ratios of the precursor (i.e., ratios of Cu/(Zn+Sn), Cu/Zn, Zn/Sn, and Cu/Sn), and sulfo-selenization conditions are the key governing parameters that need to be a concern mainly while achieving the high PCE of the TFSCs. The higher PCE can be obtained by maintaining the i-ZnO as low as 25–29 nm (minimum thickness used in the experiments). However, the predicted higher PCE with the thinner i-ZnO thickness uses just 13% of total observations in the DT model (see Figure 2a). As a result, the chances of getting very high PCE (>8.6%) are low under thinner i-ZnO even though the probability is over 79%. Further, keeping the i-ZnO thickness >29 nm (95-100 nm highest thickness used in the experiments) could also result in low PCE. The next sub-rule suggests
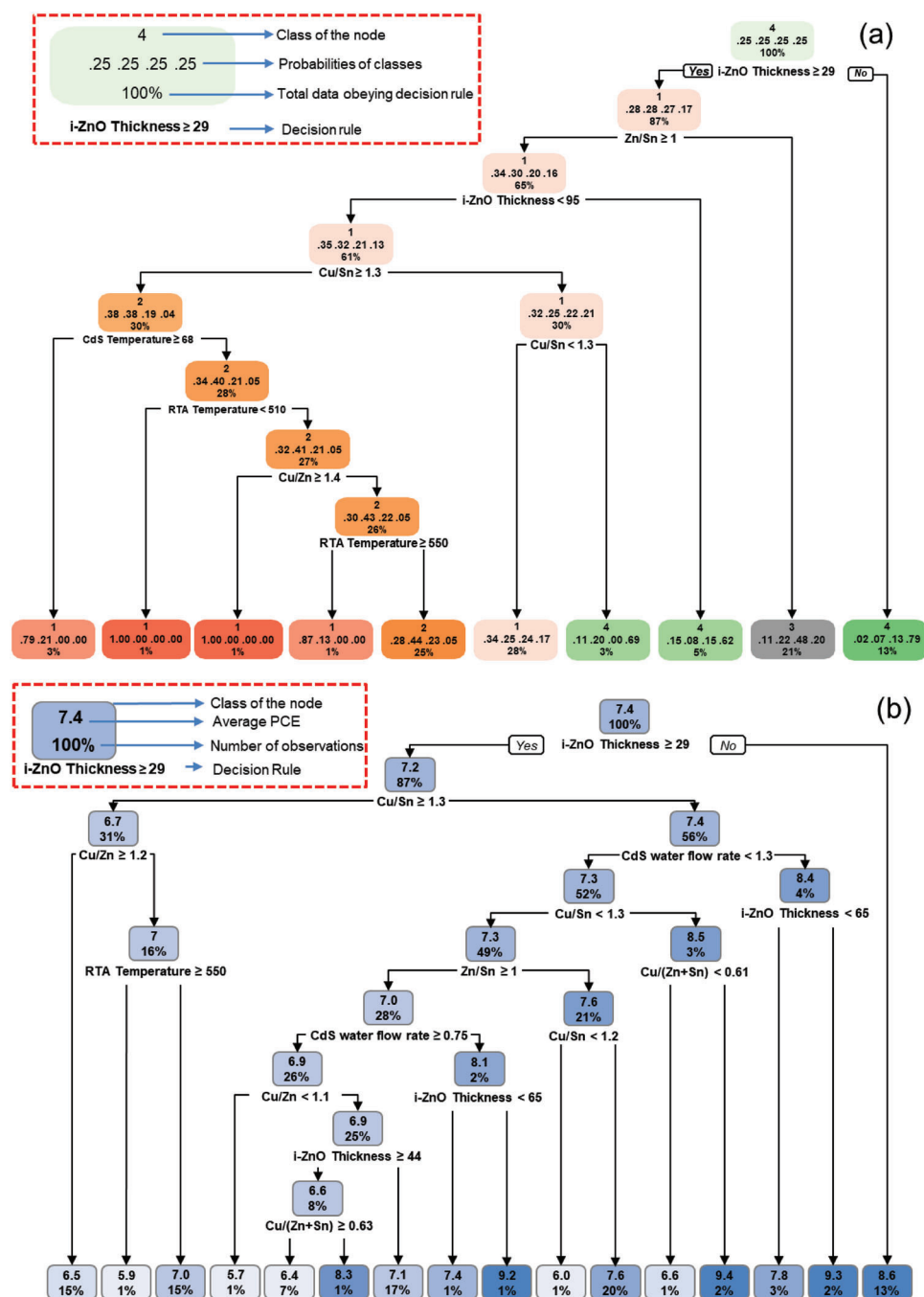
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

Figure 2. a) DT and b) CART-based models of "PCE". These models are created to identify possible rules- and heuristics based on device fabrication conditions and compositional ratios of the CZTSSe TFSCs. In the case of DT, the class of nodes has the following categories, 1: low PCE (from 5.01 to 6.53%), 2: medium PCE (from 6.53 to 7.34%), 3: high PCE (from 7.34 to 8.16%), and 4: very high PCE (from 8.16 to 10.24%). On the other hand, the CART model provides an average PEC value for each node.

that keeping the compositional ratio of Zn/Sn is >1, and the *i*-ZnO thickness <95 nm, there is a high probability of getting low PCE predicted with more than 65% of observations. On the contrary, keeping the Zn/Sn ratio close to 1 can result in more than 7% of PCE with a 21% of probability. Different experimental studies on the chalcogenide-based TFSCs have revealed the disadvantages of the thicker *i*-ZnO layer,[32–34] which are consistent with

our predicted DT and CART models. Similarly, the CART model suggests maintaining the *i*-ZnO thickness >29 nm can deliver an average PCE of ≈7.2%, whereas, at a relatively thinner thickness, a PCE of ≈8.6% can be obtained. Commonly, the highly resistive *i*-ZnO mainly acts as a secondary buffer layer that covers the shunting paths.[35] Using the too-thick *i*-ZnO layer (over 100 nm) can hamper the electron transport, which may result in low *FF*

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

and J_{SC}, and thereby poor PCE.[32–34] Further, the DT models obtained for the other device parameters such as the J_{SC} and FF of the TFSCs are shown in Figures S2 and S3 (Supporting Information). It also indicates that higher J_{SC} and FF can be achieved by keeping the lower thickness of the i-ZnO layer. The CART models further suggest relatively higher annealing pressure is required during the sulfo-selenization process (> 400 Torr) to achieve over 50% of FF and over 30 mA cm^{-2} of J_{SC} in the TFSCs shown in Figures S7 and S8 (Supporting Information). The annealing pressure mainly controls the diffusivity of sulfur (S) and selenium (Se) vapors.[36] The lower annealing pressure causes the change in the S/Se ratio, which determines the E_g of the absorber layer and thus J_{SC} and V_{OC} in the device. A similar analogy can be applied to the annealing temperature, which is recommended to be in the range of 510 to 550 °C in both DT and CART models within all device parameters shown in Figures S1–S10 (Supporting Information). In the case of the compositional ratios of the precursor, the suggested optimal Zn/Sn, Cu/Sn, and Cu/Zn compositional ratios are close to 1, 1.3, and 1.2, respectively. Since most of the high PCE devices use Cu-poor and Zn-rich compositional conditions, the excess amount of Zn (Zn/Sn >1) can cause the phase segregation of wide E_g and highly resistive Zn-based secondary phases (i.e., Zn(S,Se)); besides, the excess amount of Sn (Zn/Sn <1) can also form the conductive Sn-based secondary phases (i.e., Sn(S,Se)$_{2-x}$).[7,9] Further detailed information on the influence of the Zn/Sn content over the device performance can be found in the reports.[9,37–40] These secondary phases are known to lower the bulk conductance of the CZTSSe absorber layer, assist the higher carrier recombination rates, and provide additional R_S in the device leading to poor device performance. These above-mentioned DT and CART-based specific device fabricating conditions can be used to fabricate the highly efficient CZTSSe TFSCs.

2.2. ML-Assisted Prediction of Device Parameters for CZTSSe TFSCs

Moving to the next step, for the predictions of device parameters of the CZTSSe TFSCs, we employed two popular ML algorithms, i.e., ANN and RF. Thus, initially, the ANN algorithm was executed for the corresponding data, and adjusted R^2 (Adj. R^2) values of each ANN model were determined by linear fitting. In the present case, the ANN model for device parameters such as PCE, FF, J_{SC}, and V_{OC}, showed the Adj. R^2 values less than 0.5, suggesting that the ANN ML algorithm poorly predicts the device parameters of the CZTSSe TFSCs (Figures S11 and S12, Supporting Information). Detailed results and a discussion of the ANN ML algorithm are provided in the Supporting Information. This poor performance prediction of the ANN model could be attributed to the highly non-linear and heterogeneous dataset used in the present investigation.[41] Accordingly, we used the RF algorithm to improve the predictions and address the poor performance of the ANN algorithm.

The RF is an ensemble learning method and executes any task by creating multiple DTs, which improves the prediction property. Interestingly, the RF algorithm lowers the overfitting and training time and improves the accuracy of predictions.[42] The scatter plot related to actual and predicted device parameters of

the CZTSSe TFSCs using the RF model is shown in **Figure 3**a–d. The Adj. R^2 values of each RF model are found to be >0.91, suggesting that the RF analysis effectively predicts the device parameters of the CZTSSe TFSCs-based on the different device-fabricated process conditions and compositional ratios of the precursor. Please see Note-1 which explains the uncertainty of the input data points of the chemical composition data and PV parameters. The confusion matrix related to categorical output properties of the CZTSSe TFSCs-based on the RF approach by considering all device parameters is shown in Figure S13 (Supporting Information). The accuracy and misclassification factors of the confusion matrix are found to be >0.98 and <0.010, respectively. The feature space of the present study was reduced by identifying the correlation between all input parameters. Figure S14 (Supporting Information) represents the panel plot of all input parameters and compositional ratios. The device-fabricated process conditions and compositional ratios of the precursor are placed diagonally in the matrix. The upper half of the matrix represents Pearson's correlation coefficient between all features, whereas; its pictorial representation is shown in the lower half of the matrix. The panel plots show that the device fabricated process conditions have a very weak correlation with each other (Figure S14a, Supporting Information), and different compositional ratios of precursor have a good correlation with each other, which is shown in the extended panel plots for compositional ratios of the precursor (Figure S14b, Supporting Information). The more positive correlation coefficient value for Cu/(Zn+Sn) with respect to Cu/Zn (+0.81) in the panel plot reveals the increase in the Cu/(Zn+Sn) ratio strongly results from the incase in Cu/Zn ratio than the Cu/Sn and Zn/Sn ratios. In contrast, the more negative correlation coefficient values for the Cu/Zn ratio with respect to Zn/Sn (-0.82) suggest the decrease in Cu/Zn ratio observed due to the increase in the Zn/Sn ratio.

Here, we can clearly identify that the ANN algorithm fails to efficiently build the ML model, on the other hand, the RF algorithm gives good prediction accuracy for device parameters with high Adj. R^2 values are based on the different device-fabricated process conditions. The corresponding panel plot and confusion matrix also give acute information about the interdependency of each device fabrication step. It suggests that the change in one of the device fabrication conditions does not significantly influence other process conditions. At the same time, it also indicates that a slight change in one of the compositional ratios may change another relative composition ratio, which leads to difficulty for fine adjusting each elemental compositional ratio in the precursor materials.

2.3. Prediction of Device Parameters Based on Optimal Features

The statistical panel plots only tell us about the correlation among different process conditions and compositional ratios, but they cannot clearly show the effect of these parameters on the target output device parameters. A feature importance score can be a good way to identify which device fabricated process condition, and compositional ratio affects the device's performance. Furthermore, the feature importance score provides scientific interpretations and/or physical intuition of ML results.[43] Many ML studies recommended that the small
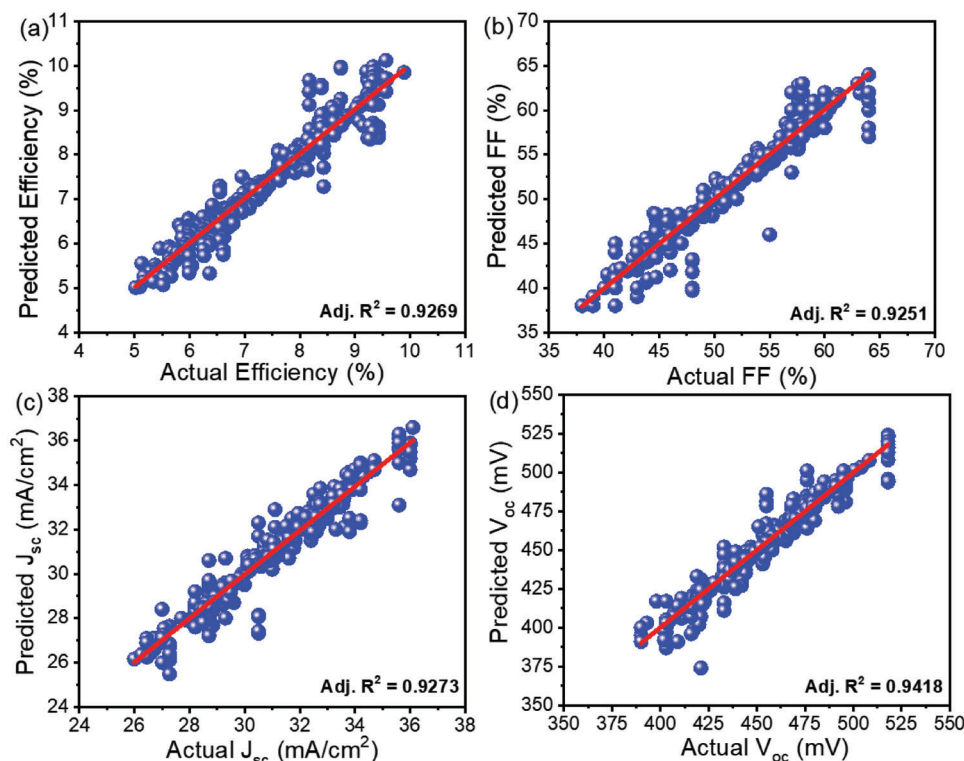
Figure 3. Predictions of a) PCE, b) *FF*, c) J_{SC}, and d) V_{OC} of the CZTSSe TFSCs using RF approach based on all input parameters (device fabricated process conditions and compositional ratios of precursor).

number of input parameters can improve the prediction quality.[44,45] In this regard, identifying and selecting the best possible input parameters and features can provide accurate predictions, lower computation costs, and improve the time required to make decisions.[46] **Figure 4** represents the feature importance score related to different output parameters of the CZTSSe TFSCs, calculated using the gradient boosting (GB) ML algorithm. With the use of a mean decrease in accuracy or Gini score, we rank the importance of features with the help of the GB algorithm.[47] The mean decrease accuracy shows how much the model's accuracy declines if a certain variable is removed. So, the greater the mean decrease accuracy or Gini score, the more significant that variable in the model. As seen in Figure 4a, the Zn/Sn ratio is the most important feature among all device parameters, whereas potassium cyanide (KCN) etching time is the least important feature of the CZTSSe TFSCs. In a nutshell, the compositional ratios and window layer conditions are the most important features, and they can directly affect the PCE of CZTSSe TFSCs. A similar trend is observed for *FF*, J_{SC}, and V_{OC}, of the CZTSSe TFSCs, as shown in Figure 4a–d, respectively.

The compositional ratios of the precursor and window layer conditions dominate the feature importance space of the CZTSSe TFSCs. Though, it should be noted that the *i*-ZnO thickness has a relatively higher score than AZO temperature and RTA temperature process among major device parameters, thus only compositional ratio and *i*-ZnO thickness are taken into account for further predictions. It would be interesting to see how these limited features can predict the different device parameters of the

CZTSSe TFSCs. Accordingly, the RF prediction results based on only compositional ratios of the precursor and *i*-ZnO thickness are shown in **Figure 5**. These results suggested that a limited set of input parameters (here; compositional ratios of the absorber and i-ZnO thickness) can provide comparable predictions over all input parameter-based models. It should be noted that the Adj. R^2 values of *FF* and V_{OC} models based on a limited set of input parameters are small as compared to all input parameter-based models. However, a limited set of input parameters-based models provides lower computational cost, time, and complexity as compared to all input parameter-based models. Most data points remained close to the fitting line, and all observations are placed diagonally in the confusion matrix (Figure S15, Supporting Information). The accuracy of the limited features-based RF models was found to be 1 (PCE), 1 (*FF*), 0.99 (J_{SC}), and 1 (V_{OC}), whereas the misclassification factor was found to be 0 (PCE), 0 (*FF*), 0 (J_{SC}), and 0 (V_{OC}). Both these performance metrics are better than all features-based RF prediction models.

For comparison, we mainly predicted the device parameters for CZTSSe TFSCs by taking only compositional ratios of the precursor as input parameters of the RF algorithm. The prediction results and confusion matrix are shown in Figures S16 and S17 (Supporting Information), respectively. Interestingly, it shows that the RF algorithm provides good prediction results when we employ the feature importance-based optimized set of input variables than panel plots-based variables. The comparative performance of all RF models is shown in Figure S18 (Supporting Information). Considering these factors, it is advisable to use
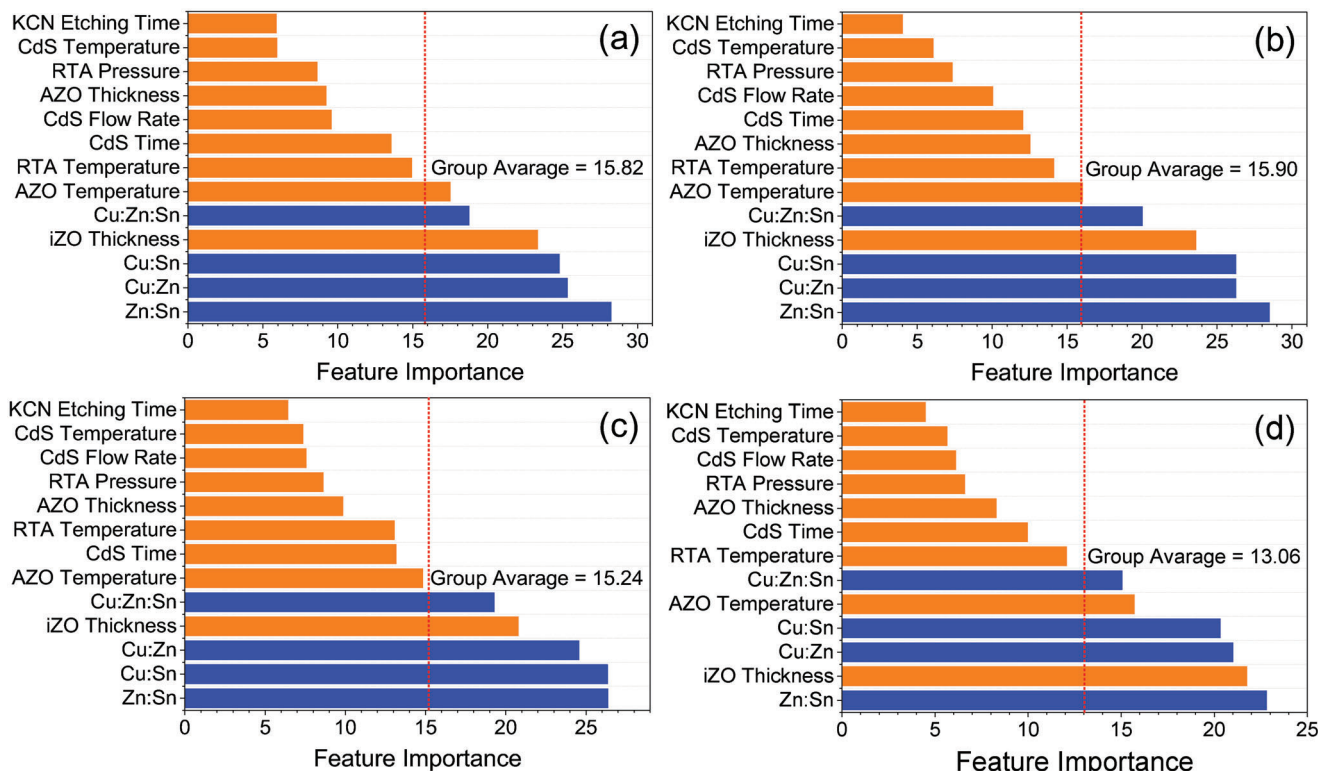
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

Figure 4. Feature importance scores of a) PCE, b) *FF*, c) J_{SC}, and d) V_{OC} of the CZTSSe TFSCs. The blue color indicates the importance of composition ratios, and the orange indicates other process conditions. The group average is calculated to identify the threshold feature importance scores. The features with a feature importance score greater than the group threshold value can be considered important for that particular output parameter.
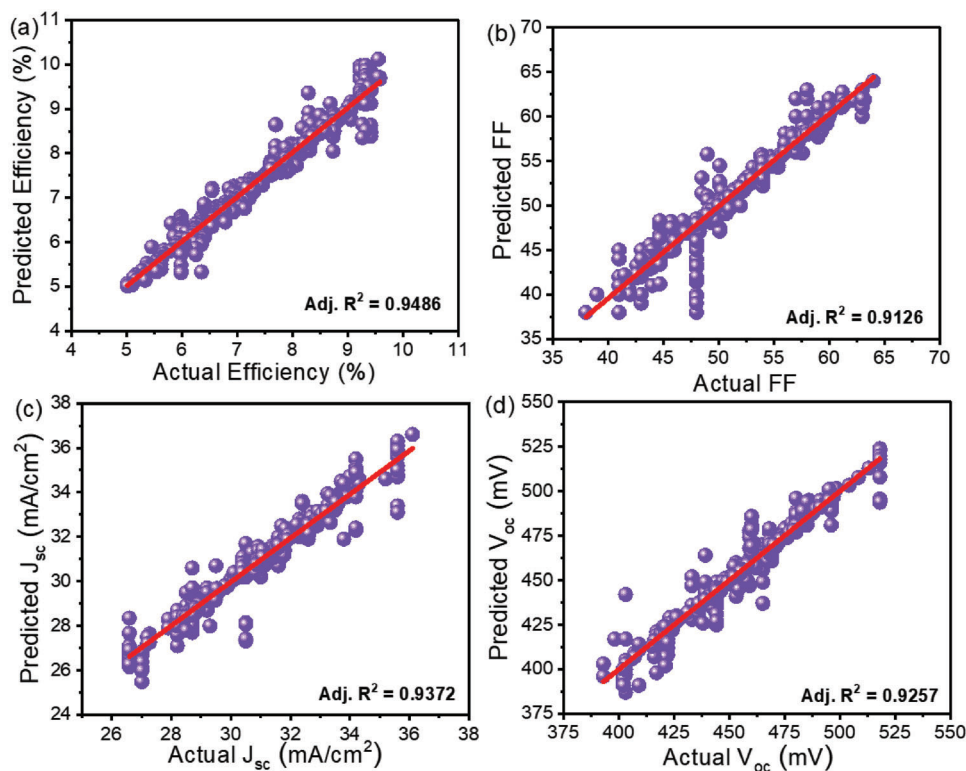

Figure 5. RF algorithm-based predictions of a) PCE, b) *FF*, c) J_{SC}, and d) V_{OC} of the CZTSSe TFSCs based on compositional ratios and *i*-ZnO thickness as input parameters.
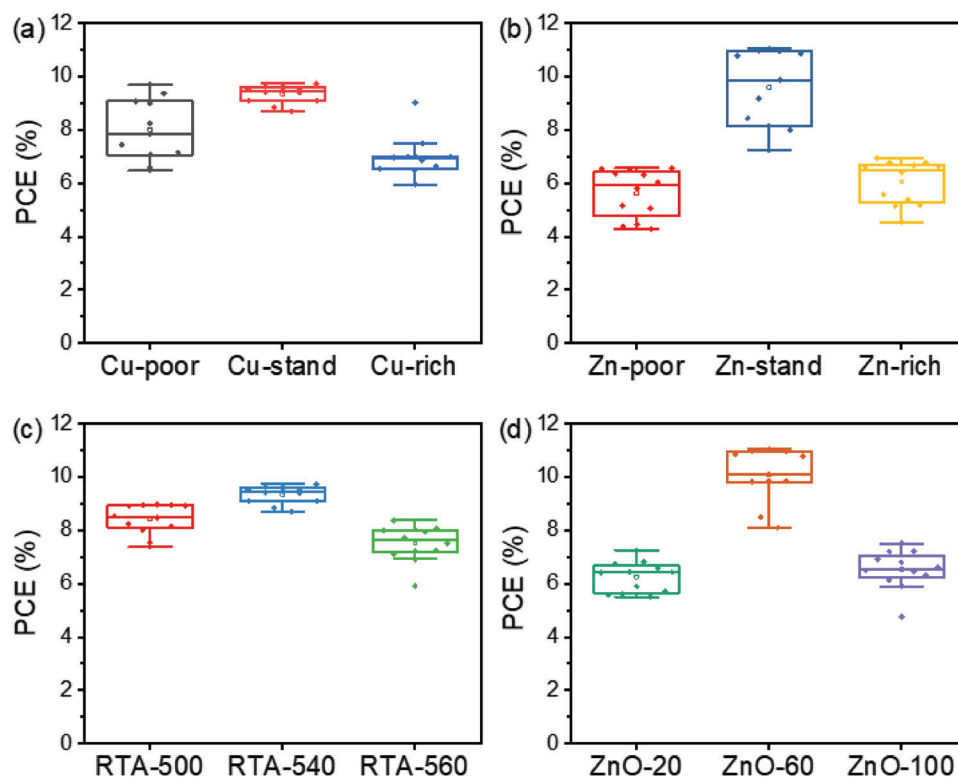
Figure 6. Statistical box plots for PCE in a) Cu series, b) Zn series, c) RTA temperature, and d) *i*-ZnO conditions. Each box plot consists of PCE data of 10–12 CZTSSe TFSCs. Note: the composition ratios for Cu series: Cu-poor (Cu/Sn ~1.25), Cu-stand (Cu/Sn ~1.30), and Cu-rich (Cu/Sn ~1.35); Zn series: Zn-poor (Zn/Sn ~0.9), Zn-stand (Zn/Sn ~1.0) and Zn-rich (Zn/Sn ~1.1); RTA temperature is the samples sulfo-selenized at 500, 540, and 560 °C; *i*-ZnO conditions are the device fabricated with 20, 60, and 100 nm *i*-ZnO thickness.

an optimized set of input features to train, test, and predict the device parameters to lower the cost and time of computations with reasonable predictability.

2.4. Practical Implementation and Verification of an ML Predicted Conditions

Based on DT, CART, and different ML conditions, we identified the 3 key governing factors such as compositional ratios of the precursor, *i*-ZnO thickness, and sulfo-selenization temperature. To validate the DT and CART rules, we selected the optimal conditions as well as the critical limits of each fabrication process step, designed the series of experiments, and fabricated CZTSSe TFSCs. For example, DT and CART rules predicted the optimal Zn/Sn ratio close to 1, so we performed the Zn composition series by fabrication of 3 CZTSSe TFSCs group with lower (Zn/Sn <1), higher (Zn/Sn >1), and optimum conditions Zn/Sn = 1 with PCE as a target property, the experimental results for similar i) compositional ratios (Zn/Sn = 0.9–1.1 and Cu/Sn = 1.25–1.35), ii) *i*-ZnO thickness (25, 60, and 100 nm) and RTA annealing temperature (500, 540, and 560 °C) are depicted in **Figure 6** and discussed as below.

As explained earlier, the CZTSSe TFSCs with different composition conditions were prepared and samples having poor and rich compositional ratios than optimal conditions are named Cu-poor/Zn-poor and Cu-rich/Zn-rich, respectively. As expected

from the ML-predicted optimal compositional conditions, the CZTSSe devices showed relatively lower average PCE (as low as 6%), at the same time at an optimal compositional ratio of Zn/Sn = 1.0 ± 0.5, Cu/Sn = 1.3 ± 0.5 Cu/Zn = 1.2 ± 0.5 and Cu/Zn+Sn = 0.63 ± 0.5 showed the average PCE over 9.5%. After keeping the ML predicted optimal compositional conditions, the sulfo-selenization temperature in RTA was varied. It was found that the device fabricated with a higher sulfo-selenization temperature (560 °C) has a much lower average PCE (≤7.3%) than the absorber prepared at a lower (500 °C) and optimal annealing temperature (540 °C). At an optimal sulfo-selenizaion temperature of 540 °C, a PCE close to 10% was obtained. In the final stage, we found that along with the absorber layer, the TCO layer conditions are as important as the absorber layer preparation conditions. Hence, the devices fabricated with thinner (<30 nm) and thicker (≥100) *i*-ZnO thickness revealed much lower PCE than the optimal *i*-ZnO thickness (60 nm). Based on the optimized DT and CART conditions, and well-trained RF models, we achieved the highest average PCE of over 10.1% (champion device ≈11.0%). The representative EQE and E_g of the CZTSSe device were determined from dEQE/dλ shown in Figure S19 (Supporting Information). It shows a device E_g close to 1.1 eV along with ≈90% absorption in the visible region. Our study suggests that combining a sustainable ML algorithm with experimental data could reduce the additional number of experiments and accelerate the decision-making process while fabricating high-efficiency devices.
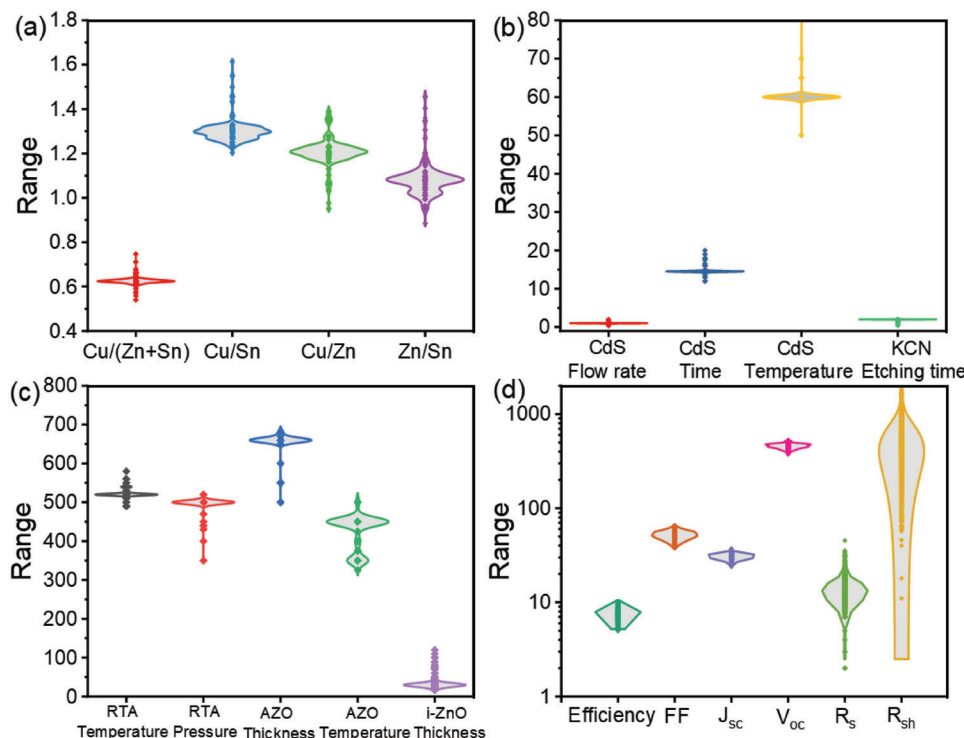
**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

Figure 7. Distributions of the a) different metal composition ratios, b) buffer layer deposition conditions, c) absorber annealing and window layer deposition conditions, and d) electrical parameters of the CZTSSe TFSCs.

3. Conclusion

With an experimental-oriented ML approach employed for >1300 CZTSSe TFSC devices, we first identified the complex relationship between device fabrication conditions and the optoelectronic properties of CZTSSe TFSCs. The DT and CART models successfully provide the hidden set of rules for the fabrication of CZTSSe TFSCs. Specifically, the unsupervised CART model provides more detailed decision rules than DT. In the case of predictive models, the RF accurately predicts the different optoelectronic properties of CZTSSe TFSCs than ANN. Based on DT, CART, and feature importance score model outcomes, the compositional ratios of the precursor, *i*-ZnO thickness, and sulfoselenization temperature were identified as the 3 key governing factors in the fabrication of TFSCs. The AZO deposition temperature can also be considered as one of the governing factors as it also dominates the feature importance score except for J_{SC}. These key governing factors validated experimentally provide excellent optoelectronic properties, further assisting the improvement in device PCE ≥11.0%. It infers that ML-assisted device fabrication can be a powerful tool to understand and accelerate the discovery of highly efficient other organic and/or inorganic-based TFSCs.

4. Experimental Section

Fabrication of CZTSSe TFSCs: The tri-layer stacked metallic precursors (Cu/Sn/Zn) were deposited on a Mo-coated soda-lime glass substrate by the direct current (DC) sputtering method at room temperature.[39] Further, the metallic precursors were soft-annealed under an Ar atmosphere

and subsequently annealed in a graphite box with a chamber-type rapid thermal annealing (RTA) system under a chalcogen vapor atmosphere.[40] After the annealing process, the chamber was allowed to cool down naturally, and annealed thin films were removed at room temperature. The annealed thin films were treated with KCN to remove the secondary phases from the absorber surface. The *n*-type CdS layer was deposited by chemical bath deposition techniques to make a *p-n* junction. The *i*-ZnO and AZO layers were deposited by radio frequency (RF) sputtering. Finally, to fabricate the TFSC devices, Al metallic grid was prepared by DC sputter Al/AZO/i-ZnO/CdS/CZTSSe/Mo/glass structure.[39,40,48,49] Further detailed information regarding each experimental stage is provided in the Supporting Information.

Dataset Preparation and ML Analysis Process: The CZTSSe TFSCs was fabricated with different process conditions and recorded their optoelectronic properties to create a dataset. It should be noted that, while fabricating the CZTSSe devices, additional strategies such as interface passivation, post-deposition heat treatment, or cation doping had not been employed. During the data collection process, the CZTSSe TFSCs having over 5% PCE were used for the ML analysis and predictions. Experimental data points considered were i) different compositional ratios such as Cu/(Zn+Sn), Cu/Sn, Cu/Zn, and Zn/Sn, CdS deposition times, CdS deposition temperature, reaction solution flow rate in the chemical bath deposited CdS bath (CdS flow rate), KCN etching time, RTA process temperature, RTA chamber pressure, RTA annealing time, AZO thickness, AZO deposition temperature, and *i*-ZnO thickness, ii) device output parameters such as PCE, V_{OC}, J_{SC}, FF, R_S, and R_{Sh}. The device fabrication process conditions were used as input and different device parameters were used as output to build the different ML models. The diversity or distribution plots of the device fabricated process conditions, compositional ratios, and device parameters properties of the CZTSSe TFSCs are shown in **Figure 7**a–d, respectively. It shows the heterogeneous or non-uniform distribution of data. Such a heterogeneous dataset helps build stronger ML models that discover hidden rules and predict the conditions for highly efficient TFSCs through different ML algorithms.

Computational Details: At the outset, DT and CART ML algorithms were employed to identify and discover a set of rules and heuristics from the experimental dataset. Distinct device fabrication process conditions and compositional ratios were used as input data to build DT and CART models. Both models provide quantitative decision rules for each device parameter, which could be used to optimize and fabricate highly efficient CZTSSe TFSCs. Similarly, ANN and RF ML algorithms were also tested to predict device parameters. During this investigation, it was observed that the prediction quality of ANN and RF models was not satisfactory. Hence, a correlation analysis was performed and quantified the feature importance score by employing a gradient-boosting ML algorithm among different device fabrication processes, annealing conditions, and compositional ratios. They provide the key governing factors (device fabrication step) that determine the device's performance. Based on the above results, the device parameters of the highly efficient CZTSSe TFSCs were predicted using the RF algorithm and utilized important features. For all the ML processes, the source code of all algorithms was created using R studio (R version 3.6.2).

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1] U. V. Ghorpade, M. P. Suryawanshi, M. A. Green, T. Wu, X. Hao, K. M. Ryan, *Chem. Rev.* **2023**, *123*, 327.

[2] X. Liu, Y. Feng, H. Cui, F. Liu, X. Hao, G. Conibeer, D. B. Mitzi, M. Green, *Prog. Photovoltaics* **2016**, *24*, 879.

[3] "Best Research-Cell Efficiency Chart, Photovoltaic Research, NREL," https://www.nrel.gov/pv/cell-efficiency.html (accessed: March, 2023).

[4] M. A. Green, E. D. Dunlop, G. Siefer, M. Yoshita, N. Kopidakis, K. Bothe, X. Hao, *Prog. Photovoltaics* **2023**, *31*, 3.

[5] G. Larramona, C. Choné, D. Meissner, K. Ernits, P. Bras, Y. Ren, R. Martín-Salinas, J. L. Rodríguez-Villatoro, B. Vermang, G. Brammertz, *J. Phys. Condens Matter* **2020**, *2*, 024009.

[6] A. Wang, M. He, M. A. Green, K. Sun, X. Hao, *Adv. Energy Mater.* **2023**, *13*, 2203046.

[7] K. S. Gour, V. C. Karade, P. T. Babar, J. S. Park, D. M. Lee, V. N. Singh, J. H. Kim, *Sol. RRL* **2021**, *5*, 2000815.

[8] M. He, C. Yan, J. Li, M. P. Suryawanshi, J. Kim, M. A. Green, X. Hao, *Adv. Sci.* **2021**, *8*, 2004313.

[9] S. Schorr, G. Gurieva, M. Guc, M. Dimitrievska, A. Pérez-Rodríguez, V. Izquierdo-Roca, C. S. Schnohr, J. Kim, W. Jo, J. M. Merino, *J. Phys. Condens Matter* **2019**, *2*, 012002.

[10] M. Chadel, M. M. Bouzaki, A. Chadel, M. Aillerie, B. Benyoucef, *J. Phys. Conf. Ser.* **2017**, *879*, 012006.

[11] S. Zhang, N. D. Pham, T. Tesfamichael, J. Bell, H. Wang, *Sustainable Mater. Technol.* **2018**, *18*, 00078.

[12] Y. Wu, J. Guo, R. Sun, J. Min, *npj Comput. Mater.* **2020**, *6*, 120.

[13] F. Li, X. Peng, Z. Wang, Y. Zhou, Y. Wu, M. Jiang, M. Xu, *Energy Environ. Mater.* **2019**, *2*, 280.

[14] J. Cai, X. Chu, K. Xu, H. Li, J. Wei, *Nanoscale Adv.* **2020**, *2*, 3115.

[15] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T. Y. J. Han, *npj Comput. Mater.* **2022**, *8*, 204.

[16] S. M. Moosavi, K. M. Jablonka, B. Smit, *J. Am. Chem. Soc.* **2020**, *142*, 20273.

[17] J. W. P. Hsu, W. Xu, *Matter* **2022**, *5*, 1334.

[18] Z. Liu, N. Rolston, A. C. Flick, T. W. Colburn, Z. Ren, R. H. Dauskardt, T. Buonassisi, *Joule* **2022**, *6*, 834.

[19] N. Majeed, M. Saladina, M. Krompiec, S. Greedy, C. Deibel, R. C. I. MacKenzie, *Adv. Funct. Mater.* **2020**, *30*, 1907259.

[20] K. Kranthiraja, A. Saeki, *ACS Appl. Mater. Interfaces* **2022**, *14*, 28936.

[21] E. A. J. Abadi, H. Sahu, S. M. Javadpour, M. Goharimanesh, *Mater. Today Energy* **2022**, *25*, 100969.

[22] C. Zhu, W. Liu, Y. Li, X. Huo, H. Li, K. Guo, B. Qiao, S. Zhao, Z. Xu, H. Zhao, D. Song, *Sol. Energy* **2021**, *228*, 45.

[23] X. Li, Z. Hou, S. Gao, Y. Zeng, J. Ao, Z. Zhou, B. Da, W. Liu, Y. Sun, Y. Zhang, *Sol. RRL* **2018**, *2*, 1800198.

[24] O. Al-Sabana, S. O. Abdellatif, *Optoelectron. Lett.* **2022**, *18*, 148.

[25] Z.-W. Zhao, Y. Geng, A. Troisi, H. Ma, *Adv. Intell. Syst.* **2022**, *4*, 2100261.

[26] M. Srivastava, J. M. Howard, T. Gong, M. R. Sousa Dias, M. S. Leite, *J. Phys. Chem. Lett.* **2021**, *12*, 7866.

[27] Q. Tao, P. Xu, M. Li, W. Lu, *npj Comput. Mater.* **2021**, *7*, 23.

[28] V. M. le Corre, T. S. Sherkar, M. Koopmans, L. J. A. Koster, *Cell Rep Phys Sci* **2021**, *2*, 100346.

[29] E. Grau-Luque, I. Anefnaf, N. Benhaddou, R. Fonoll-Rubio, I. Becerril-Romero, S. Aazou, E. Saucedo, Z. Sekkat, A. Perez-Rodriguez, V. Izquierdo-Roca, M. Guc, *J. Mater. Chem. A* **2021**, *9*, 10466.

[30] I. H. Sarker, *SN Comput. Sci.* **2021**, *2*, 160.

[31] W. Hu, R. R. P. Singh, R. T. Scalettar, *Phys. Rev. E* **2017**, *95*, 062122.

[32] R. K. Yadav, P. S. Pawar, K. R. E. Neerugatti, R. Nandi, J. Y. Cho, J. Heo, *Curr. Appl. Phys.* **2021**, *31*, 232.

[33] S. Alhammadi, H. Park, W. K. Kim, *Materials* **2019**, *12*, 1365.

[34] M. A. Shafi, A. Bouich, K. Fradi, J. M. Guaita, L. Khan, B. Mari, *Optik (Stuttg)* **2022**, *258*, 168854.

[35] B. L. Williams, V. Zardetto, B. Kniknie, M. A. Verheijen, W. M. M. Kessels, M. Creatore, *Sol. Energy Mater. Sol. Cells* **2016**, *157*, 798.

**ADVANCED
SCIENCE NEWS**

www.advancedsciencenews.com

**ADVANCED
FUNCTIONAL
MATERIALS**

www.afm-journal.de

[36] M. G. Gang, S. W. Shin, M. P. Suryawanshi, U. V. Ghorpade, Z. Song, J. S. Jang, J. H. Yun, H. Cheong, Y. Yan, J. H. Kim, *J. Phys. Chem. Lett.* **2018**, *9*, 4555.

[37] M. Azzouzi, A. Cabas-Vidani, S. G. Haass, J. A. Röhr, Y. E. Romanyuk, A. N. Tiwari, J. Nelson, *J. Phys. Chem. Lett.* **2019**, *10*, 2829.

[38] J. Li, S. Y. Kim, D. Nam, X. Liu, J. H. Kim, H. Cheong, W. Liu, H. Li, Y. Sun, Y. Zhang, *Sol. Energy Mater. Sol. Cells* **2017**, *159*, 447.

[39] V. C. Karade, E. Choi, M. G. Gang, H. Yoo, A. C. Lokhande, P. T. Babar, J. S. Jang, J. Seidel, J. S. Yun, J. S. Park, J. H. Kim, *ACS Appl. Mater. Interfaces* **2021**, *13*, 429.

[40] V. C. Karade, M. P. Suryawanshi, J. S. Jang, K. S. Gour, S. Jang, J. Park, J. H. Kim, S. W. Shin, *J. Mater. Chem. A* **2022**, *10*, 8466.

[41] A. Pasini, *J. Thorac. Dis.* **2015**, *7*, 953.

[42] M. H. Lee, *Org. Electron.* **2020**, *76*, 105465.

[43] M. H. Lee, *Adv. Intell. Syst.* **2020**, *2*, 1900108.

[44] H. Sahu, H. Ma, *J. Phys. Chem. Lett.* **2019**, *10*, 7277.

[45] H. Sahu, W. Rao, A. Troisi, H. Ma, *Adv. Energy Mater.* **2018**, *8*, 1801032.

[46] J. Im, S. Lee, T. W. Ko, H. W. Kim, Y. K. Hyon, H. Chang, *npj Comput. Mater.* **2019**, *5*, 37.

[47] H. Han, X. Guo, H. Yu, 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 219–224, doi: https://doi.org/10.1109/ICSESS.2016.7883053.

[48] K. S. Gour, V. C. Karade, M. Lee, J. S. Jang, E. Jo, P. Babar, H. Shim, J. S. Yun, J. Park, J. H. Kim, *ACS Appl. Energy Mater.* **2022**, *5*, 2024.

[49] M. C. Baek, J. S. Jang, M. P. Suryawanshi, V. C. Karade, J. Kim, M. He, S. W. Park, J. H. Kim, S. W. Shin, *J. Alloys Compd.* **2023**, *935*, 167993.