

Article

Unraveling the Effect of Compositional Ratios on the Kesterite Thin-Film Solar Cells Using Machine Learning Techniques

Vijay C. Karade ^{1,2}, Santosh S. Sutar ³, Jun Sung Jang ², Kuldeep Singh Gour ⁴, Seung Wook Shin ⁵, Mahesh P. Suryawanshi ⁶, Rajanish K. Kamat ^{7,8}, Tukaram D. Dongale ^{9,*}, Jin Hyeok Kim ^{2,*} and Jae Ho Yun ^{1,*}

- ¹ Department of Energy Engineering, Korea Institute of Energy Technology (KENTECH), Naju 58330, Republic of Korea; vijaykarade@kentech.ac.kr
- ² Optoelectronics Convergence Research Center, Department of Materials Science and Engineering, Chonnam National University, Gwangju 61186, Republic of Korea; jjst115@jnu.ac.kr
- ³ Yashwantrao Chavan School of Rural Development, Shivaji University, Kolhapur 416004, India; sss.ycsrd@unishivaji.ac.in
- ⁴ Surface Engineering Group, Advanced Materials & Processes Division, CSIR-National Metallurgical Laboratory, Jamshedpur 831007, India; gour.ks@nmlindia.org
- ⁵ Future Agricultural Research Division, Rural Research Institute, Korea Rural Community Corporation, Ansan-si 15634, Republic of Korea; swshin1211@gmail.com
- ⁶ School of Photovoltaic and Renewable Energy Engineering, University of New South Wales, Sydney, NSW 2052, Australia; m.suryawanshi@unsw.edu.au
- ⁷ Department of Electronics, Shivaji University, Kolhapur 416004, India; rkk_eln@unishivaji.ac.in
- ⁸ The Institute of Science, Homi Bhabha State University, 15 Madam Cama Road, Mumbai 400032, India
- ⁹ Computational Electronics and Nanoscience Research Laboratory, School of Nanoscience and Biotechnology, Shivaji University, Kolhapur 416004, India
- * Correspondence: tdd.snst@unishivaji.ac.in (T.D.D.); jinhyeok@chonnam.ac.kr (J.H.K.); jhyun@kentech.ac.kr (J.H.Y.)



Citation: Karade, V.C.; Sutar, S.S.; Jang, J.S.; Gour, K.S.; Shin, S.W.; Suryawanshi, M.P.; Kamat, R.K.; Dongale, T.D.; Kim, J.H.; Yun, J.H. Unraveling the Effect of Compositional Ratios on the Kesterite Thin-Film Solar Cells Using Machine Learning Techniques. *Crystals* **2023**, *13*, 1581. <https://doi.org/10.3390/cryst13111581>

Academic Editor: Dezheng Xue

Received: 13 October 2023

Revised: 2 November 2023

Accepted: 9 November 2023

Published: 12 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In the Kesterite family, the $\text{Cu}_2\text{ZnSn}(\text{S,Se})_4$ (CZTSSe) thin-film solar cells (TFSCs) have demonstrated the highest device efficiency with non-stoichiometric cation composition ratios. These composition ratios have a strong influence on the structural, optical, and electrical properties of the CZTSSe absorber layer. So, in this work, a machine learning (ML) approach is employed to evaluate effect composition ratio on the device parameters of CZTSSe TFSCs. In particular, the bi-metallic ratios like Cu/Sn, Zn/Sn, Cu/Zn, and overall Cu/(Zn+Sn) cation composition ratio are investigated. To achieve this, different machine learning algorithms, such as decision trees (DTs) and classification and regression trees (CARTs), are used. In addition, the output performance parameters of CZTSSe TFSCs are predicted by both continuous and categorical approaches. Artificial neural networks (ANN) and XGBoost (XGB) algorithms are employed for the continuous approach. On the other hand, support vector machine and k-nearest neighbor's algorithms are also used for the categorical approach. Through the analysis, it is observed that the DT and CART algorithms provided a critical composition range well suited for the fabrication of highly efficient CZTSSe TFSCs, while the XGB and ANN showed better prediction accuracy among the tested algorithms. The present work offers valuable guidance towards the integration of the ML approach with experimental studies in the field of TFSCs.

Keywords: CZTSSe; thin-film solar cells; machine learning; compositional ratio; prediction

1. Introduction

The earth-abundant element-based kesterite materials $\text{Cu}_2\text{ZnSn}(\text{S,Se})_4$ (CZTSSe) are gaining attention in chalcogenide-based thin-film solar cells (TFSCs) [1–4]. The recent developments in the device fabrication process offered a record power conversion efficiency (PCE) of 14.9% [5], though it is still lower than the other chalcogenide-based TFSCs. The higher open circuit voltage (V_{oc}) loss in the kesterite is one of the prime reasons for

it [6,7]. This high V_{oc} loss originates from potential and bandgap fluctuations, and these fluctuations occur due to the presence of a high density of defects [8–11]. More commonly, the highly efficient CZTSSe devices exhibit non-stoichiometric Zn-rich and Cu-poor conditions [12]. According to the first principle calculations, it is now well established that the precursor composition plays a vital role in the formation of these detrimental antisite defects, their defect clusters, and secondary phases [13–16]. Therefore, to control the intrinsic defect density in the CZTSSe absorber layer, diverse sets of experiments and deep insight analyses need to be performed. However, it requires extensive human effort and can lead to the consumption of large amounts of resources. Consequently, controlling intrinsic defect density without external element doping and the formation of secondary phases with a suitable composition ratio remained a challenge.

Machine learning (ML) has found promising applications in photovoltaic (PV) technology, offering opportunities to enhance PCE, reduce costs, and optimize performance [17,18]. ML algorithms can analyze large datasets of solar cells, such as device parameters, materials properties, and diverse device fabrication conditions, and can identify key factors that affect device performance [17,19]. By uncovering complex patterns and relationships, ML can guide the design and selection of materials in each layer of highly efficient solar cells. Moreover, ML can accelerate the discovery of new materials with desirable properties required for solar cells [20]. By predicting material properties and performance through ML models, researchers can narrow the search for optimal materials, saving time and resources. Similarly, in kesterite-based TFSCs, a large amount of composition data can be analyzed, and a suitable compositional window can be established via the ML technique, saving energy and resources.

In the literature, various ML algorithms have been used for solar cell-based studies [21–23]. Among different ML algorithms, decision trees (DTs), Random Forest (RF), Artificial Neural Networks (ANN), classification and regression trees (CARTs), XGBoost (XGB), k-nearest neighbors (KNN), and support vector machines are the most popular algorithms. They have been widely studied due to their lower complexity, lower computational cost, and good model accuracy. Kumar et al. [24] predicted the bandgap of TiO_2 photoanode in dye-sensitized solar cells using DT, KNN, and RF techniques. Zhu et al. [25] predicted the key governing factors that influence the device performance of CIGS solar cells using RF, GB, and ANN algorithms, and correlation studies. Moreover, different interface passivation materials suitable for perovskite [26], structures/composition analysis of perovskite thin films determining electrical properties [27], and the design of new perovskite materials [28] were also achieved with these ML techniques.

In this work, to understand the effect of cation composition ratios related to Cu, Zn, and Sn on the device parameters of CZTSSe TFSCs, different ML algorithms were used. To predict the device performance, ANN and XGB algorithms were used for the continuous prediction approach. At the same time, the KNN and SVM algorithms were also used to predict the performance with the help of a confusion matrix for the categorical prediction approach. For the present study, we examined data points from 1300 devices. DT and CART approaches provided the best suitable composition windows for the device parameters in highly efficient CZTSSe devices. Moreover, various ANN, XGB, SVM, and KNN algorithms showed their suitability through prediction accuracy during predicting CZTSSe device parameters.

2. Materials and Methods

2.1. Fabrication of CZTSSe TFSCs and Construction of Database

The detailed fabrication workflow of the CZTSSe TFSCs and precursor fabrication process is described in our previous work [19]. The diverse sets of precursor composition ratios were prepared through the deposition of Zn, Sn, and Cu metal precursors sequentially via a DC sputtering system on Mo-coated soda lime glass substrates. Each metal target with high purity produced by Taewon Science Co., Ltd. (i-TASCO), Seoul, Republic of Korea, was used. During the sputtering process, the 30 W power was applied to each target size of

3 inches. Further, based on the preoptimized deposition rate, the deposition time varied from 0.5 h to 1.0 h. For example, to obtain the precursor with different Zn composition ratios, Zn deposition time was achieved while keeping the Cu and Sn deposition time constant. In one batch, a maximum of 9 samples having size $2.5 \times 2.5 \text{ cm}^2$ were coated, which shows less than 1% composition deviation. During the deposition, substrates were constantly rotated at 5 rpm, and base pressure was maintained at around 8 mTorr. The overall precursor thickness of Cu/Sn/Zn layers over the Mo substrate was maintained at nearly 700 nm. Further, the thin films were soft-annealed in an inert (Ar) atmosphere for 0.5 h in the tube furnace and then sulfo-selenized in the rapid thermal annealing system to get the CZTSSe absorber layer. After the successful fabrication of the CZTSSe absorber layer, the n-type CdS buffer layer was deposited via a chemical bath deposition method. Subsequently, a window layer (*i*-ZnO and Al-ZnO(AZO)) was also deposited through RF sputtering. Finally, the Al top grid having an active device area of 0.30 cm^2 was deposited via DC sputtering to complete the device. The final device structure was maintained to SLG/Mo/CZTSSe/CdS/*i*-ZnO/AZO/Al via a device fabrication process established in the laboratory.

2.2. Characterizations

The formation of the CZTSSe phase in fabricated absorber thin films was confirmed through high-resolution X-ray diffraction (XRD, Philips, Amsterdam, The Netherlands). The diffraction measurements were performed from a 2θ value of $5\text{--}80^\circ$ with a step size of 0.5° . The surface morphologies and cross-sectional view of CZTSSe thin films were analyzed via a field emission scanning electron microscope having a resolution of 0.6 nm (FE-SEM, ZEISS Gemini 500 + EDS (Oxford), Jena, Germany). The J-V curves for TFSCs were measured with a solar simulator (Wacom, WXS-155S-L2, Yamaguchi, Japan) at air mass 1.5 G conditions. The external quantum efficiency (EQE) spectra for the device were obtained from 300–1300 nm with a sept size of 10 nm with CEP-25BX (Bunkou Keiki Co., Ltd., CEP-25BX, Tokyo, Japan). An X-ray fluorescence (XRF) (Axios Minerals, Almelo, The Netherlands) was used to determine the composition ratios of the Cu/Sn/Zn precursor thin films operated at 4 KW and calibrated with Internal standard.

2.3. Computational Details

The device parameters like short circuit current (J_{SC}), V_{OC} , fill factor, (FF), PCE, series resistance (R_S), and shunt resistance (R_{SH}) are considered as target/output data, while cation composition ratio was considered input data. Before the DT algorithm use, the target property data were divided into four classes, namely low, medium, high, and very high (Table S1 in Supplementary Materials). On the other hand, the CARTs can handle continuous data; therefore classification of the data was not performed. With DT and CART algorithms, the best possible decision rules were created. Meanwhile, the prediction was performed with ANN, XGB, SVM, and KNN algorithms. The R studio (R version 3.6.2) was used to develop the source code for all ML procedures.

3. Results and Discussion

The CZTSSe absorber material fabricated in this work is a compound semiconductor material, i.e., $\text{Cu}_2\text{ZnSn}(\text{S},\text{Se})_4$ consisting of metallic elements Cu, Zn, and Sn as cations, while S and Se are in an anionic form. It is well established that S and Se and their corresponding anions ratio ($\text{S}/(\text{S}+\text{Se})$ or $\text{Se}/(\text{S}+\text{Se})$) determine the band gap of the prepared CZTSSe absorber layer. Whereas the relative metal cation composition ratio of Cu, Zn, and Sn determines the material's stoichiometry, phase purity, secondary phase formation, and defect density. It eventually affects the material's optical and electronic properties. Here, the $\text{S}/(\text{S}+\text{Se})$ ratio in the CZTSSe absorber layer was fixed to $\sim 1\%$ throughout our device fabrication process as it provides high efficiency, so we performed the targeted study on metal cation composition ratio optimization and its effect. The ML-guided fabrication strategy of high-performance CZTSSe TFSCs involves identifying potential rules and

heuristics based on compositional ratios *viz* Cu/(Zn+Sn), Cu/Sn, Cu/Zn, and Zn/Sn [19]. Thus, the DT and CART algorithms were used to determine the optimal compositional ratios. Since DTs come under a supervised ML algorithm category, they require information about different classes [29]. Accordingly, for PCE, four classes: low PCE (5.01 to 6.53%), medium PCE (6.53 to 7.34%), high PCE (7.34 to 8.16%), and very high PCE (8.16 to 10.24%) were made. The classes for the other device parameters are specified in supporting Table S1. Figure 1 shows the DT model obtained after executing the algorithm for the target PCE property of CZTSSe TFSCs. It is important to note that this DT model is specific to PCE, and similar DT models for other properties like J_{SC} , V_{OC} , FF, R_S , and R_{SH} are shown in the supporting information file (Figures S1–S5 in Supplementary Materials).

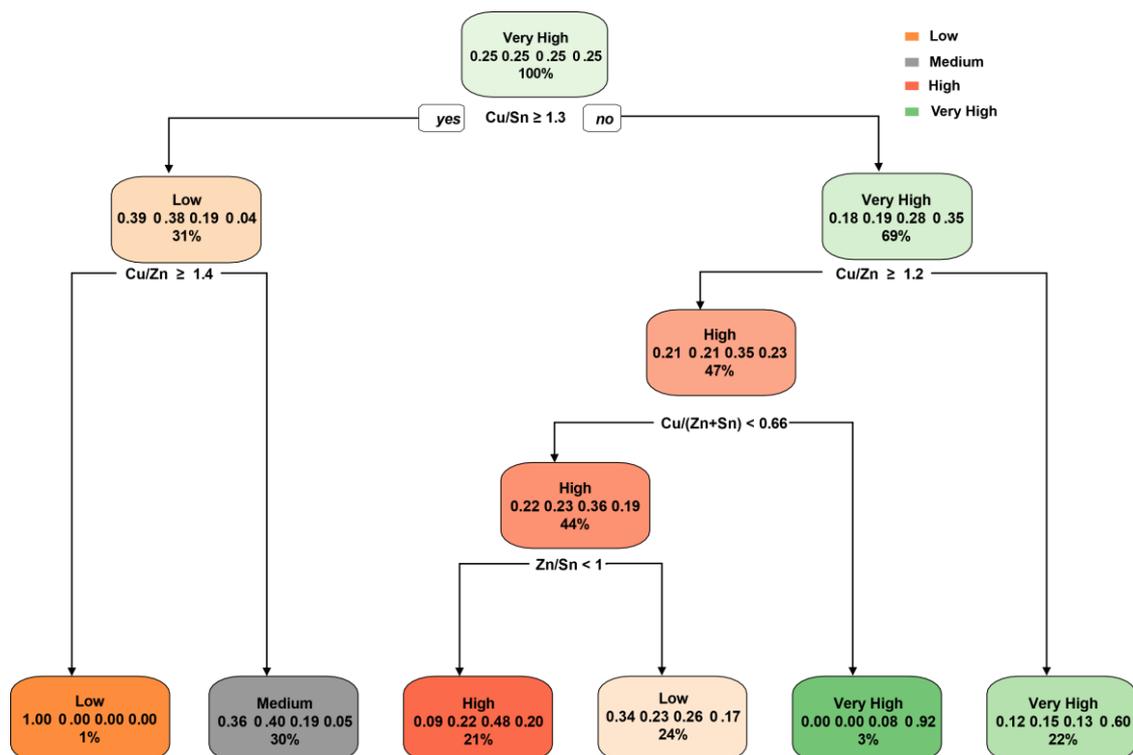


Figure 1. The DT output model aimed with PCE as a target property for CZTSSe TFSCs. Each node in the DT model contains specific details, such as the decision rule associated with that node, the class of the node, the number of observations used, and probabilities of classes. The light saffron, gray, dark saffron, and Green colors shows low, medium, high and very high PCE, respectively.

The sets of rules are generated through DTs (Figure 1) and can be understood as follows. The DT output obtained for PCE as a target property gets divided into two sub-nodes and establishes the first decision rule. (i) If the Cu/Sn ratio is ≥ 1.3 , the PCE of the CZTSSe TFSCs tends to be low, while if it is less than 1.3, then it can show a very high PCE [1,30]. The probability of observing low and very high PCE is 0.39 and 0.35, respectively. The subsequent step involves the second rule: (ii) if the Cu/Zn composition ratio is ≥ 1.4 , it corresponds to a low PCE with a probability of 1; otherwise, a medium PCE can be expected with a probability of 0.40. The third rule in the decision-making process involves the Cu/Zn composition ratio. (iii) If the Cu/Zn ratio is ≥ 1.2 , then the DT predicts a high PCE with a probability of 0.35; otherwise, it anticipates achieving a very high PCE with a probability of 0.60. Further fourth subrule suggests, (iv) if the Cu/(Zn+Sn) ratio is < 0.66 , a high PCE can be observed with a probability of 0.36. On the other hand, if it is > 0.66 , a very high PCE can be obtained for the CZTSSe TFSCs with a probability of 0.92. Finally, the fifth subrule suggests, that (v) if the Zn/Sn ratio is < 1 , a high PCE may be achieved with 0.48 of probability. Otherwise, a lower PCE can be obtained with 0.34 of probability. The light saffron (low), gray (medium), dark saffron (high), and Green (very high) colors

shows PCE trend in DT. Similar rules can be evaluated using the same methodology for the other DT models (FF, J_{SC} , V_{OC} , R_s , and R_{sh} ; Figures S1–S5, respectively).

In the subsequent phase of our work, a decision model for CZTSSe TFSCs using the CART algorithm was constructed as shown in Figure 2. The CART is a supervised ML algorithm, primarily utilized for both classification and regression tasks. So, the CART algorithm is used to predict and classify the values of continuous properties based on input features. At each node, the algorithm selects the best feature and corresponding threshold to partition the data into homogeneous subgroups with respect to the target properties. Following the execution of the CART model it shows a division of the root node into two segments. The corresponding CART rules can be interpreted as: (i) when the Cu/Sn ratio is ≥ 1.3 (Yes), an average PCE of 6.7 can be achieved, otherwise, 7.6% PCE can be achieved. (ii) If the Cu/Zn ratio is ≥ 1.2 (given that Cu/Sn ≥ 1.3), 6.5% of average PCE may achieved or a relatively higher PCE of 7.0% may achieved. Other than the Cu/Sn and Cu/Zn ratio, further critical composition ratio can be determined with other subrules: (iii) if the Cu/Sn is not ≥ 1.3 (No), then 7.4% and 8.2% PCE can be achieved with Cu/Zn ≥ 1.2 and Cu/Zn < 1.2 , respectively. (iv) If the Zn/Sn is ≥ 1 and Cu/Zn is < 1.2 , then a maximum PCE of $\sim 7\%$ can be achieved, while 7.8% can be achieved with Cu/Sn > 1.3 . This subrule-iv has a very low-node probability ($< 10\%$), thus this class of nodes can be ignored while making critical decisions. Similarly, it can be applied to other nodes too. (v) If the Zn/Sn < 1 , Cu/(Zn+Sn) < 0.66 , and Cu/Sn > 1.2 then 7.6% PCE can be obtained. (vi) If the Cu/Zn < 1.2 , Zn/Sn > 1.1 , and Zn/Sn ≥ 1.1 then 7.5% of PCE can be achieved or 8.8% can be obtained. Briefly, more than 8% of PCE can be reached in the CZTSSe TFSCs with the following composition ratios: Cu/Zn < 1.2 , Zn/Sn ~ 1.0 – 1.1 . From the light to dark blue color shows increaseing PCE trend in CART. Similar procedures can be employed to understand subsequent rules. The CART models for other properties like FF, J_{SC} , V_{OC} , R_s , and R_{sh} are shown in the supporting information file (Figures S6–S10 in Supplementary Materials). Overall, both the DT and CART output models for PCE as a target property suggest that the Cu/Sn ratio higher than 1.3 and Cu/Zn higher than 1.2 are not helpful in breaking the saturated PCE of CZTSSe TFSCs. At the same time, slightly poor ratio conditions provide high PCE. Similar ratios also determine the other device parameters. The Zn/Sn ratio of about ~ 1.0 was optimal in all scenarios with high PCE and R_{sh} . Interestingly, the overall metal composition ratio, i.e., Cu/(Zn+Sn), has minimal impact on PCE, while the binary intermetallic composition ratios such as Cu/Zn, Cu/Sn, and Zn/Sn are more sensitive to the PCE and other device parameters.

In the next stage, ANN, XGB, KNN, and SVM algorithms were used to predict the device parameters of CZTSSe TFSCs depending on compositional ratios. In particular, the ANN and XGB ML algorithms were used to predict the continuous output properties. On the other hand, the SVM and KNN algorithms were used to predict the categorical output properties. At first, we applied the ANN algorithm to the relevant data and determined the predicted values. The adjusted R^2 (Adj. R^2) values for each ANN model were obtained through linear fitting to the experimental and predicted data. The predicted and experimental values of CZTSSe TFSCs based on the ANN model are shown in Figure 3. The detailed ANN structure of the various device parameters is shown in Figure S11. The linear fitting to the scatter plots exhibited adjusted R^2 values below 0.27. It indicates that the performance of the ANN models in prediction was unsatisfactory. The ANN model's inadequate predictive performance can be ascribed to the extremely heterogeneous dataset. In response, the XGB algorithm was employed to enhance predictions and rectify the unsatisfactory performance exhibited by the ANN algorithm. The prediction results of the PCE, FF, J_{SC} , V_{OC} , R_s , and R_{sh} obtained using the XGB algorithm are depicted in Figure 4. The adj. R^2 values for each XGB model were better than the ANN models. It can be observed that the XGB predicts the device parameters better than the ANN.

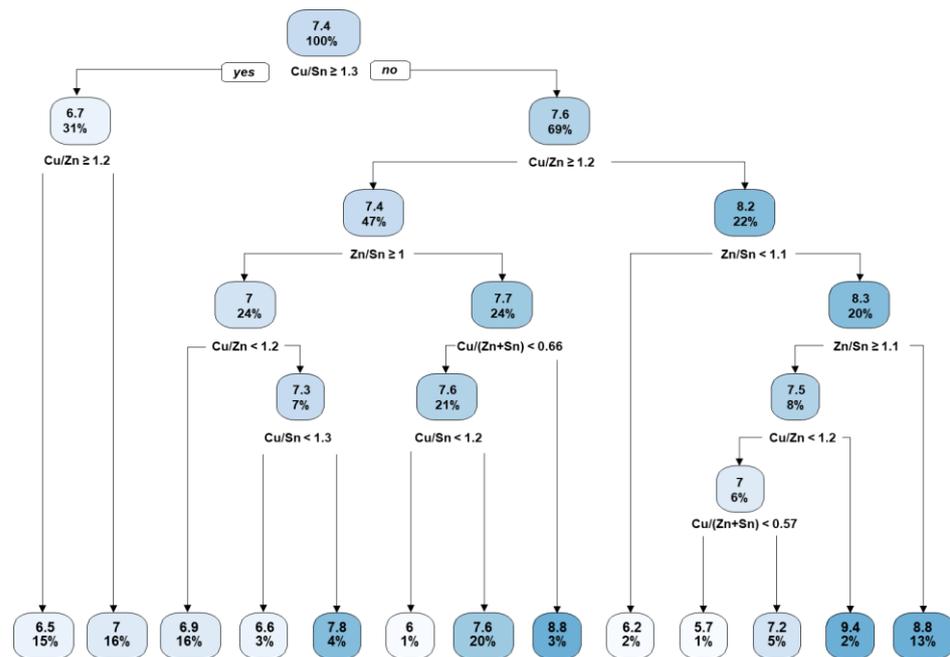


Figure 2. The CART output model aimed with PCE as a target property for CZTSSe TFSCs. Each node in the CART model consists of a node, and the average target value for each node and the percentage of data follows the rule of that node. The light blue and dark blue color shows lower and higher PCE in CART.

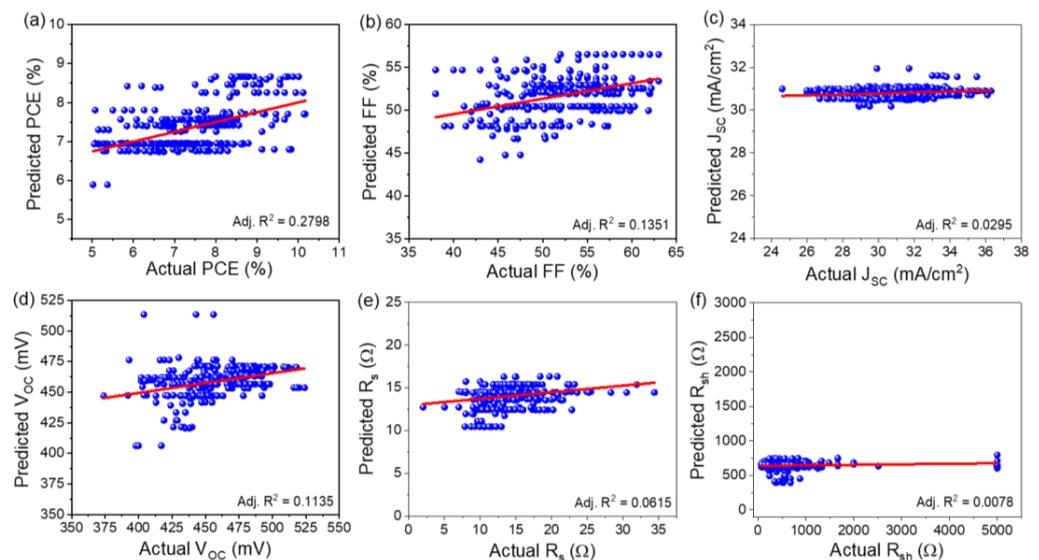


Figure 3. ANN prediction of (a) PCE, (b) FF, (c) J_{SC} , (d) V_{OC} , (e) R_s , and (f) R_{sh} of the CZTSSe TFSCs for compositional ratios.

The predictions of output performance parameters during the continuous approach are not satisfactory. Thus, we grouped the PCE, FF, J_{SC} , V_{OC} , R_s , and R_{sh} of CZTSSe TFSCs into four classes (1 to 4) and tried to predict these properties using SVM and KNN algorithms based on a categorical approach. The SVM and KNN are supervised ML algorithms, and they are generally used for classification tasks [31]. Figure 5 represents the SVM-based confusion matrix of CZTSSe TFSCs. For this, compositional ratios were taken as input parameters. In this case, the accuracy of the PCE, FF, J_{SC} , V_{OC} , R_s , and R_{sh} confusion matrix is found to be 37.44%, 33.42%, 32.91%, 42.21%, 33.17%, and 21.36%, respectively. Figure 6 represents the KNN confusion matrix of CZTSSe TFSCs. For KNN, the accuracy of the PCE, FF, J_{SC} , V_{OC} , R_s , and R_{sh} -based confusion matrix was found to be 46.79%, 45.28%, 47.92%,

54.15%, 43.58%, and 34.90%, respectively. These results suggested that the prediction results of the KNN algorithm are better than the SVM algorithm. Overall, the KNN outperforms the SVM in predicting the categorical output properties of CZTSSe TFSCs. Alternatively, the XGB predicts the device parameters better than the ANN.

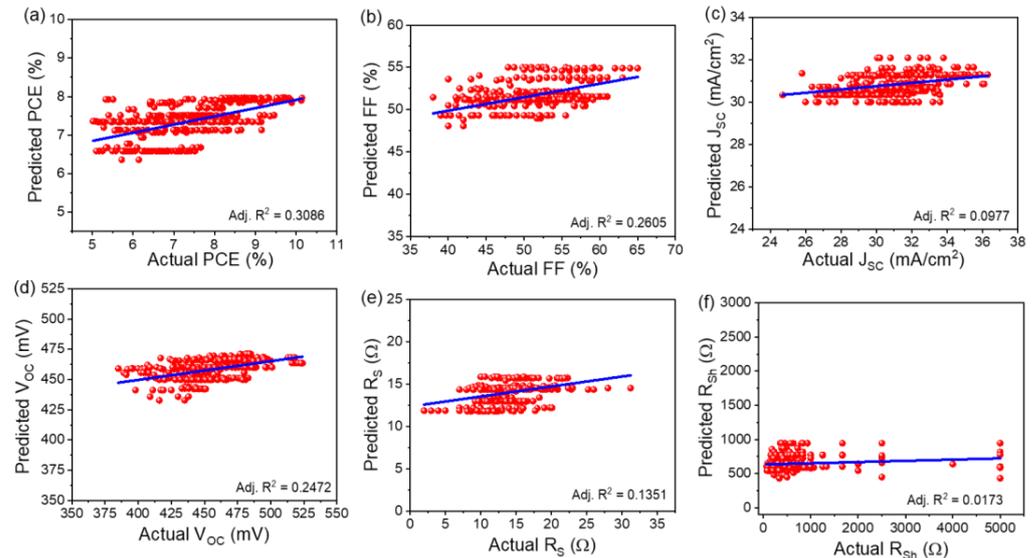


Figure 4. XGB prediction of (a) PCE, (b) FF, (c) J_{sc}, (d) V_{oc}, (e) R_s, and (f) R_{sh} of the CZTSSe TFSCs based on compositional ratios.

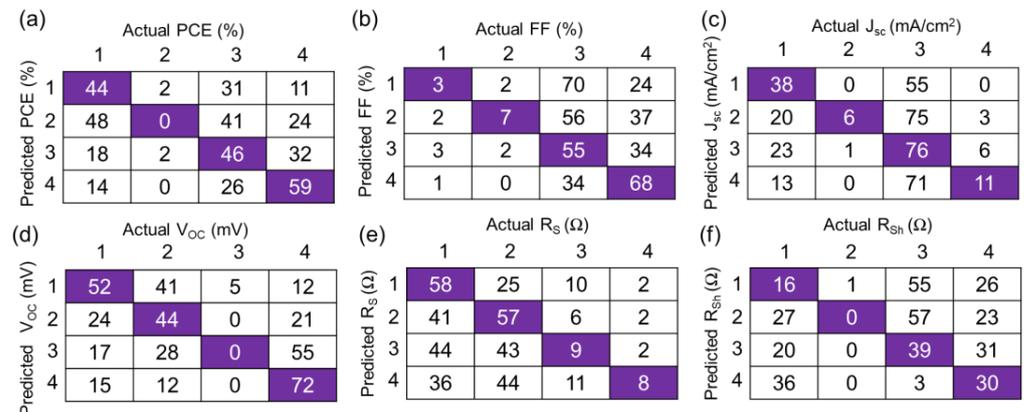


Figure 5. SVM-based confusion matrix of the (a) PCE, (b) FF, (c) J_{sc}, (d) V_{oc}, (e) R_s, and (f) R_{sh} of the CZTSSe TFSCs. Classes: 1—Low, 2—Medium, 3—High, 4—Very High.

Based on the DT and CART rule, the CZTSSe device was fabricated with an optimal composition ratio of Cu/Zn = 1.15–1.20, Cu/Sn = 1.25–1.30, Zn/Sn = 0.95–1.0, and Cu/(Zn+Sn) = 0.57–0.62. The crystal structure and morphology of corresponding representative CZTSSe thin films were analyzed via XRD and FESEM, respectively (Figure S12 in Supplementary Materials). The CZTSSe thin film showed the formation of crystalline thin films with characteristic CZTSSe peaks around 2θ value of 27.27°, 45.31°, and 53.69° corresponding to the (112), (204), and (312) planes, respectively (Figure S12a). The FE-SEM analysis showed the formation of compact and large grains with an absorber thickness of ~1.65 μm (Figure S12b). Figure 7a shows the current-voltage (J-V) characteristics measured under standard AM 1.5 G illumination test conditions. The device fabricated with a ML-optimized composition range exhibited the average device PCE of 8.61% with J_{sc}, V_{oc}, and FF, ~33.61 mA/cm², ~479 mV, and ~53.51%, respectively (Table S2), while the best sample exhibited 8.89% of PCE. The external quantum efficiency measured for the best device exhibited nearly 80% photoresponse in the visible region, as shown in Figure 7b. In addition, the bandgap values estimated for a similar device were ~1.09 eV. The present

investigation showed DT and CART rules determined from ML well matched with the practical investigation.

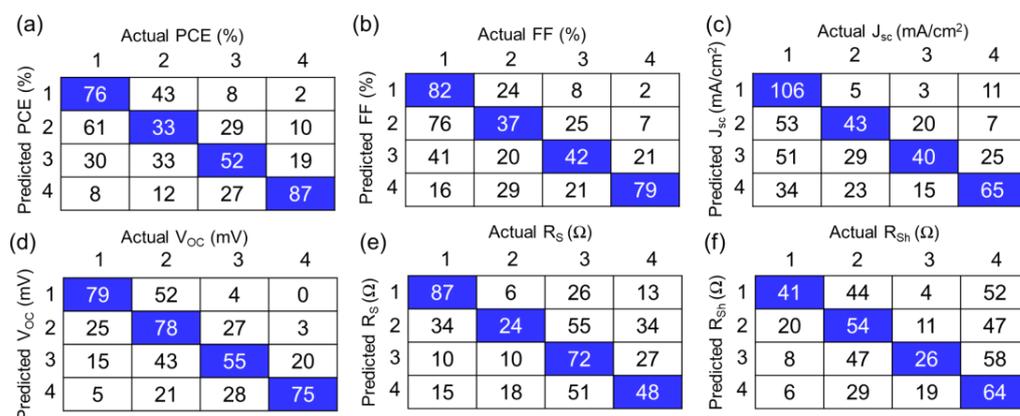


Figure 6. KNN-based confusion matrix of the (a) PCE, (b) FF, (c) J_{sc}, (d) V_{oc}, (e) R_s, and (f) R_{sh} of the CZTSSe TFSCs. Classes: 1—Low, 2—Medium, 3—High, 4—Very High.

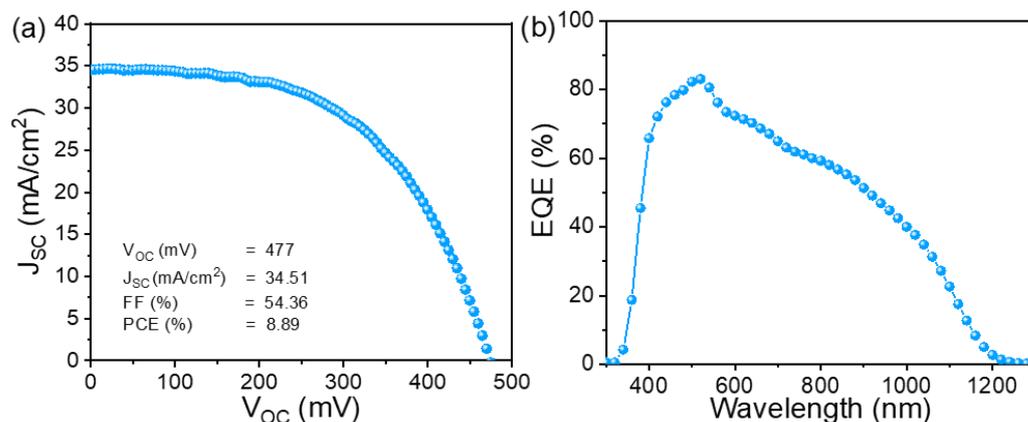


Figure 7. (a) J-V characteristics, and (b) EQE response of the CZTSSe TFSC.

Overall, it can be realized that the integration of the ML approach with experimental studies can provide critical composition windows that could assist the fabrication of high-efficiency solar cells. Moreover, it also revealed the composition windows where the device performance could decrease. Yet, the ML algorithm has its strengths and weaknesses, such as continuity in input data which limits the algorithm selection, the number of observations that decides the model reliability, and prediction accuracy counts for output result validation. Therefore, it is necessary to cautiously analyze the specific data requirements for a specific algorithm to fully explore the strength of ML in TFSCs.

4. Conclusions

In summary, various ML algorithms were utilized to understand the effect of compositional ratios (Cu/Zn, Cu/Sn, Cu/Zn+Sn, and Zn/Sn) on the CZTSSe TFSCs output device parameters (PCE, FF, J_{sc}, V_{oc}, R_s, and R_{sh}). The DT and CART algorithms provided hidden sets of rules for the fabrication of highly efficient CZTSSe TFSCs. It was observed that the bi-metallic composition ratios, such as Cu/Zn, Cu/Sn, and Zn/Sn, are more sensitive towards device properties, while overall Cu/(Zn+Sn) is less. The Cu/Zn and Cu/Sn ratios > 1.2 (i.e., Cu-rich conditions) were found to be unfavorable during achieving high PCE. In contrast, the Zn/Sn ratio ~1 was optimal. Further, the output performance of the CZTSSe TFSCs was predicted using continuous and categorical prediction approaches. In the case of the continuous approach, the XGB algorithm shows a better prediction of the device parameters than the ANN algorithm; on the other hand, the KNN predicts the categorical output properties better than the SVM. Among all tested prediction-based

algorithms, the KNN provides better results than other algorithms. The present work provides guidelines for the possible application of different algorithms in the ML-integrated fabrication of TFSCs and insights to improve the PCE of devices.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cryst13111581/s1>, Table S1. Categories of the different device parameters from the $\text{Cu}_2\text{ZnSn}(\text{S,Se})_4$ (CZTSSe) TFSCs; Figure S1: DT model for the FF of the CZTSSe TFSCs; Figure S2: DT model for the J_{SC} of the CZTSSe TFSCs; Figure S3: DT model for the V_{OC} of the CZTSSe TFSCs; Figure S4: DT model for the R_{S} of the CZTSSe TFSCs; Figure S5: DT model for the R_{Sh} of the CZTSSe TFSCs; Figure S6: CART model for the FF of the CZTSSe TFSCs; Figure S7: CART model for the J_{SC} of the CZTSSe TFSCs; Figure S8: CART model for the V_{OC} of the CZTSSe TFSCs; Figure S9: CART model for the R_{S} of the CZTSSe TFSCs; Figure S10: CART model of the R_{Sh} of the CZTSSe TFSCs; Figure S11: ANN structure related to (a) PCE, (b) FF, (c) J_{SC} , (d) V_{OC} , (e) R_{S} , and (f) R_{Sh} of the CZTSSe TFSCs; Figure S12: (a) XRD of CZTSSe absorber thin film, (b) FESEM surface morphology, inset cross-sectional view; Table S2: Experimental photovoltaic parameters of the CZTSSe TFSCs under standard test conditions.

Author Contributions: Conceptualization, methodology, formal analysis, investigation, writing—original draft preparation, V.C.K., J.S.J. and S.S.S.; data curation, M.P.S. and S.W.S.; writing—review and editing, K.S.G.; supervision, T.D.D., R.K.K., J.H.K. and J.H.Y.; project administration, J.H.K. and J.H.Y.; funding acquisition, J.H.K. and J.H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Research Foundation of Korea (NRF) (grant no. 2023R1A2C2003612) and was financially supported by the Korea Institute of Energy Research (KIER) (grant no. C2-2403). Santosh S. Sutar is thankful to Shivaji University, Kolhapur, MS, India, for providing the research grant through the “Research Strengthening Scheme” (Ref: SU/C&U.D. Section/3/888(A)). This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (2018R1A6A1A03024334).

Data Availability Statement: Data can be made available upon reasonable request.

Acknowledgments: The authors would like to thank Korean Basic Science Institute (KBSI) for the SEM and XRD analysis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, J.; Xu, X.; Wu, H.; Wang, J.; Lou, L.; Yin, K.; Gong, Y.; Shi, J.; Luo, Y.; Li, D.; et al. Control of the Phase Evolution of Kesterite by Tuning of the Selenium Partial Pressure for Solar Cells with 13.8% Certified Efficiency. *Nat. Energy* **2023**, *8*, 526–535. [[CrossRef](#)]
2. Li, Y.; Wei, H.; Cui, C.; Wang, X.; Shao, Z.; Pang, S.; Cui, G. CZTSSe Solar Cells: Insights into Interface Engineering. *J. Mater. Chem. A Mater.* **2023**, *11*, 4836–4849. [[CrossRef](#)]
3. Li, J.; Huang, J.; Ma, F.; Sun, H.; Cong, J.; Privat, K.; Webster, R.F.; Cheong, S.; Yao, Y.; Chin, R.L.; et al. Unveiling Microscopic Carrier Loss Mechanisms in 12% Efficient $\text{Cu}_2\text{ZnSnSe}_4$ Solar Cells. *Nat. Energy* **2022**, *7*, 754–764. [[CrossRef](#)]
4. Park, H.K.; Cho, Y.; Kim, J.; Kim, S.; Kim, S.; Kim, J.; Yang, K.J.; Kim, D.H.; Kang, J.K.; Jo, W. Flexible Kesterite Thin-Film Solar Cells under Stress. *npj Flex. Electron.* **2022**, *6*, 91. [[CrossRef](#)]
5. National Renewable Energy Laboratory-Best Research-Cell Efficiency Chart. Available online: <https://www.nrel.gov/pv/cell-efficiency.html> (accessed on 17 October 2023).
6. Shockley, W.; Queisser, H.J. Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *J. Appl. Phys.* **1961**, *32*, 510–519. [[CrossRef](#)]
7. Di Bartolomeo, A.; Goubard, F.; Boerasu, I.; Stefan Vasile, B. Current Status of the Open-Circuit Voltage of Kesterite CZTS Absorber Layers for Photovoltaic Applications—Part I, a Review. *Materials* **2022**, *15*, 8427. [[CrossRef](#)]
8. Gong, Y.; Zhang, Y.; Zhu, Q.; Zhou, Y.; Qiu, R.; Niu, C.; Yan, W.; Huang, W.; Xin, H. Identifying the Origin of the Voc Deficit of Kesterite Solar Cells from the Two Grain Growth Mechanisms Induced by Sn^{2+} and Sn^{4+} Precursors in DMSO Solution. *Energy Environ. Sci.* **2021**, *14*, 2369–2380. [[CrossRef](#)]
9. Azzouzi, M.; Cabas-Vidani, A.; Haass, S.G.; Röhr, J.A.; Romanyuk, Y.E.; Tiwari, A.N.; Nelson, J. Analysis of the Voltage Losses in CZTSSe Solar Cells of Varying Sn Content. *J. Phys. Chem. Lett.* **2019**, *10*, 2829–2835. [[CrossRef](#)]
10. Wei, H.; Li, Y.; Cui, C.; Wang, X.; Shao, Z.; Pang, S.; Cui, G. Defect Suppression for High-Efficiency Kesterite CZTSSe Solar Cells: Advances and Prospects. *Chem. Eng. J.* **2023**, *462*, 142121. [[CrossRef](#)]

11. Liu, F.; Wu, S.; Zhang, Y.; Hao, X.; Ding, L. Advances in Kesterite $\text{Cu}_2\text{ZnSn}(\text{S}, \text{Se})_4$ Solar Cells. *Sci. Bull.* **2020**, *65*, 698–701. [[CrossRef](#)]
12. Guo, J.; Ao, J.; Zhang, Y. A Critical Review on Rational Composition Engineering in Kesterite Photovoltaic Devices: Self-Regulation and Mutual Synergy. *J. Mater. Chem. A* **2023**, *11*, 16494–16518. [[CrossRef](#)]
13. Kumar, M.; Dubey, A.; Adhikari, N.; Venkatesan, S.; Qiao, Q. Strategic Review of Secondary Phases, Defects and Defect-Complexes in Kesterite CZTS–Se Solar Cells. *Energy Env. Sci.* **2015**, *8*, 3134–3159. [[CrossRef](#)]
14. Schorr, S.; Gurieva, G.; Guc, M.; Dimitrievska, M.; Pérez-Rodríguez, A.; Izquierdo-Roca, V.; Schnohr, C.S.; Kim, J.; Jo, W.; Merino, J.M. Point Defects, Compositional Fluctuations, and Secondary Phases in Non-Stoichiometric Kesterites. *J. Phys. Energy* **2020**, *2*, 012002. [[CrossRef](#)]
15. Maeda, T.; Nakamura, S.; Wada, T. First Principles Calculations of Defect Formation in In-Free Photovoltaic Semiconductors $\text{Cu}_2\text{ZnSnS}_4$ and $\text{Cu}_2\text{ZnSnSe}_4$. *Jpn. J. Appl. Phys.* **2011**, *50*, 04DP07. [[CrossRef](#)]
16. Xu, P.; Chen, S.; Huang, B.; Xiang, H.J.; Gong, X.G.; Wei, S.H. Stability and Electronic Structure of $\text{Cu}_2\text{ZnSnS}_4$ Surfaces: First-Principles Study. *Phys. Rev. B Condens. Matter Mater. Phys.* **2013**, *88*, 045427. [[CrossRef](#)]
17. Liu, Y.; Tan, X.; Liang, J.; Han, H.; Xiang, P.; Yan, W. Machine Learning for Perovskite Solar Cells and Component Materials: Key Technologies and Prospects. *Adv. Funct. Mater.* **2023**, *33*, 2214271. [[CrossRef](#)]
18. Yilmaz, B.; Yıldırım, R. Critical Review of Machine Learning Applications in Perovskite Solar Research. *Nano Energy* **2021**, *80*, 105546. [[CrossRef](#)]
19. Karade, V.C.; Sutar, S.S.; Shin, S.W.; Suryawanshi, M.P.; Jang, J.S.; Gour, K.S.; Kamat, R.K.; Yun, J.H.; Dongale, T.D.; Kim, J.H. Machine Learning Assisted Analysis, Prediction, and Fabrication of High-Efficiency CZTSSe Thin Film Solar Cells. *Adv. Funct. Mater.* **2023**, *33*, 202303459. [[CrossRef](#)]
20. Wu, Y.; Guo, J.; Sun, R.; Min, J. Machine Learning for Accelerating the Discovery of High-Performance Donor/Acceptor Pairs in Non-Fullerene Organic Solar Cells. *npj Comput. Mater.* **2020**, *6*, 120. [[CrossRef](#)]
21. Mahmood, A.; Wang, J.-L. Machine Learning for High Performance Organic Solar Cells: Current Scenario and Future Prospects. *Energy Env. Sci.* **2021**, *14*, 90–105. [[CrossRef](#)]
22. Li, F.; Peng, X.; Wang, Z.; Zhou, Y.; Wu, Y.; Jiang, M.; Xu, M. Machine Learning (ML)-Assisted Design and Fabrication for Solar Cells. *Energy Environ. Mater.* **2019**, *2*, 280–291. [[CrossRef](#)]
23. Malhotra, P.; Khandelwal, K.; Biswas, S.; Chen, F.-C.; Sharma, G.D. Opportunities and Challenges for Machine Learning to Select Combination of Donor and Acceptor Materials for Efficient Organic Solar Cells. *J. Mater. Chem. C Mater.* **2022**, *10*, 17781–17811. [[CrossRef](#)]
24. Kumar, C.; Patra, S.N. Prediction of Bandgap of Undoped TiO_2 for Dye-Sensitized Solar Cell Photoanode. *Appl. Sol. Energy* **2022**, *58*, 482–489. [[CrossRef](#)]
25. Zhu, C.; Liu, W.; Li, Y.; Huo, X.; Li, H.; Guo, K.; Qiao, B.; Zhao, S.; Xu, Z.; Zhao, H.; et al. Key Factors Governing the Device Performance of CIGS Solar Cells: Insights from Machine Learning. *Sol. Energy* **2021**, *228*, 45–52. [[CrossRef](#)]
26. Liu, W.; Lu, Y.; Wei, D.; Huo, X.; Huang, X.; Li, Y.; Meng, J.; Zhao, S.; Qiao, B.; Liang, Z.; et al. Screening Interface Passivation Materials Intelligently through Machine Learning for Highly Efficient Perovskite Solar Cells. *J. Mater. Chem. A* **2022**, *10*, 17782–17789. [[CrossRef](#)]
27. Priya, P.; Aluru, N.R. Accelerated Design and Discovery of Perovskites with High Conductivity for Energy Applications through Machine Learning. *npj Comput. Mater.* **2021**, *7*, 90. [[CrossRef](#)]
28. Li, Y.; Lu, Y.; Huo, X.; Wei, D.; Meng, J.; Dong, J.; Qiao, B.; Zhao, S.; Xu, Z.; Song, D. Bandgap Tuning Strategy by Cations and Halide Ions of Lead Halide Perovskites Learned from Machine Learning. *RSC Adv.* **2021**, *11*, 15688–15694. [[CrossRef](#)]
29. Omer, Z.M.; Shareef, H. Comparison of Decision Tree Based Ensemble Methods for Prediction of Photovoltaic Maximum Current. *Energy Convers. Manag. X* **2022**, *16*, 100333. [[CrossRef](#)]
30. Lou, L.; Wang, J.; Yin, K.; Meng, F.; Xu, X.; Zhou, J.; Wu, H.; Shi, J.; Luo, Y.; Li, D.; et al. Crown Ether-Assisted Colloidal ZnO Window Layer Engineering for Efficient Kesterite $(\text{Ag,Cu})_2\text{ZnSn}(\text{S,Se})_4$ Solar Cells. *ACS Energy Lett.* **2023**, *8*, 3775–3783. [[CrossRef](#)]
31. Demidova, L.A. Two-Stage Hybrid Data Classifiers Based on SVM and KNN Algorithms. *Symmetry* **2021**, *13*, 615. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.