## PAPER

Check for updates

# Predicting the photon energy of quasi-2D lead halide perovskites from the precursor composition through machine learning†

Wei Wang,‡[ab] Yueqiao Li,‡[ab] Ang Zou,[ab] Haochen Shi,[ab] Xiaofeng Huang,[ab] Yaoyao Li,[ab] Dong Wei,[c] Bo Qiao, [iD][ab] Suling Zhao, [iD][ab] Zheng Xu[ab] and Dandan Song [iD] *[ab]

Quasi-2D perovskites with the general formula of $L_2A_{n-1}Pb_nX_{3n+1}$ (L = organic spacer cation, A = small organic cation or inorganic cation, X = halide ion, and $n \leq 5$) are an emerging kind of luminescent material. Their emission color can be easily tuned by their composition and $n$ value. Accurate prediction of the photon energy before experiments is essential but unpractical based on present studies. Herein, we use machine learning (ML) to explore the quantitative relationship between the photon energies of quasi-2D perovskite materials and their precursor compositions. The random forest (RF) model presents high accuracy in prediction with a root mean square error (RMSE) of ~0.05 eV on a test set. By feature importance analysis, the composition of the A-site cation is found to be a critical factor affecting the photon energy. Moreover, it is also found that the phase impurity greatly lowers the photon energy and needs to be minimized. Furthermore, the RF model predicts the compositions of quasi-2D perovskites with high photon energies for blue emission. These results highlight the advantage of machine learning in predicting the properties of quasi-2D perovskites before experiments and also providing color tuning directions for experiments.

## 1. Introduction

Lead halide perovskites are a kind of miraculous material, which delivers unique performance in optoelectronic devices.[1–4] Quasi-2D perovskites with the general formula of $L_2A_{n-1}Pb_nX_{3n+1}$ (L = organic spacer cation, A = organic or inorganic cation, X = halide ion, and $n \leq 5$), exhibit higher material stability compared to their 3D counterparts.[5–8] Color tuning is essential to their application in light-emitting diodes (LEDs) and other optoelectronic devices. In particular, pushing the photon energy of the quasi-2D perovskites to the blue scale is gaining increasing importance for LEDs.

The photon energy of quasi-2D perovskites can be tuned from deep red to blue by adjusting halide ions and the $n$ value. Their physical origins are well explored by previous work,[9–17] which provides general directions for color tuning. However, a large number of trial-and-error experiments are still needed to

fabricate perovskites with the desired emission color, which requires lots of materials and workforce. Hence, it is critical to accurately predict the photon energy before carrying out experiments which will significantly accelerate the development of quasi-2D perovskites with expected emission colors.

The machine-learning (ML) approach is a scientific model that can efficiently learn from existing results and is gaining increasing attention in material exploration.[18–22] With the assistance of ML, researchers can explore a large number of new materials (such as lead-free perovskites),[23,24] develop efficient solar cells,[19,21,22,27] *etc.* Previously, using ML algorithms, we successfully predicted the bandgap of 3D lead halide perovskites from their compositions and proposed possible compositions of mixed halide perovskites, which can be used in tandem solar cells.[24] Marchenko *et al.* established a database of 2D perovskites; they employed ML algorithms to predict the bandgap of 2D perovskites from their compositions, $n$ values, crystal structures, and so on.[25] The following work by Wan *et al.* utilized this database and ML algorithms to accurately predict the bandgap of 2D perovskite through molecular graphic descriptors.[26] These pioneering studies reveal the power of ML in exploring material and device properties. Here, it is also possible to explore the quantitative relationship between the photon energy of quasi-2D perovskites and the governing factors.

[a]Key Laboratory of Luminescence and Optical Information, Beijing Jiaotong University, Ministry of Education, Beijing 100044, China. E-mail: ddsong@bjtu.edu.cn

[b]Institute of Optoelectronics Technology, Beijing Jiaotong University, Beijing 100044, China

[c]College of Physics and Energy, Fujian Normal University, Fuzhou, 350117, China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d2na00052k

‡ Wei Wang and Yueqiao Li contributed equally to this work.

Hence, in this work, the ML approach is employed to predict the photon energies of quasi-2D perovskites. To make the prediction more referable for experiments, we use the precursor compositions instead of the calculated material compositions as the input features and identify the relationship between the precursor compositions and the resultant photon energies. A dataset was established by collecting the reported experimental data, which covers a large range of quasi-2D perovskites, including a variety of large organic cations and small cations. In particular, a series of quasi-2D perovskites with photon energies exceeding 2.58 eV for blue LEDs are predicted, which provide essential guidance for experimental screening.

## 2. Results and discussion

The type of organic spacer cation, A-site cation ratio (cesium (Cs), formamidinium (FA), methylammonium (MA)), halide ion ratio, molar ratio of $Pb^{2+}$ to the organic spacer cation (abbreviated as P2L) and molar ratio of $Pb^{2+}$ to the organic cation (abbreviated as P2A) in the precursor solution are taken into account. Here, we use the ion ratios in the precursor solution instead of the $n$ value because the molar ratios utilized in the precursor sometimes differ from the stoichiometric ratio provided by their formula of $L_2A_{n-1}Pb_nX_{3n+1}$. We employ the ion ratios in the precursor solution since they are significantly more straightforward in the experimental design. As the green to blue color range is particularly important for luminous materials, we exclusively analyze Br-/Cl-based quasi-2D perovskites and exclude I-based perovskites due to the narrow bandgaps of I-based perovskites, which make them suitable for green to red emission. 6 input features were screened for ML: the Cs/(Cs + FA + MA) ratio, the FA/(Cs + FA + MA) ratio, the Br/(Br + Cl) ratio, and P2L, P2A and XLogP3 values of the bromide salt based on the organic spacer cation. XLogP3 predicts the log $P$ value of a query compound by starting with the known log $P$ value of a reference compound,[28] whereas log $P$ is a crucial parameter for characterizing lipophilicity. A higher XLogP3 value makes the compound more difficult to dissolve in a polar solvent (usually dimethyl sulfoxide (DMSO) and/or dimethylformamide (DMF) for quasi-2D perovskite precursors). XLogP3 values were obtained from a website.[29] The photon energy, calculated from the emission peak wavelength (most of them are electroluminescence wavelengths), represents the output performance. We didn't use the dominant absorption peak to calculate the photon energy, as the absorption energy is typically larger than the photon energy, i.e., the Stokes shift.

To build a dataset, we searched for studies that reported the photon energy of quasi-2D perovskites and collected approximately 300 data points. The dataset was then cleaned using the following rules: (1) data points with P2A larger than 2.5 ($n$ value larger than 5) were removed, as they generally reflect the properties of 3D perovskites; (2) data points with missing data for one or more input features were removed; (3) duplicate data points from different literature studies were removed; (4) for the data points with same input features but different photon energies, we reserved the data point with the reported largest value. For $PEA_2Cs_{n-1}Pb_nBr_{3n+1}$ with P2L = 2.5, for example, we

reserved the data point with the highest energy of 2.451 eV. As a result, we obtained 106 data points, listed in Table S1.† These data points include Cl-, Cl/Br mixed, Br-based perovskites with different organic spacer cations and A-site cations (Cs, FA, and MA). The maximum photon energy of quasi-2D perovskites is 2.840 eV, and the smallest value is 2.280 eV.

Firstly, we use a correlation matrix to learn the linear correlations between the features and the photon energy. As displayed in Fig. 1, the photon energy (abbreviated as PE in Fig. 1) of the perovskites has a non-negligible correlation with the halide anions, the cations, and the ratios of the organic spacer cation and small cation to $Pb^{2+}$. As expected, the photon energy shows a positive correlation with the Cl ratio (=Cl/(Br + Cl), abbreviated as Cl) and a negative correlation with the Br ratio (=Br/(Br + Cl), abbreviated as Br). Among the three types of A-site cations (MA, FA, and Cs), the photon energy exhibits a negative correlation with the FA ratio (=FA/(Cs + FA + MA), abbreviated as FA), while it is positively changed with the Cs ratio (=Cs/(Cs + FA + MA), abbreviated as Cs). The photon energy shows a negative correlation with P2L, as expected. It shows a positive correlation with P2A, implying that a higher concentration of $Pb^{2+}$, which refers to organic cations, aids in reaching higher photon energy.

To clearly show the relationship between the factors and the device performance, the statistics of the photon energy values changing with different factors are plotted, as shown in Fig. 2.

### (a) The molar ratios of $Pb^{2+}$ to the organic spacer cation (P2L) and to the A-site cation (P2A) in the precursor solution

P2L, which determines the $n$ value in the perovskite with the formula of $L_2A_{n-1}Pb_nX_{3n+1}$, is widely acknowledged as one of the dominant factors governing photon energy. Due to the quantum confinement effect and the dielectric effect, a small P2L leads to high photon energy. However, it is not always accurate in experimental results. As shown in Fig. 2a, low photon energy can be found across the P2L scale from 1 to 2.5, corresponding to $n$ values ranging from 2 to 5. The deviations of

| | XLOGP3 | Cs | FA | MA | P2L | P2A | Cl | Br | PE |
|---|---|---|---|---|---|---|---|---|---|
| **PE** | 0.19 | 0.61 | -0.58 | -0.2 | -0.49 | 0.22 | 0.36 | -0.36 | 1 |
| **Br** | -0.25 | -0.32 | 0.2 | 0.21 | 0.03 | 0.28 | -1 | 1 | -0.36 |
| **Cl** | 0.25 | 0.32 | -0.2 | -0.21 | -0.03 | -0.28 | 1 | -1 | 0.36 |
| **P2A** | -0.01 | -0.01 | 0.07 | -0.06 | -0.44 | 1 | -0.28 | 0.28 | 0.22 |
| **P2L** | 0.03 | -0.14 | 0.25 | -0.08 | 1 | -0.44 | -0.03 | 0.03 | -0.49 |
| **MA** | -0.37 | -0.64 | -0.18 | 1 | -0.08 | -0.06 | -0.21 | 0.21 | -0.2 |
| **FA** | 0.04 | -0.65 | 1 | -0.18 | 0.25 | 0.07 | -0.2 | 0.2 | -0.58 |
| **Cs** | 0.25 | 1 | -0.65 | -0.64 | -0.14 | -0.01 | 0.32 | -0.32 | 0.61 |
| **XLOGP3** | 1 | 0.25 | 0.04 | -0.37 | 0.03 | -0.01 | 0.25 | -0.25 | 0.19 |

**Fig. 1** Correlation matrix of the screened factors and the photon energy of quasi-2D perovskites. The numbers at the intersection of two factors on the figure represent their Pearson coefficients: the sign of the values indicates a positive (+) or negative (−) correlation, while the magnitude indicates a strong (large) or weak (small) correlation.
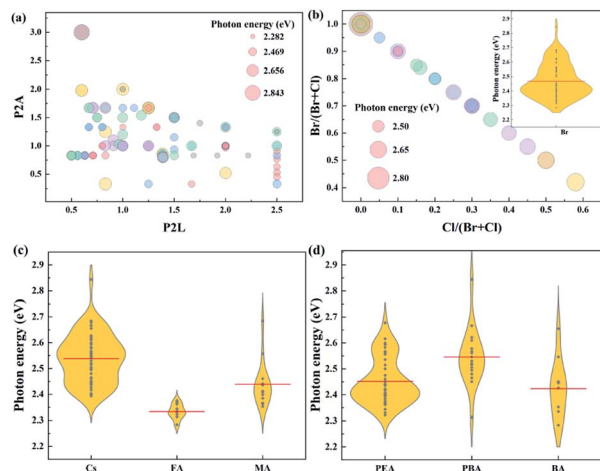
Fig. 2 Effects of different factors on the photon energy: (a) the molar ratios of $Pb^{2+}$ to the organic spacer cation (P2L) and A-site cation (P2A) in the precursor solution, (b) halide ion ratios, (c) A-site cation ratios, and (d) organic spacer cation. The photon energy determines the bubble size in (a and b). The red lines in the violin plots in (b–d) represent the corresponding mean value of the photon energy. The breadth of the violin at a given photon energy is positively related to the size of the data at that energy.

the experimental results from the theory are probably induced by the phase impurities. In quasi-2D perovskites, phase impurities with different $n$ values compared to those of the stoichiometric condition also form in the film. Large $n$ phases have low photon energies and contribute significantly to electroluminescence and hence they may present the main emission color of the device. The formation of phase impurities is difficult to control, posing challenges to the fabrication of blue devices based on low $n$ phases. On the other hand, P2A, which is equal to $n/(n-1)$ under the stoichiometric condition according to the formula of $L_2A_{n-1}Pb_nX_{3n+1}$, is sometimes smaller than 1 in the precursor solution, corresponding to the condition of excess $Pb^{2+}$. The relationship between P2A and the photon energy cannot be deduced directly from their distributions shown in Fig. 2a.

## (b) Effects of halide ion ratios and A-site cations

The effects of halide ion ratios and A-site cation ratios on the bandgap and emission photon energy of the perovskite have been thoroughly investigated in 3D perovskites. It is proved that the observed bandgap changes induced by halide ions are determined by the electronic states of the halide anion, i.e., from Cl to Br, and the valence band composition changes from 3p to 4p with a monotonic decrease in electron binding energy.[30] The changes in the bandgap caused by A-site cations are generated by their various cation sizes,[31,32] which tilt the $PbX_6$ octahedra to different extents. The photon energy changes positively with the bandgap. As shown in Fig. 2b and c, in general, the high ratio of Cl and the small Cs cation increase the photon energy, while a higher ratio of a large FA cation decreases the photon energy.

## (c) Effect of organic spacer cation

Three of the most frequently used organic spacer cations are phenylethylamine (PEA), 4-phenylbutylamine (PBA) and butylamine (BA). As shown in Fig. 2d, generally speaking, PBA leads to high PE, while BA leads to low PE. Theoretically, the organic spacer cation influences the bandgap of 2D perovskites via the dielectric confinement effect. Density functional theory calculations show that the bandgap of EA-based 2D perovskites (2.12 eV) is slightly smaller than that of PEA-based 2D perovskites (2.28 eV) due to the weaker confinement of EA with a larger dielectric constant.[33] In practice, the organic spacer cation may also affect the formation of different $n$ phases, leading to a change in photon energy. As different organic spacer cations have different solvent solubilities, the dynamic equilibrium state during the crystallization process is probably distinct, resulting in the formation of diverse $n$ phases. In this context, we introduce a descriptor, XLogP3, to describe the solubility of the organic spacer cations. A substance with a high XLogP3 value has low solubility in water. The XLogP3 values for PEABr, PBABr and BABr are 2.37, 3.37 and 1.94, respectively. The correlation between XLogP3 and the photon energy is also listed in Fig. 1, which shows a remarkable correlation (0.19).

4 algorithms were used for ML, including linear regression (LR), neural network (NN), random forest (RF) and extreme gradient boosting (XGBoost). LR is simple and facile to establish manually, and we previously proved that it was effective in predicting the bandgap of 3D perovskites.[24] The input features for the ML algorithms are the Cs/(Cs + FA + MA) ratio, FA/(Cs + FA + MA) ratio, Br/(Br + Cl) ratio, the molar ratios of $Pb^{2+}$ to an organic spacer cation (abbreviated as P2L) and of $Pb^{2+}$ to an organic cation (abbreviated as P2A) in the precursor solution, and XLogP3 value. The output is photon energy. 10-fold cross-validation was employed to optimize the hyperparameters. The dataset was randomly divided into 7 : 3 halves for training (the training set) and testing (the test set). The performances of the algorithms are evaluated using the root mean square error (RMSE) and Pearson's coefficient ($r$ value). An algorithm with a lower RMSE and a higher $r$ value predicts with higher accuracy and reliability.

Table 1 shows the performance of the different algorithms $n$ predicting the photon energy. Fig. 3a shows the comparison of the actual photon energies and the predicted values by the different algorithms. A low RMSE (≤0.07 eV) is achieved by all algorithms, indicating a high accuracy of these algorithms in prediction. Moreover, the $r$ value is higher than 0.70 for all

**Table 1** Different ML algorithms' performance in photon energy prediction of quasi-2D perovskites from their precursor solutions

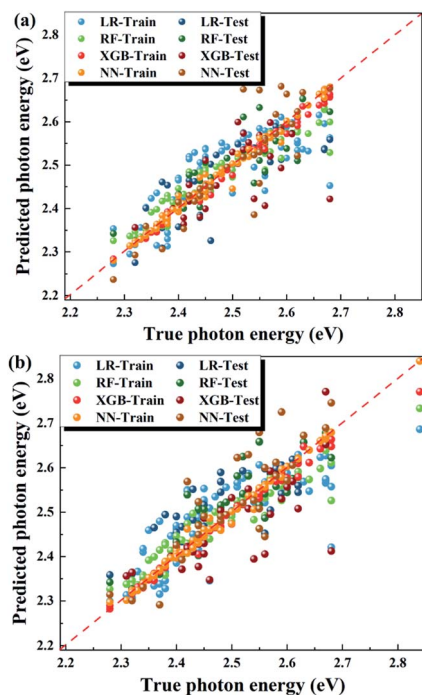| ML algorithms | Training set | | Test set | |
|---|---|---|---|---|
| | RMSE (eV) | $r$ Value | RMSE (eV) | $r$ Value |
| LR | 0.065 | 0.80 | 0.066 | 0.76 |
| RF | 0.038 | 0.94 | 0.047 | 0.92 |
| XGBoost | 0.012 | 1.00 | 0.069 | 0.71 |
| NN | 0.013 | 0.99 | 0.059 | 0.88 |

**Fig. 3** Comparison of the predicted photon energies by different algorithms and the true values of quasi-2D perovskites based on the dataset excluding the outlier with an energy of 2.844 eV (a) and the entire dataset (b), respectively. The red dashed line represents the condition in which the predicted value equals the experimental value. XGB stands for XGBoost.

algorithms, which means that the predicted values and the experimental values have a strong linear correlation. Comparing these 4 algorithms, RF performs best, with the smallest RMSE on the test set and a high $r$ value, larger than 0.9. We also calculated the average RMSE on the test set based on 10 executions, and the RF model still shows the lowest RMSE. LR is not as accurate as it is in predicting the bandgap of 3D perovskites in our previous work, because the quasi-2D perovskite is more complex, and LR is too simple to process it. This also indicates that the photon energy of the quasi-2D perovskite does not linearly depend on the screened factors. NN and XGBoost algorithms deliver excellent performance on the training set and a relatively poor performance on the test set, implying that their models are over-fitted due to the limited dataset size. The RF model is not over-fitted and exhibits its advantages in complex systems such as those in this work.

The accuracy of the algorithms highly depends on data screening. Fig. 3b depicts the prediction performance on the dataset, including the outlier with a photon energy of 2.844 eV. The predicted values deviate further from the true values than those shown in Fig. 3a based on the dataset without the outlier. Generally, higher photon energies are underestimated, whereas lower photon energies are overestimated in Fig. 3b. For the LR model, the RMSEs for the training and test sets are 0.068 and 0.067 eV, respectively, which are higher than the values (0.065 and 0.066 eV) listed in Table 1. For the RF algorithm, the RMSEs for the training and test sets are 0.040 and 0.050 eV,

respectively, which are also higher than the corresponding values listed in Table 1. These results illustrate the importance of the smart screening of the reported experimental dataset.

The RF model can assess the significance of the input features (shown in Fig. 4). The Cs ratio, rather than the P2L or Br ratio, is the most critical factor affecting photon energy. The Br ratio is also important, while P2A and XLogP3 are less important. Considering the extreme importance of the Cs ratio and its positive correlation with photon energy, these results imply that using Cs as the major A-site cation is crucial for obtaining high photon energy. In addition, it also implies that the reported quasi-2D perovskites suffer from phase impurity, as P2L is not as essential to determining photon energy as expected.

As the RF model predicts the photon energy more accurately, we use it to screen quasi-2D perovskites with desired compositions and photon energies. Here, MA-free quasi-2D perovskites tend to achieve high material stability. The settings for the precursor compositions are listed in Table S3,† leading to the generation of nearly 40 000 data points. These quasi-2D perovskites are Cs-dominated (Cs ratio in the range of 0.75–1.00) and have small $n$ values (P2L in the range of 0.6–1.6). The RF model, trained by the experimental dataset listed in Table S1,† was used to predict their photon energies.

The predicted photon energies of the perovskites vary on the scale of 2.478–2.628 eV, with the emission peak wavelengths varying from 471 nm to 500 nm. For blue emission, photon energies larger than 2.58 eV (wavelength shorter than 480 nm) are required. Fig. 5a shows the statistics of the factors enabling the photon energies to meet this requirement. As can be seen, a high Cs ratio (0.90–1.00) is essential to get blue emission. Other factors have identical ranges as those of the settings. Moreover, it is found that the factors interplay with each other. For example, for a Cs ratio of 0.90, the Br ratio should be lower than 0.50 to get blue emission, whereas for a Cs ratio of 1.00, pure Br-based perovskites can also achieve blue emission, which is in accordance with the experimental results.[34] Two aspects can explain the importance of the Cs ratio in getting high photon energy: small radius size[35] and low solubility[36] in the typically used polar solvents DMSO/DMF. The small radius size of Cs tilts the PbX$_6$ octahedra, generating the large bandgap as discussed before.[35,37] The lower solubility of Cs$^+$ compared to that of FA$^+$ and MA$^+$ (ref. 34) makes it participate in the initial
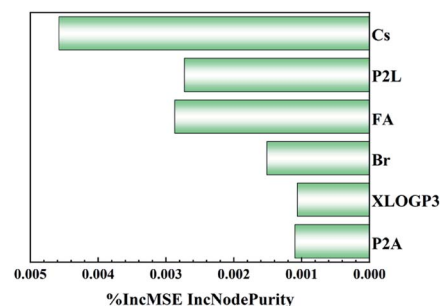


**Fig. 4** Feature importance of the input features on photon energy presented by the RF algorithm.
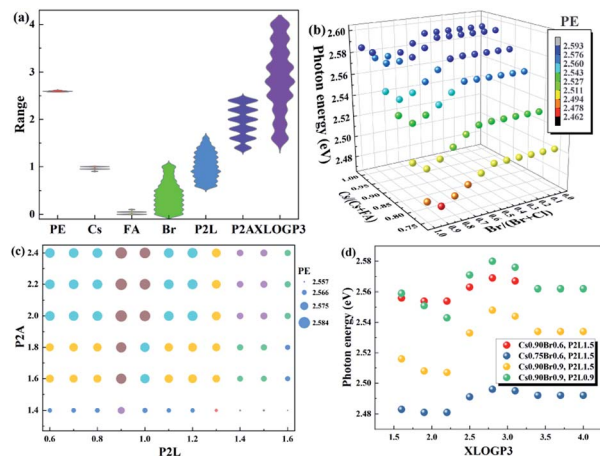
**Fig. 5** (a) Ranges of different factors for blue emission and (b–d) graphs of the predicted photon energies of the quasi-2D perovskites with different factors by the RF model: (b) the ratio of halide ions, (c) P2A and P2L, and (d) XLogP3.

nucleating process together with large spacer cations with low solubility, which can form moderate $n$ phases. Hence, the phase impurity can be reduced, bringing the photon energies closer to the expected values.

To clearly show the effects of the different factors on the photon energy of quasi-2D perovskites, the graphs of the predicted values and the factors are shown in Fig. 5b–d. Fig. 5b shows the variation of the photon energy with the ratios of Br and Cs, while P2L, P2A and XLogP3 are fixed to be 1.5, 1.6, and 2.80, respectively. Obviously, increasing the Cs ratio from 0.75 to 1.00 leads to an increase in photon energy from 2.50 eV to 2.58 eV. A decrease in the Br ratio (an increase in the Cl ratio) leads to a complex change in the photon energy. Pure Br-based perovskites have relatively higher photon energy than $Br_{0.9}Cl_{0.1}$-based ones; in Br–Cl mixed perovskites, the photon energy first increases with decreasing Br ratio, and then displays a slight chance. Despite the fact that this trend is contrary to theoretical knowledge, it is in accordance with the reported experimental results shown in Fig. 2b. The RF model learned the role of the Br ratio from the data in Fig. 2b, $i.e.$; decreasing the Br ratio does not always increase the photon energy. Hence, it assigned the corresponding weight to the Br ratio. In comparison, increasing the Cs ratio was discovered to be effective in increasing the photon energy, and the RF model assigned a large weight to the Cs ratio. As a consequence, increasing the Cs ratio efficiently increases the photon energy. These findings further demonstrate the RF model's ability to handle complex relationships.

Decreasing P2L generally increases photon energy, as shown in Fig. 5c, which is to be expected. However, the change is minor, $i.e.$, decreasing P2L from 1.5 to 1.0 (corresponding to $n$ decreasing from 3 to 2) leads to an increase of less than 0.02 eV. Though decreasing P2L is also thought to be efficient in increasing photon energy,[14,38,39] the formation of large $n$ phases with low photon energies produces the main emission peak arising from large $n$ phases. Hence, the RF model discovered that decreasing P2L did not effectively enhance photon energy.

Additionally, modifying XLogP3 also leads to a change in photon energy, as shown in Fig. 5d. Increasing XLogP3 from 1.9 to 2.8 induces the largest increase of 0.04 eV in the photon energy. This is in line with the observations from the experimental results in Fig. 2d. As P2L and other factors are controlled to be the same, this increase can be ascribed to the reduction of large $n$ phases with low photon energy.

Based on these results and the evidence in the literature, it can be summarized that the photon energy of quasi-2D perovskites is largely affected by the formation of large $n$ phases, $i.e.$, the phase impurities. Hence, suppressing phase impurity is critical for getting blue emission. From the correlation between the photon energy and the XLogP3 value, it can be speculated that the solubility of the organic spacer cation in the solvents is probably an important factor that needs to be considered to suppress the phase impurity. From these results, it can also be observed that a high Cs ratio is essential to obtain blue emission.

## 3. Methods

The $R$ tool was employed as the platform for machine learning. Linear regression (LR), neural network (NN), random forest (RF) and extreme gradient boosting (XGBoost) algorithms were used based on $glm$, $neuralnet$, $randomForest$ and $xgboost$ functions, respectively. The performance of the algorithms was evaluated using the root mean square error (RMSE) and Pearson's coefficient ($r$ value) on the test set. Here,

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(X_i - Y_i)^2}{n}}$$

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}$$

$X_i$, $Y_i$, $\bar{X}$, $\bar{Y}$, and $n$ represent the $i^{th}$ value of the experimental dataset, the $i^{th}$ value of the predicted dataset, the mean value of the experimental dataset, the mean value of the predicted dataset, and the number of the dataset points, respectively. The ratio of the test set is 0.3. To train the ML algorithms, we use 10-fold cross-validation to optimize the hyperparameters, which divides the datasets into 10 parts (90% data points for training and 10% for validation) and does the learning 10 times. The model having the lowest average RMSE on the validation sets was screened for testing on the test set and further use. In detail, the NN model had 3 hidden layers, which had 8, 6 and 4 neurons, respectively; the tree number in the RF model was 5000. The maximum depth and the number of rounds in the XGBoost model were 10 and 40, respectively.

## 4. Conclusions

In summary, this work proves that the photon energy of quasi-2D perovskites can be well predicted by machine learning based

on their precursor compositions. The random forest algorithm outperforms the other three algorithms, which presents an RMSE of ~0.05 eV on the test set. According to the random forest model's evaluation of feature relevance, the ratio of the Cs cation is found to be critical to getting high photon energy. Moreover, it is also shown that increasing the ratio of the organic spacer cation to lead ion is not so effective in increasing the photon energy. These results show that the experimental results have a large deviation from the theoretical speculations, while the data-driven analysis is critical to providing guidance for experiments. The results also suggest that the phase impurities should be suppressed to get deep blue emission. In addition, the solubility of organic spacer cations in precursor solvents is better to be considered to suppress the formation of phase impurities. Furthermore, a series of quasi-2D perovskites for blue emission are predicted by the random forest model. These results reveal that machine learning is an efficient tool to explore the underlying relations between the fabricating conditions and the properties of quasi-2D perovskites, which will greatly reduce the time and material cost of experimental material screening.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

## Notes and references

1 M. A. Green, A. Ho-Baillie and H. J. Snaith, *Nat. Photonics*, 2014, **8**, 506–514.

2 L. N. Quan, B. P. Rand, R. H. Friend, S. G. Mhaisalkar, T. W. Lee and E. H. Sargent, *Chem. Rev.*, 2019, **119**, 7444–7477.

3 Y. Cao, N. N. Wang, H. Tian, J. S. Guo, Y. Q. Wei, H. Chen, Y. F. Miao, W. Zou, K. Pan, Y. R. He, H. Cao, Y. Ke, M. M. Xu, Y. Wang, M. Yang, K. Du, Z. W. Fu, D. C. Kong, D. X. Dai, Y. Z. Jin, G. Q. Li, H. Li, Q. M. Peng, J. P. Wang and W. Huang, *Nature*, 2018, **562**, 249–253.

4 X. L. Yang, X. W. Zhang, J. X. Deng, Z. M. Chu, Q. Jiang, J. H. Meng, P. Y. Wang, L. Q. Zhang, Z. G. Yin and J. B. You, *Nat. Commun.*, 2018, **9**, 8.

5 H. H. Tsai, W. Y. Nie, J. C. Blancon, C. C. S. Toumpos, R. Asadpour, B. Harutyunyan, A. J. Neukirch, R. Verduzco, J. J. Crochet, S. Tretiak, L. Pedesseau, J. Even, M. A. Alam, G. Gupta, J. Lou, P. M. Ajayan, M. J. Bedzyk, M. G. Kanatzidis and A. D. Mohite, *Nature*, 2016, **536**, 312–316.

6 D. H. Cao, C. C. Stoumpos, O. K. Farha, J. T. Hupp and M. G. Kanatzidis, *J. Am. Chem. Soc.*, 2015, **137**, 7843–7850.

7 I. C. Smith, E. T. Hoke, D. Solis-Ibarra, M. D. McGehee and H. I. Karunadasa, *Angew. Chem., Int. Ed. Engl.*, 2014, **53**, 11232–11235.

8 Z. M. Chu, Q. F. Ye, Y. Zhao, F. Ma, Z. G. Yin, X. W. Zhang and J. B. You, *Adv. Mater.*, 2021, **33**, 9.

9 K. H. Wang, Y. D. Peng, J. Ge, S. L. Jiang, B. S. Zhu, J. S. Yao, Y. C. Yin, J. N. Yang, Q. Zhang and H. B. Yao, *ACS Photonics*, 2019, **6**, 667–676.

10 P. Vashishtha, M. Ng, S. B. Shivarudraiah and J. E. Halpert, *Chem. Mater.*, 2018, **31**, 83–89.

11 Y. F. Ng, B. Febriansyah, N. F. Jamaludin, D. Giovanni, N. Yantara, X. Y. Chin, Y. B. Tay, T. C. Sum, S. Mhaisalkar and N. Mathews, *Chem. Mater.*, 2020, **32**, 8097–8105.

12 M. Worku, Q. He, L. J. Xu, J. Hong, R. X. Yang, L. Z. Tan and B. Ma, *ACS Appl. Mater. Interfaces*, 2020, **12**, 45056–45063.

13 Y. Z. Jiang, C. C. Qin, M. H. Cui, T. W. He, K. K. Liu, Y. M. Huang, M. H. Luo, L. Zhang, H. Y. Xu, S. S. Li, J. L. Wei, Z. Y. Liu, H. H. Wang, W. Kim, M. J. Yuan and J. Chen, *Nat. Commun.*, 2019, **10**, 9.

14 Z. C. Li, Z. M. Chen, Y. C. Yang, Q. F. Xue, H. L. Yip and Y. Cao, *Nat. Commun.*, 2019, **10**, 1027.

15 L. Yang, Y. Zhang, J. M. Ma, P. Chen, Y. Yu and M. Shao, *ACS Energy Lett.*, 2021, **6**, 2386–2394.

16 F. S. Ma, J. W. Li, W. Z. Li, N. Lin, L. D. Wang and J. Qiao, *Chem. Sci.*, 2017, **8**, 800–805.

17 M. V. Kovalenko, *30th Annual Conference of the IEEE Photonics Society*, Florida, US, 2017.

18 Y. Q. Song, J. Lindsay, Y. Zhao, A. Nasiri, S. Y. Louis, J. Ling, M. Hu and J. J. Hu, *Comput. Mater. Sci.*, 2020, **183**, 109792.

19 J. X. Li, B. Pradhan, S. Gaur and J. Thomas, *Adv. Energy Mater.*, 2019, **9**, 10.

20 N. T. P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi, T. Buonassisi and S. Sun, *Nat. Commun.*, 2020, **11**, 4172.

21 N. Majeed, M. Saladina, M. Krompiec, S. Greedy, C. Deibel and R. C. I. MacKenzie, *Adv. Funct. Mater.*, 2019, **30**, 1907259.

22 H. Sahu, W. Rao, A. Troisi and H. Ma, *Adv. Energy Mater.*, 2018, **8**, 9.

23 S. H. Lu, Q. H. Zhou, Y. X. Ouyang, Y. L. Guo, Q. Li and J. L. Wang, *Nat. Commun.*, 2018, **9**, 3405.

24 Y. Y. Li, Y. Lu, X. M. Huo, D. Wei, J. Meng, J. Dong, B. Qiao, S. L. Zhao, Z. Xu and D. D. Song, *RSC Adv.*, 2021, **11**, 15688–15694.

25 E. I. Marchenko, S. A. Fateev, A. A. Petrov, V. V. Korolev, A. Mitrofanov, A. V. Petrov, E. A. Goodilin and A. B. Tarasov, *Chem. Mater.*, 2020, **32**, 7383–7388.

26 Z. Y. Wan, Q. D. Wang, D. C. Liu and J. H. Liang, *New J. Chem.*, 2021, **45**, 9427–9433.

27 C. W. Zhu, W. Liu, Y. Y. Li, X. M. Huo, H. T. Li, K. Guo, B. Qiao, S. L. Zhao, Z. Xu, H. G. Zhao and D. D. Song, *Sol. Energy*, 2021, **228**, 45–52.

28 T. J. Cheng, Y. Zhao, X. Li, F. Lin, Y. Xu, X. L. Zhang, Y. Li, R. X. Wang and L. H. Lai, *J. Chem. Inf. Model.*, 2007, **47**, 2140–2148.

29 *SwissADME*, http://www.swissadme.ch/, accessed May 2021.

30  D. Y. Luo, R. Su, W. Zhang, Q. H. Gong and R. Zhu, *Nat. Rev. Mater.*, 2019, **5**, 44–60.

31  A. Walsh, *J. Phys. Chem. C*, 2015, **119**, 5755–5760.

32  C. Y. Yi, J. S. Luo, S. Meloni, A. Boziki, N. Ashari-Astani, C. Gratzel, S. M. Zakeeruddin, U. Rothlisberger and M. Gratzel, *Energy Environ. Sci.*, 2016, **9**, 656–662.

33  B. Cheng, T. Y. Li, P. Maity, P. C. Wei, D. Nordlund, K. T. Ho, D. H. Lien, C. H. Lin, R. Z. Liang, X. H. Miao, I. A. Ajia, J. Yin, D. Sokaras, A. Javey, I. S. Roqan, O. F. Mohammed and J. H. He, *Commun. Phys.*, 2018, **1**, 8.

34  Y. Q. Shang, G. Li, W. M. Liu and Z. J. Ning, *Adv. Funct. Mater.*, 2018, **28**, 1801193.

35  C. H. Lu, G. V. Biesold, Y. J. Liu, Z. T. Kang and Z. Q. Lin, *Chem. Soc. Rev.*, 2020, **49**, 4953–5007.

36  N. Yantara, N. F. Jamaludin, B. Febriansyah, D. Giovanni, A. Bruno, C. Soci, T. C. Sum, S. Mhaisalkar and N. Mathews, *ACS Energy Lett.*, 2020, **5**, 1593–1600.

37  Q. A. Akkerman, G. Raino, M. V. Kovalenko and L. Manna, *Nat. Mater.*, 2018, **17**, 394–405.

38  Y. Han, S. Park, C. Kim, M. Lee and I. Hwang, *Nanoscale*, 2019, **11**, 3546–3556; P. Y. Pang, G. R. Jin, C. Liang, B. Z. Wang, W. Xiang, D. L. Zhang, J. W. Xu, W. Hong, Z. W. Xiao, L. Wang, G. C. Xing, J. S. Chen and D. G. Ma, *ACS Nano*, 2020, **14**, 11420–11430.

39  J. Xing, Y. B. Zhao, M. Askerka, L. N. Quan, X. W. Gong, W. J. Zhao, J. X. Zhao, H. R. Tan, G. K. Long, L. Gao, Z. Y. Yang, O. Voznyy, J. Tang, Z. H. Lu, Q. H. Xiong and E. H. Sargent, *Nat. Commun.*, 2018, **9**, 3541.