Physical Education, Health and Social Sciences

https://journal-of-social-education.org E-ISSN: 2958-5996

P-ISSN: 2958-5988

Predicting Band Gaps of 2D Phosphorene-Isoelectronic materials Using Machine Learning

Tayyeba Idrees[1], Sehar Saleem[1], Saeed Rasheed[2]

[1] Department of Physics, University of Agriculture, Faisalabad, Pakistan
Email: ranaidrees6565@gmail.com
[1] Department of Physics, University of Agriculture, Faisalabad, Pakistan
Email: seharsaleem873@gmail.com
[2] Department of Computer Science, University of Agriculture, Faisalabad, Pakistan
Email: saeed.rasheed0211@gmail.com

DOI: https://doi.org/10.63163/jpehss.v4i2.1442

Abstract

It also opened new possibilities of being used in electronic and optoelectronic ways in future since the two dimensional (2D) materials have been discovered. Among these materials, the phosphorene and their isoelectric counterparts are those materials that attracted the attention of people due to their adjustable electronic properties and high quality bandgap properties. However, the conventional methods of calculation such as Density Functional Theory (DFT) physically determining the gap is slow and expensive, particularly with large material libraries being screened. The framework is a machine-learning model that makes accurate and efficient prediction which bandgaps of 2D phosphorene-isoelectronic materials (PINs) are possible. A set of structural, electronic, and physicochemical properties of various 2D materials was used to develop and test various machine learning techniques, especially Random Forest Regression, Support Vector Regression (SVR), and Gradient Boosting algorithms. To determine the most important material properties that affect the bandgap values, feature engineering and feature selection techniques were implemented. The accuracy of the machine learning models developed was assessed based on commonly used regression tests, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). A set of statistical measures used to evaluate the prediction accuracy, reliability and generalizability of the models. The results obtained in the present work show the machine learning methods are very effective and can predict the materials bandgap properties with a very good accuracy. Moreover, the proposed models are able to substantially decrease the computational effort and time needed for traditional first-principles calculations, enabling practical applications in materials screening and discovery for larger systems. The work in this study demonstrates the power of data-driven strategies to unlock the rapid identification and design of 2D semiconductors with engineered electronic properties. The primary areas of interest include: 2D Materials, Phosphorene-Isoelectronic Materials, Bandgap Prediction, Machine Learning, Density Functional Theory (DFT), Materials Informatics and Semiconductor Design.

Introduction:

Recent research on two-dimensional (2D) materials has not only changed the face of materials science, but also usurped a new array of platforms with which to construct new high performance electronic, optoelectronic, and energy devices. Researchers have looked at many new types of 2D materials with distinct atomic structure and electronic properties since the discovery of graphene.

Among these, the few- or single layer black phosphorus (a polymorph of phosphorus) has attracted a special interest because of the direct and anisotropically tunable band gap, high mobility and its anisotropic electronic behavior. Nevertheless, phosphorene presents potential for application in a transistor, in a material for a photodetector, as a sensor and as a flexible electronic device. Over the last several years, scientists have considered and created a dozen so-called iso-electronic materials that are “phosphorene”. They share similar electronic characteristics as is the case with phosphorene which have different structural and chemical characteristics. They can be such as group IV-VI and group III-VII compound monolayers, and are sometimes found to have semiconducting characteristics within a wide band gap values. Proper knowledge of band gap of those materials is highly important as the band gap has a significant influence on both the electrical conductivity, optical absorption and the performance of a device.

The traditional way of calculating the bandgaps is from the first principles calculation of the computational atomic and quantum chemistry theories such as DFT. It is true that although it is efficient in electronic properties study of a material it can be extremely computation- and time-intensive when a large number of candidate materials are used. The number of possible 2D materials continues to grow, and there is a need to find more predictive methods for these materials in an efficient and rapid manner. Machine learning (ML) is now an extremely efficient approach to speed up materials discovery and discover both complex relationships between material descriptors and target properties. While quantum-mechanical methods can calculate material properties, these calculations can be expensive and difficult to perform per material, machine-learned models can learn trends from the experimental data available and quickly and accurately estimate material properties. However, recent progress in materials informatics and machine learning has allowed the prediction of material properties with more efficiency and with less reliance on expensive simulations. Machine learning techniques are used in this work to estimate the bandgap energies of isoelectronic 2D materials based on phosphorene. A number of structural, chemical, and electronic features were used to build models to establish accurate relationships between material properties and bandgap values. The study was aimed at assessing the predictive power of various machine learning algorithms, as well as pinpointing the most important factors influencing the bandgap properties, and establishing a robust framework for the fast screening of novel 2D semiconductor materials. The results reveal the potential of data-driven methods to speed up the discovery of materials and parameters in the design of advanced electronic and optoelectronic devices.

Methodology:

DFT, within the projector-augmented wave (PAW) approach implemented in the Vienna Ab Initio Simulation Package (VASP) [30] was used to calculate the electronic structure of the systems. Throughout the simulations periodic boundary conditions were used, and a vacuum spacing of ≥ 15 Å between the adjacent periodic images was taken into account to prevent interaction between the periodic images. The Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional [31] was used in conjunction with the conjugate gradient algorithm [32] for the optimization of the structure. For all calculations an energy cut-off of 400 eV was used. The optimized structures were deemed converged if the Hellmann–Feynman forces on each atom were less than 0.01 eV/Å. The screened hybrid Heyd–Scuseria–Ernzerhof (HSE) functional [33,34] was used in single-shot calculations on the PBE-relaxed structures to get more accurate estimates of the electronic bandgap and the positions of the band edges. The standard mixing parameter (0.25) and range-separation parameter (0.2 Å^{-1}) were used. The Brillouin zone was sampled with dense Monkhorst–Pack k-point meshes with a minimum number of k points equal to $10 \times 10 \times 1$ in the primitive unit cells [35].

Machine learning methods:

This is then analyzed with the machine learning models obtained from the machine learning package scikit-learn [36] to obtain the obtained DFT results. Supervised learning methods can be used to find the connection between target electronic properties and predictors. A good predictive model is highly sensitive to the regression model that we use, the predictors that we select and the quality of our data. The selection of appropriate predictors (regression models which best represent data set) are very critical in ensuring that the regression models have the predictive power and accuracy to the given data set. In order to predict the electronic properties of the materials studied, several sets of descriptors were explored, namely descriptors from the PBE calculations and basic properties of the constituent elements. A series of machine learning algorithms was used to create relationships between these descriptors and the target electronic properties. In the present study, three different models were studied: Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR).

The initial model was a linear one, modeling the dependency of the target property on the chosen predictors. The model parameters were optimized by minimizing the difference between measured and calculated values by the least-squares optimization method. Other methods to improve the robustness of the model and to avoid overfitting included regularization based techniques such as LASSO and Ridge Regression. These methods include terms from the L1 norm and L2 norm of the regression coefficients respectively, to assist in controlling the complexity of the model and to enhance the predictive ability in high dimensional data sets.

Advanced machine learning techniques were also explored as materials properties frequently have complicated, non-linear relationships. The nonlinear interactions amongst input features and target variables were detected by using the method of Random Forest Regression (RFR) which belongs to the ensemble-learning family. The algorithm constructs multiple decision trees from random subsets of the training features and samples. Thus the final prediction is the average of the predictions produced by all of the trees, which results in greater accuracy and stability. Another benefit of Random Forest models is that they can also estimate the importance of the features so that the most significant descriptors that determine the electronic properties of materials can be identified. Support Vector Regression was also used to model nonlinear dependencies in the data. This method involves projecting the data into a higher dimensional feature space using kernel functions in order to better capture the complicated relationships in the data. The Radial Basis Function (RBF) was chosen as the kernel because it has been shown to have excellent properties for capturing nonlinear patterns in material-property data. The SVR model can be used to predict the electronic properties with high accuracy using a small number of material descriptors because it learns the optimal support vectors and minimizes prediction errors. These machine learning algorithms offer a holistic solution to the problem of understanding factors that affect the electronic properties and creating reliable predictive models of advanced semiconductor materials.

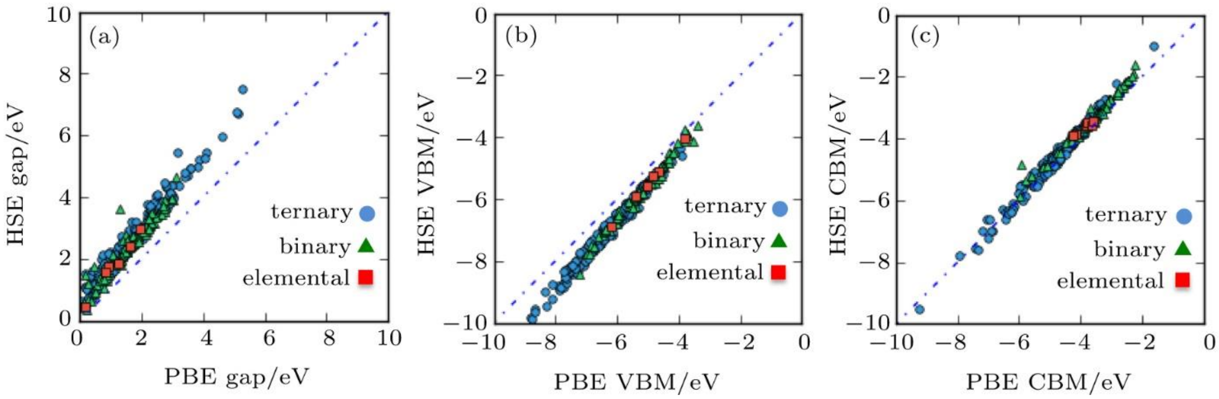
Table 1. Three different sets of predictors have been employed for machine learning models predicting electronic band gaps and band alignments.

Target property	**Predictors set-I**	**Predictors set-II**	**Predictors set-III**
E_g (HSE)	E_g (PBE)	elements signatures	E_g (PBE), elements signatures
VBM (HSE)	VBM (PBE)	elements signatures	VBM (PBE), elements signatures
CBM (HSE)	CBM (PBE)	elements signatures	CBM (PBE), elements signatures

Choosing the right input features is as crucial as choosing the right machine learning algorithm to get the accurate and reliable prediction. Previously, the number of descriptors used in the feature space was large and dimensionality reduction techniques were used afterwards. This can enhance the efficiency of computation, but can also make the physical meaning of each feature hard to discern. The usefulness of the DFT-derived properties for prediction was compared to the fundamental chemical and physical properties of the constituents in the present study. In order to accomplish this goal three different sets of predictors were built, as summarized in Table 1. The first set (Set-I) comprised only electronic properties extracted from PBE calculations such as the bandgap, valence band maximum (VBM) and conduction band minimum (CBM). The second set (Set-II) included elemental descriptors like electron affinity, electronegativity, electronegativity difference between cationic and anionic species, ionization energy and atomic mass. The third set (Set-III) was a blend of the features of Set-I and Set-II. Set-I requires initial DFT calculations, but gives direct electronic information related to the target properties. This is, in contrast, Set-II, which is based only on easily available elemental parameters, without the need for computationally demanding electronic structure calculations. The purpose of Set-III was to examine the possibility of further improving predictive accuracy by merging both types of information.

Electronic properties (pbe, HSE):

The calculated electronic properties such as the fundamental bandgap, valence band maximum (VBM) and conduction band minimum (CBM) are shown in Figure 2. The results of PBE and HSE calculations show that there is a strong linear correlation between the two approaches. While hybrid functionals are often used to obtain more accurate bandgap values than those of PBE functionals, they are considerably more expensive. The bandgap results in Figure 2(a) suggest that the HSE bandgaps tend to grow in a near linear fashion with the PBE bandgap for the materials studied. This observation opens the possibility of using lower cost PBE calculations as good descriptors to estimate HSE-level electronic properties using machine learning models.



Moreover, there is a linear correlation even when materials are classified as per their chemical composition and structural properties. Despite having different elements, different groups of isoelectronic materials (different colored regions in Figure 2(a)) exhibit similar trends. Such consistent relationships underscore the promise of using the PBE-derived electronic parameters and elemental descriptors as reliable predictors to machine-learning estimate advanced electronic properties. There is a good correspondence between the PBE and HSE results for both valence and conduction band edge positions. The ionization energy of a material is directly related to the valence band maximum. It is now known that the ionization energy of a material can often be reasonably measured experimentally and is directly related to the valence band maximum. The general trend in the HSE method is to predict lower VBM values than in the PBE method as seen in Figure 2(b), but the correlation between the two methods is still very linear. The discovery indicates that PBE parameters could be useful for interpolating the more expensive HSE parameters and maybe even be predictive of experimental electronic properties. A similar linear behaviour is also found for the conduction band minima, as shown in Figure 2(c). Despite the discrepancy between CBMs estimated by the HSE and the CBMs predicted by PBE, there is an overall correlation between the two. This difference between the two methods is larger for the valence band maxima than it is for the conduction band minima. These findings suggest that PBE calculations can offer useful information to predict electronic properties calculated by the HSE scheme, and can be usefully applied in data-driven studies of materials.

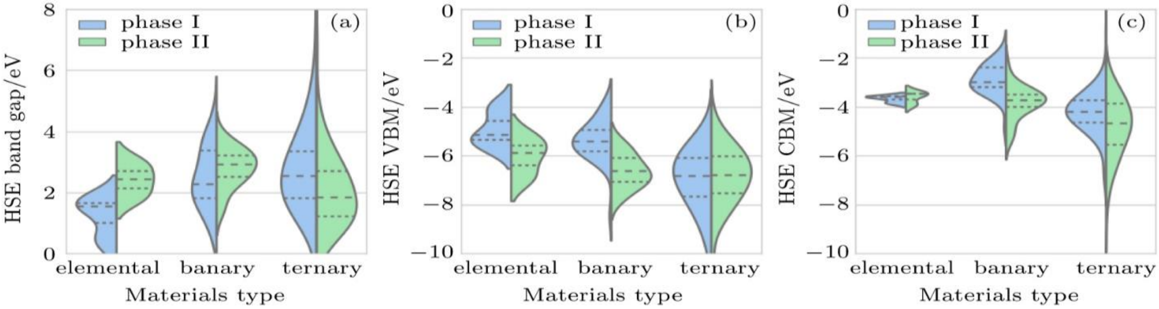


Fig. 3. The distribution of the band gap value (a), VBM (b) and CBM (c) vary with the kind of material. The areas of blue colour are revealed to be phase-I materials and the areas which are green, those of phase-II. The median of the data is the longest dashed line.

Machines-based predictive models:

The results from all the calculations have been reported using two levels of the theory: DFT—PBE, and S-CR-HSE (Screened Consistent Range-separated Hybrid Functional) theory. The first is easily calculated and cannot model the minute fundamental band gap of regular semiconductors (type diamond or semiconductors with no localised d or f electrons) and the second can model the values and alignments of the band-gap of regular semiconductors accurately, but is too expensive to be used to screen large volumes of useful functional materials. Thus it would be desirable to develop computational efficient techniques that would be extremely accurate as well. Machine learning models such as Linear Regression (LR), Random Forest Regression (RFR) and Support Vector Regression (SVR) were trained, with the help of the predictor sets introduced in Section 2.2, to predict the electronic properties calculated with the HSE method. The ultimate goal was to develop predictive relationships that could be used as efficient alternatives to computationally more demanding HSE calculations and even help to estimate the electronic properties as measured in the experiments.

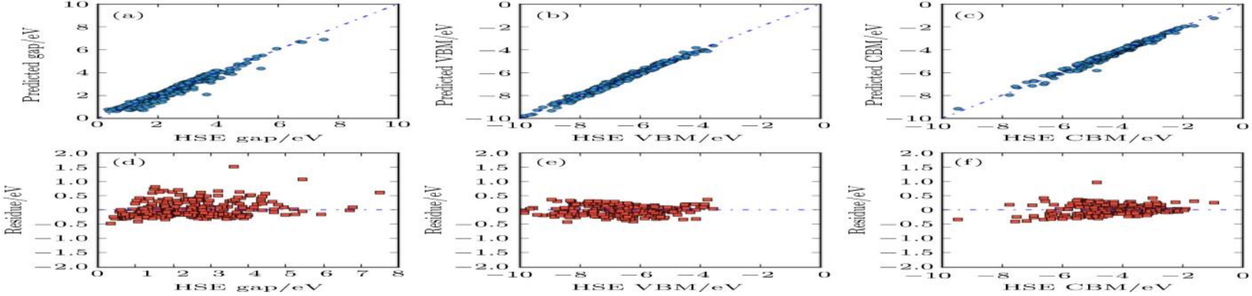
Set-I predictors:

Due to the strong linear relationship between the PBE- and HSE-based bandgaps as discussed in Section 3, the PBE-derived electronic properties are effective predictors of the HSE-level electronic properties. For Set-I, only the bandgap values calculated using PBE were used as the input features. Nearly linear relationship was observed between the predictor and target variables, so the Linear Regression was thought to be most appropriate for modeling this feature set. Higher order nonlinear techniques like Random Forest Regression and Support Vector Regression were not held to offer any great benefits, given the low dimensionality and simplicity of the predictor space. The accuracy of the Linear Regression model was checked by comparing the values from the HSE model with the values from the Linear Regression model for the validation data set, as shown in Figure 4(a). The good agreement between the predicted and actual bandgap values indicates that the information in the PBE bandgaps is adequate to accurately estimate the HSE bandgaps. This result shows that it is possible to use the computationally inexpensive DFT-PBE calculations as substitutes for higher-accuracy electronic properties, which can be predicted via machine learning methods.

$$E_{\mathrm{g}}^{\mathrm{HSE}} = 1.21E_{\mathrm{g}}^{\mathrm{PBE}} + 0.52\mathrm{eV}.$$

The difference between the calculated and theoretical values of band-gap, also known as error scores, are presented in Fig. 4(d) and the majority of the scores are located within the energy window [0.5 eV, −0.5 eV] with an outstanding score of prediction. However, just 2 data points fall outside that range of difference as it is more than 1.0 eV. They will not make much impact on our regression model though they they may be outlying cases. However, there are other error measures for the predictive model including the root mean square error (RMSE) and mean absolute percent error (MAPE) that can be calculated between the predictive model and the small prediction error of 0.25 eV and 10.67% respectively, shown in the Table. Another value was the accuracy of the model around which the result is displayed, CO2 result with high accuracy is displayed amongst all the measurements (5).

Fig. 4. Panels (a), (b) and (c) show the comparison of the HSE calculated and predicted (using PBE-derived parameters from Set-I) values of the bandgap, the VBM and the CBM, respectively. The dotted line at 45° indicates an agreement between the predicted and calculated results. The good agreement between the data points and the line suggests that the LR model well reproduces the HSE calculated electronic properties. The residual plots further illustrate the model's ability to predict the data, showing only slight differences between the predicted and actual values. The results obtained also validate the predictive ability of the electronic properties obtained from PBE in the estimation of electronic properties at the HSE level.



Similarly, the only feature of VBMPBE in set-I predictors space is used to predict the positions in the VBMHSE. As seen in the fig. 4(b) the VBMHSE values predicted by the VBMHSE of the validation sets were in the good correlation with the desired values of the VBMHSE. All the residues [Fig. 4(e)] are in the [−0.5 eV, 0.5 eV] energy range. The fact that the VBM predictive model is more linear than the band gap, also contributes to the fact that the prediction errors are smaller as given in Table. 2. The following formula can be used to predict VBMHSE using the results from VBMBM:

$$\mathrm{VBM}^{\mathrm{HSE}} = 1.15\mathrm{VBM}^{\mathrm{PBE}} + 0.23\mathrm{eV}.$$

For example, the model of CBMPBE can be predicted by LR method using CBMBHE, and the accuracy of the model is shown below (Figs.). 4(c) and 4(f) In case of CBM, the HSE / PBE results are related to each other.

$$\mathrm{CBM}^{\mathrm{HSE}} = 1.07\mathrm{CBM}^{\mathrm{PBE}} + 0.51\mathrm{eV}.$$

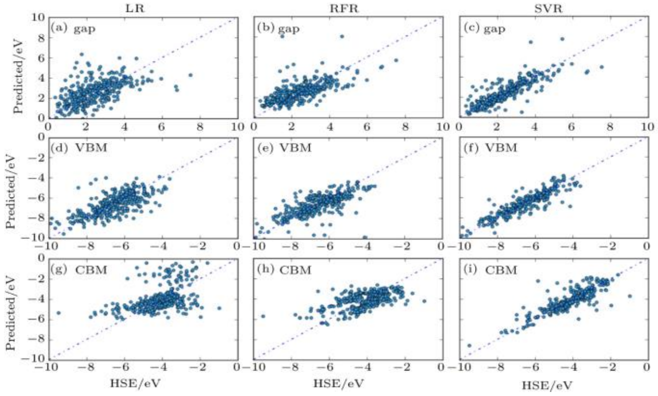
The three linear models above have been cross validated with samples of 2D materials that were selected at random from our data set of 2D materials and should thus be representative to materials within the family that are isoelectronic with the members of the family.

Table 2. Prediction errors of band gaps in the LR, RFR, and SVR models.

Regr. Methods	Pred. sets	Band gap RMSE/eV	Band gap MAPE/%	VBM RMSE/eV	VBM MAPE/%	CBM RMSE/eV	CBM MAPE/%
LR	set-I	0.25	10.67	0.15	1.85	0.14	2.53
	set-II	0.87	35.07	0.88	10.3	0.8	16.03
	set-III	0.15	5.55	0.09	1.04	0.09	1.56
RFR	set-I	–	–	–	–	–	–
	set-II	0.7	26.37	0.67	7.23	0.57	10.22
	set-III	0.25	7.44	0.18	1.75	0.18	2.64
SVR	set-I	–	–	–	–	–	–
	set-II	0.57	16.8	0.49	4.83	0.43	7.07
	set-III	0.13	4.93	0.08	0.96	0.09	1.65

(a) Set-II Predictors

Ideally, an effective machine learning model should be based on properties of the elements that are easily accessible, rather than electronic properties calculated with expensive DFT calculations. This would greatly ease the process of materials screening and would speed up and make the prediction of electronic properties more practical. To explore this idea, predictor set-II was created from only the basic features of the elements that make up the prediction, as mentioned in Subsection 2.2. This feature set was then used to evaluate the predictive power of Linear Regression (LR), Random Forest Regression (RFR) and Support Vector Regression (SVR). These models are shown to be effective by comparing the predicted values with the calculated electronic properties in Figure 5. The results show that the performance of the Linear Regression model is not as good if only the elemental descriptors are used as input features. This diminishes the accuracy as demonstrated by relatively high prediction errors such as RMSE of 0.87 eV and MAPE of 35.07% as shown in Table 2. Despite all this, the model results are comparatively better for the prediction of valence band maxima (VBMs) and conduction band minima (CBMs) with MAPE values of 10.30% and 16.03%, respectively. Regularized regression techniques like Ridge and LASSO regression were also used to see if the apparent errors were attributable to the complexity of the model. Both methods, however, gave no significant improvement in the accuracy of prediction and thus it is concluded that the limitations are due more to the information content of the elemental descriptors selected than to the over-fitting of the regression model.



The results of three machine learning models (Linear Regression (LR), Random Forest Regression (RFR), and Support Vector Regression (SVR)) for predicting HSE-calculated electronic properties are shown in Figure 5. The elements themselves constituted the sole set of predictors for the training of the models (Set-II). All models were validated and developed with 10-fold cross-validation, which is reported as the prediction of reliable models. The figure is filled by rows of different electronic properties and columns of individual machine learning algorithms. The calculated HSE values, plotted against the predicted, are compared in each plot with the dashed line which has a 45° slope, representing the perfect agreement between prediction and calculation. The LR model shows that the elemental descriptors are not linearly related to the target electronic properties, as evidenced by its relatively poor performance. This makes it impossible to capture all the underlying patterns in the data using simple linear models. More sophisticated nonlinear methods, like RFR and SVR, achieve much better predictive accuracy, however. These models work better in learning the complex feature–target relationships and thus make predictions very close to the value calculated by the HSE. The results demonstrate the significance of considering nonlinear interactions when predictive models, based only on elemental properties as input features, are used to forecast electronic properties.

As shown, SVR is determined to have the lowest RMSEs for the band gaps (0.57 eV), VBMs (0.49 eV) and the CBMs (0.43 eV) which is almost half the RMSEs of the LR model. It has to be noted that although the performance of the SVR model developed here is not as high as that of LR, when the DFT-PBE results are used as features, the attraction of the SVR model, is its simplicity of the feature space, without the need to resort to any DFT calculation. The VBMs' order and CBMs' order by the feature importance is different from the order for the band-gap predictors: The average mass does not appear to be as important as the band-gap predictors. The most crucial parameter affecting the CBMs is affinity parameter and the difference in electronegativity of the cation-anion is important to determine the VBMs.

Set-III Predictors

Machine learning models can be used to predict the performance of the system, further improvement in the predictive performance is possible by combining descriptors from Set-I and Set-II to form a larger feature space, termed as Set-III. The combined predictor set contains both DFT-PBE derived electronic properties as well as fundamental elemental properties, thus providing more information relevant to the target electronic properties. The fact that PBE calculations alone offer useful predictive information about the HSE results can be considered an instance of model stacking. This can be a very effective way to exploit the power of several information sources, and typically results in better prediction accuracy, more robust models, and better overall performance than models using only one set of features.

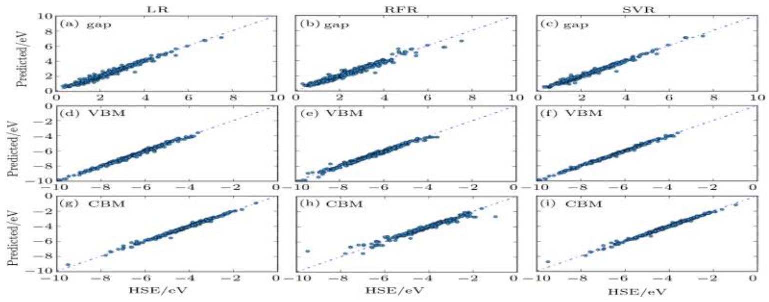


Fig. 6. The calculated HSE data have been compared to the predicted fundamental band gaps and bands alignments by LR, RFR and SVR models along with the interesting PBE results and the choice of predictors (predictor set-III) used as constitutive elements. Only the outcome of the predictions in the validation sets (10-fold cross validated) are visualised. The columns and rows of the subfigures show the different (selected) electronic properties and different selected predictive model, respectively. The dashed lines are line of sight such that prediction values are same as the computed values of the HSE.

Discussion

Correct prediction of the bandgap depends on the correct estimation of the valence band maximum (VBM) and the conduction band minimum (CBM). When only a small number of descriptors is available, this goal can be difficult to achieve because they may not accurately represent all of the factors that affect both electronic states. However, the prediction of VBM or CBM is usually easier to predict and may be possible with reasonable accuracy with available descriptors. Incorporation of a more complex feature space (such as interaction terms, derived descriptors or other material properties) might enhance the predictive power of the models. The advanced feature engineering techniques were not studied in the present work, however. By comparing the performance of various machine learning models (model selection), we have drawn the general principle of model selection. RMSE and MAPE are used as the models' accuracy metrics for the LR, RFR and SVR. In the light of these three models, it is observed that irrespective of the dimension of the feature space to be used (set-II or set-III) the SVR model is the best accurate model in the present study. Specifically, the knowledge of the elemental information alone is greatly favourabler for the SVR, since it is likely to produce a lot of features, than the other two approaches (i.e., predicting desired materials properties). Therefore it is recommended to be applied for SVR in case there are no previous DFT-PBE results. But, when numbers are provided for DFT-PBE, it is probably best to start with a good model, which is LR. This makes it possible to implement a simple analytical model for the relationship between the DFT-PBE values and features of the DFT-PBE model, and thus a simple prediction of the target values.

For different variations of the target properties the difference in performance of each target property of one machine learning model is identified. Whereas VBM and CBM are strongly related to the band gap, the value of VBM and CBM in almost all cases RMSE and MAPE is less than that of the band gap. The inhomogeneity of accuracy is related with the following: Correct prediction of the bandgap depends on the correct estimation of the valence band maximum (VBM) and the conduction band minimum (CBM). When only a small number of descriptors is available, this goal can be difficult to achieve because they may not accurately represent all of the factors that affect both electronic states. However, the prediction of VBM or CBM is usually easier to predict and may be possible with reasonable accuracy with available descriptors. Incorporation of a more complex feature space (such as interaction terms, derived descriptors or other material properties) might enhance the predictive power of the models.

The materials highlighted in this work are only a few of the wide variety of materials that can be devised with the proposed isoelectronic materials design strategy. The trained machine learning models, especially the Support Vector Regression (SVR) model trained with Set-II descriptors, give a suitable framework for predicting the electronic properties of other members of this material family. These predictions are largely based on the elemental characteristics, which means that they can quickly and efficiently be used to screen a vast number of materials with minimal computational effort. With this power they can be used as valuable tools to discover and design new semiconducting materials for future electronic and optoelectronic applications.

Phase-I can be stabilized with materials like CaSe by alloying with compounds like SnSe, for example. The machine learning models developed in this study are also applicable to the prediction of electronic properties of such alloys. Prediction of improved accuracy can be obtained by using weighted averages of the properties of the elements of the alloy. The trained models are therefore a practical and computationally efficient method for estimating bandgaps and band-edge alignments of the two-dimensional materials of wide range. This is a very useful feature for screening potential candidates for possible applications in electronics, optoelectronics and photocatalysis without extensive first-principles calculations.

Conclusions

The family of two-dimensional (2D) semiconductors isoelectronic with phosphorene was studied using machine-learning algorithms. An extensive analysis of the electronic properties of these materials, such as the fundamental band gaps and band edge alignments, was carried out to assess their potential for advanced electronic applications. The crystal structure is the same for all the materials studied, but the bandgap varies greatly among them, from almost metallic to wide bandgap semiconducting. A machine learning model was trained to precisely predict the bandgap energies and the positions of valence band maximum (VBM) and conduction band minimum (CBM). The Support Vector Regression (SVR) algorithm showed the best predictive performance with RMS errors of below 0.15 eV for VBM and CBM predictions with PBE-derived descriptors, as well as with elemental descriptors. In addition, the developed models were readily applicable to more complex material systems, such as ternary alloys and quaternary compounds, greatly improving their usefulness for a wider class of 2D semiconductors. They reveals the capacity of machine learning for speedy identification and characterization of materials on economy scale, as it is an efficient and useful technique for rational design, discovery and screening materials. It can save time and power in exploring materials aiding next generation's electronic development and other devices.

References

Zhu, Z., Dong, B., Guo, H., Yang, T. and Zhang, Z., 2020. Fundamental band gap and alignment of two-dimensional semiconductors explored by machine learning. Chinese Physics B, 29(4), p.046101.

Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O. and Walsh, A., 2018. Machine learning for molecular and materials science. Nature, 559(7715), pp.547–555.

Ward, L., Agrawal, A., Choudhary, A. and Wolverton, C., 2016. A general-purpose machine learning framework for predicting properties of inorganic materials. Npj Computational Materials, 2(1), pp.1–7.

Pilania, G., Mannodi-Kanakkithodi, A. and Ramprasad, R., 2016. Machine learning bandgaps of double perovskites. Scientific Reports, 6, p.19375.

Na, G.S., Jang, S., Lee, Y.L. and Chang, H., 2020. Tuplewise material representation based machine learning for accurate band gap prediction. Journal of Physical Chemistry A, 124(50), pp.10616–10623.

Borlido, P., Schmidt, J., Huran, A.W., Tran, F., Marques, M.A.L. and Botti, S., 2020. Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning. npj Computational Materials, 6(96).

Zhang, Y., Xu, W., Liu, G., Zhang, Z., Zhu, J. and Li, M., 2021. Bandgap prediction of two-dimensional materials using machine learning. PLOS ONE, 16(8), e0255637.

Alibagheri, S., Mortazavi, B. and Rabczuk, T., 2021. Predicting the electronic and structural properties of two-dimensional materials using machine learning. Computers, Materials & Continua, 67(1), pp.1287–1300.

Wu, L., Xiao, Y., Ghosh, M., Zhou, Q. and Hao, Q., 2020. Machine learning prediction for bandgaps of inorganic materials. ES Materials and Manufacturing, 9, pp.34–39.

Huang, Y., Yu, C., Chen, W., Liu, Y., Li, C., Niu, C., Wang, F. and Jia, Y., 2019. Band gap and band alignment prediction of nitride-based semiconductors using machine learning. Journal of Materials Chemistry C, 7, pp.3238–3245