JCEC
THE JOURNAL OF ENGINEERING AND EXACT SCIENCES

OPEN ACCESS

The Journal of Engineering and Exact Sciences – jCEC, Vol. 09 N. 09 (2023)
journal homepage: https://periodicos.ufv.br/jcec
eISSN: 2527-1075
ISSN: 2446-9416

Perovskite Solar Cell: Chemical Composition and Bandgap Energy via Machine Learning

Célula solar de perovskita: composição química e energia de bandgap via aprendizado de máquina

Article Info:
Article history: Received 2023-09*et al*-01 / Accepted 2023-12-20 / Available online 2023-12-28
doi: 10.18540/jcecvl9iss9pp17804

Filipi França dos Santos
ORCID: https://orcid.org/0000-0003-4891-2538
Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo, RJ, Brazil
E-mail: filipi.santos@iprj.uerj.br
Kelly Cristine da Silveira
ORCID: https://orcid.org/0000-0002-6055-6778
Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo, RJ, Brazil
E-mail: kelly.dasilveira@iprj.uerj.br
Gesiane Mendonça Ferreira
ORCID: https://orcid.org/0000-0001-9338-1217
Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo, RJ, Brazil
E-mail: gmferreira@iprj.uerj.br
Daniella Herdi Cariello
ORCID: https://orcid.org/0000-0002-0869-4911
Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo, RJ, Brazil
E-mail: dherdi@iprj.uerj.br
Mônica Calixto de Andrade
ORCID: https://orcid.org/0000-0001-6530-0651
Universidade do Estado do Rio de Janeiro, Instituto Politécnico, Nova Friburgo, RJ, Brazil
E-mail: mcalixto@iprj.uerj.br

Resumo

O crescimento exponencial nas publicações e aplicações das células fotovoltaicas de perovskita destaca sua relevância na conversão de energia e na mitigação das emissões de carbono. No período de 2009 a 2023, a eficiência dessas células evoluiu significativamente, passando de 3,9% para 25,7%. A capacidade adaptativa das estruturas perovskitas para a absorção do espectro solar e o deslocamento de corrente é fortemente influenciada pela energia da banda de gap, idealmente situada entre 1,3 e 1,7 eV. Embora diversas composições de perovskita possam atingir essa faixa de energia, as sínteses continuam sendo empíricas, apresentando desafios para a viabilidade experimental. Nesse contexto, a utilização de bancos de dados experimentais, fornecidos por pesquisadores globais, emerge como uma abordagem eficaz para acelerar e viabilizar a pesquisa das estruturas perovskitas destinadas a células fotovoltaicas. Este estudo empregou o banco de dados da plataforma *MaterialsZone* para alimentar algoritmos de aprendizado de máquina, concentrando-se nas técnicas de Máquina de Vetores de Suporte (SVM) e Floresta Aleatória (RF) para a predição de energia da banda de gap em uma composição específica de perovskita. Ao direcionar os experimentos de síntese para composições particulares, orientadas pelas predições dos modelos, é possível alcançar a energia da banda de gap desejada de maneira eficiente. Esse enfoque resulta em avanços mais rápidos na pesquisa, reduzindo os custos associados à síntese de perovskitas. O modelo RF apresentou um erro percentual médio de 5,13%, desvio padrão do erro percentual de

6,99%, e Erro Quadrático Médio (RMSE) de 0,119. Por outro lado, o SVM registrou um erro percentual médio de 4,05%, desvio padrão do erro percentual de 6,45%, e RMSE de 0,881. Os modelos desenvolvidos não apenas demonstram uma alta capacidade preditiva, mas também fundamentam o entendimento da relação entre a composição química e os valores de energia da banda de gap das perovskitas. Ao empregar algoritmos de aprendizado de máquina, este trabalho abre caminho para otimizações direcionadas, e ainda, impulsiona avanços substanciais na fabricação de células fotovoltaicas baseadas em perovskita.
Palavras-chave: Perovskita. Células fotovoltaicas. Bandgap. Máquinas de Vetores de Suporte (SVM). Floresta Aleatória (RF)

Abstract
The exponential growth in publications and applications of perovskite photovoltaic cells highlights their significance in energy conversion and carbon emissions mitigation. From 2009 to 2023, the efficiency of these cells has significantly increased from 3.9% to 25.7%. The adaptive capacity of perovskite structures for solar spectrum absorption and current displacement is strongly influenced by the bandgap energy, ideally situated between 1.3 and 1.7 eV. Although various perovskite compositions can potentially attain this energy range, the synthesis methodologies remain empirically driven, presenting challenges to experimental viability. In this context, leveraging experimental databases provided by global researchers emerges as an effective approach to expedite and enable research on perovskite structures for photovoltaic cells. This study utilized the comprehensive MaterialsZone database to feed machine learning algorithms, focusing on Support Vector Machine (SVM) and Random Forest (RF) methodologies to predict the bandgap energy in a targeted perovskite composition. By conducting synthesis experiments towards specific compositions guided by model predictions, it becomes feasible to efficiently achieve the desired bandgap energy. Such a strategy not only accelerates research progress but also serves to curtail costs associated with the synthesis of perovskite materials. The RF model exhibited an average percentage error of 5.13%, a standard deviation of the percentage error of 6.99%, and a Root Mean Square Error (RMSE) of 0.119. In contrast, the SVM model recorded an average percentage error of 4.05%, a standard deviation of the percentage error of 6.45%, and RMSE of 0.881. These developed models not only demonstrate high predictive capacity but also contribute substantively to the comprehension of the intricate relationship between the chemical composition and bandgap energy values of perovskites. By deploying machine learning algorithms, this work paves the way for targeted optimizations and considerable strides in the manufacturing of perovskite-based photovoltaic cells.
Keywords: Perovskite. Photovoltaic cells. Bandgap. Support Vector Machines (SVM). Random Forest (RF)

1. Introduction

Organic-inorganic halide-based solar cells, known as Perovskite Solar Cells (PSCs), have shown significant progress in energy conversion. This advancement is partially due to their ability to tune their bandgap energy, which ideally ranges between 1.3 and 1.7 eV (Park *et al*., 2016; Hui *et al*., 2023). With low-cost constituent materials and simplified fabrication processes for deposition and thin-film production, there is a challenge in surpassing already well-established technologies in terms of longevity (Fu *et al*., 2022). The interest around organic-inorganic halide-based perovskites, with the general formula ABX3, lies not only in their intrinsic properties but also in the diversity of their compositions, offering a broad spectrum of possibilities (Prochowicz *et al*., 2019). However, the search for the optimized composition, resulting in an ideal bandgap energy, is challenging given the countless possible combinations. Given the scale of this task, conventional experimentation approaches have their limits, along with associated high costs. Computational intelligence, with an emphasis on machine learning, has gained considerable attention in the discovery of new materials. Advanced algorithms can make fast and accurate predictions, allowing the identification of material compositions with desired properties more efficiently and economically (Schmidt *et al*., 2019;

Butler *et al*., 2018; Shi *et al*., 2018; Da Silveira *et al*., 2023; Hui *et al*., 2023; Li *et al*., 2023). While the use of machine learning in predicting material properties is not novel, its application specifically in predicting the bandgap energy of perovskites remains emergent and holds great potential for optimizing the performance of photovoltaic devices. In this study, predictive models utilizing machine learning approaches were implemented to determine the bandgap energy of perovskites based on their chemical composition. Grounded in data from the MaterialsZone platform (Jacobsson *et al*., 2022), this work aims not only to prevent the synthesis of nonviable materials or those with low response for photovoltaic applications but also to identify and prioritize the synthesis of perovskites that computational modeling suggests as ideal. For this, the Random Forest (RF) and Support Vector Machine (SVM) algorithms were employed, and a comparative analysis of their effectiveness was conducted. Additionally, all explored compositions underwent rigorous screening, preprocessing, and extraction of pertinent attributes.

2. Methodology

To offer practical insights into the design of perovskite materials, this research follows a comprehensive workflow. It encompasses initial screening and data preparation, followed by feature extraction, which involves the extraction of elementary properties. Subsequently, the process includes model implementation and optimization.

2.1 Database

The database from the MaterialsZone platform was employed. A consolidated database available to the scientific community, focusing on perovskites used in solar cells (Jacobsson *et al*., 2022), provides comprehensive insights into the current state of these materials. The dataset explored in this study includes information about 43,239 perovskites.

2.1 Screening and Data Preprocessing

A screening and preprocessing procedure was conducted on the initial dataset derived from MaterialsZone to ensure the quality and relevance of the data used in this investigation. Initially, the dataset encompassed 43,239 perovskites. The screening process is illustrated in Figure 1. In the initial data screening phase, the removal of duplicates resulted in a reduction in the total number of perovskites to 43,098. The following step was characterized by the removal of perovskites with identical composition and bandgap energy, which differed only in specific characteristics of solar cell fabrication and characterization. These differences included the deposition method, total solar cell area, additive concentration, surface roughness, and thickness of the solar cell's perovskite layer. This refinement resulted in a set of 2,237 uniquely composed perovskites. In the next step, perovskites not adhering to the ABX3 general configuration were discarded, updating the total dataset to 2,004 perovskites. The fourth step involved excluding perovskites lacking bandgap energy, bringing the count down to 865. The penultimate phase of preprocessing aimed to restrict the dataset solely to inorganic and organic perovskites containing simple organic cations extensively recognized in the photovoltaic materials literature (Fu *et al*., 2022; Park, 2015). Only formamidinium and methylammonium-based organic perovskites were retained. This step resulted in 672 perovskites. In the final stage, the dataset was narrowed down to include only the composition and respective bandgap energies of the perovskites, ensuring direct alignment with the specific objectives of this study. This procedure is crucial to ensure that the implemented models are trained with perovskites relevant to the study's proposal, which involves determining the bandgap energy based on their chemical composition.

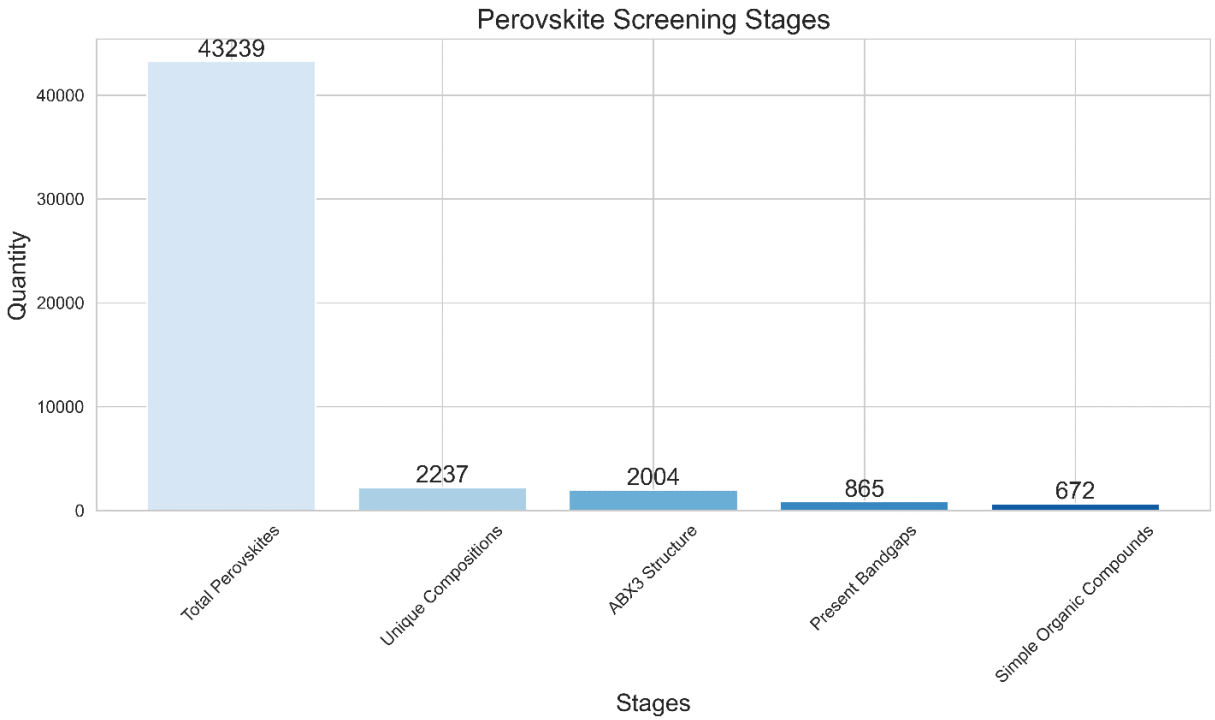

Figure 1 – Graphical representation of screening stages and the remaining number of perovskites after each step.

2.3 Feature Extraction

After refinement, the database now encompasses 672 perovskites, distinguished by two principal attributes: composition and bandgap energy. Subsequently, a meticulous analysis of these compositions is undertaken to unveil intrinsic attributes linked to each chemical composition. The feature extraction step plays a central role in the data preprocessing for machine learning, converting specific properties of each perovskite's composition into quantitative values interpretable by the algorithms. Consequently, the selected attributes are based on the fundamental characteristics of the elements and the composition of the perovskites. Figure 2 provides a succinct representation of these attributes and their correlations. A detailed description of the extraction of the 149 attributes would exceed the scope of a concise visualization.

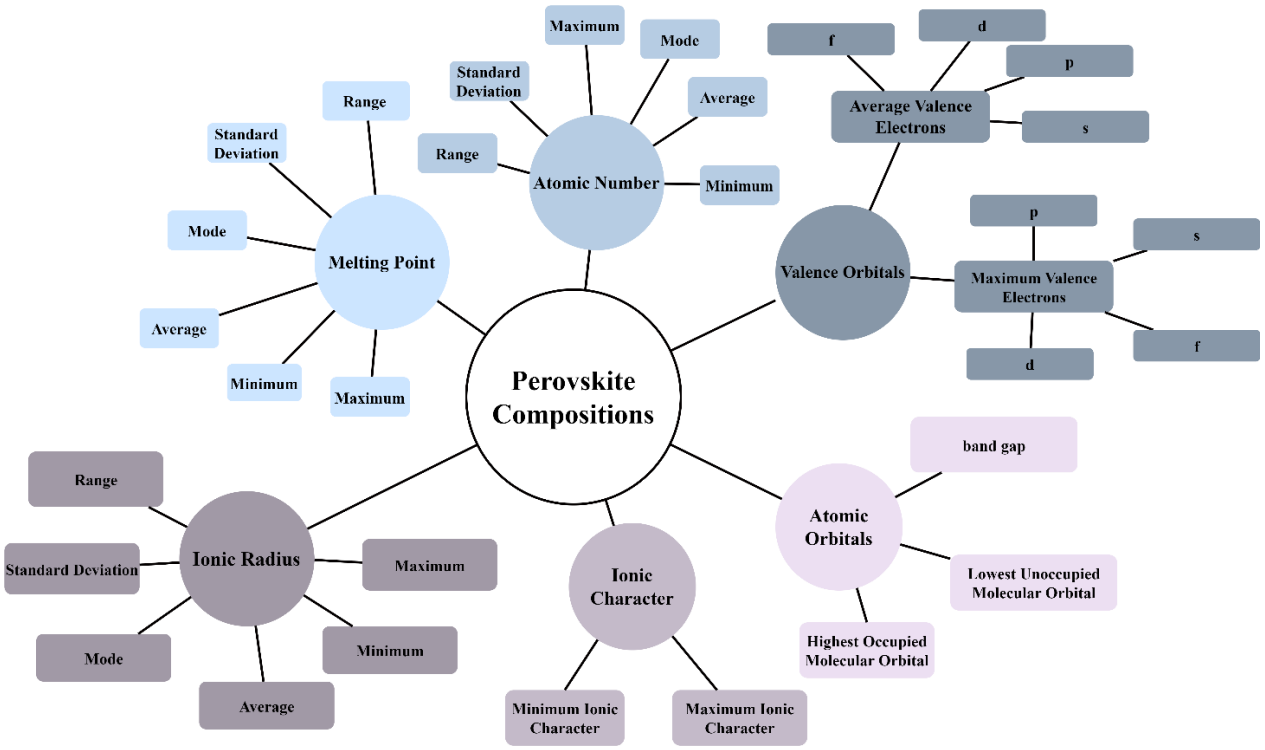

Figure 2 – Diagram illustrating intrinsic properties extraction from perovskite compositions.

Elementary property - Magpie method: A comprehensive examination of perovskite attributes is essential for a nuanced understanding of their chemical and physical characteristics. Leveraging the Materials-Agnostic Platform for Informatics and Exploration (Magpie), a method well-established and validated within the scientific community (Ward *et al.*, 2016; Zhang *et al.*, 2018), was applied to extract attributes associated with perovskite compositions in the designated database. Utilizing computational tools, specifically *MatMiner* and *Pymatgen* in the Python programming language, Magpie systematically conducts a thorough analysis of the chemical compositions of perovskites, resulting in the extraction of a comprehensive collection of attributes that capture the intrinsic properties of these materials. The implementation of this methodology facilitates the transformation of chemical compositions into an extensive numerical dataset, thereby enabling a more insightful analysis through the application of machine learning algorithms. Consequently, this method delineates 22 distinct attributes for each composition, covering essential physical and chemical properties, covering atomic mass, atomic radius, electronegativity, and melting point. Statistical parameters, including minimum, maximum, mean, mode, standard deviation, and range were determined.

Extraction of valence orbital characteristics: Characteristics related to valence electrons were extracted to gather insights into the electronic properties of perovskites. These data include the average and fraction of electrons in the *s*, *p*, *d*, and *f* orbitals. Such indicators have been extensively employed in the literature to elucidate the chemical and electronic properties of compounds (Jain *et al.*, 2013).

Characteristics of atomic orbitals: This investigation delves into the characteristics of atomic orbitals, focusing on the energetic attributes of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), which provides a comprehension of the electronic properties inherent to perovskites. These parameters are extracted to unveil insights into the electronic structure of perovskite materials. The energy difference between these orbitals has a significant interest, as it plays a crucial role in determining the electronic and optical behavior of materials (Ward *et al.*, 2018).

Ionic properties: Attributes representing the specific ionic properties of perovskites were extracted, addressing both maximum and average ionic character values. The term "ionic character" delineates the extent to which an atom tends to donate or accept electrons, serving as a fundamental indicator for evaluating the physical and chemical properties of perovskites. A higher ionic character suggests a pronounced tendency towards the formation of ionic-character chemical bonds, while lower values indicate more pronounced covalent characteristics. The importance of these characteristics in material properties has been extensively discussed in the literature, emphasizing their relevance in the context of perovskite studies (Meng *et al*., 2016).

Stoichiometry metrics: In this phase, a set of attributes was extracted on established metrics defining the stoichiometry of the composition, with a specific focus on Norm-0, Norm-2, Norm-3, Norm-5, Norm-7, and Norm-10. These metrics aim to provide an in-depth understanding of the relative proportions of different elements comprising the perovskite. These stoichiometry patterns are essential for decoding the inherent complexity of the compositions and have demonstrated their significance in prior studies in the field of material science (Jha *et al*., 2018).

With the completion of this attribute extraction stage, a comprehensive representation of the perovskite compositions was achieved. The combination of these attributes provides a broad panorama of the fundamental characteristics of these materials, paving the way for more precise modeling of their properties.

2.4 Implementation of Models

Random Forest (RF) and Support Vector Machine (SVM): For predicting the bandgap energy of perovskites, two machine learning algorithms were selected, taking into consideration the complexity and the extensive number of attributes present in the database, which comprises 672 perovskites and 149 extracted attributes for each composition. The chosen algorithms were Random Forest (RF) and Support Vector Machine (SVM). The models were implemented using the *scikit-learn* library, a common choice for regression problems (Pedregosa *et al*., 2011). The Random Forest algorithm is particularly suited for handling high-dimensional data and can model complex non-linear relationships (Keller *et al*., 2019). Moreover, Random Forest is known for its robustness and flexibility, being less susceptible to overfitting compared to other algorithms (Shah *et al*., 2020; Da Silveira *et al*., 2023; Santos *et al*., 2023). Support Vector Machine, in turn, has been recognized as a robust tool in machine learning due to its ability to handle large dimensional spaces and its effectiveness in finding optimized separation margins. This characteristic aligns with the present study involving a substantial dimensionality (149 attributes). The effectiveness of SVM has been extensively documented across applications ranging from pattern recognition to complex regression problems, making it a pertinent choice for modeling bandgap energy in perovskites.

Data splitting: The dataset was separated into inputs, comprising the extracted attributes, and output, represented by the bandgap energy of the materials - a central variable targeted for estimation by the model. To ensure a meticulous and robust evaluation of the model's performance, the 4-fold cross-validation technique, a form of *k-fold cross-validation*, was employed for both training and validation. Cross-validation is a well-established and extensively used technique in machine learning, recommended in several studies for complex analyses involving multiple attributes (James *et al*., 2013; Arlot & Celisse, 2010). This technique provides a more reliable estimate of the model's ability to generalize to new data, as it employs different data splits for training and testing, thereby mitigating the risk of overfitting to the training data (Hastie, Tibshirani & Friedman, 2009).

Hyperparameter optimization: To optimize the hyperparameters of the models employing Random Forest and Support Vector Machine for the regression problem, a random search strategy was adopted. Unlike a comprehensive grid search, this technique assesses random values within a predefined sample of hyperparameter combinations, offering significant computational resource

savings. For both algorithms under evaluation, 100 iterations were conducted, and each set of hyperparameters underwent evaluation using 4-fold cross-validation to ensure robustness in the selections made.

In the case of the Random Forest, the range of considered trees spanned from 50 to 5000, with increments of 20. The number of attributes randomly selected for each tree in the forest was determined by the square root of the total number of available attributes. The maximum depth of the trees ranged from 10 to 40, in intervals of 10, including the option of no depth limit as well. The minimum samples for splitting were explored from 2 to 10, while the minimum samples in a leaf node varied between 1 and 4.

For the Support Vector Machine model, the investigated hyperparameters included the regularization parameter, with values ranging from 0.01 to 1000; the tolerance margin parameter, varying from 0.001 to 1; and different kernel functions, such as linear, polynomial, radial basis function, and sigmoid. In the case of the polynomial function, the degrees considered were 2, 3, and 4. Additionally, the coefficient for the polynomial, radial basis function, and sigmoid kernels were analyzed, with both categorized and automatic values. The evaluation metric adopted for the models was the mean absolute percentage error.

3. Discussion of results

This study explores the optimization of hyperparameters and assesses the performance of Random Forest and Support Vector Machine models in predicting perovskite bandgap energy. Optimal hyperparameters for both models were systematically determined using a random search strategy, allowing for a thorough exploration of the parameter space. Afterwards, utilizing statistical metrics, a comparative analysis is presented to evaluate their predictive capabilities. In this section, the information has been systematically categorized to improve clarity and understanding of the results.

3.1 Hyperparameter Optimization Results of the Models

For the optimization process, the random search strategy was employed. Specifically concerning the Random Forest, the optimized hyperparameters were a total of 1890 estimator trees, a maximum depth of 20, combined with a minimum of 3 samples to perform a split and requiring at least one sample in each leaf node. This information was obtained after evaluation within the predefined hyperparameter range. Regarding the SVM model, the optimization search delineated the subsequent hyperparameters: the radial basis function kernel paired with a categorized coefficient. Additionally, the established tolerance margin was 0.01, with a degree of 4 for the kernel function, and a regularization parameter set at 10.

3.2 Distribution of Bandgap Prediction Errors in Perovskites by the Random Forest Model

The application of the Random Forest model to predict the bandgap of 672 perovskites involved the utilization of the cross-validation technique. This approach ensures that the prediction for each perovskite is conducted without the model having access to its experimental bandgap value, thereby providing an authentic assessment of the model's predictive capacity.

The model's performance was initially assessed using the Root Mean Square Error (RMSE), yielding a value of 0.120. This metric expresses the model's precision in estimating the bandgap of perovskites, indicating favorable predictive accuracy. Additionally, the Mean Absolute Percentage Error exhibited a rate of 3.92%. This outcome, coupled with a standard deviation of 5.65%, demonstrates the model's stable and reliable performance across various samples.

Illustrated in Figure 3, a histogram elucidates the distribution of prediction errors. It shows that 252 perovskites had errors ranging from 0 to 1.5%, 365 displayed errors between 1.5 to 10%, 33 manifested errors between 10 to 20%, and 20 perovskites were characterized by errors exceeding 20%.

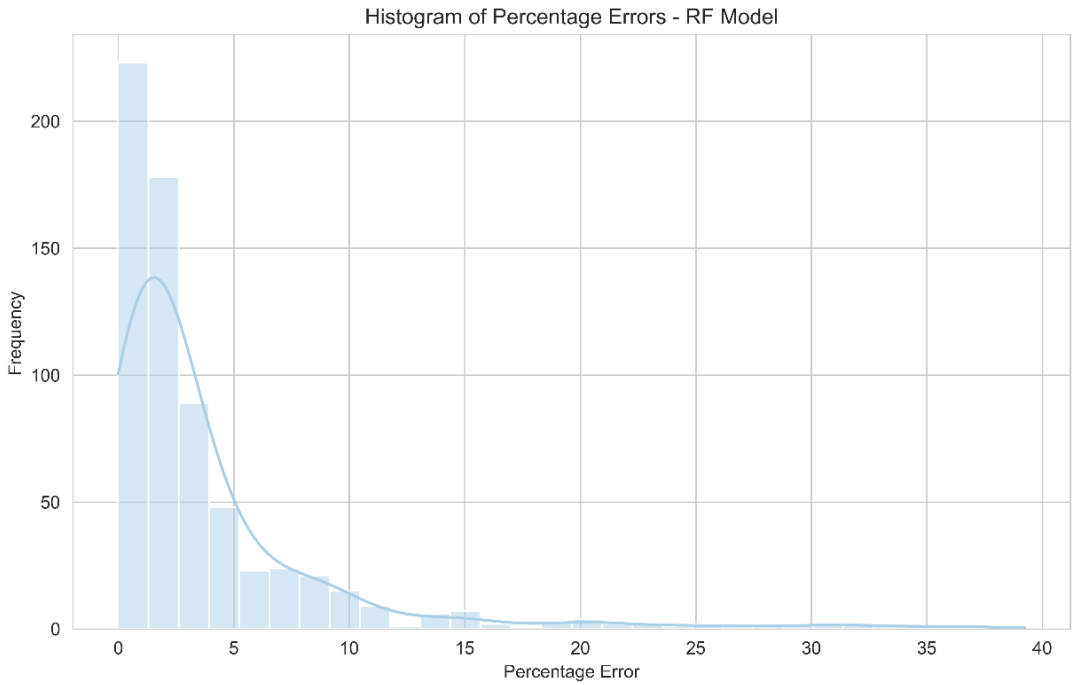

Figure 3 – Distribution of bandgap prediction errors in perovskites using Random Forest algorithm.

To comprehensively illustrate the predictions across all 672 perovskites, a scatter plot was constructed, as shown in Figure 4. This graphical representation displays the relationship between the experimental bandgap values of the perovskites (X-axis) and the corresponding values predicted by the Random Forest model (Y-axis). The visual presentation facilitates the examination of the alignment between the model's predictions and the actual experimental data for all analyzed samples.

In the chart, in an ideal scenario where the model exhibits complete accuracy, each point representing an individual perovskite would perfectly align along the dotted line. This line serves as a symbolic representation of a perfect correlation, where the predicted values precisely match the experimental data. While the overall trend of the predictions follows this ideal line, indicative of the model's commendable performance, certain discrepancies are discernible, suggesting variations in predictions for specific samples.

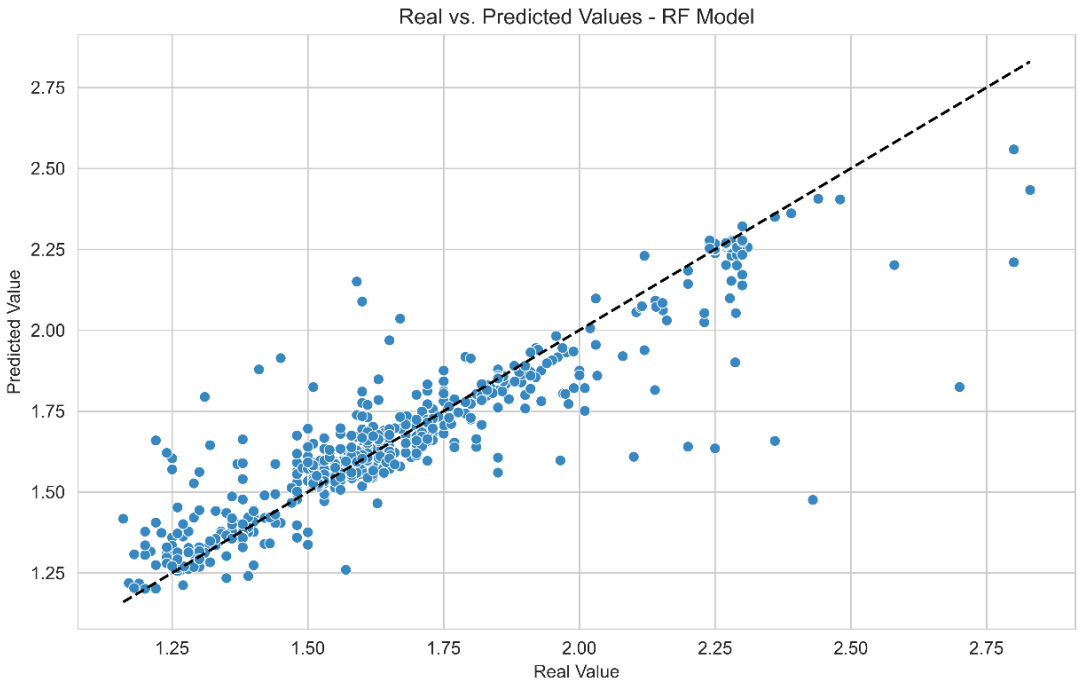

Figure 4 – Comparison of experimental and predicted bandgap values in perovskites: scatter plot generated by Random Forest model.

3.3 Analysis of the SVM Model's Efficiency in Predicting the Bandgap of Perovskites

The application of the Support Vector Machine (SVM) model to predict the bandgap of 672 perovskites involved the implementation of the cross-validation method, similar to the Random Forest model. This approach ensures impartiality and accuracy in the model's predictive outcomes.

Following the optimization of hyperparameters, the SVM model attained a Root Mean Square Error (RMSE) of 0.131. While this result is marginally higher than the RMSE of 0.120 observed in the Random Forest approach, it still signifies commendable precision in bandgap estimation. As illustrated in the histogram in Figure 5, the SVM model exhibited significant performance, especially in the lowest error range. Remarkably, the model accurately categorized 301 out of 672 perovskites with errors ranging from 0 to 1.5%, showcasing heightened precision within this specific range and surpassing the Random Forest model in this regard. Additionally, the histogram reveals that 304 perovskites manifested errors between 1.5 to 10%, 42 perovskites registered errors in the range of 10 to 20%, and 25 perovskites exhibited errors surpassing 20%. This error distribution underscores the SVM model's ability to generate highly accurate predictions, although revealing a substantial number of predictions with larger errors.

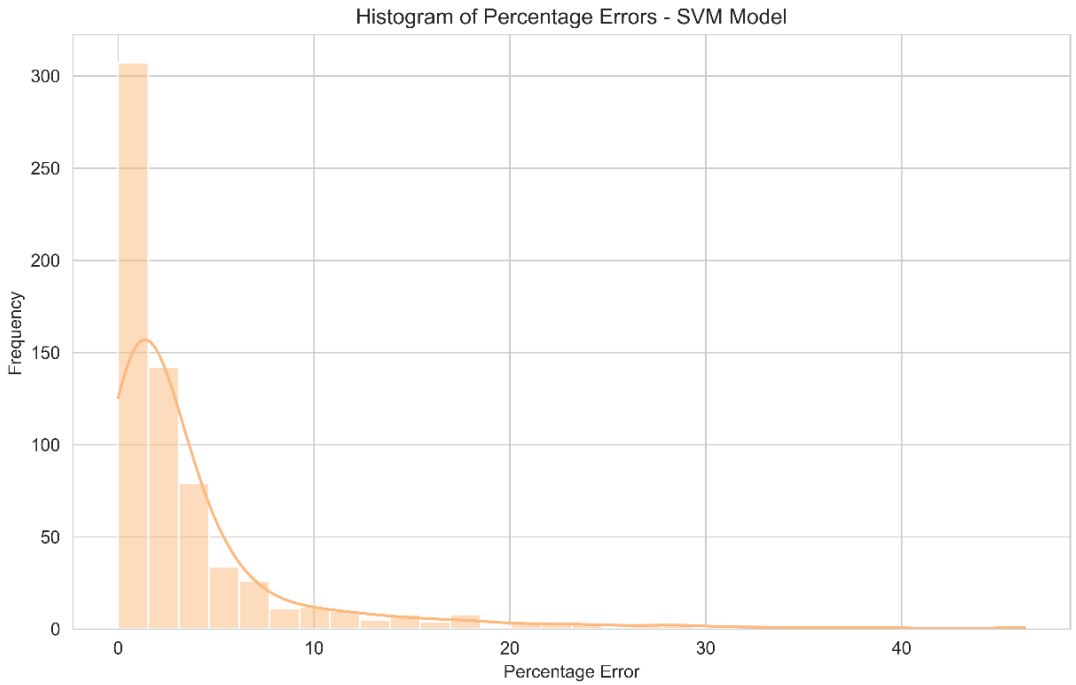

Figure 5 – Distribution of bandgap prediction errors in perovskites using Support Vector Machine algorithm.

The Mean Absolute Percentage Error (MAPE) for the SVM model registered at 3.93%, with a standard deviation of 6.24%. In contrast, the Random Forest exhibited a MAPE of 3.92% with a standard deviation of 5.65%. This comparison implies that while the SVM demonstrates comparable accuracy, it tends to display greater variability in its errors. This attribute is illustrated in Figure 5, presenting a histogram of the error distribution. The histogram highlights that the SVM, despite having more outliers compared to the Random Forest, succeeded in producing a higher number of predictions with elevated precision. This suggests that when the SVM is accurate, its prediction tends to be more precise, but in cases of error, the deviations are more pronounced.

Moreover, Figure 6 presents the scatter plot of the SVM model. Analogous to the one utilized in the Random Forest analysis, this chart provides a clear and visual representation of the relationships between experimental data (X-axis) and model-predicted values (Y-axis) for the entire sample set. Without delving into a reiterated explanation of the chart's concept, it is crucial to emphasize its role as a succinct and efficacious illustration of the SVM's prediction outcomes. This underscores the overall efficacy of the model, notwithstanding some variations and the presence of outliers.

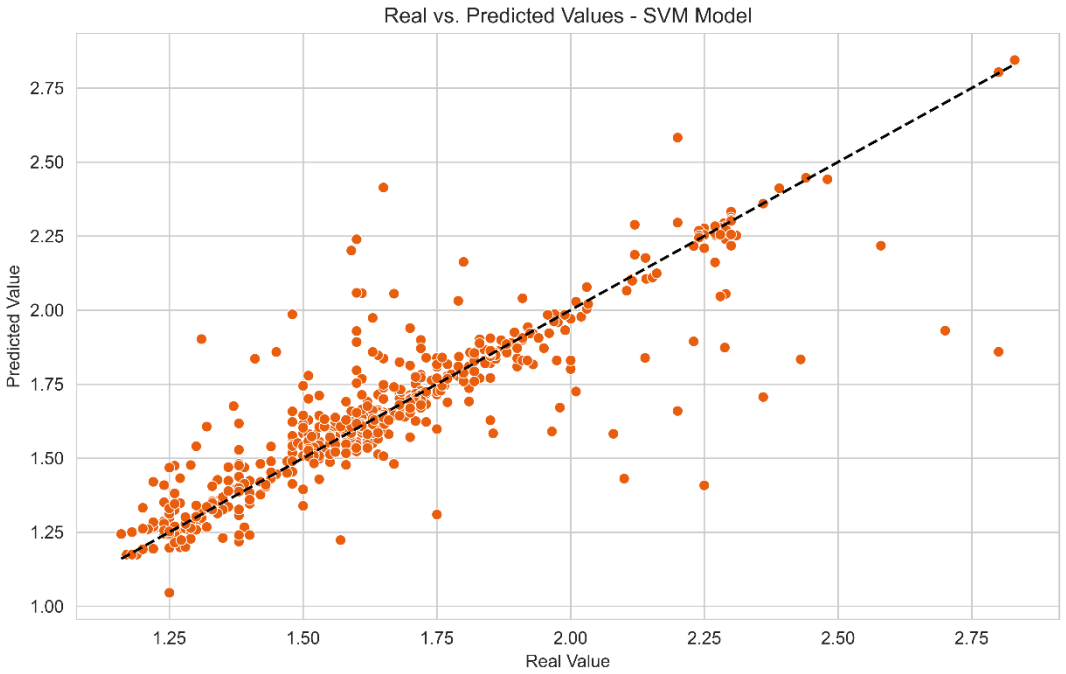

Figure 6 – Comparison of experimental and predicted bandgap values in perovskites: scatter plot generated by Support Vector Machine model.

Scatter plots and histograms were concurrently plotted on a shared scale, presented side by side in Figure 7, to conduct a comprehensive and visual comparative analysis of the Random Forest and Support Vector Machine models. This arrangement allows an instant visual appraisal of the distinctions and similarities between the two methodologies in predicting the bandgap of perovskites. The side-by-side histograms emphasize the distribution of prediction errors, while the scatter plots demonstrate the correlation between the experimental values and those predicted by each model. This visual representation facilitates the understanding of the specific nuances of each approach, offering a direct and intuitive comparative analysis of their efficacies.

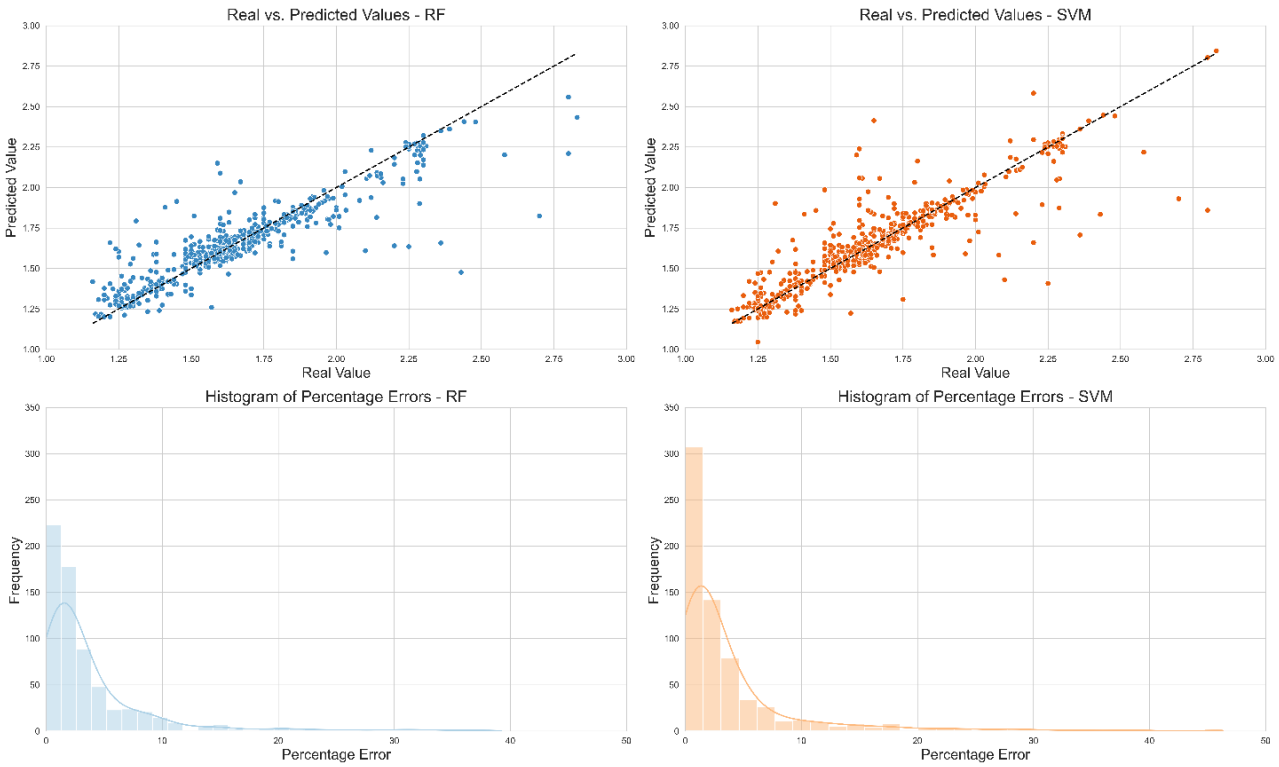

Figure 7 – Comparative visual analysis of Random Forest and Support Vector machine models for bandgap prediction in perovskites: histograms and scatter plots.

4. Conclusions

This investigation successfully predicted the bandgap energy of perovskites using machine learning algorithms, specifically Random Forest and Support Vector Machine. The Random Forest algorithm demonstrated superiority based on the Root Mean Square Error metric, while the Support Vector Machine showed notable accuracy in most predictions, with some significant deviations in specific situations. Challenges were observed for certain perovskites with unique attributes and limited representation in the dataset, emphasizing the need for further studies to understand these cases and refine predictive models. The anticipation of perovskite bandgap energy through machine learning in this work introduces a robust computational approach with significant promise for guiding chemical composition selection, enhancing synthesis efficiency, and advancing photovoltaic cell research.

Acknowledgments

The present study was conducted with the support of the Coordination for the Improvement of Higher Education Personnel Brazil (CAPES), funding code 001. The authors also express their gratitude for the financial support received from the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro (FAPERJ) and the National Council for Scientific and Technological Development (CNPq).

References

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statist. Surv., 4, 40-79. https://doi.org/10.1214/09-SS054

Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 559(7715), 547-555. https://doi.org/10.1038/s41586-018-0337-2

Da Silveira, K. C., Siqueira, M. H. S., Gama, J. M. R., Gois, J. N., Toledo, C. F. M., & Silva Neto, A. J. (2023). A Comparison of Machine Learning Approaches in Predicting Viscosity for

Partially Hydrolyzed Polyacrylamide Derivatives. VETOR-Revista de Ciências Exatas e Engenharias, 33(1), 2-12. https://doi.org/10.14295/vetor.v33i1.15157

Fu, C., Gu, Z., Tang, Y., Xiao, Q., Zhang, S., Zhang, Y., & Song, Y. (2022). From structural design to functional construction: amine molecules in high-performance formamidinium-based perovskite solar cells. Angewandte Chemie International Edition, 61(19). https://doi.org/10.1002/anie.202117067

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: Data mining, inference, and prediction (Vol. 2). New York: Springer. https://doi.org/10.1007/978-0-387-84858-7

Hui, Z., Wang, M., Yin, X., & Yue, Y. (2023). Machine learning for perovskite solar cell design. Computational Materials Science, 226, 112215. https://doi.org/10.1016/j.commatsci.2023.112215

Jacobsson, T. J., Hultqvist, A., García-Fernández, A., Anand, A., Al-Ashouri, A., Hagfeldt, A., ... & Unger, E. (2022). An open-access database and analysis tool for perovskite solar cells based on the FAIR data principles. Nature Energy, 7(1), 107-115. https://doi.org/10.1038/s41560-021-00941-3

Jain, A., Ong, S.P., Hautier, G., Chen, W., Richards, W.D., Dacek, S., Choudhary, A., & Ceder, G. (2013). Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. APL Materials, 1(1), 011002. https://doi.org/10.1063/1.4812323

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). New York: Springer. https://doi.org/10.1007/978-1-4614-7138-7

Jha, D., Ward, L., Paul, A., Liao, W.-C., Choudhary, A., Agrawal, A., Richey, J. (2018). ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition. Scientific Reports, 8(1), 17593. https://doi.org/10.1038/s41598-018-35934-y

Keller, C. A., & Evans, M. J. (2019). Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10. Geoscientific Model Development, 12(3), 1209-1225. https://doi.org/10.5194/gmd-12-1209-2019

Li, W., Hu, J., Chen, Z., Jiang, H., Wu, J., Meng, X., Fang, X., Lin, J., Ma, X., Yang, T., & Cheng, P. (2023). Performance prediction and optimization of perovskite solar cells based on the Bayesian approach. Solar Energy, 262, 111853. https://doi.org/10.1016/j.solener.2023.111853

Meng, L., You, J., Guo, T.F., & Yang, Y. (2016). Recent advances in the inverted planar structure of perovskite solar cells. Accounts of Chemical Research, 49(1), 155-165. https://doi.org/10.1021/acs.accounts.5b00404

Park, N.G., Grätzel, M., Miyasaka, T., Zhu, K., & Emery, K. (2016). Towards stable and commercially available perovskite solar cells. Nature Energy, 1(11), 16152. https://doi.org/10.1038/nenergy.2016.152

Park, N.-G. (2015). Perovskite solar cells: An emerging photovoltaic technology. Materials Today, 18(2), 65-72. https://doi.org/10.1016/j.mattod.2014.07.007

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Prochowicz, D., Runjhun, R., Tavakoli, M. M., Yadav, P., Saski, M., Alanazi, A. Q., ... & Grätzel, M. (2019). Engineering of perovskite materials based on formamidinium and cesium hybridization for high-efficiency solar cells. Chemistry of Materials, 31(5), 1620-1627. https://doi.org/10.1021/acs.chemmater.8b04871

Santos, F. F., Da Silveira, K. C., Carrielo, D. H., Ferreira, G. M., Domingues, G. D. M. B., & Andrade, M. C. (2023). Evaluation of the Thermogravimetric Profile of Hybrid Cellulose Acetate Membranes using Machine Learning Approaches. VETOR-Revista de Ciências Exatas e Engenharias, 33(1), 51-59. https://doi.org/10.14295/vetor.v33i1.15167

Schmidt, J., Marques, M.R., Botti, S., & Marques, M.A. (2019). Recent advances and applications of machine learning in solid-state materials science. npj Computational Materials, 5, 83. https://doi.org/10.1038/s41524-019-0221-0

Shah, K., Patel, H., Sanghvi, D., & Shah, M. (2020). A comparative analysis of logistic regression, random forest and KNN models for the text classification. Augmented Human Research, 5, 16. https://doi.org/10.1007/s41133-020-00032-0

Shi, Z., Li, S., Li, Y., Ji, H., Li, X., Wu, D., ... & Du, G. (2018). Strategy of solution-processed all-inorganic heterostructure for humidity/temperature-stable perovskite quantum dot light-emitting diodes. ACS Nano, 12(2), 1462-1472. https://doi.org/10.1021/acsnano.7b07856

Ward, L., Agrawal, A., Choudhary, A., & Wolverton, C. (2016). A general-purpose machine learning framework for predicting properties of inorganic materials. npj Computational Materials, 2(1), 16028. https://doi.org/10.1038/npjcompumats.2016.28

Ward, L., Dunn, A., Faghaninia, A., Zimmermann, N. E., Bajaj, S., Wang, Q., ... & Jain, A. (2018). Matminer: An open-source toolkit for materials data mining. Computational Materials Science, 152, 60-69. https://doi.org/10.1016/j.commatsci.2018.05.018

Zhang, Y., & Ling, C. (2018). A strategy to apply machine learning to small datasets in materials science. npj Computational Materials, 4(1), 25. https://doi.org/10.1038/s41524-018-0081-z