ICWIP 2023
8th International Conference on Women in Physics
Online Mode
10-14 July 2023

Machine Learning accelerated Prediction of Bandgap of Cubic Perovskites

Ponkiya Zarna D. [1], *Sattigeri Raghottam M.* [2], *Jha Prafulla K.* [1]

[1]*Department of Physics, Faculty of Science, The Maharaja Sayajirao University of Baroda, Vadodara, Gujarat, India-390002.*
[2]*International Research Centre MagTop, Institute of Physics, Polish Academy of Sciences, aleja/Lotnikow 32/46, Warsaw, Poland 02-668*

Introduction

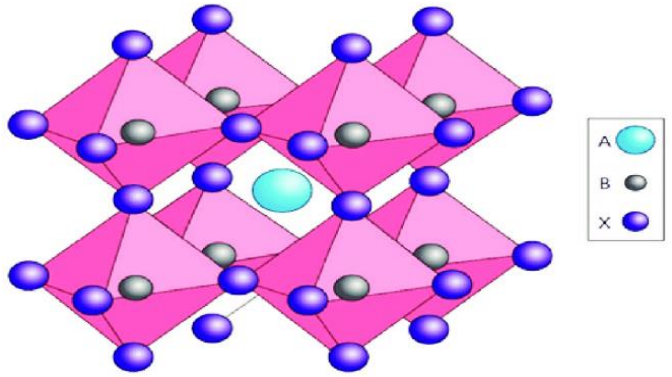
- Why Machine Learning ?
- Perovskites Materials : Simplicitly of structure, flexibility of composition, suitable band structure[1]

Traditional Methodology: Trial & error, Continual synthesis and characterization

Computational material simulation methods : DFT, Monte Carlo simulation, and MD

Data driven - Machine Learning (ML) methods

- Bandgap determines material's electronic and optical properties and plays a crucial role to decide its use in Photocatalysis[2], photovoltaics[3].
- Here, we predicted Bandgap of cubic perovskites(ABX_3) by 5 different Graph Neural Network based ML models only from crystal structure.
- We found CGCNN model optimally predicts the bandgap with the average RMSE of ~ 0.39 eV and R^2-value > 0.90, which is comparable to first principle study(Theoretical) calculations.
- Crystal structure ➔ Bandgap ➔ Application in various fields



ABX_3:
- A and B: cations of different sizes (A being larger than B)
- X: anion (X : O, S, N, F)

Figure: cubic perovskites ABX_3 structure (space group $Pm\bar{3}m$) (Source: https://alchetron.com/Perovskite-%28structure%29)

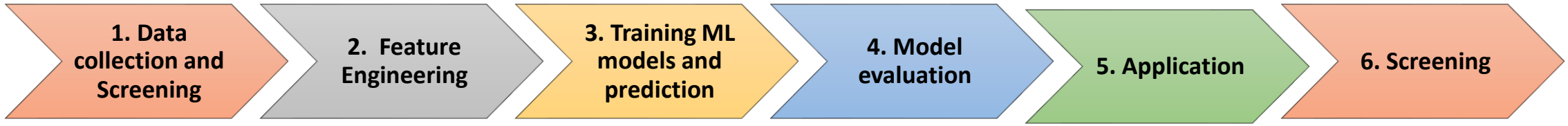
Methodology

Target property = f (Materials) = f (ACS)

Bandgap = *f* (Crystal Structure)

↑

ML Regression



(ML workflow)

1. Data collection :

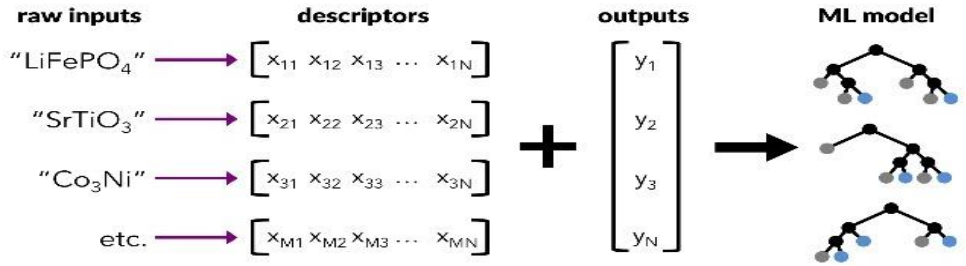
Training Data : 18,000 Castelli perovskites data [4,5]

↓ Screening

735 perovskites with non-zero direct bandgap

↓ Screening

Data of crystal structure and bandgap



2. Feature engineering :

- Descriptors : Crystal structure as input feature and Bandgap as output feature
- 80% Training set - 20% Testing set

Methodology

3. Training: by MatDeepLearn package [6, 7]

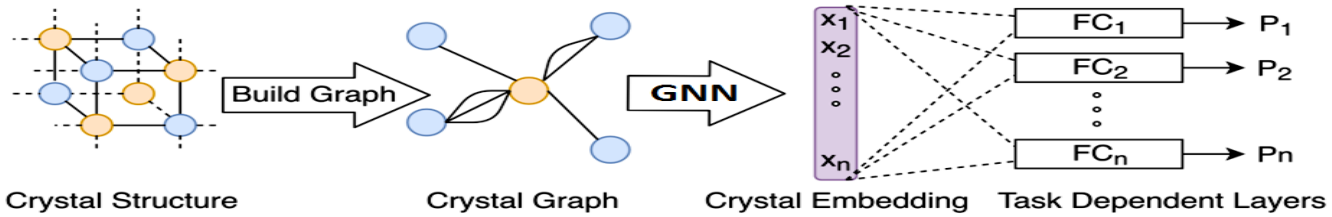


Figure: Working of Graph Neural Network (Source: https://arxiv.org/pdf/1811.05660)

Table 1. List of the five ML Models used and optimized Hyperparameters

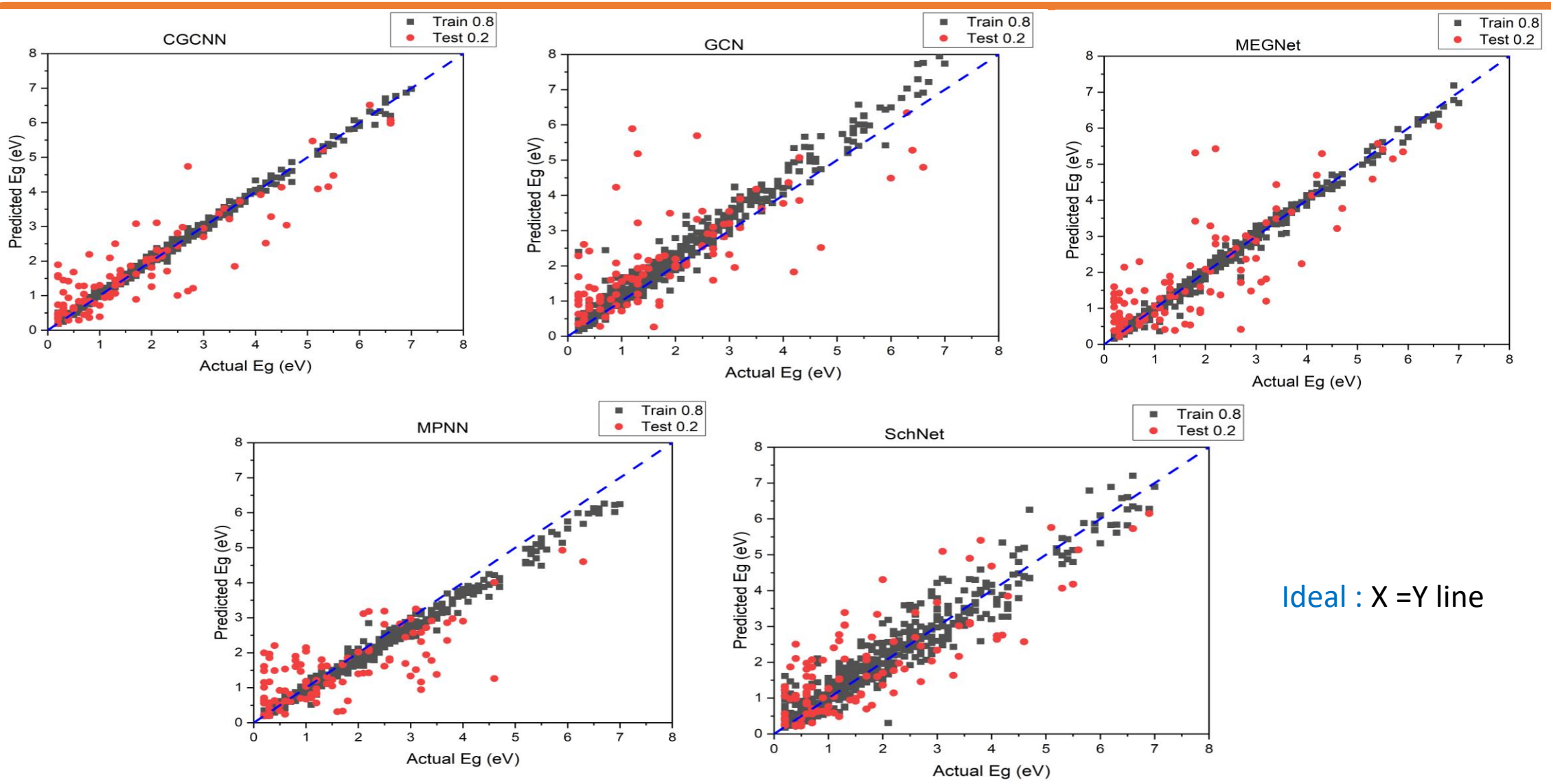
No.	Model	Hyperparameters
1	CGCNN	"dim1": 100, "dim2": 150, "conv_count": 4, "fc_count": 1, "pool": "global_mean_pool", "lr": 0.002, "batch_size": 100, "epochs": 250
2	SchNet	"dim1": 100, "dim2": 100, "dim3": 150, "conv_count": 4, "fc_count": 1, "pool": "global_mean_pool", "lr": 0.0005, "batch_size": 100, "epochs": 250
3	MPNN	"dim1": 100, "dim2": 100, "dim3": 100, "conv_count": 4, "fc_count": 1, "pool": "global_mean_pool", "lr": 0.001, "batch_size": 100, "epochs": 250
4	MEGNet	"dim1": 100, "dim2": 100, "dim3": 100, "conv_count": 4, "fc_count": 1, "pool": "global_mean_pool", "lr": 0.0005, "batch_size": 100, "epochs": 250
5	GCN	"dim1": 100, "dim2": 150, "conv_count": 4, "fc_count": 1, "pool": "global_mean_pool", "lr": 0.002, "batch_size": 100, "epochs": 250

4. Model evaluation :

- Generalization : 20 – Fold Cross Validation (CV)
- Hyperparameter optimization by grid search
- Accuracy and stability are assessed by the Evaluation Metrics:

1. Plots between predicted vs actual Bandgap values
2. Correlation Co-efficient (R^2)
3. Root Mean Squared Error (RMSE)

Results



Ideal : X =Y line

Figure: Predicted vs Actual Bandgap (Eg) by different ML models

Results

Ideal Values : RMSE : (minimum ~ 0) R^2-score : (~ 1)

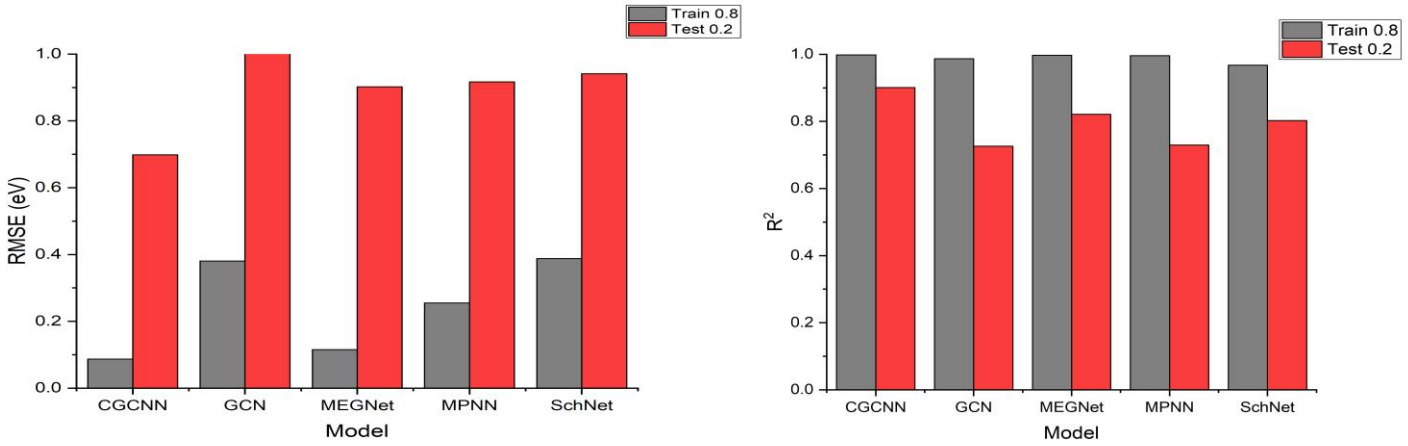


Figure: RMSE and R^2-Score for different ML models

Optimal Configuration: **CGCNN** model having minimum train/test RMSE of **0.08 / 0.69 eV** and highest R^2 **(0.99/0.91)**

Conclusion:

The correlation between crystal structure and bandgap has been established by the ML algorithm and provided an alternate and fast way to DFT calculations to directly predict the bandgap of perovskite materials using only the crystal structure as a feature with good precision (RMSE of 0.08 / 0.69 eV and R^2 (0.99/0.91)). By predicting Bandgap using the explored CGCNN model, we can screen large libraries of perovskite materials for potential use in Photocatalysis, photovoltaics or luminescence within a negligible time.

References

1. Q. Tao et al. In: npj Computational Materials. 7 (1 2021), p. 23.
2. T. Bligaard et al. In: Chemical bonding at surfaces and interfaces, Elsevier. (2008) p. 255
3. R. Olivares-Amaya et al. In: Energy Environ. Sci. 4 (12 2011), p. 4849.
4. I. E. Castelli et al. In: *Energy & Environmental Science*, 5 (10 2012), p. 9034.
5. I. E. Castelli et al. In: Castelli Perovskites Data. *figshare. Dataset*, ***https://doi.org/10.6084/m9.figshare.7215323.v1*** (2012)
6. V. Fung et al. In: *npj Computational Materials*, 7 (1 2021), p. 84
7. T. Xie et al. In: Physical Review Letters. 120 (14 2018), p. 145301.

Acknowledgements

ZDP acknowledges the ASEAN-SERB Project (CRG/2019/001292), Government of India, for the opportunity and financial assistance.

Thank You