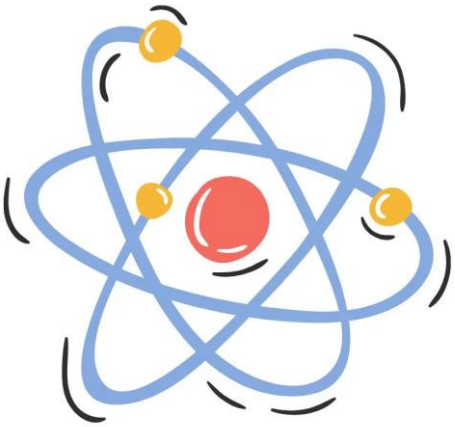

JOURNAL OF DYNAMICS AND CONTROL

VOLUME 8 ISSUE 12

HARNESSING OF DEEP LEARNING AND MACHINE LEARNING ALGORITHMS TO PREDICT THE CRYSTAL STRUCTURE

Mahalekshmi A[1], Heren Chellam[2]

[1]Research Scholar, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, TamilNadu, India

[2]Assistant Professor, Department of Computer Science, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, TamilNadu, India

HARNESSING OF DEEP LEARNING AND MACHINE LEARNING ALGORITHMS TO PREDICT THE CRYSTAL STRUCTURE

Mahalekshmi A[1], Heren Chellam[2]

[1]*Research Scholar, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, TamilNadu, India*

[2]*Assistant Professor, Department of Computer Science, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Tirunelveli, TamilNadu, India*

Abstract: *Crystal structure prediction is the most challenging problem for material science and the chemical field. Nowadays, there are many methods have been developed in machine learning and deep learning methods to predict the crystal structure with less computational time. In this paper, we combine the machine learning and deep learning methods to predict the crystal structure. We propose the Deep-Machine Learning methods to classify the crystal structure. In this work, to identify the four crystal structures. They are orthorhombic, cubic, tetragonal, and rhombohedral with the best accuracy. To harness the deep-machine learning methods to classify the crystal structure, it gives the best accuracy and less time complexity. Corr-ANB model gives the 87% accuracy of predict the crystal structure in 10-fold cross validation. Our method works very fast, and it is inexpensive.*

Keywords: **Crystal structure prediction, machine learning, deep learning, deep-machine learning.**

1. Introduction

Crystal structure prediction in an arbitrary chemical composition remains an unsolved problem in solid-state physics. Theoretically, a local optimisation problem for a potential energy surface specified on the space of atomic coordinates can be solved to identify stable or metastable structures made up of atomic or molecule assemblies. In theory, a local optimisation problem for an energy surface that can be specified on the space of atomic coordinates can be solved to identify stable or metastable structures made by atomic or molecule assemblies. The repeated computing of first-principles potential energy surfaces is a key component of computational crystal structure prediction (CSP), such as density functional theory (DFT) calculations. In the course of these computations, the random search, simulated annealing, basin-hopping, minima hopping, evolutionary algorithm (EA), particle swarm optimization (PSO), Bayesian optimization (BO), and look ahead based on quadratic approximation (LAQA) have been employed as the optimization methods.

More recently, there has been a lot of interest in machine-learning inter atomic potentials since they significantly speed up optimisation by avoiding the laborious ab initio computations. These methods can be classified as sequential search and batch selection. Sequential search methods such as EA and PSO investigate the potential energy surface's local or global minimum. These techniques repeatedly apply ab initio structural optimisation to the now obtained materials by iteratively modifying a current set of material crystalline forms using a predetermined set of genetic modifications. From a predetermined collection of crystals, batch selection techniques like BO and LAQA find more favourable structures with lower expected energy by using surrogate models. Both situations require the use of a crystal structure generator to produce a suitable set of starting structures.

There are many methods have been developed in Machine Learning and Deep Learning to predict the crystal structure. Recently, Yong Zhao et al.,[12, 2020] propose and evaluate machine-learning algorithms for determining the structure type of materials, given only by their compositions. The crystal system and space group prediction of materials is achieved by combining random forest (RF) and multiple layer perceptron (MLP) neural network models using three different feature types: Magpie, atom vector, and one-hot encoding (atom frequency). SMOTE(Synthetic Minority Over-Sampling) is used to reduce the impact of unbalanced data sets. According to their findings, MLP with atom frequency features is the best algorithm to predict structural polymorphisms, whereas RF with Magpie characteristics performs better than other algorithms for binary and space groups and multiclass prediction of crystal systems. Samples of the materials have been collected from the Materials Project.

Here, we propose a procedure to classify crystal structures. Deep learning and machine learning methods have been successfully used to predict crystal structures. In particular, we aim to harness these methods to improve the results of crystal structure prediction.

Our contribution can be summarized as follows:

- We propose combining deep learning and machine learning methods to address the crystal structure prediction problem.
- We evaluated our algorithm and found that it can predict crystal structures with high accuracy, which further leads to successful predictions of their crystal structures.
- We developed deep learning and machine learning methods for the classification of crystal structure predictions.

2. Literature Survey

Gallo-Bueno et al[4, 2022] proposed to use unsupervised machine learning to classify crystal structures based on structural distortion for the Li-argyrodite solid-state electrolyte case study. They found that when used to train a variety of popular unsupervised outlier identification methods, Steinhardt order parameters are highly accurate descriptors of the cubic argyrodite structure. They have shown that DFT-optimized structures may be effectively classified based on the degree of distortion with respect to an archetypal standard by combining SOP-based geometrical descriptors with certain traditional unsupervised outlier detection algorithms.

Sungwon Kim et al.,[9,2020] proposed employ a coordinate-based crystal representation modelled upon point clouds to generate crystal structures using a machine learning generative adversarial network (GAN). In order to identify possible photoanode materials, a proposed generative HTVS (high-throughput virtual screening) framework for creating novel Mg−Mn−O ternary combinations found 23 new crystal compounds with a band gap and adequate stability in an aquatic environment.

Guanjian Cheng et al.,[24, 2022] reported a machine-learning approach for crystal structure prediction, in which a graph network (GN) is employed to establish a correlation model between the crystal structure and formation enthalpies in the specified database, and the search for the crystal structure with the lowest formation enthalpy is speeds up using an optimisation algorithm (OA). They implemented two benchmark databases, such as the open quantum materials database (OQMD) and Matbench (MatB), and three OAs, such as a random searching (RAS), particle-swarm optimization (PSO) and Bayesian optimisation, which can predict crystal structures at a specific periodic cell atom count.

Abeer Abdulaziz Alarfaj et al[2, 2022] proposed a Feature Fusion Deep Learning Model aimed at detecting defects in crystal structures. Their model leverages a novel dense neural network (DNN) using a multitasking approach, validated on a dataset containing 16,000 crystal structures, of which 30% were highly defective. The multitask DNN achieved an impressive 97% accuracy and 96% precision. Additionally, the study compares the computational time for one epoch of training between their model and a leading alternative, highlighting its efficiency. However, the model's reliance on large datasets and the lengthy training times are noted constraints. Future extensions of this work may involve deploying the model in clinical settings for LECD crystal structure analysis in real-time applications.

Satvik Lolla et al.,[6, 2022] proposed a semi-supervised deep-learning approach for crystal structure classification. It was used to address the problem of identifying the space group and Bravais lattice of inorganic crystals. The semi-supervised generative deep learning model described can be trained on both labeled data, such as diffraction patterns with corresponding crystal structures, and unlabeled data, which includes diffraction patterns without this information. In this work, powder diffraction patterns are classified into all 14 Bravais lattices and 144 space groups which covers more crystal classes. It was used to solve the problem of identifying the space group and Bravais lattice of inorganic crystals.

Jianjun Hu et al.,[21, 2023] proposed Deep learning based prediction of contact maps and crystal structures of inorganic materials. They propose AlphaCrystal, a crystal structure prediction method that uses genetic algorithms to reconstruct the 3D structure following predicting the atomic contact map of a given material using a deep residual neural network model. In this study, the MP database is utilized to train and test the contact map predictor. The training dataset is obtained from the Materials Project through the Pymatgen API. They take 1136 known materials in the dataset as the test set and use the remaining 10219 samples as the training set for training our deep neural network model for contact map prediction.

3. Methodology

The main aims of this work to build the Deep-ML method is classify the different crystal structures of Perovskites materials.

In this work the following workflow is used:

(i) Database construction – to develop database system.
(ii) Predicting Framework.
(iii) Hyper parameter optimization – DeepML model tunning.

Figure 1 shows the prediction framework for target properties of crystal structures material.

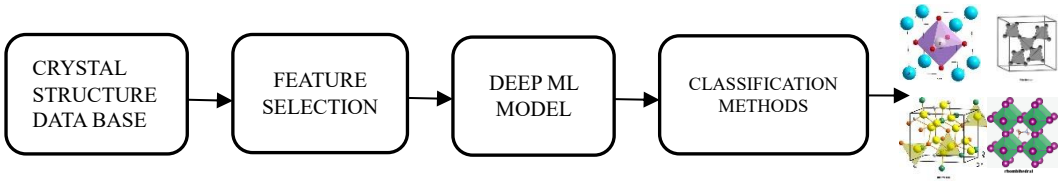

Figure 1. Workflow of the Project

3.1 Database Construction

In our work, we utilized experimental datasets for perovskite materials, comprising over 5,000 entries. There are no null or unknown values in this database. The main goal of this study is to develop an efficient model to categorize perovskite materials by classifying their crystal structure. Table 1 shows the features and labels for the collected datasets. Features were selected based on the crystallographic properties of the perovskite dataset as they relate to the output variables. The same input features are used for crystal structure classification, though the output labels vary between classification models.

Table 1: Current Database Features and Labels

Feature List	Label for classification models
r(AXII)	Cubic
r(AVI)	Rhombohedral
EN(A)	Tetragonal
EN(B)	Orthorhombic
l(A-O)	
l(B-O)	
ΔENR	
t_G	

3.2 Prediction Framework

Feature selection is an important step in building a machine learning model, as it helps identify the best possible set of features. This classification model is proposed for detecting perovskite structures, and the dataset may contain recurring features. The model performs feature encoding on a dataset of over 5,000 entries to predict perovskite crystal structures. The dataset is split into training and validation sets, with 20% of the data reserved for validation.

3.2.1 Cross Validation Test

The dataset is divided into 10 folds for cross-validation to estimate the model's performance on unseen data. A 10-fold cross-validation experiment was conducted to evaluate the performance of the Deep-Machine learning models. The training dataset was split randomly into 10 distinct subsets. For each model, training and testing were performed 10 times, using one unique subset for testing and the remaining nine for training. The model's performance was reported as the average performance across all 10 sets. A binary classification model is proposed for crystal structure prediction. If a crystal structure is determined based on the t_G values, with a value between 0.8 and 1.19, it indicates whether the crystal structure is predictable. This value is set as the target feature for this model. We applied one-hot encoding to the "Lowest Distortion" feature to evaluate its effect on crystal structure prediction.

3.2.2 Feature correlation

The dataset may contain recurring features. The degree and direction of the relationship between two variables can be identified via the correlation coefficient.

It helps identify and eliminate redundant features while quantifying the relationships between features by assigning a value between -1 and 1. The correlation's direction and strength are both represented by this value. The Pearson correlation coefficient matrix contains various features, with higher values indicating stronger correlations and lower values suggesting weaker relationships between features. A correlation of 1 represents a perfect correlation between two features. For example, the valence of elements A and B shows a high correlation, which reflects the balance in the chemical structure of the crystal. Features with smaller values are less important and may be removed from the dataset. However, all features play a crucial role in building an effective crystal structure prediction model.

3.3. Hyper parameter optimization – DeepML method tunning

In this study, machine learning methods such as SVM, NB, KNN, DT, LR, XGBoost, and RF, as well as deep learning methods like AlexNet, ResNet, LSTM, and SGD, are used to predict crystal structures. In previous work, the seven machine learning methods were combined with the AlexNet deep learning method because AlexNet provides the highest accuracy. Therefore, we combined these methods, resulting in the best accuracy. In this work, we removed duplicate values and reduced the space to improve the predictive accuracy of the classification methods. We propose that, in this work, Corr-ANB provides the highest accuracy for classifying crystal structure predictions.

Pseudocode for Corr-ANB Algorithm:

```
from sklearn.naive_bayes import GaussianNB

    Predictions = rf_classifier # Use the combined model for predictions.

    rf_classifier = GaussianNB()

    history = rf_classifier.fit(features_train, y_train).

    anticipate(features_test)
    Round(accuracy_score(predictions,Y_test)*100,2)

    print(rfacc)=rfacc
    Print(clf) = classification_report(predictions,Y_test)

feature_extractor = tf.keras.Model(outputs=model.layers[-1].output, inputs=model.input)

print(f"Accuracy:  {round(acc[1]*100,2)}%\n")
```

4. Result and Discussion

This section primarily examines the performance of the developed models for classifying various crystal structures. The Deep Learning and Machine Learning amethods, including AlexNet, RF, SVM, NB, DT, XGB, KNN, and LR, were used for crystal structure prediction. Before training, the dataset was divided into training and testing sets, with 80% used for training and 20% for testing. Model parameters were then optimized through k-fold cross-validation and hyperparameter tuning. Correlation is used to evaluate the strength and direction of relationships between variables, with correlation coefficients ranging from -1 to +1. The Corr_ANB method, known for its classification capabilities, was found to outperform other models, achieving the highest accuracy in predicting and classifying crystal structures.

Feature selection plays a crucial role in the performance of different methods, particularly in machine learning and deep learning, where the choice of optimal features is key. Building a reliable predictive model relies on identifying and selecting relevant features that effectively represent the target variables. To achieve this, the Pearson correlation coefficient matrix is used to identify the most important variables. Additionally, computational feature selection methods are employed to remove irrelevant and redundant features. The Corr_ANB method has proven effective in selecting features for all models. Proper feature selection helps reduce dataset dimensionality and enhances the performance of machine learning algorithms.

The NB, SVM, and LR methods achieve the lowest accuracy, while DT, KNN, RF, and XGBoost gives the highest accuracy. Among all, the deep learning method AlexNet provides the best accuracy. To assess its robustness, we applied k-fold cross-validation. This method divides the data into k equally sized subsets. The model is trained on k-1 subsets and tested on the remaining one. Finally, the overall accuracy is calculated. The following Table 2 shows the accuracy of prediction crystal structure.

Table 2: Proposed Methods For Predicting The Crystal Structure

Methods		Accuracy	Precision	Recall	F1-Score
Machine Learning	**Deep Learning**				
RF	AlexNet	77	82	84	83
XGB		76	86	90	88
DT		65	83	78	80
Corr_ANB(Proposed)		**87**	91	85	88
LR		81	80	78	86
KNN		71	88	84	86
SVM		78	73	80	78

The goal of these modeling systems is to reduce the cost of predicting crystal structures. Another key aspect of this work is comparing the effectiveness of all these models in predicting crystal structures.

The below Figure 2 shows the accuracy of the two models for both the training and validation sets. The results show that the validation set achieves the best performance in predicting the crystal structure.

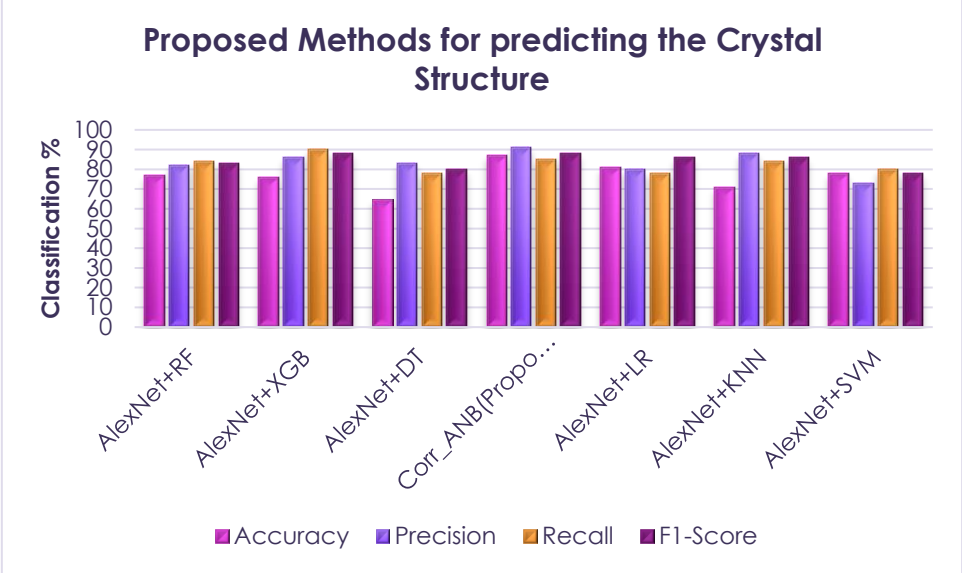

Figure 2. Accuracy Chart for The Proposed Deep-ML Methods

In this study, we combined the AlexNet deep learning amethod with seven machine learning methods. While the NB algorithm initially yields the lowest accuracy, when combined with AlexNet, it achieves the highest accuracy of 87%. As a result, our initial classification model showed an improvement over the performance of previous studies.

4. Conclusion

In this paper, we focus on classifying crystal structures for prediction. We proposed a Crystal Structure Prediction (CSP) method that combines machine learning and deep learning techniques to predict the crystal structure. The goal is to predict the stable structure for a given composition by referencing existing crystal structures in a database.

Various machine learning methods, including RF, SVM, NB, DT, XGB, KNN, and LR, along with the deep learning method AlexNet, were used to predict the crystal structure. In our approach, we combine both deep learning and machine learning methods to achieve the best results. Because of utilizing 10-fold cross-validation, our method outperforms previous approaches. Our work significantly enhances the speed, accuracy, precision, recall, f1-score and ease of crystal structure prediction. The propose method Corr_ANB method provides the best accuracy, achieving 87%. Future work will focus on further improving the formability and structural prediction results.

References

[1] Bhuvaneswari, V., Priyadharshini, M., Deepa, C., Balaji, D., Rajeshkumar, L. and Ramesh, M., 2021. "Deep learning for material synthesis and manufacturing systems: A review". Materials Today: Proceedings, 46, pp.3263-3269.

[2] Suzuki, Y., Taniai, T., Saito, K., Ushiku, Y. and Ono, K., 2022. "Self-supervised learning of materials concepts from crystal structures via deep neural networks". Machine Learning: Science and Technology, 3(4), p.045034.

[3] Kusaba, M., Liu, C. and Yoshida, R., 2022. "Crystal structure prediction with machine learning-based element substitution". Computational Materials Science, 211, p.111496.

[4] Lee, I.H. and Chang, K.J., 2021. "Crystal structure prediction in a continuous representative space". Computational Materials Science, 194, p.110436.

[5] Clements, R.J., Dickman, J., Johal, J. et al. "Roles and opportunities for machine learning in organic molecular crystal structure prediction and its applications". MRS Bulletin 47, 1054–1062 (2022). https://doi.org/10.1557/s43577-022-00434-y

[6] Yin, X. and Gounaris, C.E., 2022. "Search methods for inorganic materials crystal structure prediction". Current Opinion in Chemical Engineering, 35, p.100726.

[7] Mao, Y., Yang, H., Sheng, Y., Wang, J., Ouyang, R., Ye, C., Yang, J. and Zhang, W., 2021. "Prediction and classification of formation energies of binary compounds by machine learning: an approach without crystal structure information". ACS omega, 6(22), pp.14533-14541.

[8] Park, W.B., Chung, J., Jung, J., Sohn, K., Singh, S.P., Pyo, M., Shin, N. and Sohn, K.S., 2017. "Classification of crystal structure using a convolutional neural network". IUCrJ, 4(4), pp.486-494.

[9] Wengert, S., Csányi, G., Reuter, K. and Margraf, J.T., 2022. "A Hybrid Machine Learning Approach for Structure Stability Prediction in Molecular Co-crystal Screenings". Journal of Chemical Theory and Computation, 18(7), pp.4586-4593.

[10] Spencer, M., Eickholt, J. and Cheng, J., 2014. "A deep learning network approach to ab initio protein secondary structure prediction". IEEE/ACM transactions on computational biology and bioinformatics, 12(1), pp.103-112.

[11] Meredig, B., Agrawal, A., Kirklin, S., Saal, J.E., Doak, J.W., Thompson, A., Zhang, K., Choudhary, A. and Wolverton, C., 2014. "Combinatorial screening for new materials in unconstrained composition space with machine learning". Physical Review B, 89(9), p.094104.

[12] Williams, L., Mukherjee, A. and Rajan, K., 2020. "Deep learning based prediction of perovskite lattice parameters from Hirshfeld surface fingerprints". The Journal of Physical Chemistry Letters, 11(17), pp.7462-7468.

[13] Ziletti, A., Kumar, D., Scheffler, M. et al. "Insightful classification of crystal structures using deep learning". Nat Commun 9, 2775 (2018). https://doi.org/10.1038/s41467-018-05169-6

[14] Sungwon Kim, Juhwan Noh, Geun Ho Gu, Alan Aspuru-Guzik, and Yousung Jung "Generative Adversarial Networks for Crystal Structure Prediction" *ACS Central Science* **2020** *6* (8), 1412-1420. DOI: 10.1021/acscentsci.0c00426

[15] Cheng, G., Gong, XG. & Yin, WJ. "Crystal structure prediction by combining graph network and optimization algorithm". Nat Commun 13, 1492 (2022). https://doi.org/10.1038/s41467-022-29241-4