Project Report on Exploring the classification of Crystal Structure of Perovskite Oxide using Machine Learning and Deep Learning

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF TECHNOLOGY
IN
ENGINEERING PHYSICS

Submitted by

SAKSHAM CHECKER (2K20/EP/95)

SAKSHAM GUPTA (2K20/EP/96)

Under the supervision of

DR. DESHRAJ MEENA

DEPARTMENT OF APPLIED PHYSICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

MAY, 2024

CANDIDATE'S DECLARATION

We, hereby declare that the work which is presented in Major Project-II titled "Exploring the classification of Crystal Structure of Perovskite Oxide using Machine Learning and Deep Learning" which is submitted by us to the Department of Applied Physics, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Bachelor of Technology, is original and not copied from any source without proper citation.

The matter presented in this report has not been submitted by us for the award of any other degree of this or any other Institute. The work has been accepted and presented in peer reviewed Scopus Indexed conference with the following details:

Title of the Paper : Exploring the classification of Crystal Structure of Perovskite Oxide using Machine Learning and Deep Learning

Name of Conference : IEEE International Conference on Communication, Computer Sciences and Engineering (IC3SE-2024)

Author Names : Saksham Checker, Saksham Gupta, Dr. Deshraj Meena

Conference Dates with Venue: 9-11th May 2024, Amity University

Status of paper: Accepted and Presented

Date of acceptance: 2 May 2024

Place: Delhi — Saksham Checker

Date: 28-05-2024 — Saksham Gupta

SUPERVISOR CERTIFICATE

To the best of my knowledge, the above work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere. I, further certify that the publication and indexing information given by the students is correct.

28.5.2024

Place: Delhi — Dr. Deshraj Meena

Date: 28-05-2024 — **SUPERVISOR**

DEPARTMENT OF APPLIED PHYSICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "Exploring the classification of Crystal Structure of Perovskite Oxide using Machine Learning and Deep Learning" which is submitted by Saksham Checker, Saksham Gupta, Roll No's – 2K20/EP/95,2K20/EP/96, Department of Applied Physics ,Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Bachelor of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

28.05.2024

Place: Delhi

Dr. Deshraj Meena

Date: 28-05-2024

SUPERVISOR

DEPARTMENT OF APPLIED PHYSICS
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to Dr. Deshraj Meena for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi — Saksham Checker

Date: 28-05-2024 — Saksham Gupta

Abstract

The project explores the classification of crystal structures of perovskite oxides (ABO3) using a combination of machine learning and deep learning algorithms. Various features such as valence, ionic radius, and electronegativity were extracted from the dataset, and those deemed crucial for perovskite classification were utilized. The classification task involved categorizing the oxides into cubic, tetragonal, orthorhombic, and rhombohedral crystal structures. Our analysis encompassed the application of Kernel SVM, a support vector machine model utilizing the kernel trick for non-linear data handling, which emerged as the most accurate model with a 0.94 accuracy rate. Hyperparameter tuning using GridSearchCV further enhanced the model's performance. Additionally, we implemented data augmentation techniques, leading to improved accuracy across all models except Logistic Regression. A novel deep learning approach, Encoder-Decoder Kernel SVM (EDKSVM), was proposed, expanding the feature matrix to achieve superior performance compared to traditional machine learning methods. This study not only contributes to the advancement of perovskite classification methodologies but also showcases the potential of integrating deep learning techniques for enhanced accuracy and reliability in materials science research.

Contents

List of Tables

List of Figures

List of Symbols

Symbol	Description
β_c	Vector of coefficients (weights) corresponding to class c
$\mathbf{x}$	Vector of features (independent variables)
K	Total number of classes
$\mathbf{w}$	Weight vector (coefficients) of the hyperplane
b	Bias term (intercept)
α_i	Lagrange multipliers obtained through optimization
$f(\mathbf{x})$	Decision function of a Decision Tree
w_i	Weights associated with each feature
$\hat{y}_i$	Predicted value for the i-th sample
k	Number of folds in cross-validation
$\mathbf{x}_i$	Feature vector
y_i	Class label

Chapter 1

INTRODUCTION

1.1 Introduction

Perovskite Oxide type structures, represented by the formula ABO3, have gained significant interest due to their (i) simple synthesis process, (ii) remarkable stability owing to the abundance of elements (=90 in the periodic table) in this formula, (iii) diverse electronic properties, (iv) ease of modifying their nanostructure and chemical composition,(v) ability to maintain their stoichiometry even when doped and (vi) tunable[1, 2, 3] **??**. crystal structure . All these advantages of ABO3 oxides make them useful formany applications such as resistive gas sensors [3], solar cells [4], PEC water splitting [5], photcatylitic degradation [6], optoelectronic devices [7, 8], supercapacitors [9] and batteries [10]. It is interesting to note that the efficiency of the mentioned application is highly tunable and dependent on the crystal structure of perovskite oxides. The crystal structure plays a vital role in determining the perovskite oxides' chemical, optical, and electrical properties. Literature suggests that the structural deviations in ideal cubic perovskite result in tetragonal, orthorhombic, rhombohedral, monoclinic and triclinic crystal structures in perovskites [11].Therefore, it becomes crucial to classify the various perovskite oxides into different crystal structures. The different crystal structures are shown in Figure 1.1. Therefore, employing six machine learning models, we have conducted a comparative analysis of crystal structure classification for the enhanced and non-augmented datasets in terms of accuracy, precision, recall, and F1 score in this study.Additionally, the results indicate that the enlarged dataset performed better in the classification of crystal structures. Certainly, here's a more narrative description of each perovskite structure:

In the cubic perovskite structure, atoms arrange themselves with a remarkable symmetry, belonging to the cubic crystal system where all three axes are of equal

length and intersect at 90-degree angles. At the heart of this structure lies a central cation, typically a larger one like strontium or barium, surrounded by oxygen anions forming octahedral cages. This symmetrical arrangement fosters stability and efficient transport of ions and electrons, making cubic perovskites ideal for applications such as solid oxide fuel cells and certain solar cells.

Contrasting the cubic form, the orthorhombic perovskite structure presents a more elongated and distorted arrangement. Its crystal system, orthorhombic, features unequal axis lengths and angles other than 90 degrees. These distortions, often arising from size mismatches between cations or external influences, enhance properties like ferroelectricity and piezoelectricity, rendering orthorhombic perovskites valuable in sensors, actuators, and memory devices.

The tetragonal perovskite structure showcases elongation along one axis, resulting in a rectangular prism shape. While retaining the basic perovskite framework, this elongation introduces structural distortions that confer unique properties such as enhanced ferromagnetism and distinct optical behaviors. Tetragonal perovskites find applications in spintronics, optoelectronics, and data storage owing to these properties.

Lastly, the rhombohedral perovskite structure, belonging to the trigonal crystal system, exhibits a rhombohedral unit cell where all axes are of equal length but do not intersect at 90 degrees. This structure, often emerging from temperature-induced phase transitions or doping effects, displays multiferroic properties, concurrently exhibiting ferroelectric and magnetic orderings. Rhombohedral perovskites hold promise in advanced memory devices, sensors, and energy-efficient electronics due to their unique properties.

1.2 Motivation

This study is driven by the astounding adaptability and significance of perovskite oxides in a wide range of technological applications, from electronics to catalysis. It is essential to be able to precisely categorize their crystal structures in order to comprehend their characteristics and create new materials with specific functions. But conventional approaches to classifying crystal structures frequently depend on time-consuming experimental procedures or computationally demanding simulations. My curiosity in using machine learning techniques to create a more accurate and efficient method was piqued by this inefficiency. Our goal is to improve materials science and technology by optimizing the classification process through the use of data-driven models.

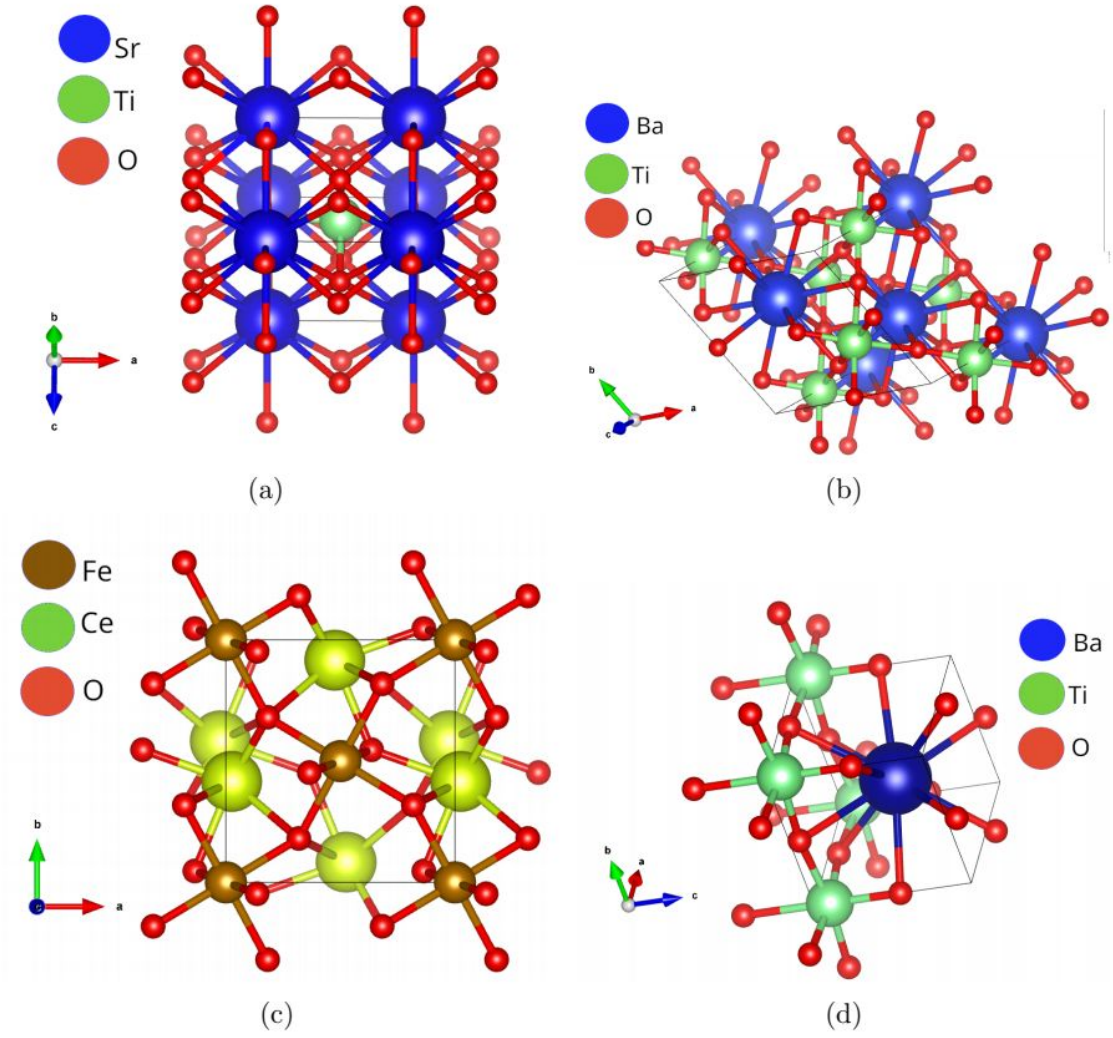


Figure 1.1: Different crystal structures of Perovskite: (a)Cubic ($SrTiO_3$), (b) Rhombohedral ($BaTiO_3$), (c)Orthorhombic ($CeFeO_3$) (d)Tetragonal ($BaTiO_3$)

1.3 Major Contributions

- Exploration of crystal structure classification of perovskite oxides (ABO3) using a blend of machine learning and deep learning algorithms.
- Application of Kernel SVM, demonstrating its superiority with a 0.94 accuracy rate in classifying perovskite crystal structures.
- Introduction of a novel deep learning approach, Encoder-Decoder Kernel SVM (EDKSVM), showcasing superior performance compared to traditional methods.
- Utilization of GridSearchCV for hyperparameter tuning, enhancing the accuracy and robustness of the classification models.
- Advancement of perovskite classification methodologies and integration of deep learning techniques for improved accuracy and reliability in materials science research.

1.4 Relevance

The relevance of this project lies in its potential to address critical challenges in materials science and engineering. Perovskite oxides, with their diverse crystal structures and unique properties, play a pivotal role in advancing various technological domains such as energy storage, electronic devices, and environmental remediation. However, accurately classifying their crystal structures is a complex and time-consuming task, hindering the rapid development and optimization of perovskite-based materials. By applying machine learning algorithms to this classification problem, we aim to streamline and automate this process, enabling faster and more precise characterization of perovskite oxides. This project's relevance extends beyond academic curiosity, offering practical solutions that can catalyze innovation and drive progress in materials research and application development.

1.5 Impact on Environment and Society

The effects of this undertaking on society and the environment are wide-ranging and significant. The potential of perovskite oxides in renewable energy technologies, like fuel cells and solar cells, has attracted a lot of attention. These technologies offer sustainable substitutes for traditional energy sources. The creation of more affordable

and effective perovskite-based devices is directly aided by the accurate classification of their crystal structures by machine learning, which has the potential to completely transform the renewable energy industry. Furthermore, this research's contributions to the field of materials science may result in greener solutions for problems like carbon capture, water purification, and pollution control. This research could have a beneficial effect on society and the environment by developing sustainable behaviors and encouraging technological advancements that benefit the environment by improving our knowledge of and capacity to engineer perovskite materials.

Chapter 2

RELATED WORKS

The field of crystal structure prediction and classification has witnessed significant advancements, with various studies employing diverse methodologies and datasets to tackle the challenges inherent in this domain. This section provides an extensive review of related works, highlighting the methodologies, datasets utilized, advantages, limitations, and areas for improvement observed in each study.

The study referenced as [12] proposed a methodology based on the Light Gradient Boosting Method for categorizing cubic, tetragonal, orthorhombic, and rhombohedral crystal systems. While their approach contributed to the categorization of crystal systems, it lacked comprehensive feature engineering, limiting its potential for further model refinement. In contrast, our study takes a different approach by providing a comparative analysis of feature importance using shapely values, thus identifying significant features crucial for model robustness.

In [13], the authors focused on developing a robust model for predicting the formability of perovskite materials and classifying crystal structures. By testing Random Forest, XGBoost, and LightGBM models, they achieved notable success, with an accuracy of 97.65

Another notable contribution comes from [14], wherein the authors employed a Genetic Algorithm-Neural Network (GA-NN) approach to classify diverse crystal structures and predict lattice parameters. Their methodology demonstrated promising results, with accuracies of 0.88 and 0.95 in crystal structure prediction and lattice parameter estimation, respectively. Furthermore, the study underscored the potential applications of their approach in optimizing renewable energy sources, thus broadening the scope of crystal structure prediction research.

Addressing data imbalance issues in perovskite oxide datasets, [15] utilized the SMOTE resampling technique and tested various machine learning methods, including KNN, Decision Trees, SVM, Random Forest, Neural Networks, Naive Bayes,

Gradient Boosting, and AdaBoost. Their comparative analysis revealed Random Forest as the most accurate model, achieving an 0.86 accuracy rate after overcoming dataset imbalance challenges. However, the study's focus was limited to perovskite oxides, leaving room for exploring other material systems.

Furthermore, a comparative study by [16] investigated the prediction of ABO3-type perovskite crystal structures using machine learning models based on a dataset comprising 5329 data points. Employing the XGBoost algorithm, they achieved an accuracy rate of 0.84, demonstrating the effectiveness of their approach. However, the study's scope was constrained by the dataset size, warranting further exploration with larger datasets encompassing diverse material compositions.

In summary, the reviewed studies collectively contribute valuable insights into the application of machine learning in classifying and predicting properties of various crystalline structures, including perovskite oxides. While each study has made significant strides in addressing specific challenges, there remains ample opportunity for further refinement and exploration, particularly in enhancing model accuracy, conducting comprehensive model evaluations, and expanding the scope of comparison across different methodologies and datasets.

Table 2.1: Comparison of Related Works

Reference	**Methodology**	**Dataset Used**	**Advantages**	**Limitations**
[12]	Light Gradient Boosting Method	Not specified	Categorization of crystal systems	Lack of feature engineering
[13]	Random Forest, XGBoost, LightGBM	Perovskite material data	Achieved high accuracy	Limited to crystal structure classification
[14]	Genetic Algorithm-Neural Network (GA-NN)	Diverse crystal structures	Predicting lattice parameters	Potential applications in renewable energy
[15]	SMOTE resampling, various ML methods	Perovskite oxide datasets	Addressing data imbalance	Limited to perovskite oxides
[16]	XGBoost algorithm	ABO3-type perovskite crystal structures	Prediction accuracy	Limited dataset size

In addition to the insights gleaned from these studies, it is imperative to acknowledge the overarching limitations observed across the reviewed literature, which provide valuable opportunities for future research and improvement:

1. **Limitations in Accuracy Improvement Opportunities:**

    - While studies such as [14] and [16] achieved reasonably high accuracies of 0.88 and 0.84, respectively, for predicting crystal structures using GA-NN and XGBoost, there remains scope for enhancing the accuracy of these models further through advanced feature engineering and algorithmic refinements.

2. **Focused Comparison and Lack of Comprehensive Model Evaluation:**

    - Comparative reports, as seen in [15] and [16], often focus on specific methods (e.g., Random Forest and XGBoost) without providing a comprehensive exploration of the performance characteristics of other machine learning models. Future studies could benefit from broader comparisons encompassing a diverse range of methodologies.

3. **Moderate Accuracy and Scope of Comparison:**

    - Despite achieving moderate accuracy rates, as evidenced in [16] with an accuracy rate of 0.84 using the XGBoost algorithm, there remains a need for more extensive comparisons with other studies and methodologies to assess the generalizability and robustness of proposed models across different material systems and datasets.

Chapter 3

METHODOLOGY

3.1 Problem Statement

Develop a classification system for perovskite oxides (ABO_3) crystal structures using machine learning and deep learning models. Identify key features such as valence, ionic radius, and electronegativity, and leverage advanced techniques like Kernel SVM and EDKSVM to achieve high classification accuracy. Explore the impact of data augmentation on model performance and propose novel approaches to enhance classification accuracy.

3.2 Dataset

We implemented the augmentation technique using the available dataset with data points to expand the dataset. As Machine Learning models performs better on large datasets, augmentation is performed. Figure 3.1 shows a correlation matrix using Spearman's rank correlation. Due to less correlation between various features, all inputs are taken independently. Each observation is described by 13 feature columns and 1 class column which identifies it to be either a cubic, tetragonal, orthorhombic, and rhombohedral structure [17]. The parameters used in this study encompass various characteristics that are crucial for understanding and predicting the properties of materials, particularly in the context of crystal structure and stability. Here's a detailed explanation of each parameter:

- $v(A)$ **and** $v(B)$:
  - $v(A)$ and $v(B)$ denote the valences of elements A and B, respectively. The valence of an element refers to the number of electrons that an atom can donate, accept, or share in forming chemical bonds.

- $r(A)$ **and** $r(B)$:
  - $r(A)$ and $r(B)$ denote the ionic radii of cations A and B, respectively. Ionic radius is the measure of the size of an ion in a crystal lattice.
- **EN**(A) **and EN**(B):
  - EN(A) and EN(B) represent the average electronegativity values of elements A and B, respectively. Electronegativity indicates the ability of an atom to attract electrons towards itself in a chemical bond.
- $l(A-O)$ **and** $l(B-O)$:
  - $l(A-O)$ and $l(B-O)$ signify the bond lengths of the A-O and B-O pairs, respectively. Bond length is the average distance between the nuclei of two bonded atoms.
- δ**ENR**:
  - δENR represents the electronegativity difference with radius. This parameter combines the difference in electronegativity values of elements A and B with their respective ionic radii.
- tG:
  - tG is the Goldschmidt tolerance factor, a dimensionless quantity used to predict the stability of crystal structures. It is calculated based on the sizes and charges of ions in the crystal lattice.
- τ:
  - τ is a new tolerance factor introduced in this study, which provides an alternative measure to assess the stability of crystal structures based on specific criteria.
- M:
  - M is the octahedral factor, a parameter used to analyze the geometry and stability of octahedral complexes in crystal structures.

These parameters collectively contribute to the characterization and understanding of the structural and chemical properties of materials, particularly in relation to their crystallographic stability and electronic structure. They are essential tools in materials science and chemistry for predicting and designing new materials with tailored properties.

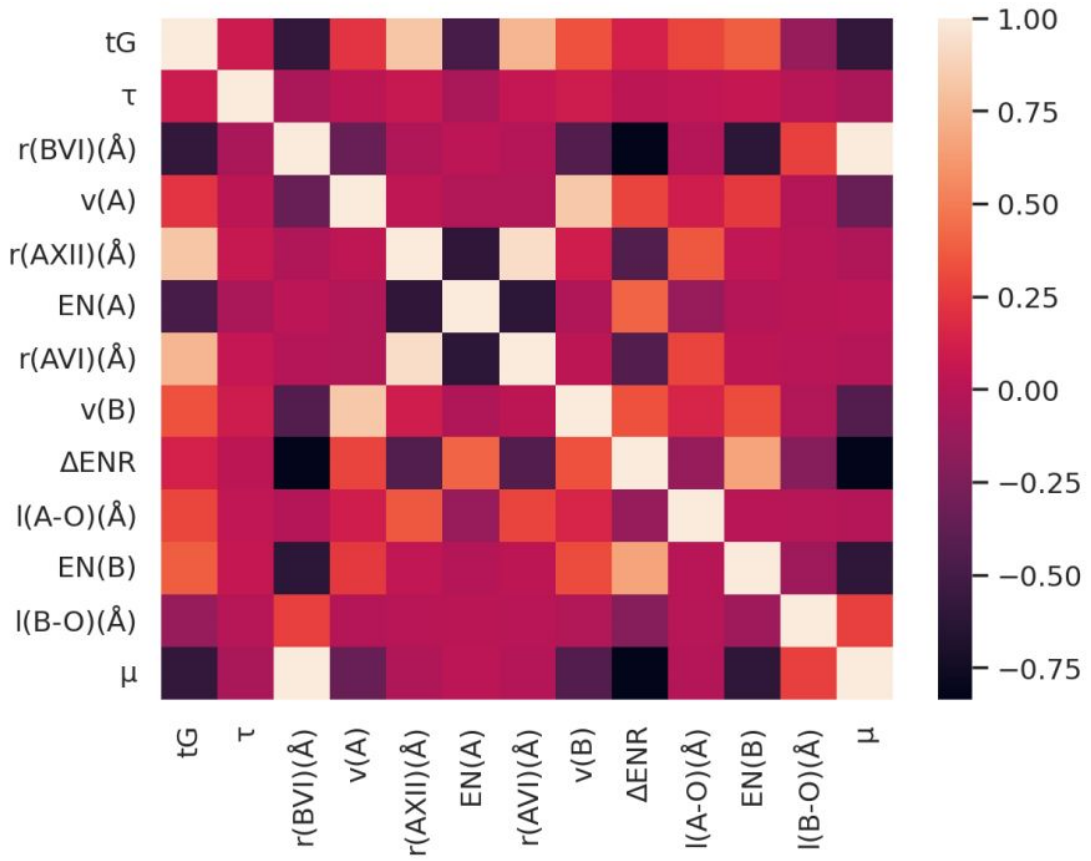


Figure 3.1: Correlation Matrix

3.3 Prelimneries

We propose two different methodologies, first having three main steps of data augmentation, a machine learning model, and hyperparameter tuning using GridSearchCV, while other using the Autoencoder approach.

3.3.1 Data Augmentation

Data Augmentation is a data preprocessing method which overcomes the underfitting issue. Overfitting means that the amount of data is less than the requirement of a complex machine learning algorithm [18]. This method creates realistic data based on the original dataset based on the trends observed in the main dataset [19, 20].

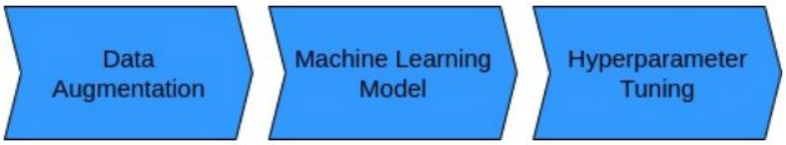


Figure 3.2: Methodology

Gaussian Noise is a distribution where most data is near the mean value, and the probability of the points being away from the mean is low. Two major components of this distribution are mean and standard deviation [21].
Following steps are performed to add Gaussian Noise based data augmentation on statistical data as shown in Figure 3.3

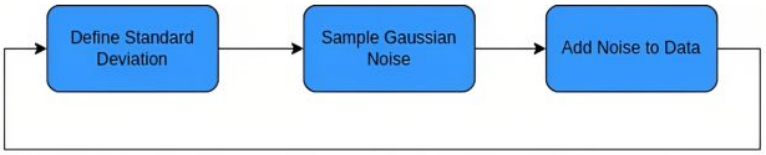


Figure 3.3: Steps of Data Augmentation

1. **Define Standard Deviation:** A standard deviation controls the amount of noise added to the values.
2. **Sample Gaussian Noise:** A noise is generated for every numerical value, keeping the mean 0 and standard deviation as specified.
3. **Add Noise to Data Points:** The sampled Gaussian noise is now added to the corresponding value in the dataset.

The above process is repeated till the required number of times.

In this study, we applied data augmentation with a standard deviation of 0.01, and the process was repeated five times.

3.3.2 Machine Learning Models

It is a technique to enable a computer to learn without being programmed while training the data by identifying the patterns and making predictions based on the same. The following models were considered in this study.

Logistic Regression

Logistic Regression stands as a cornerstone in statistical methodology, particularly adept at tackling classification conundrums across both binary and categorical domains. Its efficacy stems from the employment of the logit function, commonly recognized as the sigmoid function, which facilitates the computation of probabilities. Implicit within its workings is the assumption of a linear relationship between the input features and the output classes [22, 23, 24].

The logistic regression equation, custom-tailored for multiclass classification scenarios, is articulated as follows [23, 24]:

$$P(y_i = c|\mathbf{x}) = \frac{e^{\beta_c \cdot \mathbf{x}}}{\sum_{j=1}^{K} e^{\beta_j \cdot \mathbf{x}}}$$

Breaking down the components of this equation elucidates its inner workings:

- $P(y_i = c|\mathbf{x})$ signifies the probability of the instance $\mathbf{x}$ being assigned to class c.
- β_c represents the vector of coefficients (weights) corresponding to class c.
- $\mathbf{x}$ denotes the vector encompassing the features, or independent variables, associated with the instance.
- K stands as a pivotal parameter, denoting the total number of classes under consideration.

In essence, logistic regression emerges not just as a mere classification algorithm, but rather as a powerful framework underpinning the probabilistic estimation of class membership. Its versatility and interpretability render it indispensable in various fields, from healthcare and finance to marketing and beyond, where informed decision-making relies on accurate predictive modeling. Through its fusion of statistical rigor and computational efficiency, logistic regression continues to empower practitioners in navigating the intricacies of modern data analysis with confidence and precision[24].

Linear Support Vector Machine

The Linear Support Vector Machine (LinearSVC) emerges as a formidable supervised algorithm revered for its prowess in binary and multiclass classification tasks. Its moniker, LinearSVC, underscores its primary function in discerning patterns and boundaries within data sets. Central to its methodology is the minimization of the

hinge loss function coupled with the maximization of the margin, delineating the spatial separation between distinct classes [25].

At its heart lies the LinearSVC equation, succinctly capturing the essence of its decision boundary:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Dissecting this equation unveils the significance of its components:

- $\mathbf{w}$ symbolizes the weight vector, encapsulating the coefficients governing the orientation and inclination of the hyperplane.
- $\mathbf{x}$ denotes the feature vector, comprising the independent variables pivotal in classification.
- b represents the bias term, serving as an intercept that influences the positioning of the decision boundary.

In essence, the SVM epitomizes a blend of geometric intuition and optimization principles, culminating in a robust framework for discerning complex decision boundaries. Its utility spans across diverse domains, from image recognition and text categorization to financial forecasting and beyond, where discerning subtle patterns amidst noisy data proves paramount. Due to its ability to create clear boundaries within the data landscape, the Linear Support Vector Machine remains a fundamental tool in the repertoire of machine learning practitioners worldwide[26].

Kernel Support Vector Machine

The mechanismstands as a pivotal extension to the traditional SVM, offering a sophisticated approach to handling nonlinearities inherent in complex datasets. While the conventional SVM excels in linearly separable data, the Kernel SVM transcends this limitation by leveraging a powerful mathematical technique known as the kernel trick. This ingenious maneuver enables the SVM to operate effectively in higher-dimensional feature spaces, where linear boundaries may not suffice to delineate between classes.

The fundamental objective of Kernel SVM mirrors that of its progenitor: to identify the optimal hyperplane, known as the decision boundary, that maximizes the margin between distinct classes while minimizing classification errors. This objective is achieved through an intricate interplay of mathematical optimization principles and geometric intuition.

1. **Kernel Trick:** At the heart of Kernel SVM lies the kernel trick, a clever mathematical transformation that implicitly maps the input data into a higher-dimensional space, where nonlinear relationships become linear. This transformation allows the SVM to effectively discern intricate patterns and relationships that might be obscured in the original feature space.

2. **Maximizing Margin:** Like its predecessor, Kernel SVM seeks to maximize the margin, which refers to the distance between the decision boundary and the nearest data points, known as support vectors. By maximizing this margin, Kernel SVM ensures robust generalization and enhanced resilience to noise in the data.

3. **Reducing Hinge Loss:** In tandem with maximizing the margin, Kernel SVM endeavors to minimize the hinge loss, a convex surrogate loss function that penalizes misclassifications. By minimizing hinge loss, Kernel SVM strives to strike a delicate balance between margin maximization and error minimization, thus fostering a model that generalizes well to unseen data.

Mathematically, the optimization problem underlying Kernel SVM can be formulated as follows:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

where:

- $\mathbf{w}$ represents the weight vector (coefficients) of the hyperplane,
- $\mathbf{x}$ denotes the feature vector,
- b is the bias term (intercept).

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i Z(\mathbf{x}_i, \mathbf{x}) + b\right)$$

where:

- α_i - Lagrange multipliers obtained through the optimization process,
- y_i represents the class labels,
- $KZ\mathbf{x}_i, \mathbf{x})$ denotes the function.

In essence, Kernel SVM embodies a sophisticated fusion of mathematical elegance and computational efficiency, enabling it to tackle a wide array of classification tasks with remarkable precision and robustness. Through its adept handling of nonlinearities and intricate patterns, Kernel SVM continues to find application across diverse domains, from image recognition and bioinformatics to finance and beyond, where the delineation of complex decision boundaries is paramount.

Decision Trees

Decision Trees are a widely-used supervised learning method known for their straightforwardness and ease of interpretation. They find applications across various fields such as finance, healthcare, and marketing due to their capability to construct a tree-like model that makes decisions based on specific rules [27, 28, 29].

The construction of a decision tree involves several key steps:

1. **Recursive Partitioning:** This method involves dividing the dataset into subsets through recursive partitioning based on the values of input features. At each node, a feature and a threshold are selected to split the data into two or more homogeneous groups. This recursive process continues for each subset until specific stopping criteria are met, such as a maximum tree depth or a purity level.

2. **Determining Splitting Criteria:** For each node, the algorithm identifies the best feature to split on by maximizing information gain or minimizing impurity. Metrics like Gini impurity or entropy are commonly used to evaluate impurity. The objective is to form subsets that are as pure as possible in terms of class labels.

3. **Handling Multi-Class Classification:** Decision Trees can be adapted for multi-class classification problems, such as with the C4.5 algorithm. In these cases, nodes may split into more than two branches, each corresponding to a different class. The splitting criteria are modified to manage multiple classes, ensuring each branch effectively captures the class diversity present in the data.

Overall, Decision Trees provide a robust and interpretable means of tackling classification and regression problems. By recursively partitioning the feature space and making rule-based decisions, they offer valuable insights into data relationships, making them useful tools for both prediction and understanding [27, 28, 29].

Random Forest

Random Forest represents an extension of the Decision Trees algorithm, revered for its exceptional performance and robustness against overfitting [30, 31].

1. **Ensemble Learning:** Random Forest utilizes the concept of ensemble learning by combining multiple decision trees to enhance predictive accuracy and generalization. Each decision tree is trained independently on a different random subset of the training data, and their outputs are aggregated to form the final prediction.

2. **Bagging Aggregation:** To generate multiple subsets of the original training data, Random Forest employs bagging (bootstrap aggregating). This technique involves random sampling with replacement. Each subset is then used to train an individual decision tree, resulting in a diverse set of models.

3. **Robustness Against Overfitting:** By aggregating the predictions of numerous trees, Random Forest reduces the risk of overfitting and decreases variance. This ensemble approach improves predictive accuracy and generalization on new, unseen data.

The decision function of a Random Forest can be expressed as a weighted combination of the individual decision trees' outputs:

$$f(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} f_i(\mathbf{x})$$

Light Gradient Boosting Mechanism

Light Gradient Boosting Mechanism (LightGBM) is a powerful framework in the realm of gradient boosting, renowned for its efficiency and effectiveness [12, 32].

1. **Leaf-Wise Tree Growth:** Unlike traditional gradient boosting algorithms that adopt a level-wise tree growth strategy, LightGBM employs a leaf-wise growth strategy. In this approach, a random node is chosen for splitting that results in the largest reduction in loss. This leaf-wise strategy leads to faster convergence and superior predictive performance compared to level-wise growth.

2. **Feature Bucketing:** LightGBM incorporates feature bucketing, a technique that discretizes continuous features into discrete bins. This binning process reduces memory usage and improves computational efficiency. LightGBM utilizes a histogram approach to construct histograms for each feature and selects the best split points based on histogram statistics. By bucketing features, LightGBM can handle high-dimensional datasets more efficiently.

Mathematically, the prediction function of LightGBM can be represented as follows:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(\mathbf{x}_i)$$

where $\hat{y}_i$ is the predicted value for the i-th sample, K is the number of trees in the ensemble, and $f_k(\mathbf{x}_i)$ represents the output of the k-th tree.

In summary, LightGBM offers a powerful and efficient framework for gradient boosting, characterized by its leaf-wise tree growth strategy and feature bucketing technique. By leveraging these innovations, LightGBM achieves state-of-the-art performance on various machine learning tasks, making it a popular choice among practitioners.

3.3.3 Hyperparameter Tuning

It is a method of finding the optimal configuration of an algorithm. A configuration of a model consists of hyperparameters that can't be learned from the dataset and have to be set before training [33]. The hyperparameter tuning is to find the best values to get maximum performance on the test split of the data [34, 35].
The study uses GridSearch along with cross-validation in this study for hyperparameter tuning.The GridsearchCV process is repeated k times, where the model is trained on k-1 subsets and validated on the last set, giving the best performance based on the average of all the performances.
The steps followed in GridSearchCV are shown in Figure 3.4

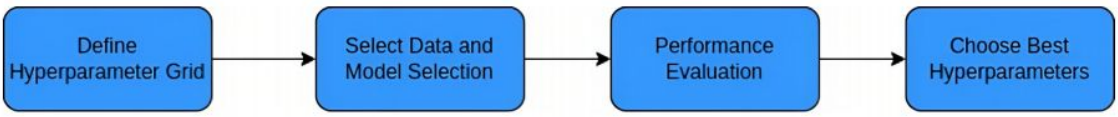


Figure 3.4: Steps of the GridSearchCV

1. **Define Hyperparameter Grid:** A grid of hyperparameters was defined initially in the form of a dictionary on which the algorithm tries all the combinations.

2. **Split Data and Model Selection:** Using the K-fold method, the data has been divided into two. Then algorithm is trained for each set of parameters.

3. **Performance Evaluation:** All the evaluation metrics were calculated and averaged for all the folds.

4. **Selecting best Hyperparameters:** Based on all the evaluation metrics, the best-performing set of hyperparameters are recorded.

3.3.4 Deep Learning Model

We devised a novel approach, the Encoder-Decoder based model (EDKSVM: Encoder-Decoder Kernel Support Vector Machine), to enhance feature representation. The extracted features were subsequently utilized for perovskite classification, as shown in Figure 3.5.

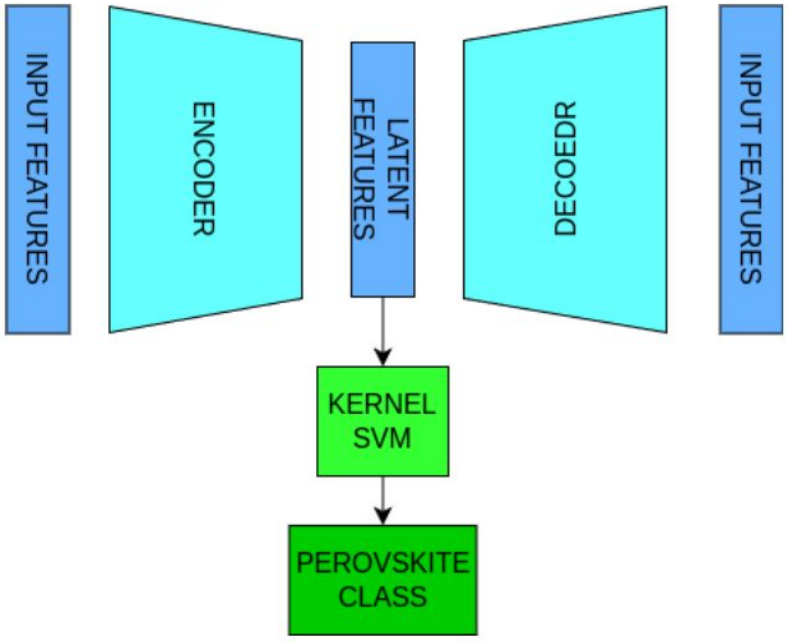


Figure 3.5: Deep Learning Approach

The Encoder-Decoder framework comprised an end-to-end structure with Deep Neural Network layers, aimed at mapping the input features of the data to reconstruct them via an encoded feature matrix. This process enabled the learning of latent representations that effectively captured complex patterns and relationships within the data.

Upon enhancing the Neural Network model, the encoded features were leveraged in conjunction with the KSVM to classify the perovskite samples. By integrating the encoded features into the SVM model, we aimed to capitalize on the complementary strengths of both techniques:the capability of the Neural Network to learn intricate representations and the discriminative strength of the SVM in high-dimensional spaces.

The expansion of features through the Encoder-Decoder model provided the Support Vector Machine classifier with enriched and detailed information gleaned from the latent representations learned by the Neural Network. This fusion of Deep Learning and traditional Machine Learning methodologies aimed to enhance the overall classification performance and robustness of the system.

Algorithm 1 EDKSVM:Encoder-Decoder Kernel Support Vector Machine

Require: Data: X_{train} (training data), X_{test} (testing data), labels (y_{train} for training)

Ensure: Extracted features: X_{train_new}, X_{test_new}

1. Define Dense Autoencoder Model:
Create a model with the following layers:
- Input layer with 13 neurons
- 3 Layered Multi layer perceptron with neurons as 13–>16–>32–>64 to get the center layer of encoding.
- 3 Layered Multi later perceptron with neurons as 64–>32–>16–>13 (to achieve reconstruction)

2. Create Feature Extractor:
Create a model using the first three layers of the autoencoder:
- Input layer with 13 neurons
- 3 Layered Multi layer perceptron with neurons as 13–>16–>32–>64 to get the center layer of encoding. (till output layer for extracted features)

3. Copy Weights to Extractor:
for $i = 1$ to L (number of layers in extractor) **do**
    Set extractor.layers[i].weights = autoencoder.layers[i].weights
end for
4. Extract Features:
$X_{train\_new} \leftarrow$ extractor.predict(X_{train})
5. Train Kernel SVM:
Train a kernel SVM classifier using the extracted features:
Train the SVM with X_{train_new} as features and y_{train} as labels.

Chapter 4

RESULTS and DISCUSSION

4.1 Experimental Setup

All the experiments were conducted on Fedora Linux Distribution Version 37 with an AMD Ryzen 4600H processing unit and NVIDIA GeForce GTX1650Ti graphical processor using a DDRAM of 8GB. The choice of operating system and hardware was made to ensure compatibility and performance optimization for the experimental setup. Fedora Linux Distribution, known for its stability and robustness, provided a reliable environment for conducting experiments. The AMD Ryzen 4600H processor, coupled with the NVIDIA GeForce GTX1650Ti graphical processor, offered significant computational power necessary for processing complex algorithms and datasets efficiently. Additionally, the 8GB DDRAM ensured sufficient memory capacity to handle large datasets and computational tasks seamlessly.

4.2 Performance evaluation matrix

We used the following metrics to compare our proposed model with other baseline models

- **Accuracy:** Accuracy is the ratio of accurately predicted to all predicted points[36, 37].
- **Precision:** It is the ratio of all predicted actual positive values to all predicted positives[36, 37].
- **Recall (Sensitivity):** Number of correctly identified outputs divided by the total number of anomalous outputs[36, 37].
- **F1 Score:** It is the weighted aggregation of precision and recall[36, 37].

4.3 Result Analysis

The dataset was divided into training and testing subsets, and various machine learning models were evaluated based on metrics including accuracy, precision, recall, and F1-score.

Table 4.1: Benchmark with and without augmentation

Method	Accuracy	Precision	Recall	F1 Score
Without Augmentation				
Logistic Regression	64%	59%	64%	59%
Linear SVC	62%	56%	63%	58%
Kernel SVM	75%	74%	75%	74%
Decision Tree	75%	73%	75%	74%
Random Forest	78%	78%	78%	77%
LightGBM	80%	79%	80%	79%
With Augmentation				
Logistic Regression	63%	58%	63%	58%
Linear SVC	63%	58%	63%	59%
Kernel SVM	94%	94%	94%	94%
Decision Tree	78%	78%	78%	77%
Random Forest	91%	91%	91%	91%
LightGBM	91%	91%	91%	91%

Table 4.1 presents the performance of machine learning models both with and without data augmentation. Without augmentation, LightGBM achieved the highest accuracy of 80%, along with corresponding precision, recall, and F1-score values of 79%. However, after data augmentation, Kernel SVM surpassed all other models with an accuracy of 94% and identical precision, recall, and F1-score values of 94%. These results highlight the effectiveness of data augmentation, particularly in improving the performance of Kernel SVM in classifying perovskite oxide crystal structures.

4.4 Comparison with State of Art Work

Table 4.2 provides a comparative analysis of the proposed method with existing literature on crystal structure classification of perovskite oxides. Notable methods from literature include Light GBM, RF, GA-NN, KNN, and XGBoost. Our proposed methods, DA+KSVM and EDKSVM, outperform all existing methods with an ac-

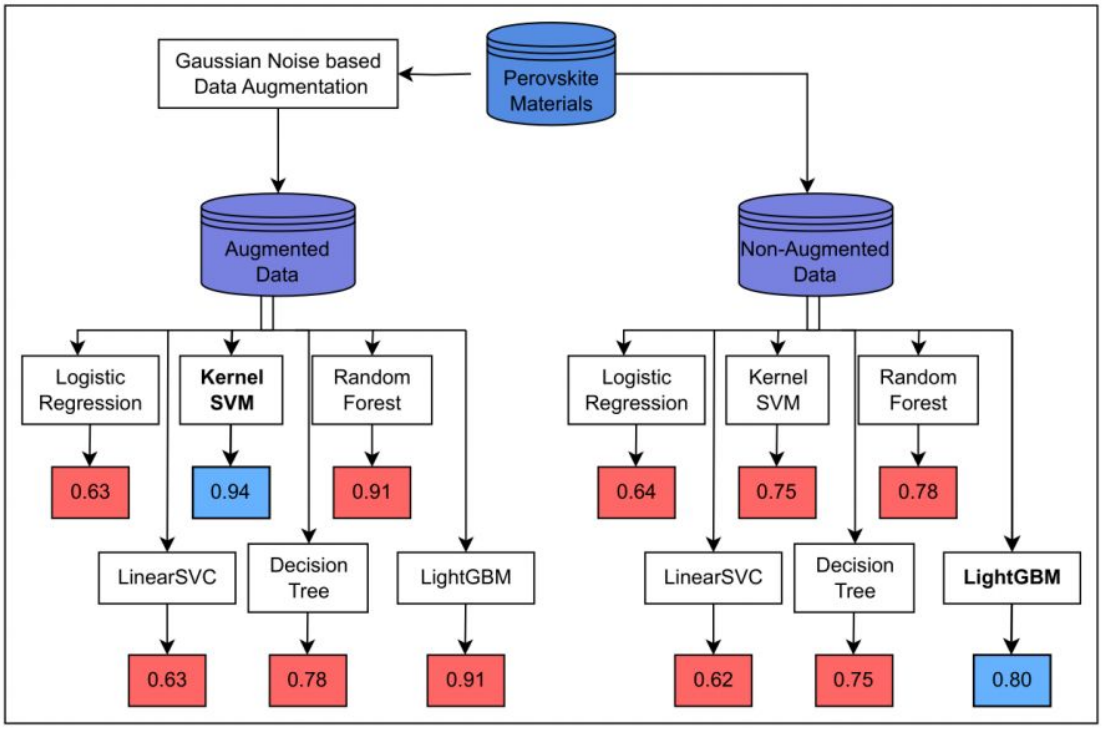


Figure 4.1: Benchmarking of Perovskite oxide classification

Table 4.2: Comparative Report of Studied Literature with Proposed Method

Method	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Light GBM [12]	90.4%	N.R.	N.R.	N.R.
RF [13]	90.56%	91.29%	90.56%	90.2%
GA-NN [14]	88%	N.R.	N.R.	N.R.
KNN [15]	84%	83%	84%	84%
XGBoost [16]	83.67%	82.10%	82.12%	82.11%
DA+KSVM	**94%**	**94%**	**94%**	**94%**
EDKSVM	**98%**	**97%**	**96%**	**97%**

curacy of 94% and 98% respectively. These results demonstrate the superiority of our approaches in accurately classifying perovskite oxide crystal structures, thereby contributing significantly to the field of materials science.

Our model,EKDSVM, results were compared to those performed on the same dataset in Table 4.2 outperforming the existing techniques with 98% of accuracy. The scores which were not reported are mentioned as N.R. in the Tables.

4.5 Discussions and Findings

To illustrate the outcomes of our experimentation, we present the confusion matrices derived from the performance of each model on the test dataset, both with and without the integration of Data Augmentation techniques, as depicted in Figure 4.2 and 4.3 respectively. In each confusion matrix, the x-axis delineates the True Labels, representing the actual classification categories, while the y-axis illustrates the False Labels, encompassing misclassifications. These labels correspond to distinct crystal structures: cubic, orthorhombic, rhombohedral, and tetragonal.

The LightGBM model emerged as the superior performer in the non-augmented dataset scenario. This can be attributed to its capability to explore a vast array of hyperparameters and leverage gradient-based optimization techniques across each boosting iteration. In this context, the model exhibited promising True Positive rates across various crystal structures. Notably, the True Positive rate for cubic perovskite stood at 0.89, indicating a strong capability in identifying this particular structure. However, the performance was comparatively weaker for tetragonal and rhombohedral structures, with True Positive rates of 0.45 and 0.36 respectively. Surprisingly, the model demonstrated decent accuracy in detecting orthorhombic structures, with

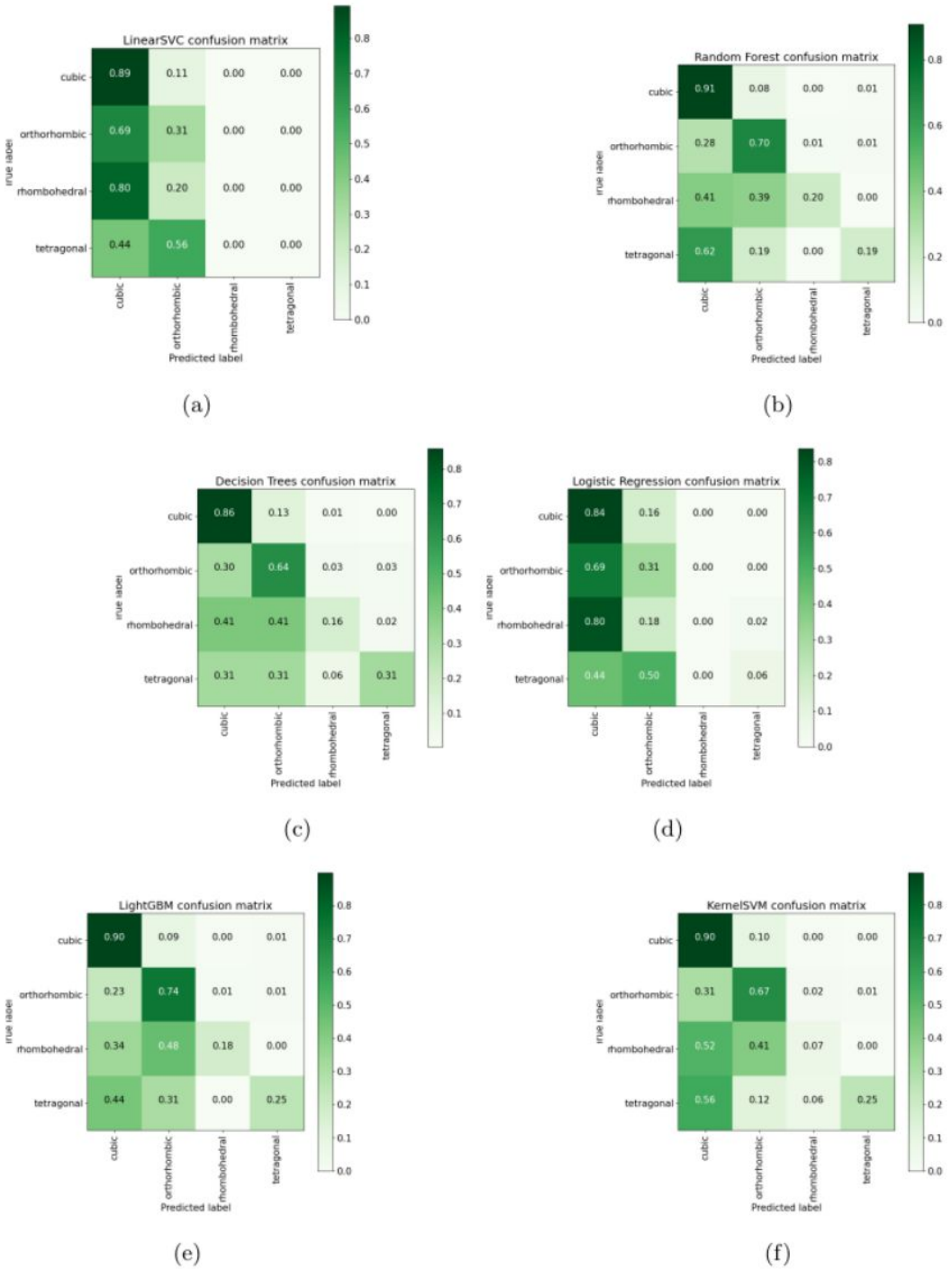


Figure 4.2: Confusion Matrix on test set without augmentation : (a) LinearSVC, (b) Random Forest, (c) Decision Tree, (d) Logistic Regression, (e) LightGBM, (f) Kernel SVM

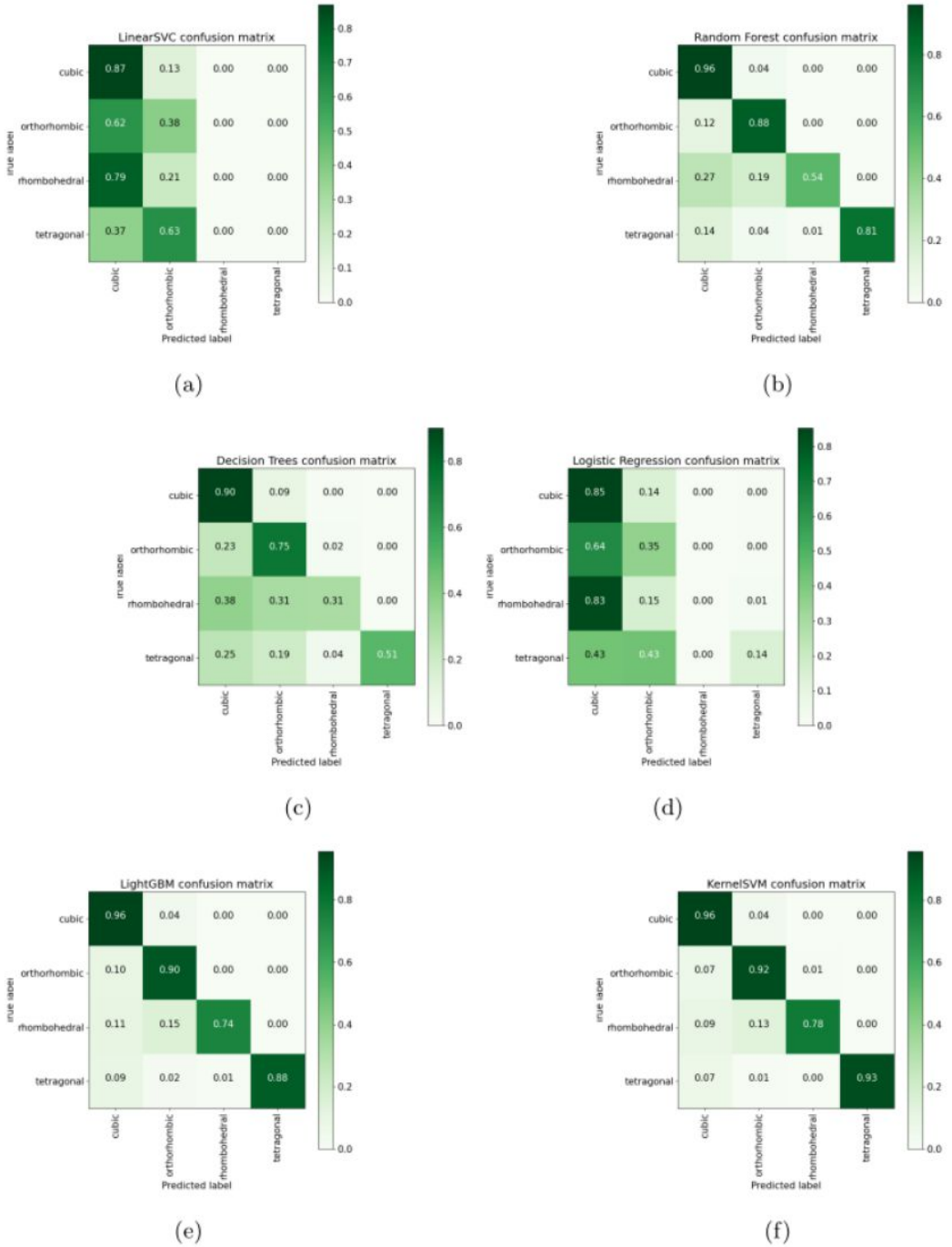


Figure 4.3: Confusion Matrix on test set with augmentation : (a) LinearSVC, (b) Random Forest, (c) Decision Tree, (d) Logistic Regression, (e) LightGBM, (f) Kernel SVM

a True Positive rate of 0.78.

To further enhance the predictive accuracy of our models, we augmented the dataset using Gaussian Noise augmentation technique. This augmentation approach yielded noticeable improvements in the performance of models, particularly evident in the case of Kernel Support Vector Machine (KernelSVM). Upon augmentation, the KernelSVM model showcased robust performance across all crystal structures, with significantly higher True Positive rates. Specifically, the True Positive rate for cubic perovskite surged to 0.96, indicative of a remarkable improvement in classification accuracy. Moreover, the model achieved commendable True Positive rates of 0.78, 0.93, and 0.92 for rhombohedral, tetragonal, and orthorhombic structures respectively. This enhanced performance can be attributed to the kernel trick employed by KernelSVM, which effectively addresses the non-linearity inherent in the dataset.

The augmentation of the dataset mitigated the risk of overfitting and resulted in more robust model performance across all tested algorithms. This trend was observed consistently across all confusion matrices, showcasing improved performance metrics for models such as LightGBM, Linear Support Vector Machine (LinearSVC), Decision Trees, and Random Forest. Conversely, the performance of the Logistic Regression model remained relatively consistent, reflecting the inherently low complexity of the model architecture.

Chapter 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

Our study presents a robust pipeline for automating the classification of perovskites into various crystal structures, namely cubic, tetrahedral, orthorhombic, and tetragonal. The pipeline involves augmenting the dataset and training multiple machine learning models, including Logistic Regression, LinearSVM, LightGBM, KernelSVM, Decision Trees, and Random Forest. Through extensive testing and training on both augmented and non-augmented datasets, we observed varying performance among the models. LightGBM emerged as a top performer on non-augmented datasets, achieving an accuracy of 0.80 due to its extensive hyperparameter tuning capabilities. On the other hand, KernelSVM excelled on the augmented dataset with an impressive 0.94 accuracy, effectively addressing overfitting issues observed in the non-augmented dataset. These findings highlight the importance of dataset augmentation and model selection in achieving accurate and reliable classification results for perovskite materials.

5.2 Limitation

Despite the progress made in automating the classification of perovskites, there are several limitations that warrant consideration. Firstly, while machine learning models have shown promising accuracy rates, there is still room for improvement in terms of achieving higher accuracy and robustness across different datasets and crystal structures. Secondly, the reliance on specific data augmentation techniques and hyperparameter tuning may introduce biases or limitations in the model's generalizability to real-world scenarios. Additionally, the choice of machine learning algorithms and their interpretability can pose challenges in understanding and validating the

results effectively. Furthermore, the limited scope of feature engineering and analysis in some studies may hinder a comprehensive understanding of the underlying factors influencing classification accuracy. These limitations underscore the need for continued research and development efforts to address these challenges and enhance the effectiveness and reliability of automated perovskite classification systems.

5.3 Potential Industrial Applications

The advancements made in automating the classification of perovskites using machine learning models have significant potential for various industrial applications. One key area of application is in the development of more efficient and cost-effective renewable energy technologies. Accurate classification of perovskite crystal structures can lead to the optimization of perovskite-based solar cells and fuel cells, improving their performance and reducing production costs. Additionally, the ability to automate this classification process opens doors for rapid materials discovery and development in other sectors such as electronics, catalysis, and environmental remediation. Industries involved in materials science, renewable energy, electronics, and advanced manufacturing stand to benefit greatly from the insights and technologies derived from this research, paving the way for innovative solutions and advancements in various industrial sectors.

5.4 Future Work

Looking ahead, our future research endeavors will focus on expanding data augmentation techniques beyond Gaussian Noise to further enhance the robustness of our model. Additionally, we plan to incorporate deep learning algorithms into our pipeline to explore their potential for improving classification performance. By leveraging advanced data augmentation methods and deep learning techniques, we aim to develop a superior-performing model that can accurately classify perovskite materials with even higher precision and reliability. This ongoing work represents our commitment to continuous improvement and innovation in the field of automated classification of crystal structures in perovskite oxides.

Bibliography

[1] A. Bhalla, R. Guo, and R. Roy, "The perovskite structure—a review of its role in ceramic science and technology," *Materials Research Innovations*, vol. 4, no. 1, pp. 3–26, 2000.

[2] J. Park, Y.-N. Wu, W. A. Saidi, B. Chorpening, and Y. Duan, "First-principles exploration of oxygen vacancy impact on electronic and optical properties of abo3 (- delta) (a = la, sr; b = cr, mn) perovskites," *Phys. Chem. Chem. Phys.*, vol. 22, pp. 27 163–27 172, 2020.

[3] K. I. Bulemo Peresi Majura, "Recent advances in abo3 perovskites: their gas-sensing performance as resistive-type gas sensors," *J. Korean Ceram. Soc*, vol. 57, no. 1, pp. 24–39, 2020.

[4] M. Thambidurai, F. Shini, P. Harikesh, N. Mathews, and C. Dang, "Highly stable and efficient planar perovskite solar cells using ternary metal oxide electron transport layers," *Journal of Power Sources*, vol. 448, p. 227362, 2020.

[5] J. Shi and L. Guo, "Abo3-based photocatalysts for water splitting," *Progress in Natural Science: Materials International*, vol. 22, no. 6, pp. 592–615, 2012.

[6] H. Zhang, X. Ji, H. Xu, R. Zhang, and H. Zhang, "Design and modification of perovskite materials for photocatalytic performance improvement," *Journal of Environmental Chemical Engineering*, vol. 11, no. 1, p. 109056, 2023.

[7] A. Alashkar, M. A. Ayoub, A. H. Al-Alami, M. A. Abdelkareem, A. G. Olabi, and A. Mdallal, "An outlook on the utilization of perovskite films in photocatalytic and optoelectronic devices," in *2023 Advances in Science and Engineering Technology International Conferences (ASET)*, Feb 2023, pp. 1–5.

[8] E. Kim, J. Kim, and K. Min, "Prediction of dielectric constants of abo3-type perovskites using machine learning and first-principles calculations," vol. 24, pp. 7050–7059, 2022.

[9] D. Meng, H. Gu, Q. Lu, Y. Zhao, G. Zhu, Y. Zhang, Q. Zhong, and Y. Bu, "Advances and perspectives for the application of perovskite oxides in supercapacitors," *Energy & Fuels*, vol. 35, no. 21, pp. 17 353–17 371, 2021.

[10] C. Sun, J. A. Alonso, and J. Bian, "Recent advances in perovskite-type oxides for energy conversion and storage applications," *Advanced Energy Materials*, vol. 11, no. 2, p. 2000459, 2021.

[11] P. M. Woodward, "Structural distortions, phase transitions, and cation ordering in the perovskite and tungsten trioxide structures," 1996.

[12] S. Behara, T. Poonawala, and T. Thomas, "Crystal structure classification in abo3 perovskites via machine learning," *Computational Materials Science*, vol. 188, p. 110191, 2021.

[13] M. U. Ahmad, A. R. Akib, M. M. S. Raihan, and A. B. Shams, "Abo3 perovskites' formability prediction and crystal structure classification using machine learning," in *2022 International Conference on Innovations in Science, Engineering and Technology (ICISET)*, 2022.

[14] S. Jarin, Y. Yuan, M. Zhang, M. Hu, M. Rana, S. Wang, and R. Knibbe, "Predicting the crystal structure and lattice parameters of the perovskite materials via different machine learning models based on basic atom properties," *Crystals*, vol. 12, no. 11, 2022.

[15] N. Ujjwal, A. Singh, and R. G. Tiwari, "Perceptive classification for crystal structure of abo3 perovskite-type oxides using machine learning," in *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, 2022, pp. 428–433.

[16] R. Priyadarshini, H. Joardar, S. K. Bisoy, and T. Badapanda, "Crystal structural prediction of perovskite materials using machine learning: A comparative study," *Solid State Communications*, vol. 361, p. 115062, 2023.

[17] A. A. Emery and C. Wolverton, "High-throughput dft calculations of formation energy, stability and oxygen vacancy formation energy of abo3 perovskites," *Scientific Data*, vol. 4, p. 170153, 10 2017.

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, 12 2019.

[19] C. Shorten, T. M. Khoshgoftaar, and B. Furht, "Text data augmentation for deep learning," *Journal of Big Data*, vol. 8, p. 101, 12 2021.

[20] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, 2016, pp. 1–6.

[21] D. A. van Dyk and X.-L. Meng, "The art of data augmentation," *Journal of Computational and Graphical Statistics*, vol. 10, pp. 1–50, 3 2001.

[22] M. P. LaValley, "Logistic regression," *Circulation*, vol. 117, pp. 2395–2399, 5 2008.

[23] T. G. Nick and K. M. Campbell, "Logistic regression," pp. 273–301, 2007.

[24] A. DeMaris, "A tutorial in logistic regression," *Journal of Marriage and the Family*, vol. 57, p. 956, 11 1995.

[25] S. Suthaharan, "Support vector machine," pp. 207–235, 2016.

[26] Y. Tang, "Deep learning using linear support vector machines," 2015.

[27] C. Kingsford and S. L. Salzberg, "What are decision trees?" *Nature Biotechnology*, vol. 26, 9 2008.

[28] L. Rokach and O. Maimon, "Decision trees," *Data Mining and Knowledge Discovery Handbook*.

[29] S. B. Kotsiantis, "Decision trees: a recent overview," *Artificial Intelligence Review*, vol. 39, 4 2013.

[30] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, pp. 31–39, 1 2017.

[31] G. Biau and E. Scornet, "A random forest guided tour," *TEST*, vol. 25, 6 2016.

[32] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30. Curran Associates, Inc., 2017.

[33] R. Bardenet, M. Brendel, B. Kégl, and M. Sebag, "Collaborative hyperparameter tuning," in *Proceedings of the 30th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, S. Dasgupta and D. McAllester, Eds., vol. 28, no. 2, Atlanta, Georgia, USA, 17–19 Jun 2013, pp. 199–207.

[34] T. Agrawal, "Hyperparameter optimization using scikit-learn," *Hyperparameter Optimization in Machine Learning*, pp. 31–51, 2021.

[35] M. Feurer and F. Hutter, "Hyperparameter optimization," pp. 3–33, 2019.

[36] N. Ketkar, *Deep Learning with Python*. Apress, 2017.

[37] Rácz, Bajusz, and Héberger, "Multi-level comparison of machine learning classifiers and their performance metrics," *Molecules*, vol. 24, p. 2811, 8 2019.

Gmail Saksham Checker <checkersaksham01@gmail.com>

IC3SE-2024 #0782- Acceptance Letter

Amity Conference <conference@gn.amity.edu> Thu, May 2, 2024 at 4:04 PM
To: Saksham Checker <checkersaksham01@gmail.com>, "saksham.gpt2001@gmail.com" <saksham.gpt2001@gmail.com>, "deshrajmeena@dtu.ac.in" <deshrajmeena@dtu.ac.in>
Cc: Chair Conference <conferencechair@gn.amity.edu>

Dear Author,

Greetings from Amity University!

We are pleased to inform that your manuscript ID: 0782 with Title: "Exploring the classification of Crystal Structure of Perovskite Oxide using Machine Learning and Deep Learning", submitted to 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE-2024) at Amity University, Greater Noida, has been accepted. Kindly consider this email as formal acceptance of this submission.

The Conference is scheduled to be held from 09th to 11th May 2024.

Registration
Please pay the registration fee for this submission, at the earliest. *Registration fee category must be chosen based on affiliation and membership of 1st author only.* For registration fee submission, please visit Registration webpage on Conference Website.

If the 1st author of the submission is a valid IEEE Member, you may avail a discount benefit in registration fee as mentioned on Registration Webpage. In case the discount is availed, kindly send us the IEEE Membership Letter/ Card for the same, failing to which, you will be required to submit the availed discounted amount.

Final Version Manuscript
After registration, you are required to submit the final version of the manuscript in IEEE Template in both .pdf and .doc/.docx (MS Word) format, which must include the suggestions/ changes as per instructions/ improvement mentioned by the Technical Program Committee/ Program Committee Members/ Reviewers in the comments, if any. These changes must be limited to 20% part of the manuscript.

The final version must be submitted to our *Publication Support Wing* (PSW) (publicationsupport@gn.amity.edu). Final version of manuscript must comply with IC3I-2023 Submission Guidelines.

Preprint
Once you will submit the final version of your manuscript to PSW, PSW will check for format of the same as per IEEE Template, will amend if required keeping the content of manuscript unchanged and get it validated/confirmed from you. After validation/ confirmation from your end, PSW will get the preprint ready at its end.

e-Copyright Transfer

After all the above formalities, you will also be receiving an email from IEEE (ECopyright@ieee.org) for e-signing of e-copyright transfer form. Transfer of copyright to IEEE is a mandatory requirement. Our PSW will also intimate you via email, once it will initiate the Copyright Transfer Process. Copyright Transfer will be a self-explanatory process. Still, if any assistance is required, you may always reach out to PSW.

In case of any query, please feel free to contact us.

Regards,

Conference Program Committee
on behalf of
Conference Chair
Primary Contact ID: conference@gn.amity.edu
Primary Contact # +91 9654 1393 99

Amity University Uttar Pradesh | Greater Noida Campus | India

Gmail Saksham Checker <checkersaksham01@gmail.com>

Payment Notification – IC3SE 2024

2 messages

noreply@amity.edu <noreply@amity.edu> Sat, May 4, 2024 at 12:30 AM
To: checkersaksham01@gmail.com

Dear **Saksham Checker**

Your payment has been successfully initiated. Your payment details are given below:-

Your Application No. : **602741**

Your Transaction Id : **318590753**

Your Amount : **6300**

Your Payment Status : **SUCCESS**

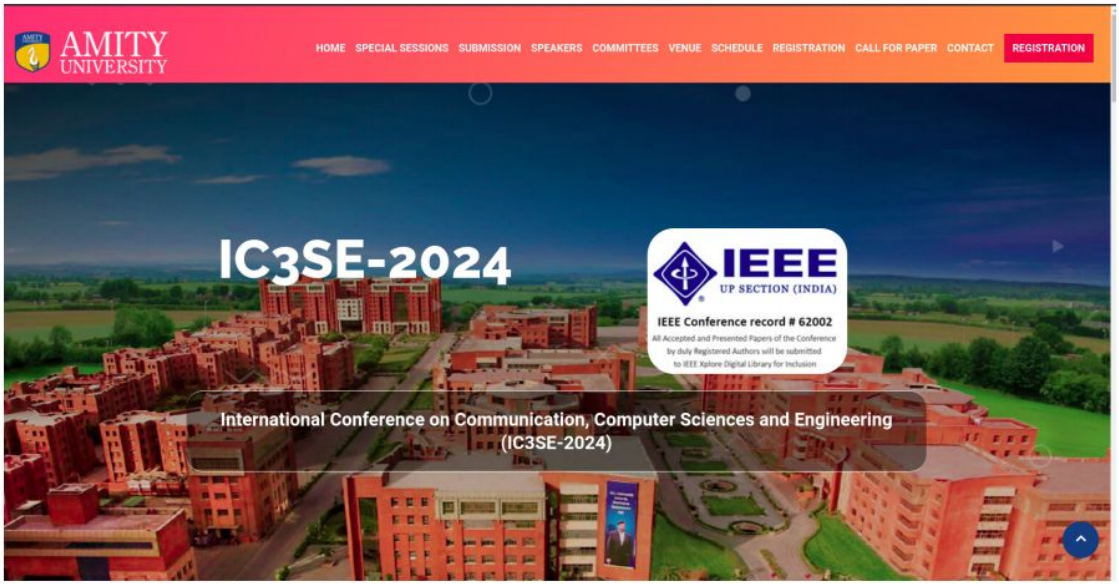
Thanks for registering yourself in IC3SE 2024

Saksham Checker <checkersaksham01@gmail.com> Sat, May 4, 2024 at 12:31 AM
To: Saksham Gupta <saksham.gpt2001@gmail.com>

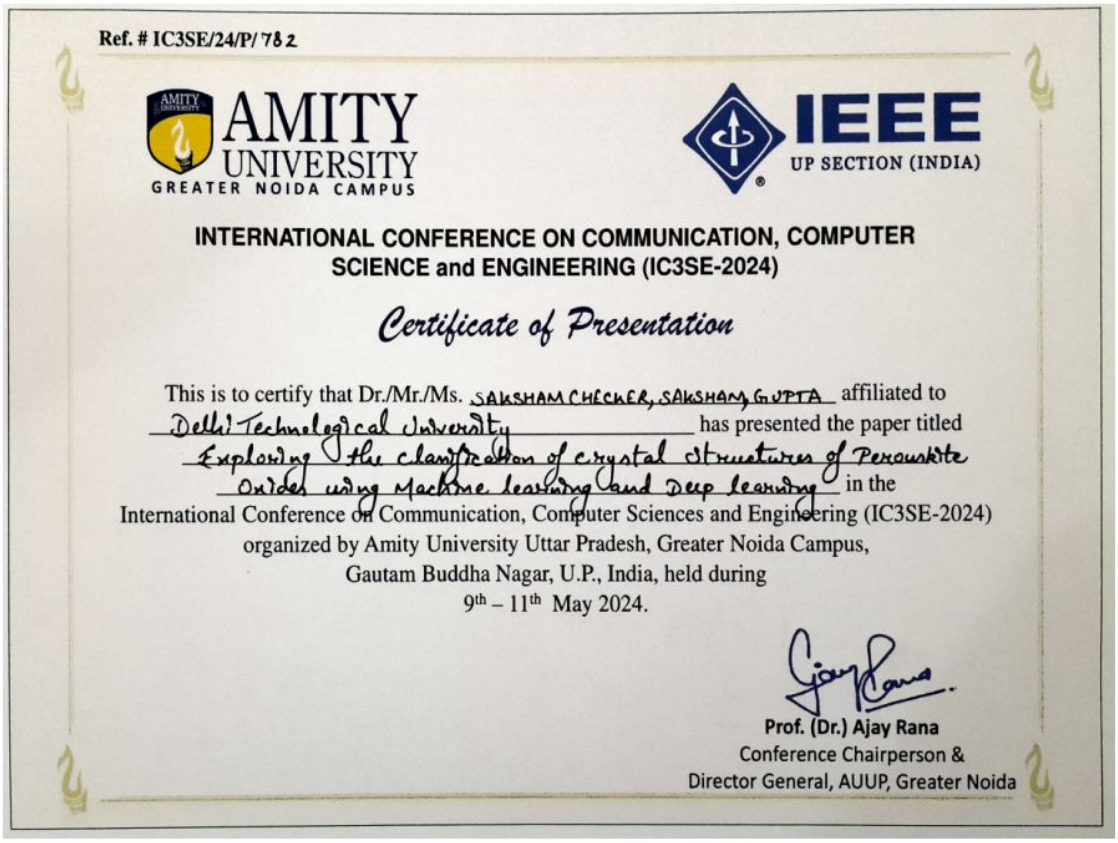
--
Warm regards,
Saksham Checker
M. +91 9013844884

[Quoted text hidden]

Proof of Scopus Indexing



Certificate of Presentation



Ref. # IC3SE/24/P/782

AMITY UNIVERSITY
GREATER NOIDA CAMPUS

IEEE
UP SECTION (INDIA)

INTERNATIONAL CONFERENCE ON COMMUNICATION, COMPUTER SCIENCE and ENGINEERING (IC3SE-2024)

Certificate of Presentation

This is to certify that Dr./Mr./Ms. SAKSHAM CHECHER, SAKSHAM GUPTA affiliated to Delhi Technological University has presented the paper titled Exploring the classification of crystal structures of Perovskite Oxides using Machine learning and Deep learning in the International Conference on Communication, Computer Sciences and Engineering (IC3SE-2024) organized by Amity University Uttar Pradesh, Greater Noida Campus, Gautam Buddha Nagar, U.P., India, held during 9th – 11th May 2024.

Prof. (Dr.) Ajay Rana
Conference Chairperson &
Director General, AUUP, Greater Noida

Pictures from the conference

PAPER NAME
Major_Project_Thesis (2)-5-50.pdf

AUTHOR
Deshraj Meena

WORD COUNT
8848 Words

CHARACTER COUNT
51957 Characters

PAGE COUNT
46 Pages

FILE SIZE
6.8MB

SUBMISSION DATE
May 21, 2024 12:08 PM GMT+5:30

REPORT DATE
May 21, 2024 12:10 PM GMT+5:30

28/05/2024

10% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 10% Internet database
- 1% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Crossref database
- Submitted Works database
- Bibliographic material
- Quoted material
- Small Matches (Less then 8 words)