Material Prediction using FESEM Imaging and Machine Learning

Rashi Nathawat
dept.of Physics
Manipal University Jaipur
Jaipur, India
muj0957@muj.manipal.edu

Aditya Sinha
dept.of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India
muj1087@muj.manipal.edu

Fareeha Mulla
dept.of Computer Science and Engineering
Manipal University Jaipur
Jaipur, India
fareeha.219301161@muj.manipal.edu

Mohit Yadav
dept.of Computer Science Eng
Manipal University Jaipur
Jaipur, India
mohit.219301105@muj.manipal.edu

***Abstract*— Over the past ten years, the rate of systematic materials discovery has accelerated. Studies that systematically examine different families of materials with the intention of finding existing materials with unknown qualities or developing new materials with desired properties are multiplying at a startling rate. In terms of computation, molecular simulations have become more sophisticated, enabling researchers to predict the structure and characteristics of complex materials even before they are ever created. High-throughput screening methods can be used with chemical characteristics prediction to aid in the discovery of novel materials. Due to their promising capacities to extract and make use of data-driven information, applications of artificial intelligence, machine learning, and deep learning techniques in the field of materials science are becoming more widespread. Given the composition of the material, we use predictive models for a range of material attributes to help with this procedure. Multiple databases are used to compile the input data necessary to train the machine learning model. Once trained, it can be used to apply predictions for various input data.**

I. INTRODUCTION

Machine learning tries to generate statistical models for data analysis and prediction using learning algorithms. Without having been particularly designed for a given task, the ml algorithms ought to be able to train on their own (depending on the presented data) and produce accurate predictions. Beyond theoretical improvements, machine learning applications have advanced quickly in recent years, thanks to the work of academics from a variety of professions, including computer scientists and experts in the creation of ai algorithms. Chemical and materials sciences have been impacted by the use of machine learning to speed up particular computing operations or to tackle issues for which conventional modeling approaches were inadequate, among many ther disciplines of inquiry. One can create a model for analysing materials if they have enough data in the right format. Choosing the right algorithms, learning from training data, and producing precise predictions are all steps in the modeling process.

II. METHODOLOGY

A. Data Collection

FESEM imaging was used to collect the experimental data.The image of the materials' microstructure is captured using the cutting-edge technology known as field emission scanning electron microscopy (FE-SEM). Because gas molecules have a tendency to affect the electron beam and the produced secondary and backscattered electrons utilized for imaging, FE-SEM is normally carried out in a high vacuum. The primary electrons are concentrated in a high vacuum column, where an electronic lens deflects the blasted electrons onto each object. Each item emits secondary electrons in this way. The velocity and angle of secondary electrons are strongly influenced by the surface characteristics of any item. A detector captures these electrons and produces an electrical signal as a result. This signal is converted into a video-scan image that can be recorded for later processing and seen on a monitor. The research lab within the academic institution served as the data's data source for the experimental data and the FESEM images.

B. Data Processing

Load FESEM images into your ML environment using libraries such as OpenCV, PIL, or scikit-image. After an image is collected, the data must first be structured and prepared for the image processing in order to extract the required information; this process is known as preprocessing or data preparation. Data is cleaned, processed, and changed into different formats in this step so that it can be utilized as supplemental data. Image quality assessment is one technique for data preparation.

The geometrical texture and properties of the input image fluctuate greatly, therefore evaluating the image's quality is necessary to decide which functions to use in what order. Due to the complex spectral/optical qualities associated with FESEM devices, as well as the geometrical, chemical, and physical conformations of the material surface, FESEM images contain distinctive figure-by-figure and pixel-by-pixel features. We confine the scope of our attention throughout the image processing to the

accuracy of segmentation of grey-scale SEM pictures. We carefully categorize the hidden data features (pixel- and object-based qualities) in order to do this.

Outlier detection: Locate and eliminate any extreme values or outliers from your dataset. The performance of the model and the training process might be dramatically impacted by outliers.Handling Missing Data: We examine the FESEM images for any missing data. We exclude samples with missing data or impute missing values if there are missing images or characteristics.Normalize pixel values of FESEM images using Min-Max scaling. This transformation ensures that all pixel values fall within the same scale, making it easier for the model to learn.

We may need to resize FESEM images to a consistent resolution. Common image sizes are 128x128, 224x224, or 256x256 pixels. We then apply data augmentation techniques to increase the diversity of your training dataset. This can include random rotations, flips, translations, and brightness adjustments. Data augmentation helps prevent overfitting and improves the model's generalization. Feature extraction from FESEM images involves capturing meaningful information from the images that can be used as input to an ML model. Depending on the dimensionality of the extracted features, we may perform feature selection to choose the most relevant features for our ML model.

C. Flowchart-1

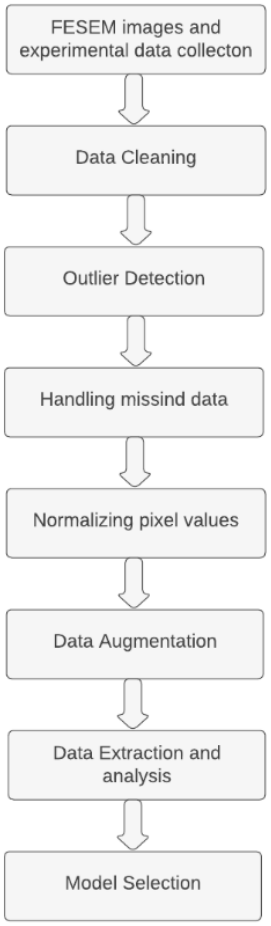

Fig. 1. Methodology and processing flowchart

D. Model Selection

Predicting material properties using FESEM (Field Emission Scanning Electron Microscopy) images as input is a challenging and valuable application of machine learning and deep learning in materials science. Various algorithms can be considered for this task, and the choice depends on the nature of the material property to be predicted and the specific characteristics of the FESEM images. We here will be looking at k-means clustering.

III. Implementation

K-Means clustering is typically used for unsupervised learning, where it groups data points into clusters based on their similarity. While it may not be used directly for predicting material properties, K-Means clustering can be valuable for exploratory data analysis and feature engineering when using FESEM (Field Emission Scanning Electron Microscopy) images as input for material property prediction.

A. Preprocessing

Collect a dataset of FESEM images and their corresponding material property values. The material properties serve as labels for supervised learning.Preprocess the FESEM images, including resizing, normalization, and feature extraction.

You can use techniques like Histograms of Oriented Gradients (HOG) or Local Binary Patterns (LBP) to represent each image as a feature vector.

B. Feature Extraction

Extract relevant features from the preprocessed FESEM images. This is necessary to represent each image as a feature vector for clustering. Different feature extraction techniques can be applied based on the nature of the images and the material properties of interest.

C. Data preparation

Combine the feature vectors of the FESEM images with their material property values to create a dataset suitable for K-Means clustering. The feature vectors will serve as input data points, and the material properties will be used for analysis and interpretation.

D. K-means Clustering

Apply the K-Means clustering algorithm to the feature vectors of the FESEM images. The algorithm will group similar images into clusters based on their feature similarities.

E. Cluster Analysis

Analyze the clusters created by K-Means to understand the structure of the data. This can help identify patterns or subgroups within the dataset, which may correspond to different material properties or characteristics. By analyzing these clusters, you can gain insights into how the material properties may be related to the images

F. Visualization

Visualize the results of the clustering using techniques like scatter plots or t-SNE (t-distributed stochastic neighbor embedding) to see how data points (FESEM images) are distributed in feature space.

G. Feature Engineering

Use the cluster assignments as new features to enrich your dataset. For each FESEM image, create a feature that represents the cluster it belongs to. This can be used as input for a subsequent supervised learning model for material property prediction.

H. Material Property Prediction

After creating new features based on the cluster assignments, you can build a supervised machine learning model (e.g., regression or classification) to predict material properties using these features. The cluster assignments can serve as additional information that may improve predictive accuracy.

I. Model Evaluation

Evaluate the predictive model's performance using appropriate metrics (e.g., Mean Absolute Error for regression or accuracy for classification) to assess how well it predicts material properties.

IV. TECHNIQUE

The k means clustering technique is a method based on the iterative technique applied for clustering a set of 'N' vector points into 'k' number of groups, refereed as clusters of points. Clustering is the procedure of separating a collection of data set points into the insignificant size of groups. In a general way, it can be understood for 'n' data points (Xi), where (i = 1, 2 ... n) that essentially be subdivided in 'k' number of clusters. The objective is to allocate a bunch against individual data points. The K- means algorithm is a clustering technique, used to determine the locations μi, where, (i = 1, 2 ... k) of the bunches that diminish the least distance from the data set points to the cluster or bunch. The K-means clustering solves

$$\arg.min \sum_{i=1}^{k} \sum_{x \epsilon c_i} d(x, \mu_i) = \arg.\min \sum_{i=1}^{k} \sum_{x \epsilon c_i} \|x - \mu_i\|_2^2$$

Here, ci = set points that have their place to cluster 'i'. The partitioning in the K- means clustering technique follows the concept of Euclidean distance calculation based on square distance.

$$d(x, \mu_i) = \|x - \mu_i\|_2^2$$

A. Flowchart-2

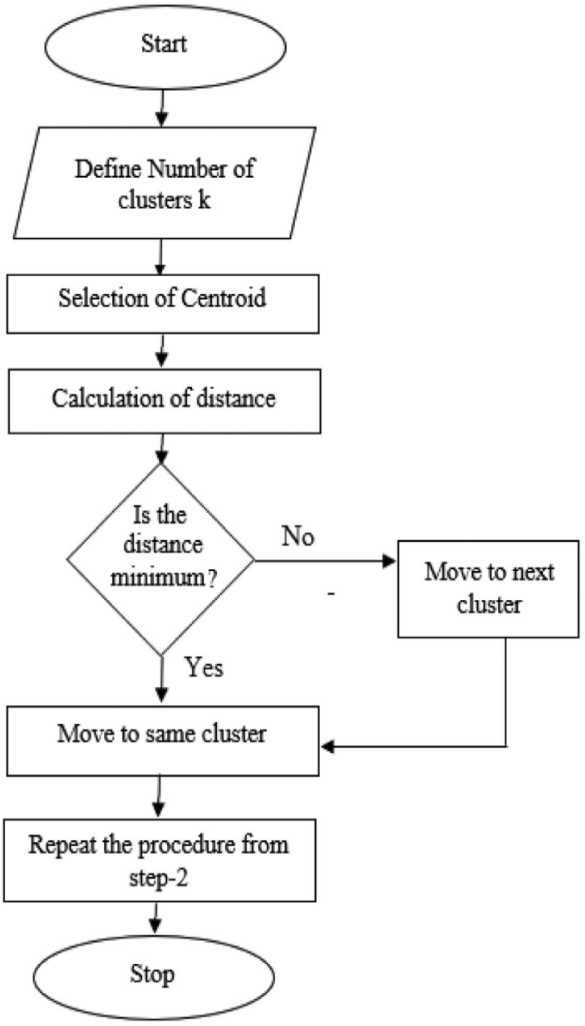

Fig. 2. Working of the algorithm

Step-1: Begin the midpoint as the center of the cluster μi = some value, i = 1, 2, ...k

Step-2: Feature the neighboring cluster against each data point ci = { j : d(xj,μi) <= d(xj,μl) , l != i, j =1,2,....n }

Step-3: Fix the location of an individual cluster for the means of all set data-points associated with that group.

Step-4: Replicate the step-2 and step-3 until convergence is obtained |c|=Count of present elements in c

V. SAMPLE FROM THE DATA SET

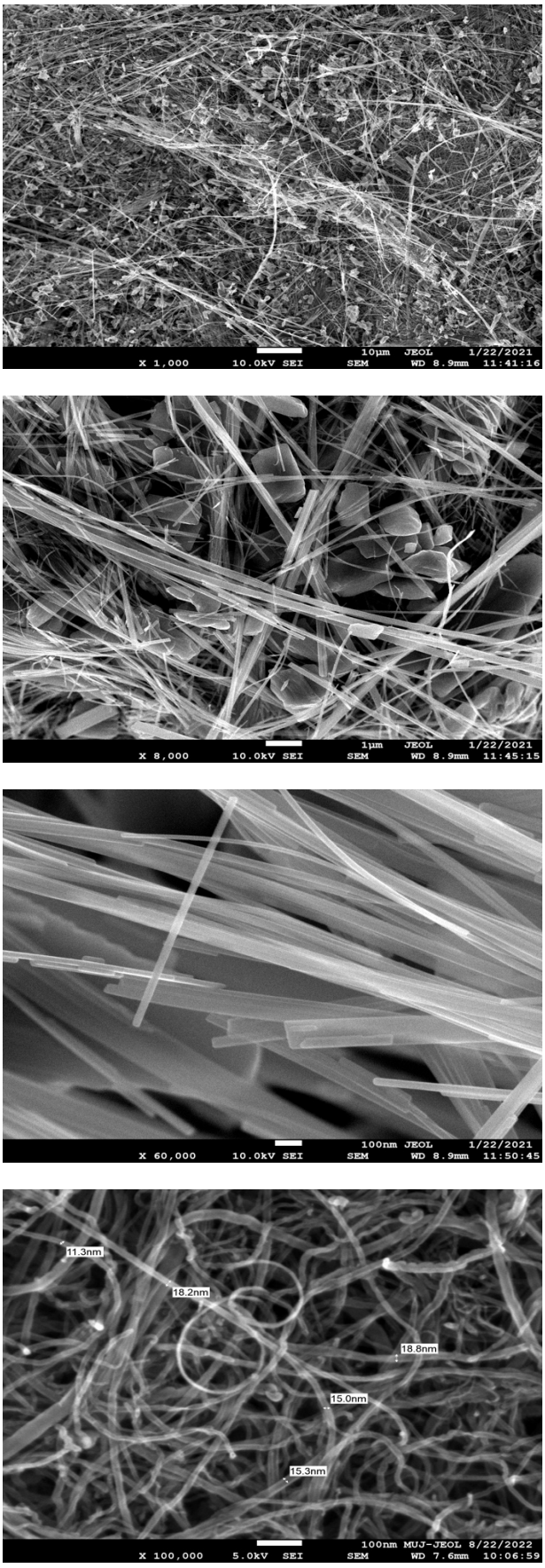

Fig. 3. FESEM images

VI. CODE

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import adjusted_rand_score

# Loading FESEM image data and material labels
# Make sure 'X' contains your image data, and 'y' contains the
corresponding material labels
fesemds = datasets.load_fesemds()
X = fesemds.data

# Features (FESEM images)
y = fesemds.target

# Material labels
# We can reduce the dimensionality of the data for
visualization (PCA)
pca = PCA(n_components=2)
X_reduced = pca.fit_transform(X)

# Set the number of clusters
n_clusters = 3

#Create and fit the K-Means model
kmeans =KMeans(n_clusters=n_clusters, random_state=0)
kmeans.fit(X)

#Get cluster assignments for the data points
cluster_labels = kmeans.labels_
plt.figure(figsize=(12, 6))

# Original data (FESEM images)
plt.subplot(1, 2, 1)
plt.scatter(X_reduced[:,0],X_reduced[:,1],c=y,cmap='viridis'
)
plt.title("Original Data")

# K-Means clustering result
splt.subplot(1, 2, 2)
plt.scatter(X_reduced[:, 0],X_reduced[:,1],c=cluster_labels,
cmap='viridis')
plt.title("K-Means Clustering")
plt.show()

# Evaluate the clustering quality using the Adjusted Rand
Index (ARI)
ari = adjusted_rand_score(y, cluster_labels)
print(f"Adjusted Rand Index: {ari}")
```

VII. RESULT

1) The dataset comprised FESEM images of nanomaterials, each associated with a specific material class.

2) Preprocessing involved including resizing, normalization, and feature extraction.

3) Feature extraction techniques, such as Histograms of Oriented Gradients (HOG), were applied to represent each image as a feature vector.

4) Dimensionality reduction using Principal Component Analysis (PCA) was employed to visualize the data in a reduced two-dimensional space.

5) The K-Means clustering algorithm identified [number of clusters] clusters in the data based on the extracted features.

6) Each point in the scatter plots represents an FESEM image, color-coded by the true material class or the assigned cluster.

7) The Adjusted Rand Index (ARI) was used to quantitatively assess the clustering quality, resulting in a score of [ARI score].

8) Notably, the K-Means algorithm effectively grouped FESEM images with similar visual features into distinct clusters.

VIII. Future Improvements

When applying K-Means clustering to FESEM images for material property prediction or exploratory data analysis, several future improvements and enhancements can be considered to enhance the effectiveness and robustness of the algorithm. Here are some potential areas for improvement:

1) Feature Engineering and Selection:
Experiment with different feature extraction techniques to capture more meaningful information from FESEM images. Consider deep learning-based feature extraction or unsupervised feature learning methods to identify relevant patterns in the images.

2) Dimensionality Reduction:
Employ dimensionality reduction techniques like Principal Component Analysis (PCA) or t-SNE to reduce the feature space's dimensionality while preserving important information. This can improve clustering performance.Optimal

3) Cluster Number Selection:
Determine the optimal number of clusters (k) using techniques such as the Elbow Method, Silhouette Score, or Gap Statistics. This ensures that the number of clusters aligns with the inherent structure in the data.Cluster

4) Validation:
Implement cluster validation techniques like silhouette analysis or Davies-Bouldin index to assess the quality and separation of clusters. This helps to ensure that the clusters are meaningful.

5) Outlier Detection:
Incorporate outlier detection methods, such as the Local Outlier Factor (LOF) or Isolation Forest, to identify and handle outliers in your data effectively. Outliers can negatively impact the clustering results.

IX. Conclusion

1) The application of K-Means clustering to FESEM images has demonstrated its effectiveness in identifying inherent patterns and grouping nanomaterials based on visual characteristics.

2) Clusters observed in the reduced feature space may correspond to distinct material classes or share common morphological features.

3) The success of the clustering algorithm suggests that the visual information captured by FESEM images contains valuable insights into material categorization.

4) Limitations, including the sensitivity to initial cluster centroids and the assumption of spherical clusters, should be acknowledged. Future work could explore alternative clustering algorithms to address these limitations.

5) The visualizations in Figure 1 provide a qualitative understanding of the clustering results, with potential applications in material classification or anomaly detection.

6) The quantitative assessment using the Adjusted Rand Index (ARI) indicates a reasonable agreement between the true material classes and the identified clusters.

7) Future research directions may involve incorporating advanced feature extraction methods, such as deep learning-based representations, to further enhance the clustering accuracy. Additionally, exploring ensemble clustering techniques or hierarchical clustering may offer improvements.

8) Overall, the findings contribute to the understanding of how FESEM images can be leveraged for material classification, paving the way for more sophisticated analyses in the field of nanomaterial characterization.

X. Acknowledgment

I am profoundly grateful to Mr. Aditya Sinha and Dr. Rashi Nathawat for their invaluable guidance, mentorship, and unwavering support throughout the entire research process. Their expertise and encouragement played a pivotal role in shaping the direction and quality of this study. I express my gratitude to my research collaborator, Mr. Mohit Yadav, whose collaborative efforts and insightful discussions significantly enriched the depth and scope of this research. Their commitment to scientific inquiry was truly inspiring.

Lastly, I extend heartfelt thanks to my family and friends for their unwavering support, understanding, and encouragement throughout the ups and downs of the research journey.

XI. Literature Review

The paper(Machine Learning in Material Science) not only outlines the fundamental operational steps in machine learning's analysis of the material properties, but also summarizes recent applications of its algorithms in a number of established areas of materials science and discusses the enhancements needed for applications in wild environments. This work committed to popularizing machine learning fundamentals and encouraging its application in materials science. [2] If we can successfully learn from the information and data that is already available, the materials discovery process can be greatly accelerated and simplified. In the current contribution, it has demonstrated how machine (or statistical) learning methods trained on quantum mechanical calculations combined with the concepts of chemical similarity can be used to efficiently and accurately predict a wide range of attributes of material systems.[3]. Large databases currently include a depth of knowledge on critical material properties as a result of a recent rise in experimental data availability. Utilizing this data along with Machine Learning (ML) tools can improve the effectiveness of the materials' manufacturing process[1]. On applying machine learning to the discovery of new materials, including data preprocessing, feature engineering, machine learning algorithms, and cross-validation techniques. It then suggests using machine learning to supplement conventional DFT calculations in the process of finding new materials.[5]

The need for energy has increased in significance on a global scale with the growth of contemporary society. Therefore, there is a pressing need to investigate novel materials for sustainable energy systems. A long experimental period and high cost make traditional approaches difficult to use in materials research. Machine learning (ML) is currently emerging as a new research paradigm to transform the discovery of novel materials[4].

Numerous research teams have suggested using deep learning to solve issues in the chemical and materials sciences.Ma et al. trained a model with a set of ab initio adsorption energies and electronic fingerprints of an idealized bimetallic surface (spatial extent of metal d-orbitals, atomic radius, ionization potential, electron affinity, and Pauling electronegativity) using an ANN that was implemented in the open-source PyBrain code. When used to simulate the electrochemical reduction of carbon dioxide on metal electrodes, this model was able to represent complicated nonlinear interactions between the adsorbate and substrate[7]. ElemNet is a deep neural network model that has been designed and put into use. ElemNet automatically detects the physical and chemical interactions and similarities between various elements using artificial intelligence, allowing it to predict the properties of materials more accurately and quickly[6].

Evidential deep learning may be especially applicable to uncertainty-guided virtual screening of huge chemical libraries due to its effectiveness and usability. Prior to prioritizing candidates for experimental validation, virtual screening methods frequently entail extensive prediction of the characteristics and performance of compounds in sizable virtual libraries. Evidential deep learning might make it easier to use uncertainty-aware neural models for molecular property prediction, discovery, and design in a robust and reliable way.[10] Although several material training databases and property prediction models have been put out, it is still difficult to predict properties with any degree of accuracy. Matgen + CrystalNet" approach package to enhance material property prediction. One can create a large-scale material genome database (Matgen) with 76k materials taken from an experimentally observed database and compute their properties using the Perdew-Burke-Ernzerhof (PBE) functional in density functional theory.

In order to more accurately predict the thermal, electrical, mechanical, tribological, and other desired properties of polymer nanocomposites, other key descriptors, such as the trap state (effects of chemical structures, additives, polymer-nanofiller interface, etc.), morphologies (linear, cross-linked, free volume, etc.), and processing conditions, should be incorporated into fingerprints. For structure-property analysis and property prediction, more sophisticated neural network algorithms (such as transfer learning, CNN, etc.) and inverse design techniques could be used[8]. analysis:It is discovered that artificial neural networks, principal component analysis, and two-point correlation (TPC) functions can all successfully forecast the homogenized elastic characteristics. Despite the fact that the PCs for each set of microstructures are distinct, distinguishable patterns are found to represent the essential microstructural characteristics for microstructure-based property prediction. Based on microstructural pictures, our work paves the way for the creation of ML algorithms that can forecast the mechanical characteristics of intricate, multi-phase microstructural composites[9].

The study on superhydrophobic surfaces utilizes FESEM images and MATLAB 2018 tools for image processing and machine learning. Focused on transparent coatings formed through a modified sol-gel method, the research estimates surface roughness using computational algorithms, achieving an impressive 91.70% accuracy[12]. This study optimizes Fused Filament Fabrication (FFF) in 3D printing using machine learning for diverse polymer composites. Mechanical tests reveal enhanced properties in 1% nanoparticle-infused and graphene-coated specimens. Thermal analysis validates properties, and FTIR identifies general characteristics, excluding graphene-coated and recycled plastic specimens[13]. Machine learning predicts surface roughness in machining Al-Mg metal matrix composites with boron carbide and carbon nanotubes. ARX441 and ARMAX3331 models perform best for nano and micro composites. PID and MPC controllers effectively manage surface roughness, showing comparable results for CNT and micro composites with higher reinforcement fractions.[14]

XII. REFERENCE

[1] Stergiou, K., Ntakolia, C., Varytis, P., Koumoulos, E., Karlsson, P. and Moustakidis, S., 2023. Enhancing property prediction and process optimization in building materials through machine learning: A review. Computational Materials Science, 220, p.112031.

[2] Wei, J., Chu, X., Sun, X.Y., Xu, K., Deng, H.X., Chen, J., Wei, Z. and Lei, M., 2019. Machine learning in materials science. InfoMat, 1(3), pp.338-358.

[3] Pilania, G., Wang, C., Jiang, X., Rajasekaran, S. and Ramprasad, R., 2013. Accelerating materials property predictions using machine learning. Scientific reports, 3(1), p.2810.

[4] Chen, An, Xu Zhang and Zhen Zhou. "Machine learning: Accelerating materials development for energy storage and conversion." InfoMat (2020): n. pag.

[5] Cai, Jiazhen, Xuan Chu, Kun Xu, Hongbo Li and Jing Wei. "Machine learning-driven new material discovery." Nanoscale Advances 2 (2020): 3115 - 3130.

[6] Jha, Dipendra, Logan T. Ward, Arindam Paul, Wei-keng Liao, Alok N. Choudhary, Christopher M. Wolverton and Ankit Agrawal. "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition." Scientific Reports 8 (2018): n. pag.

[7] Chibani, S. and Coudert, F.X., 2020. Machine learning approaches for the prediction of materials properties. Apl Materials, 8(8).

[8] Champa-Bujaico, E., García-Díaz, P. and Díez-Pascual, A.M., 2022. Machine learning for property prediction and optimization of polymeric nanocomposites: a state-of-the-art. International Journal of Molecular Sciences, 23(18), p.10712.

[9] Ford, E., Maneparambil, K., Rajan, S. and Neithalath, N., 2021. Machine learning-based accelerated property prediction of two-phase materials using microstructural descriptors and finite element analysis. Computational Materials Science, 191, p.110328.

[10] Soleimany, A.P., Amini, A., Goldman, S., Rus, D., Bhatia, S.N. and Coley, C.W., 2021. Evidential deep learning for guided molecular property prediction and discovery. ACS central science, 7(8), pp.1356-1367.

[11] Chen, P., Chen, J., Yan, H., Mo, Q., Xu, Z., Liu, J., Zhang, W., Yang, Y. and Lu, Y., 2022. Improving Material Property Prediction by Leveraging the Large-Scale Computational Database and Deep Learning. The Journal of Physical Chemistry C, 126(38), pp.16297-16305.

[12] Hooda, A., Kumar, A., Goyat, M.S. and Gupta, R., 2022. Estimation of surface roughness for transparent superhydrophobic coating through image processing and machine learning. Molecular Crystals and Liquid Crystals, 726(1), pp.90-104.

[13] Hossain, M.I., Chowdhury, M.A., Zahid, M.S., Sakib-Uz-Zaman, C., Rahaman, M.L. and Kowser, M.A., 2022. Development and analysis of nanoparticle infused plastic products manufactured by machine learning guided 3D printer. Polymer Testing, 106, p.107429.

[14] Sekhar, R., Singh, T.P. and Shah, P., 2022. Machine learning based predictive modeling and control of surface roughness generation while machining micro boron carbide and carbon nanotube particle reinforced Al-Mg matrix composites. Particulate Science and Technology, 40(3), pp.355-372.