Leveraging machine learning for the determination of optimal cerium and cobalt content in bioglass scaffold for improved angiogenesis and mechanical properties

SURGE PROJECT

ADARSH SAHU

Metallurgical and Materials Engineering

National Institute of Technology Raipur

Guide: Prof. Kantesh Balani

Date: 21 July 2024

OUTLINE

MOTIVATION :

Bioglass is a widely recognized material that has found extensive applications in bone scaffolding and dental implants. It is renowned for its exceptional bioactivity and ability to stimulate bone growth. However, it has suboptimal angiogenesis and mechanical properties. Cobalt (Co) is well known for its propensity to induce angiogenesis and osteogenesis. Similarly, Cerium (Ce) is also a prime candidate known for its antioxidant and mechanical properties. Nevertheless, there is a risk of potential cytotoxicity and altered property changes at certain doping levels. By leveraging machine learning to predict doping levels based on the required material properties, we can significantly accelerate the discovery of better applications, reduce experimental costs, and enhance the efficiency of material development processes.

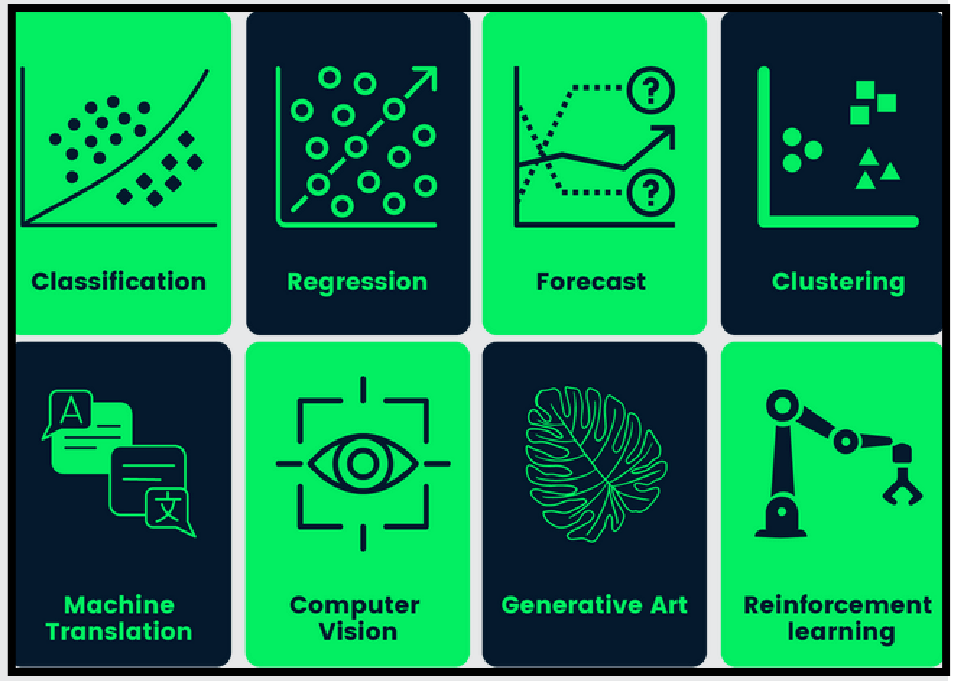
MACHINE LEARNING

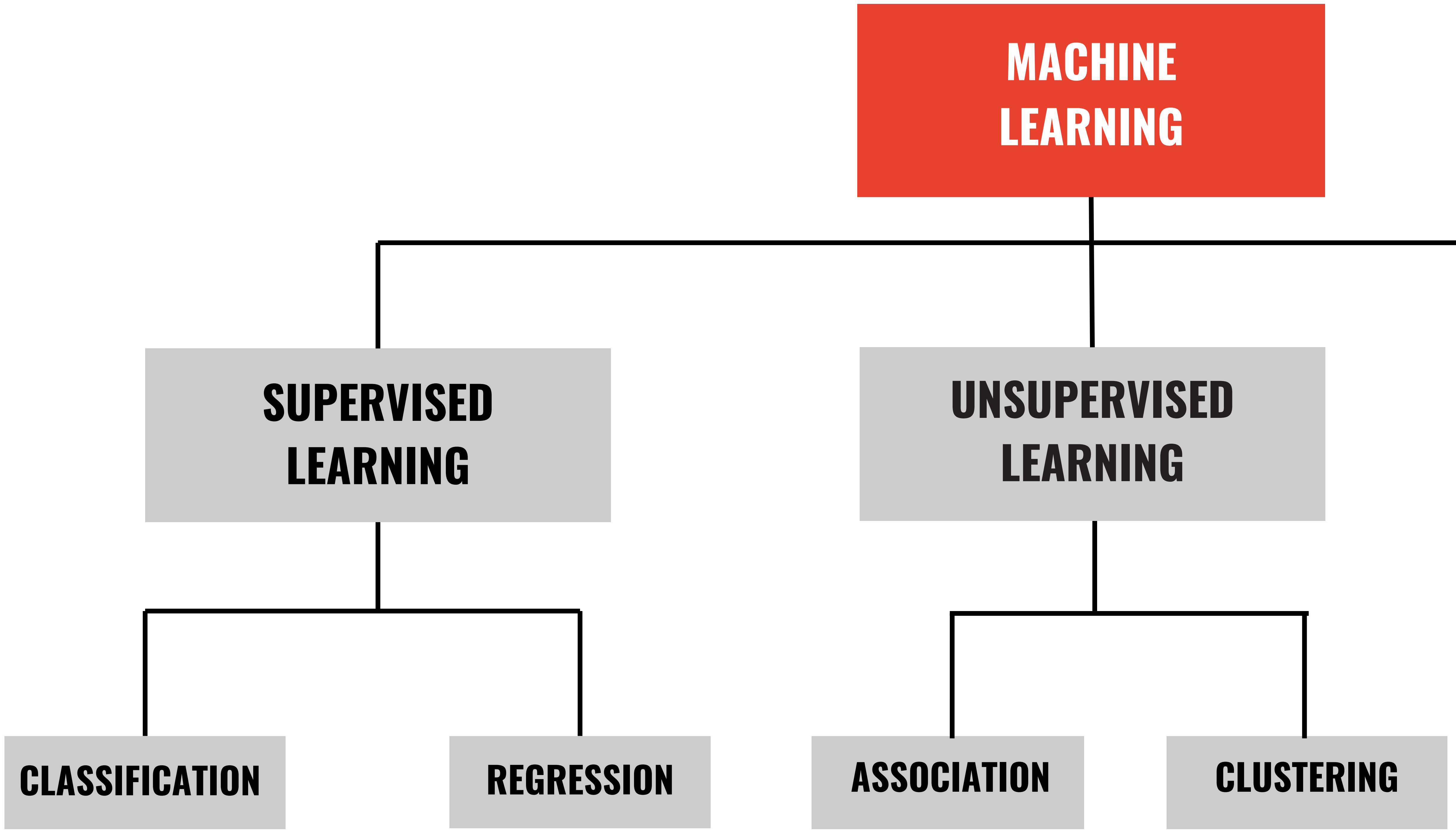
Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalize to unseen data, and thus perform tasks without explicit instructions.

APPLICATIONS:

1. Computer Vision
2. Healthcare Sector
3. Financial Services
4. Natural Language Processing

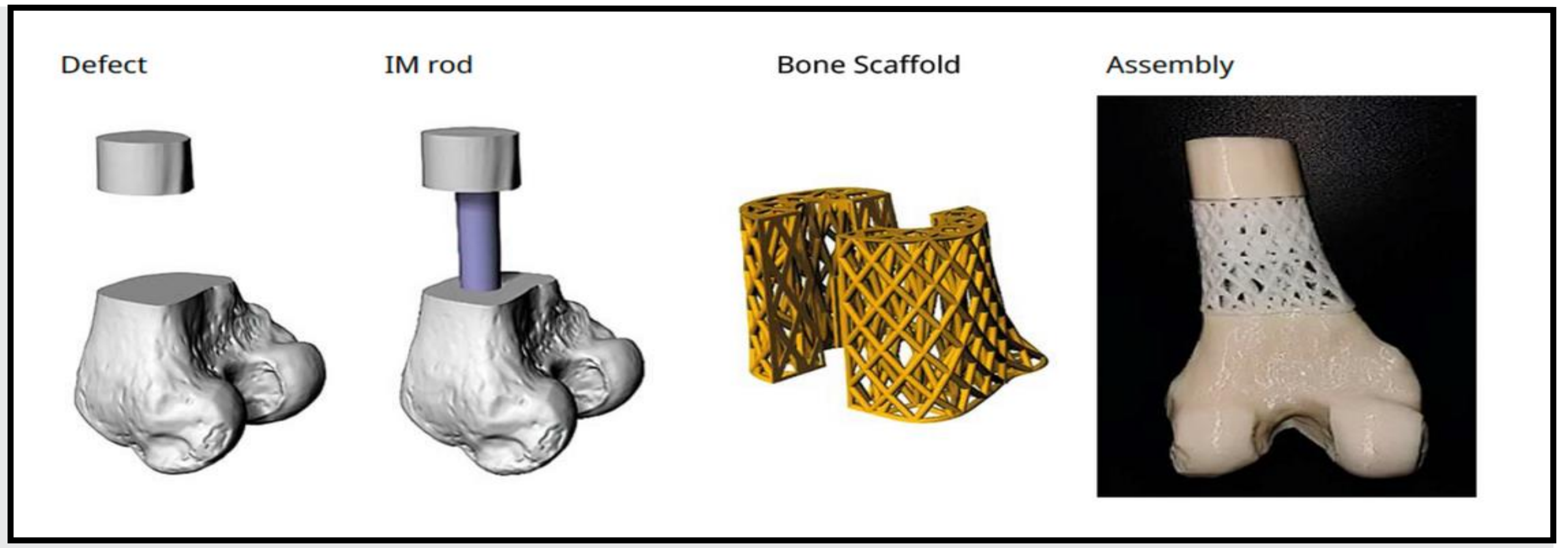
Stock Price Prediction




MACHINE LEARNING
SUPERVISED LEARNING
UNSUPERVISED LEARNING
REINFORCEMENT LEARNING
CLASSIFICATION
REGRESSION
ASSOCIATION
CLUSTERING

BIOGLASS

- Silica (SiO_2)
- Lime (CaO)
- Phosphorus pentoxide (P_2O_5)



Properties:

- Biocompatibility- signifies minimal immunological response and favorable tissue integration to provide long-term implants.
- Bioactivity -surface reactivity that promotes osteogenesis (bone formation) and angiogenesis (blood vessel formation) through the controlled release of beneficial ions.

Mechanical Properties (Eg: 45S Bioglass):

- Compressive modulus- 60GPa
- Bending Strength- 40MPa

Applications :

- Bone Grafting
- Wound Healing Dressings
- Dental Reconstruction

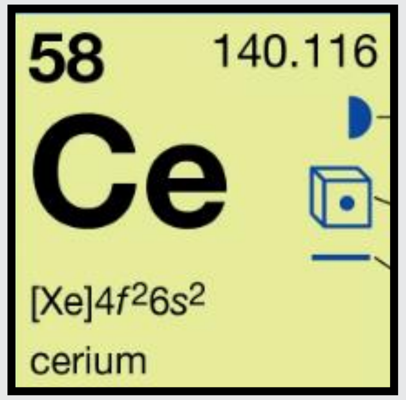
Limitations :

- Mechanical Weakness especially under tension or bending in load-bearing application; Elastic Modulus-50GPa, Flexural Strength-70GPa
- Limited Biological functionality like lower antibacterial properties, angiogenesis & osteoinductivity rates
- Potential Inflammation, can lead to complications in wound healing processes

https://images.app.goo.gl/yBsP7dZDu8HPsr9V6

Why Cerium and Cobalt doping ?

Cerium

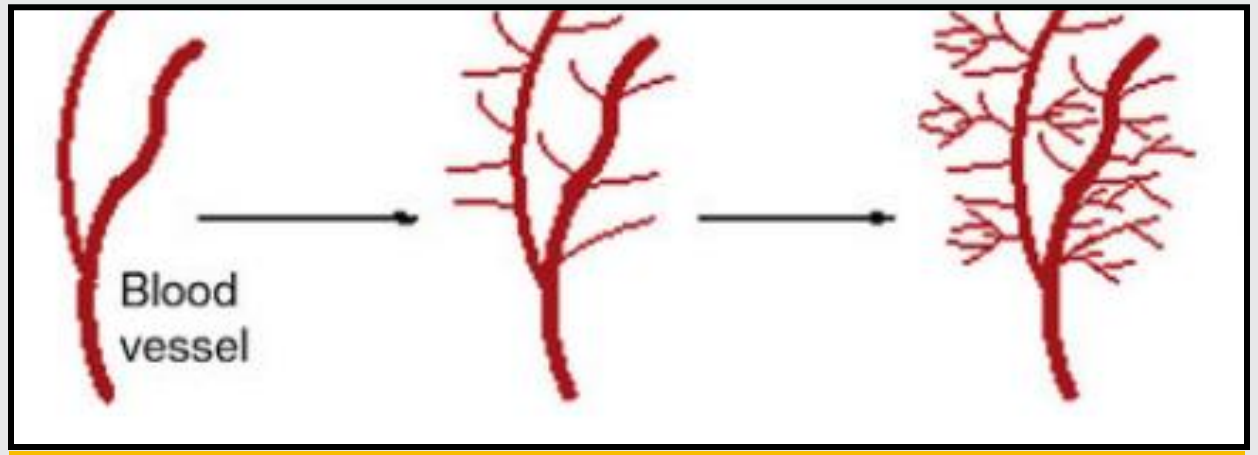


- Antioxidant Properties (Release of Ce3+, Ce4+)
- Enhanced Bioactivity (Tissue Integration)
- Antibacterial Activity (Tested against *S. Aureus and E.Coli*)
- Improves Mechanical Properties(Hardness, Fracture Toughness)

Tensile and Compressive Strength

Cobalt

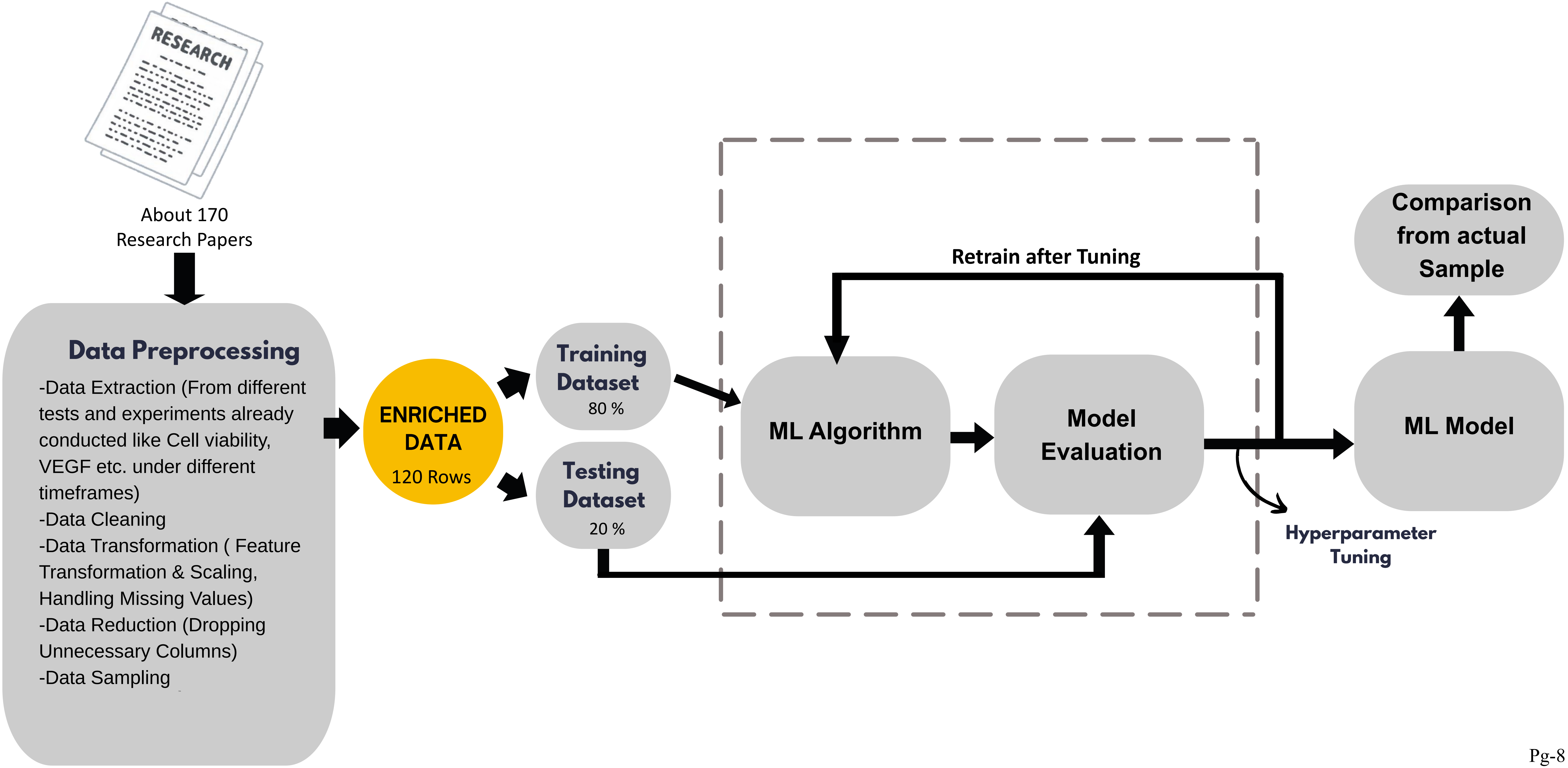
- Promotes angiogenesis (Co2+ stimulates HIF-1α)
- Enhanced Osteogenesis
- Enhancement of Mechanical Properties
- Can impart Magnetic Properties (Useful in targeted drug delivery systems, MRI compatibility of devices)

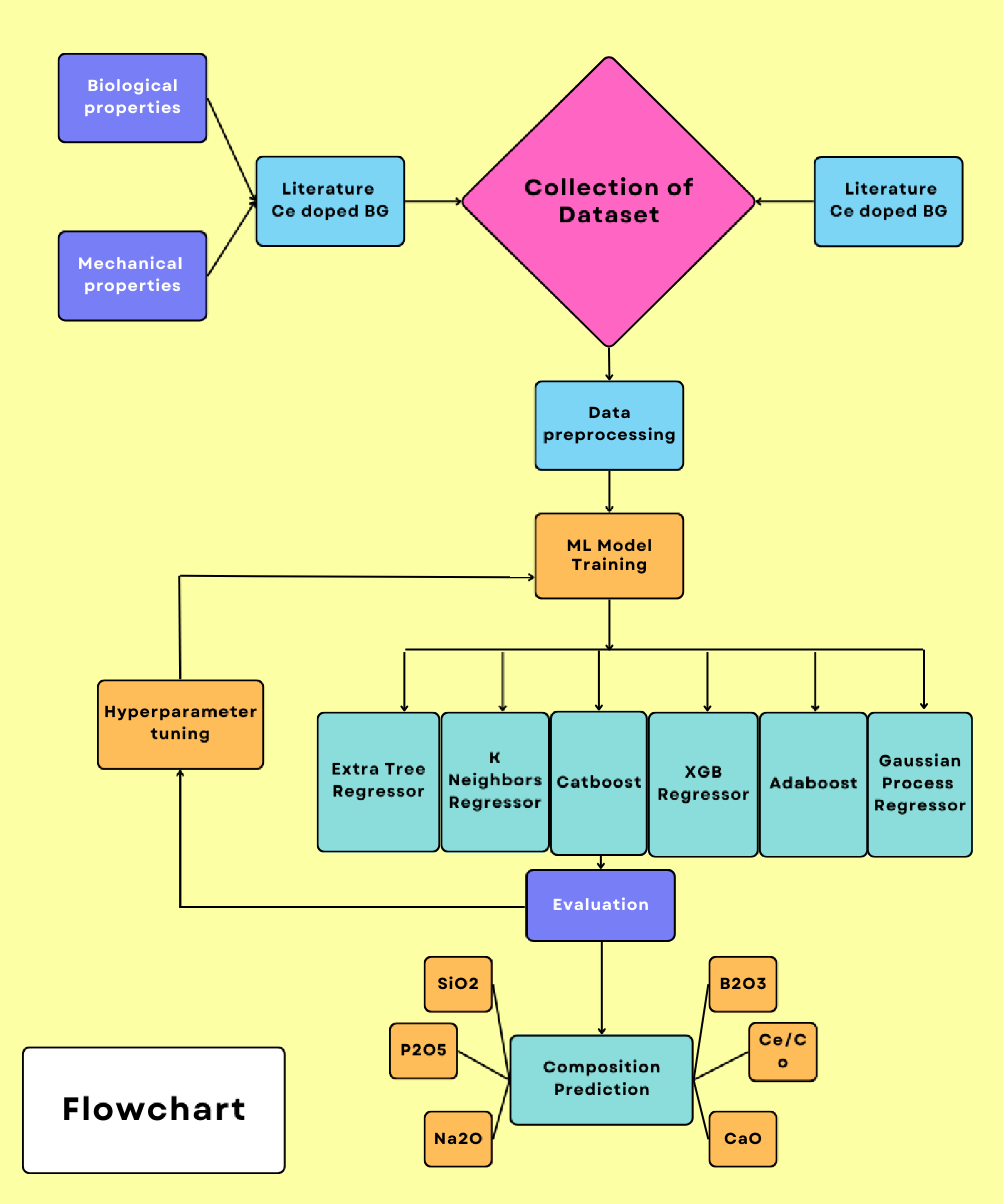


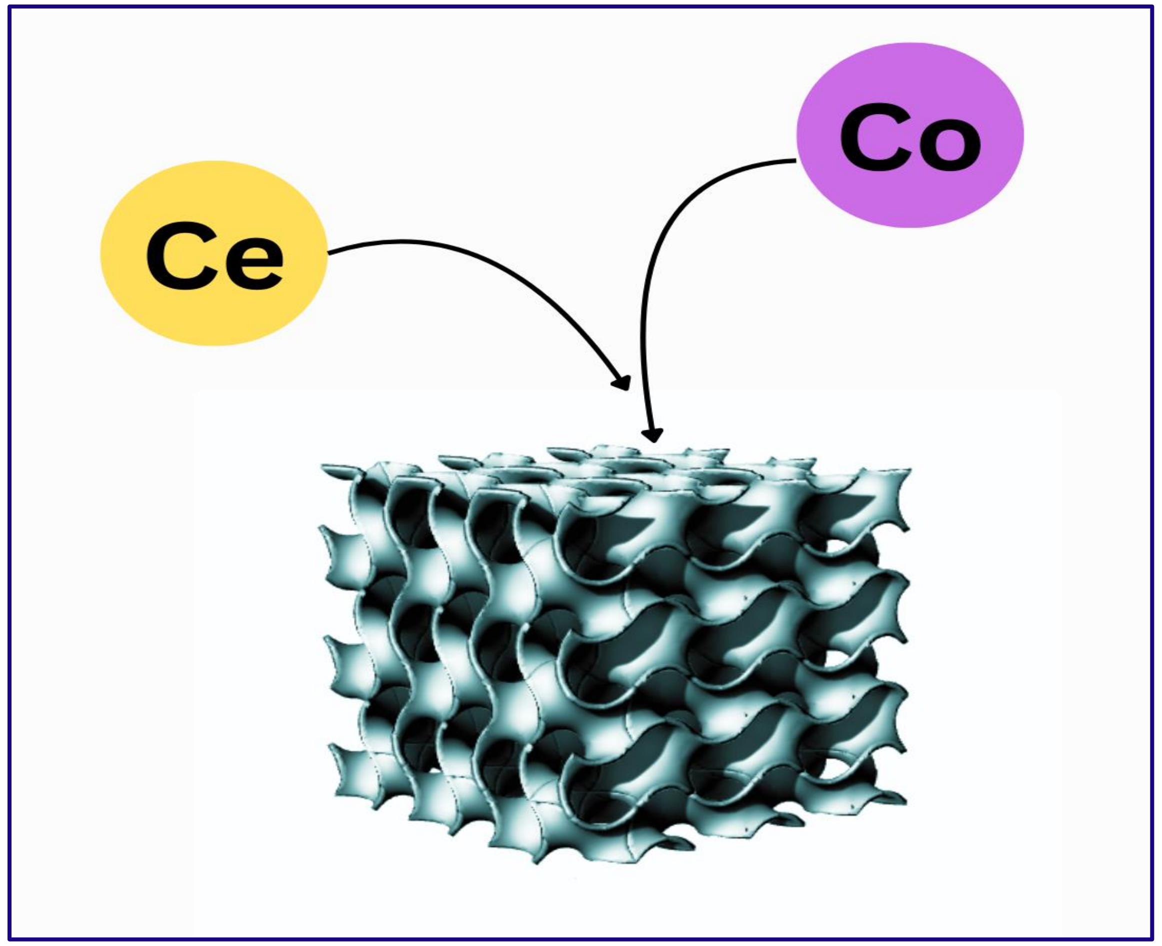
Angiogenesis

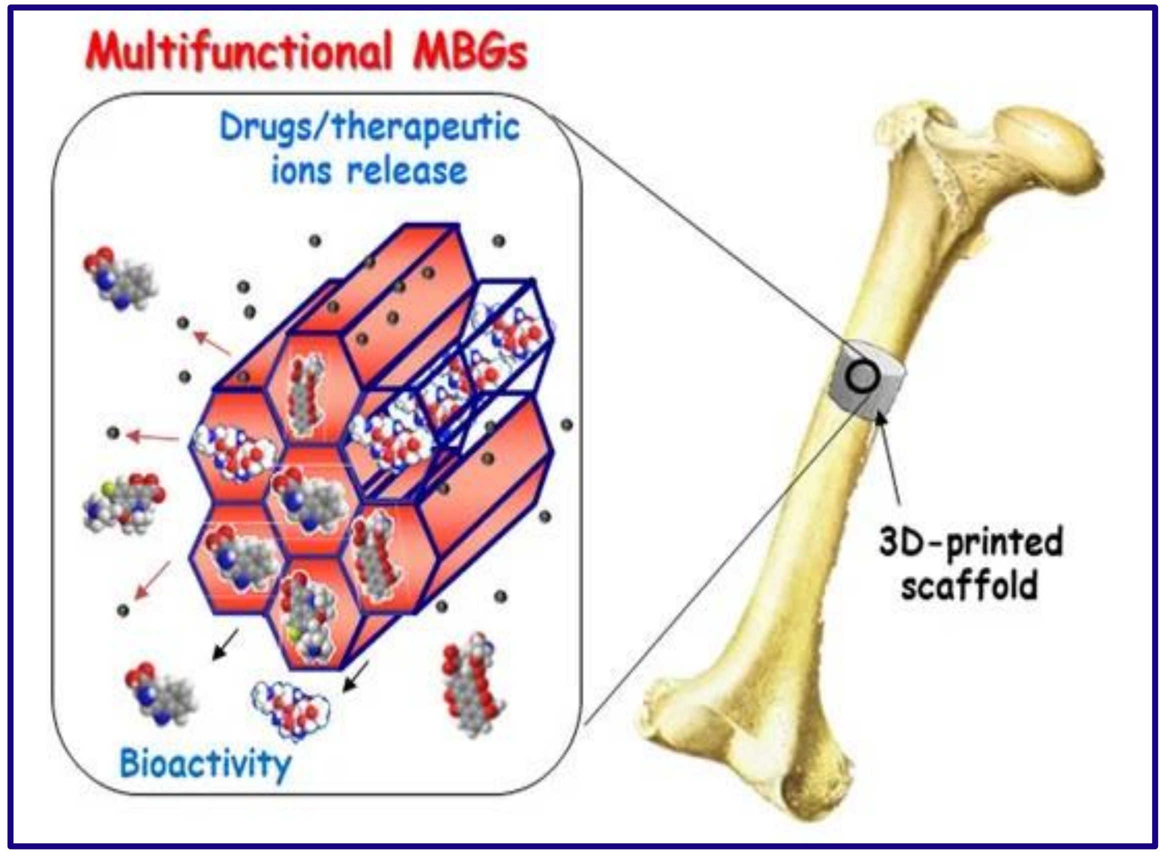
https://www.azom.com/article.aspx?ArticleID=18589 , https://www.researchgate.net/figure/Angiogenesis-is-the-physiological-

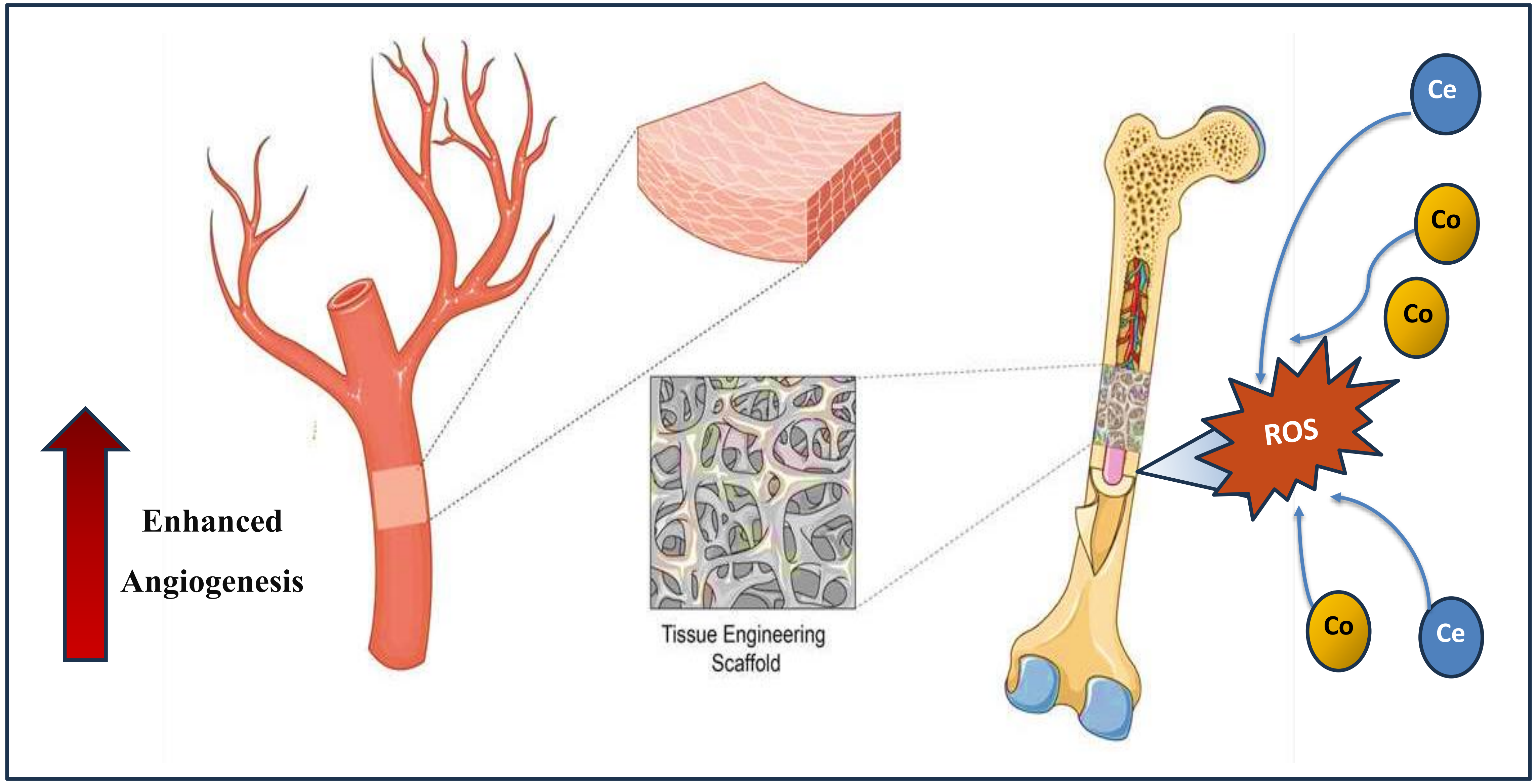
WORK PLAN

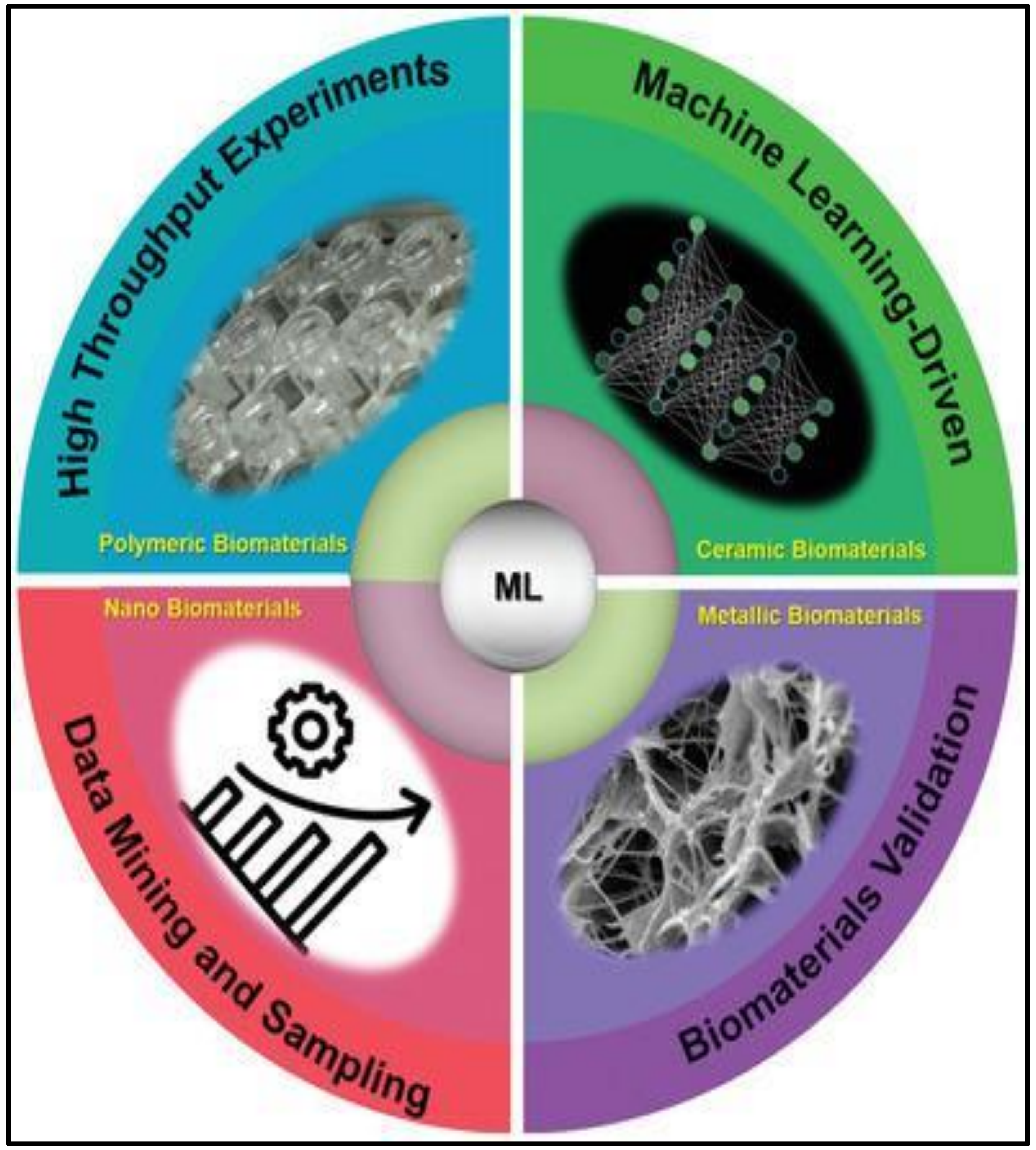

RESEARCH
About 170
Research Papers
Data Preprocessing
-Data Extraction (From different tests and experiments already conducted like Cell viability, VEGF etc. under different timeframes)
-Data Cleaning
-Data Transformation (Feature Transformation & Scaling, Handling Missing Values)
-Data Reduction (Dropping Unnecessary Columns)
-Data Sampling
ENRICHED DATA
120 Rows
Training Dataset
80 %
Testing Dataset
20 %
Retrain after Tuning
ML Algorithm
Model Evaluation
Hyperparameter Tuning
ML Model
Comparison from actual Sample


Biological properties
Mechanical properties
Literature Ce doped BG
Collection of Dataset
Literature Ce doped BG
Data preprocessing
ML Model Training
Hyperparameter tuning
Extra Tree Regressor
K Neighbors Regressor
Catboost
XGB Regressor
Adaboost
Gaussian Process Regressor
Evaluation
SiO2
P2O5
Na2O
Composition Prediction
B2O3
Ce/Co
CaO
Flowchart


Ce
Co


Multifunctional MBGs
Drugs/therapeutic ions release
Bioactivity
3D-printed scaffold


Ce
Co
Co
ROS
Co
Ce
Tissue Engineering
Scaffold
Enhanced
Angiogenesis


High Throughput Experiments
Machine Learning-Driven
Polymeric Biomaterials
Ceramic Biomaterials
ML
Nano Biomaterials
Metallic Biomaterials
Data Mining and Sampling
Biomaterials Validation

Steps and Challenges

Training parameters as Input (X):

- Cell Viability Percentage % - MTT assay (24h, 48h, 72h)
- ALP activity (U/L)(1week, 2weeks, 3weeks)
- Vascular Endothelial Growth Factor (1day, 3day, 7days)
- Mechanical Properties including

Compressive strength (MPa), Pore Size (nm), Wall Thickness(nm), Lattice Parameter (nm) , Pore Volume (cm3/g) etc.

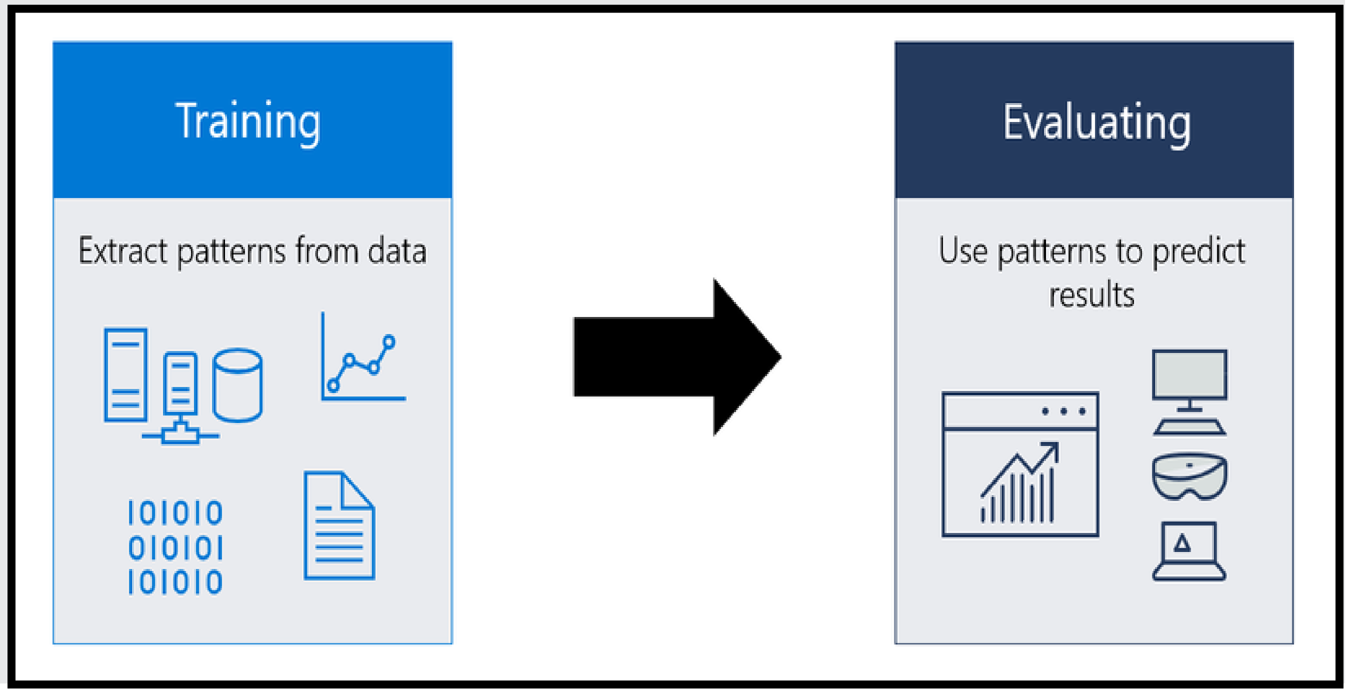
Training parameters Used as Output (Y):

Percentage composition of Ce and Co in mol% as dopants.

1.DATA PRE-PROCESSING

- Handling missing property values using Imputation Techniques
- Normalizing or Standardizing the Data in-order to scale

2.MODEL SELECTION AND TRAINING



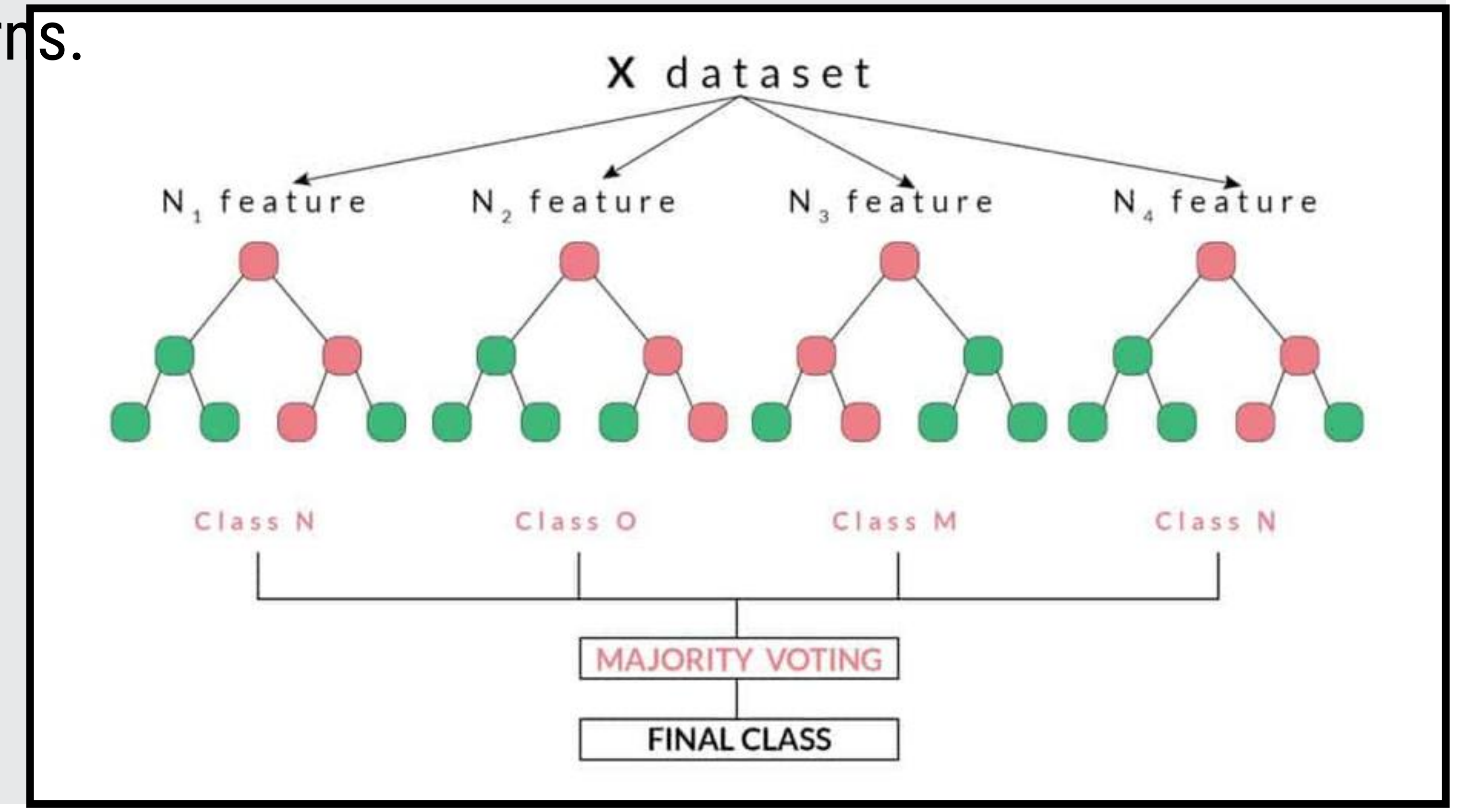
MACHINE LEARNING MODELS

As per our case scenario ML models most suitable to apply in case of multiple parameter prediction would be:

- **Multi-Output Regression Wrapper (used with other base learners)**
  Useful for extending single-output models to handle multiple outputs.
- **Random Forest Regressor (Extra Tree Regressor)**
  Robust against overfitting and can handle non-linear relationships.
- **Gaussian Process Regressor**
  Effective for small datasets and can capture complex patterns.

3.MODEL EVALUATION

- Cross- Validation Score
- MSE, R-Squared, RMSE
- Visual Analysis of Residuals
- Predicted v/s Actual Plots



https://www.azom.com/article.aspx?ArticleID=18589

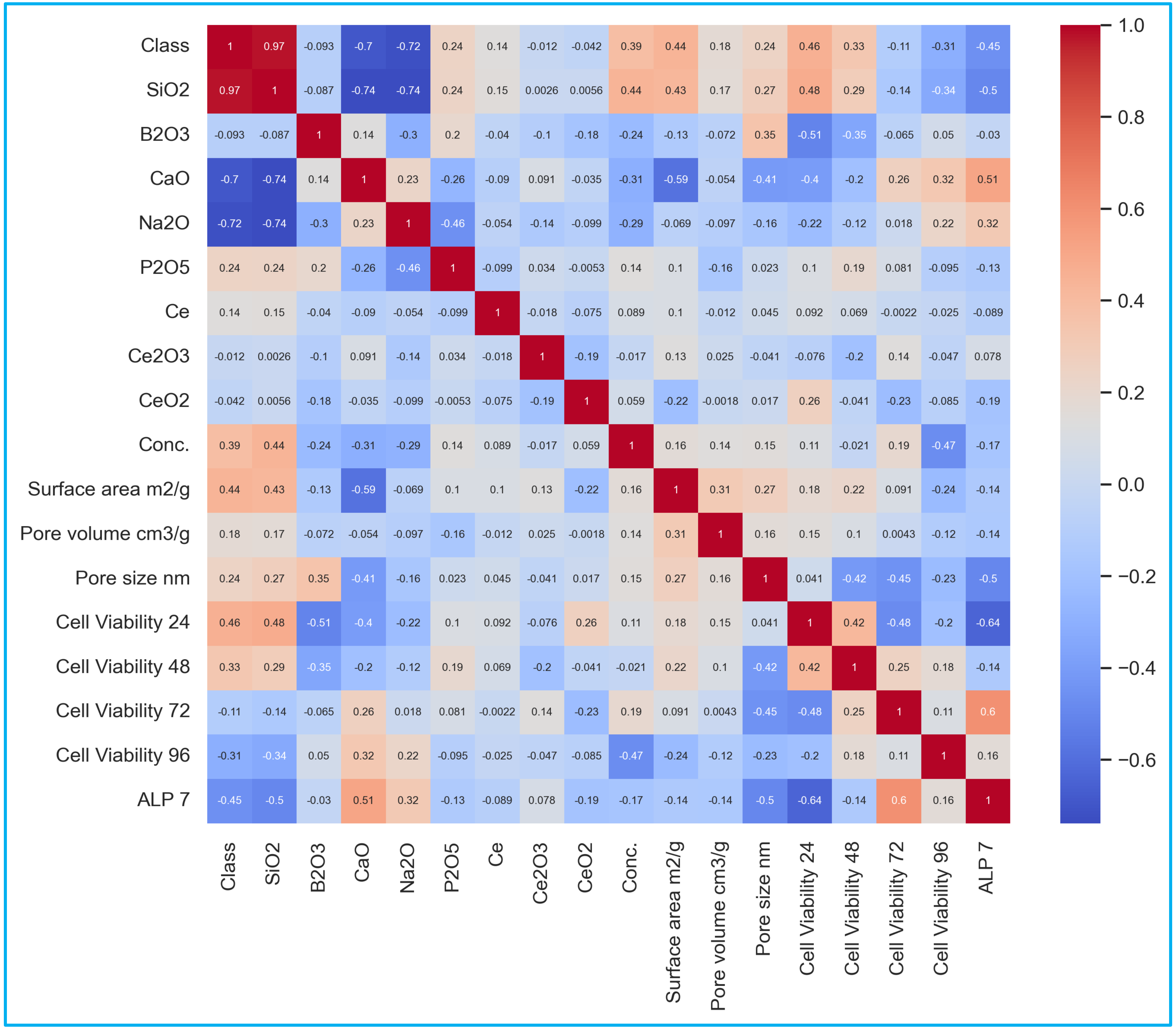
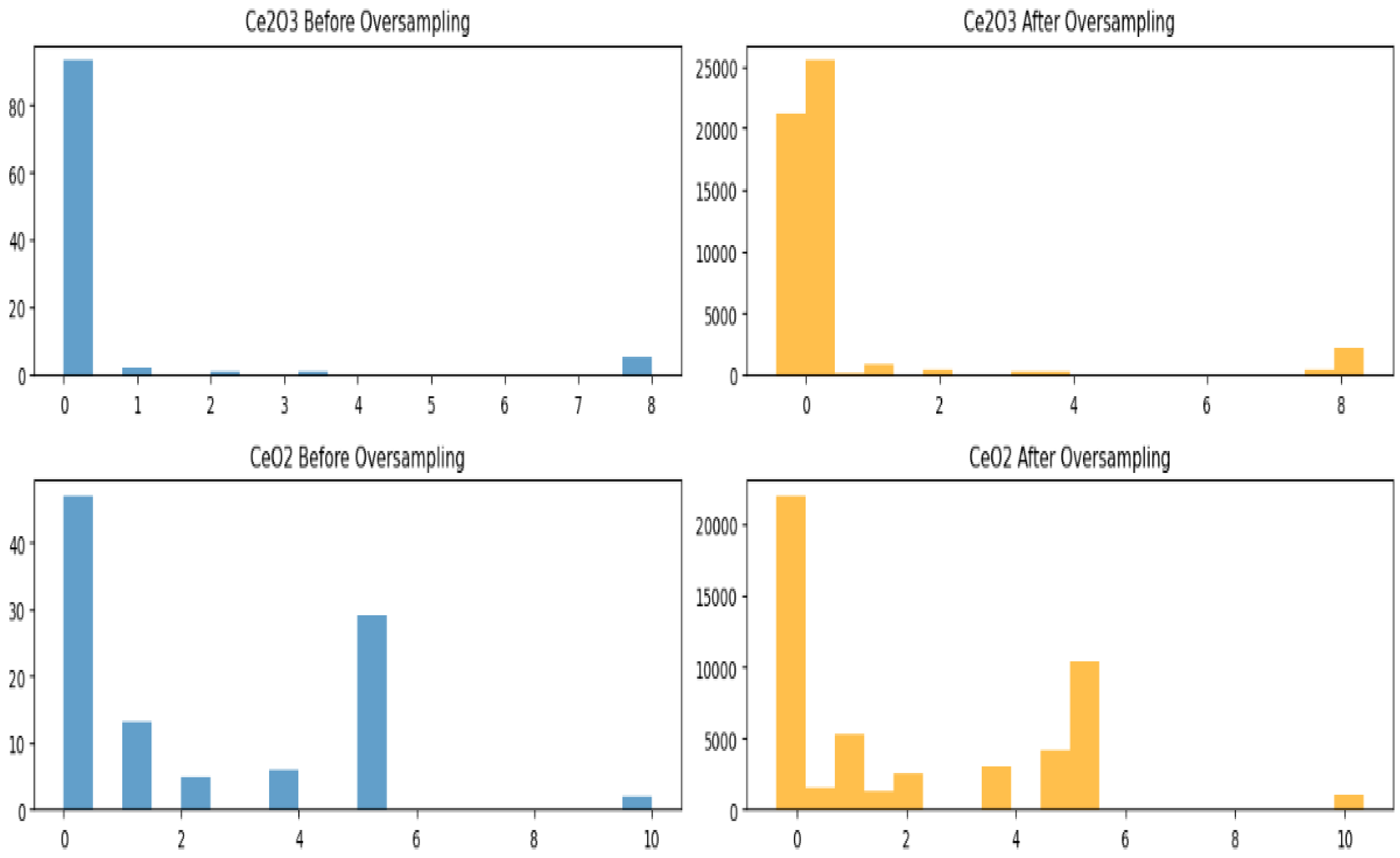


Fig.1 Correlation Heatmap for Cerium doped BG

RESULTS

- This Pearson Correlation Heatmap relates the different targets and features present in our dataset. Correlation **doesn't necessitate** causation.
- The greater the correlation value means the values are directly proportional and an increase in one quantity may lead to an increase in the next one
- Negative Correlation values represent inverse relations of the parameters with each other.
- Parameters like cell viability relates to CaO percentage.



- The bar chart plotted depicts the change in the **dataset distribution** caused by the application of the KNN imputation technique for handling the missing values present in our raw dataset.
- The attached Plots are for the change caused in distribution for **CeO2** and **Ce2O3**.

Fig.3 Data Spread Comparison Before and After Oversampling (Cerium doped)

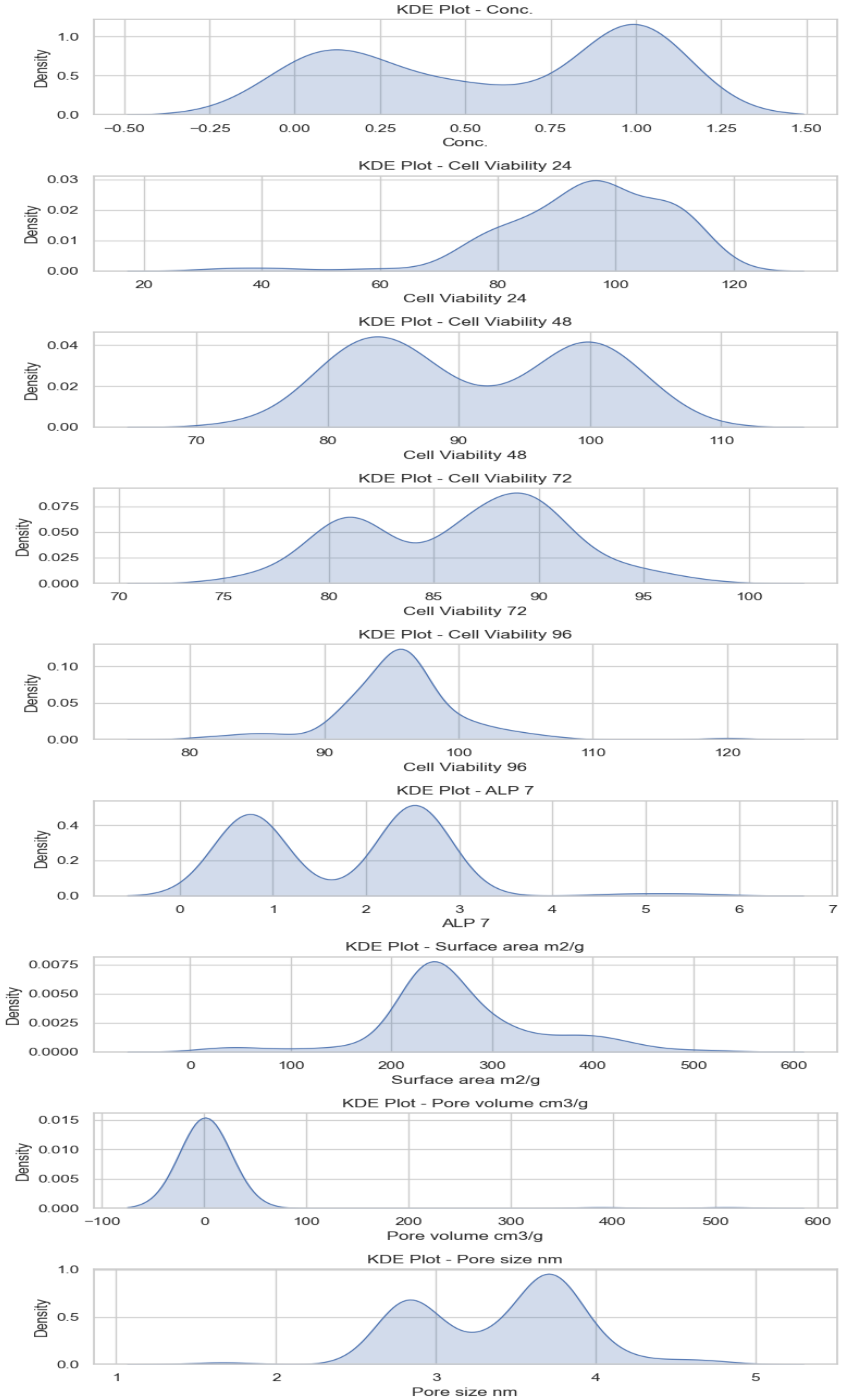


Fig.4 KDE plots for Cerium as dopant

- These Kernel Density Estimate Plot (KDE) shows the values of different **Mechanical and Biological features** present in our dataset to the number of data points available in our dataset sample.
- The curve in the plots represents a **Bimodal** distribution in most cases or a **Unimodal** distribution in some as well.

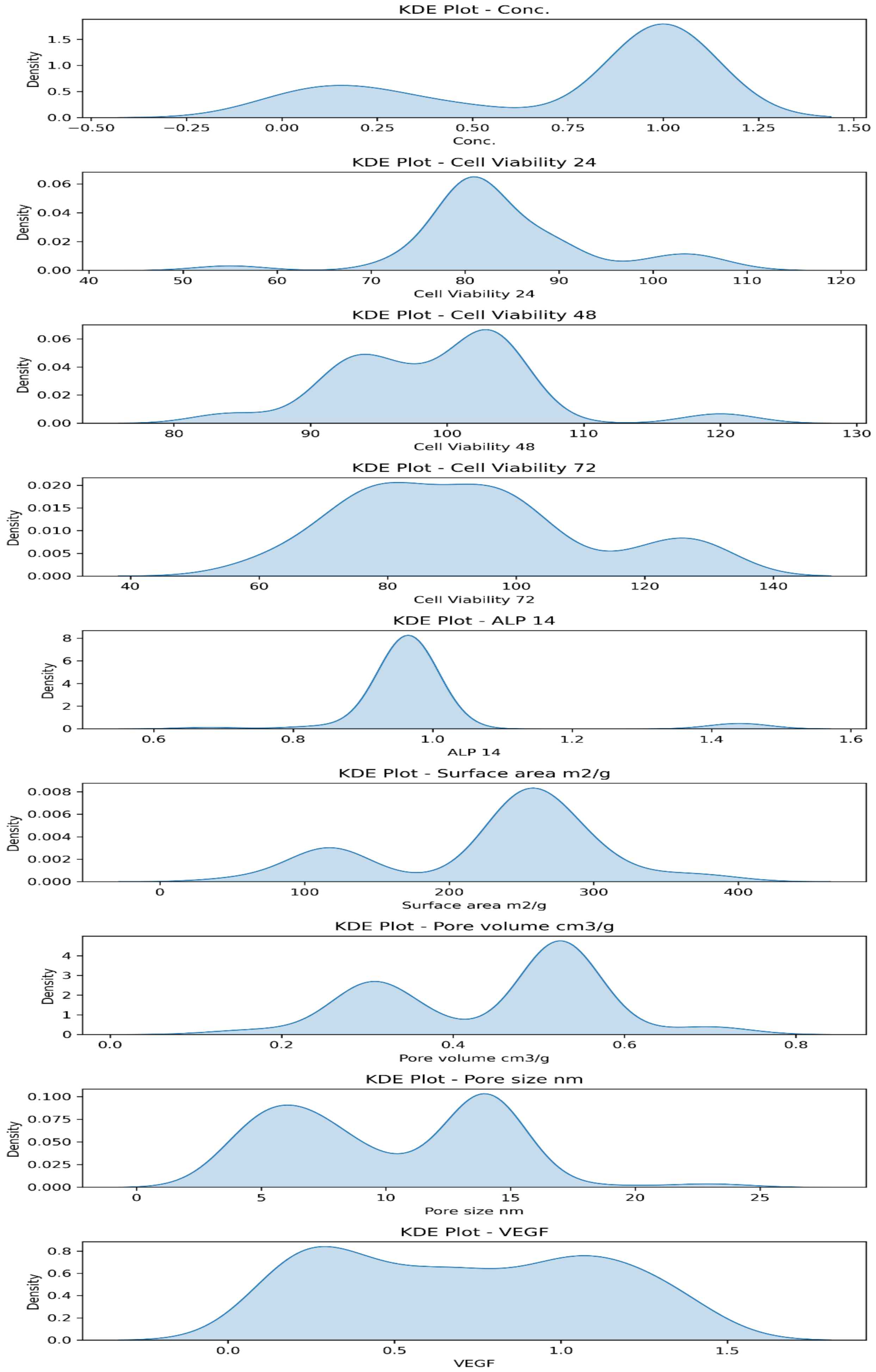


Fig.5 KDE plots for Cobalt as dopant

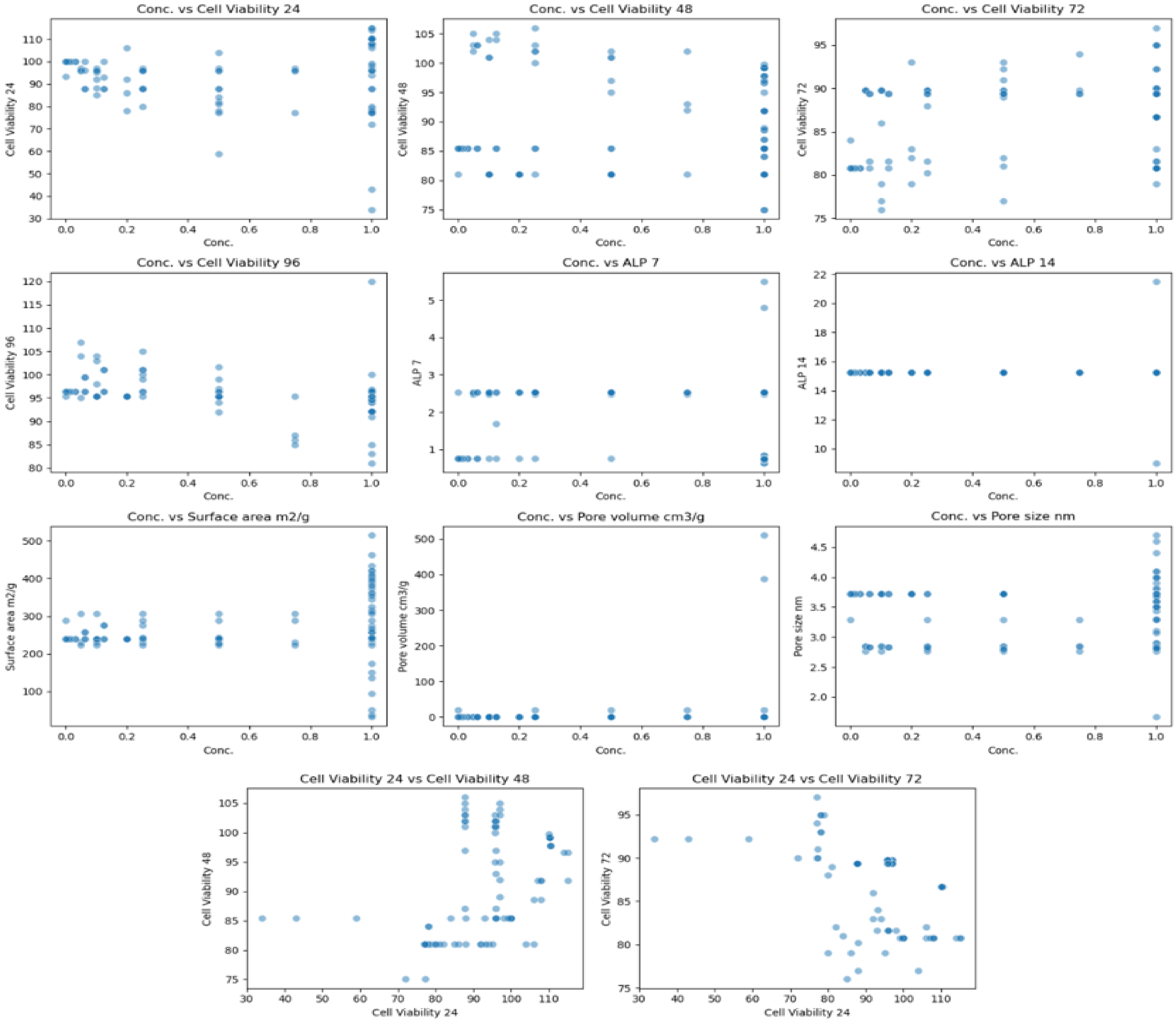


Fig.6 Data features distribution scatter plots (Cerium)

- The Alongside Scatter plot relates the sample Concentration taken for testing to Biological Properties, Concentration to Mechanical Properties and Interrelation of Biological features.

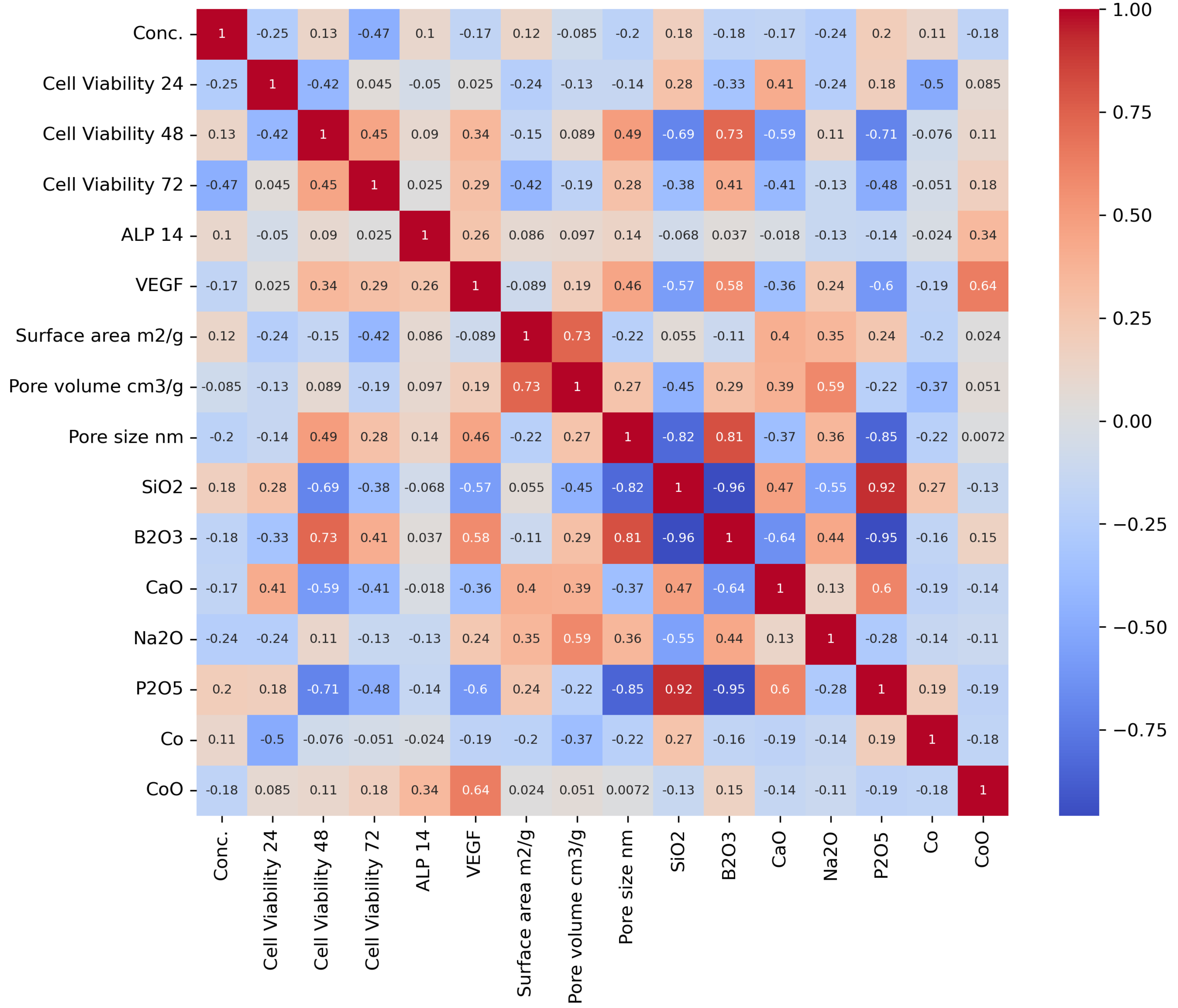


Fig.7 Correlation Heatmap for Cobalt doped BG

- This Pearson Correlation Heatmap relates the different targets and features present in our dataset. Correlation **doesn't necessitate** causation.
- The greater the correlation value means the values are directly proportional and an increase in one quantity may lead to an increase in the next one
- Negative Correlation values represent inverse relations of the parameters with each other.
- Mechanical properties show a higher relation with SiO_2 and B_2O_3.

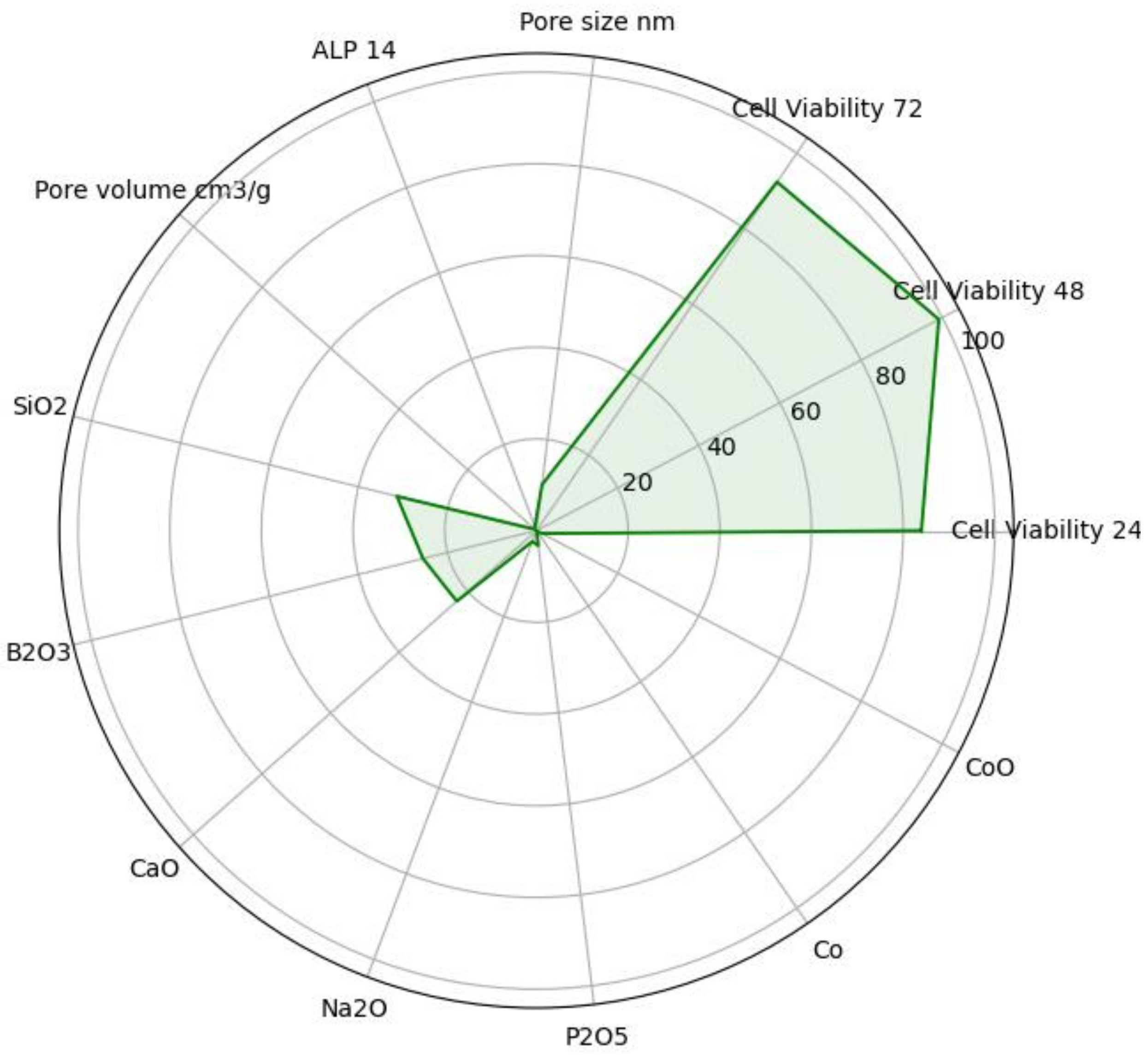


Fig.8 Radar plot (Cobalt)

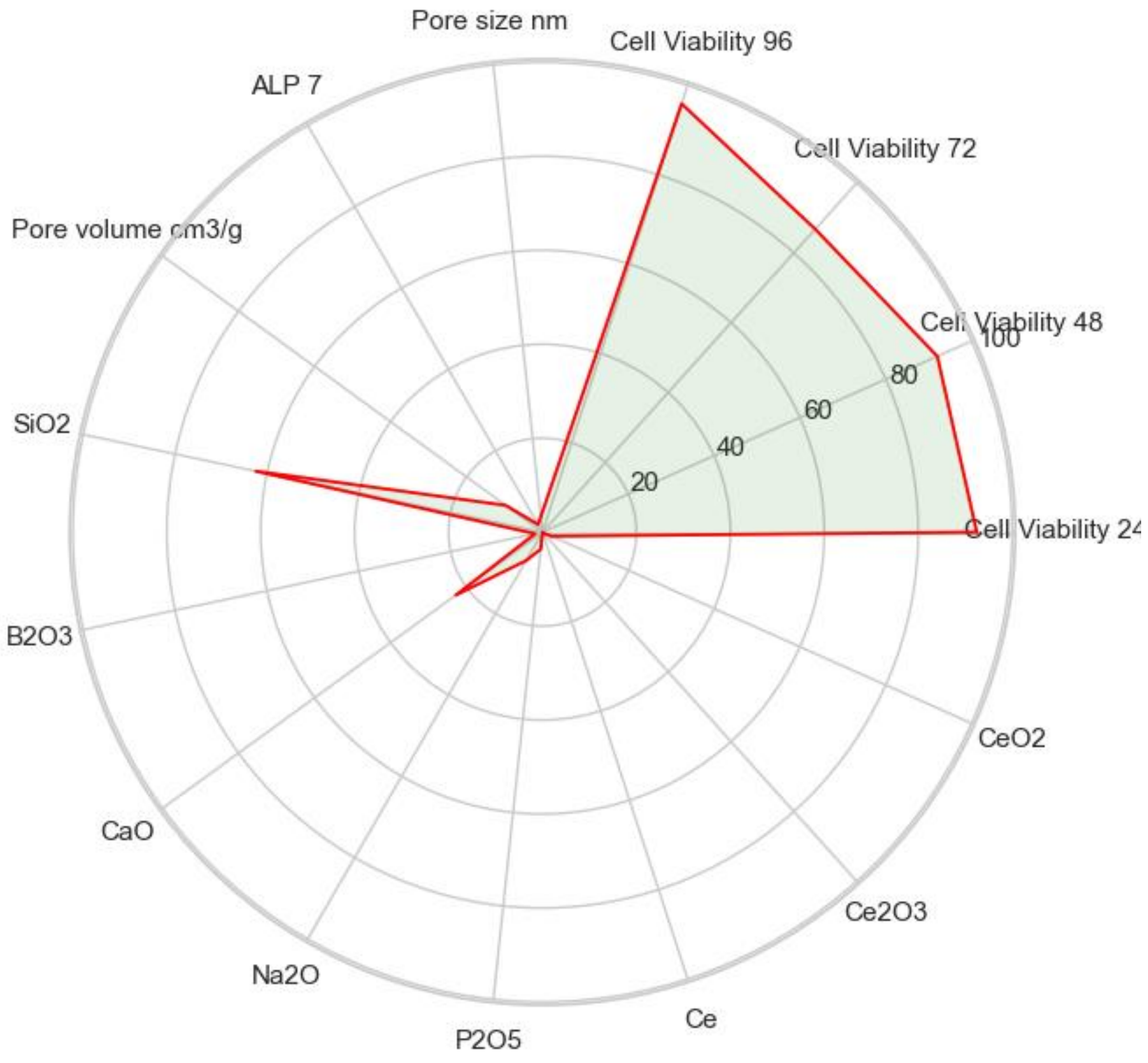


Fig.9 Radar plot (Cerium)

The Radar plots attached show the distribution of Mechanical and Biological properties along with the composition of the BG

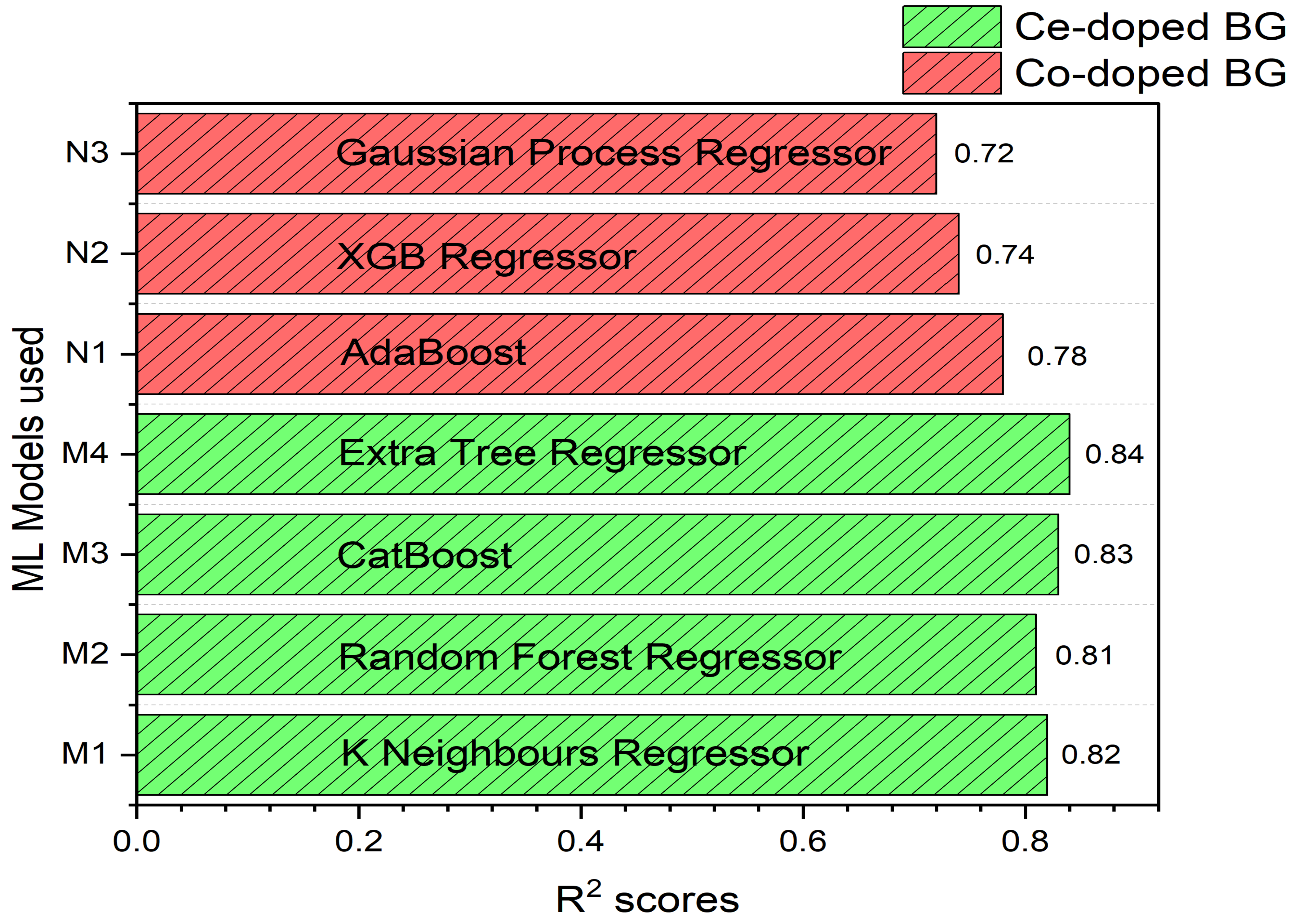
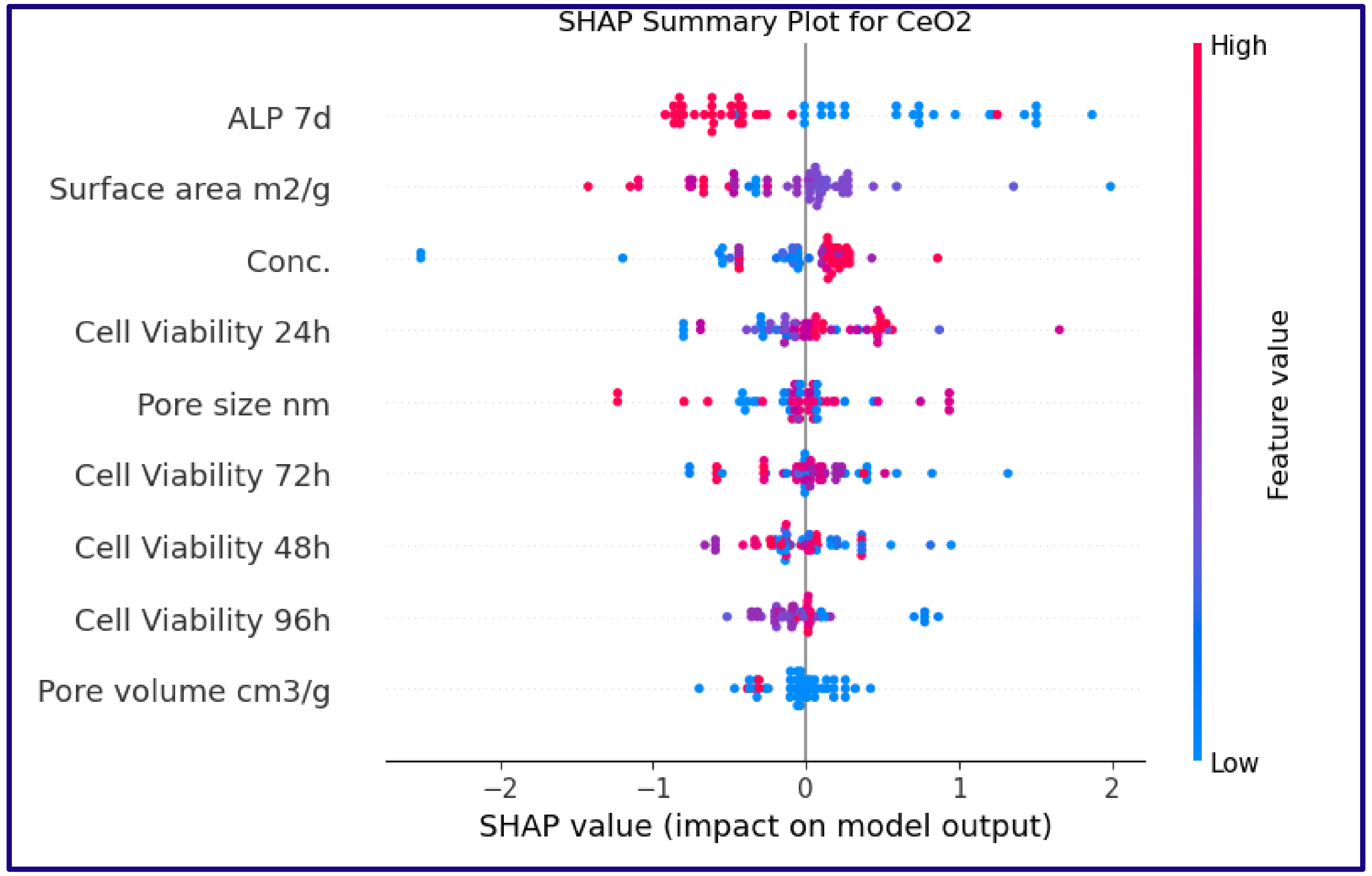
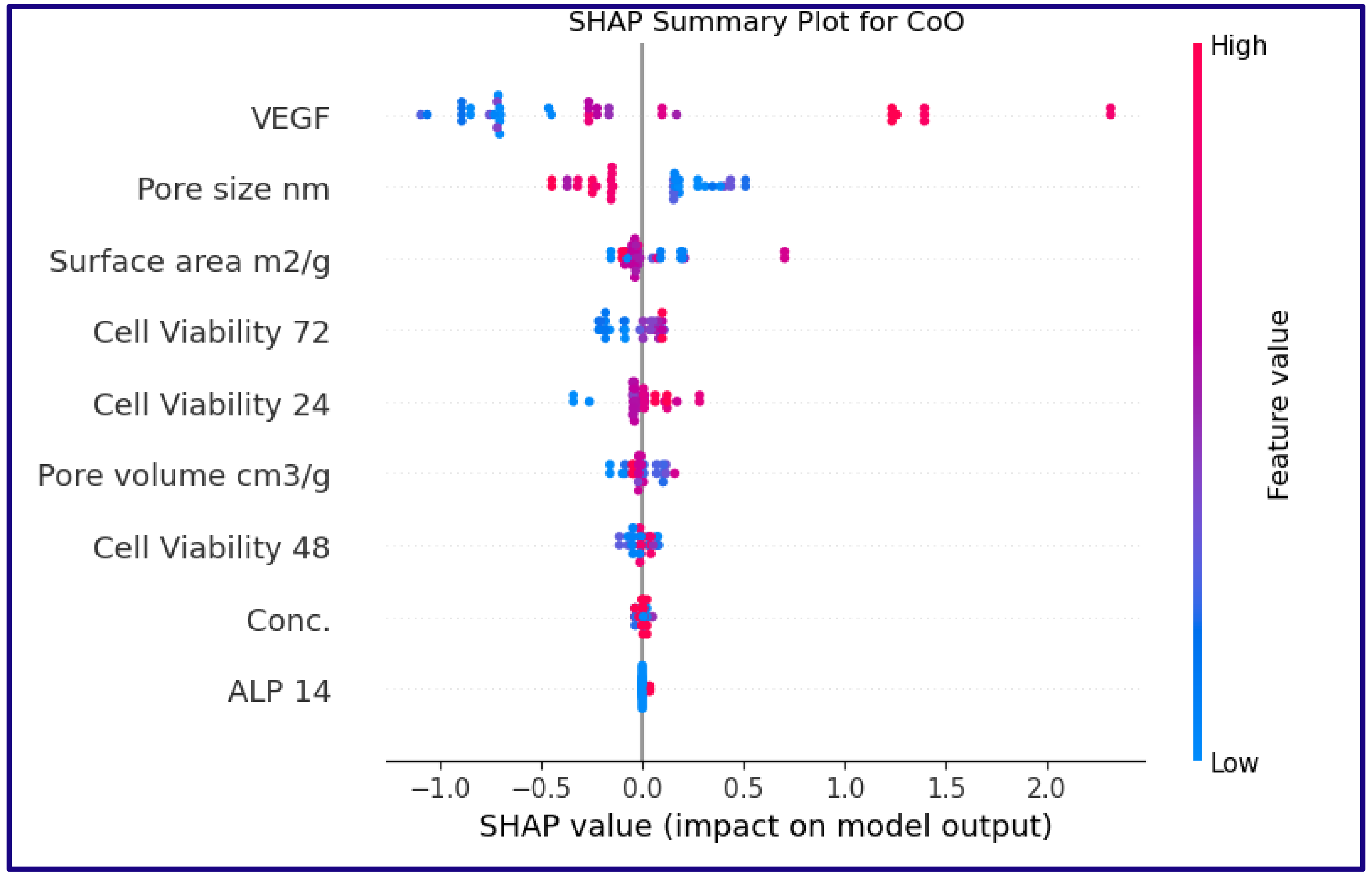


Fig 10. Graphical Representation R^2 Scores for each of the different models trained





SHAP Analysis- Ce-doped BG

SHAP Analysis- Co-doped BG

Element Category	Machine Learning Model	Overall Model R^2 Score	Mean R^2 Score across all folds
Cerium	K-nearest neighbors Regressor	0.8523	0.8245
	Random Forest Regressor	0.8113	0.7439
	Cat Boost	0.8341	0.7987
	Extra Tree regressor	0.8286	0.6506
Cobalt	AdaBoost	0.7882	0.9410
	XGB Regressor	0.7431	0.8611
	Gaussian Process Regressor	0.7298	0.9164

Out of all the multiple models trained for Cerium-doped BG K-Neighbors Regressor and Catboost came out to be the best with a R^2 Score of 0.85 and 0.83 respectively.

Multiple models trained for Cobalt-doped BG showed best results for AdaBoost model with an R^2 score of 0.78.

Table 1. Scores of different ML models trained

CONCLUSION

- The study will aim to determine the composition of the best-suited Cerium or Cobalt-doped Bioglass using the pre-existing data from the research conducted.
- The K-Neighbors Regressor for Cerium and the AdaBoost Regressor for the Cobalt-doped BG with R2 scores of 0.85 and 0.78 respectively.
- The composition of the doped material can be found using the Model with the required parameters (specific use case).
- The developed composition will exhibit enhanced features in mechanical and biological aspects.

THANK YOU