Machine learning phase prediction of high entropy nitrides

- PRIYANSHU BHATIA

Problem Statement:

High-Entropy Nitrides (HENs) are promising materials with diverse applications, but their phase classification and structural stability prediction are challenging due to their complex compositional space. This project aims to develop machine learning models to predict the structural stability and classify phases of HENs, leveraging computational datasets and advanced algorithms to accelerate material discovery and guide experimental research.

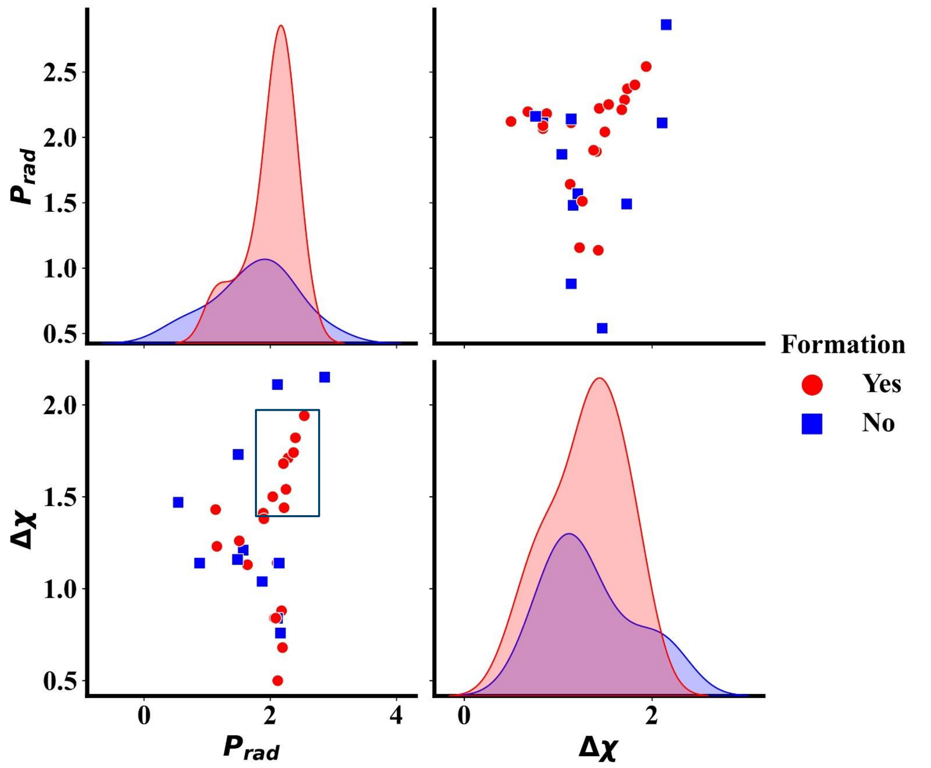
Properties

High-Entropy Nitrides (HENs) exhibit exceptional properties, such as high hardness, thermal stability, and corrosion resistance, due to their unique compositional and structural complexity. Predicting the phases and structural stability of HENs is challenging, as it depends on intricate thermodynamic and physical factors, including configurational entropy, enthalpy of formation, atomic size mismatch, and electronic structure. This project aims to develop machine learning models that leverage these properties to classify phases and predict the structural stability of HENs, enabling efficient exploration of their vast compositional space.

Objective:

1. To develop machine learning models to help in phase classification of high entropy nitrides.
2. To advance the discovery of high entropy materials by developing machine learning predictors for the structural stability of HENs

Sorting criteria selection from Individual nitride formation tendencies of MN-type



Sorting criteria for both quaternary and quinary compositions:

1.34 ≤ Xp ≤ 1.94

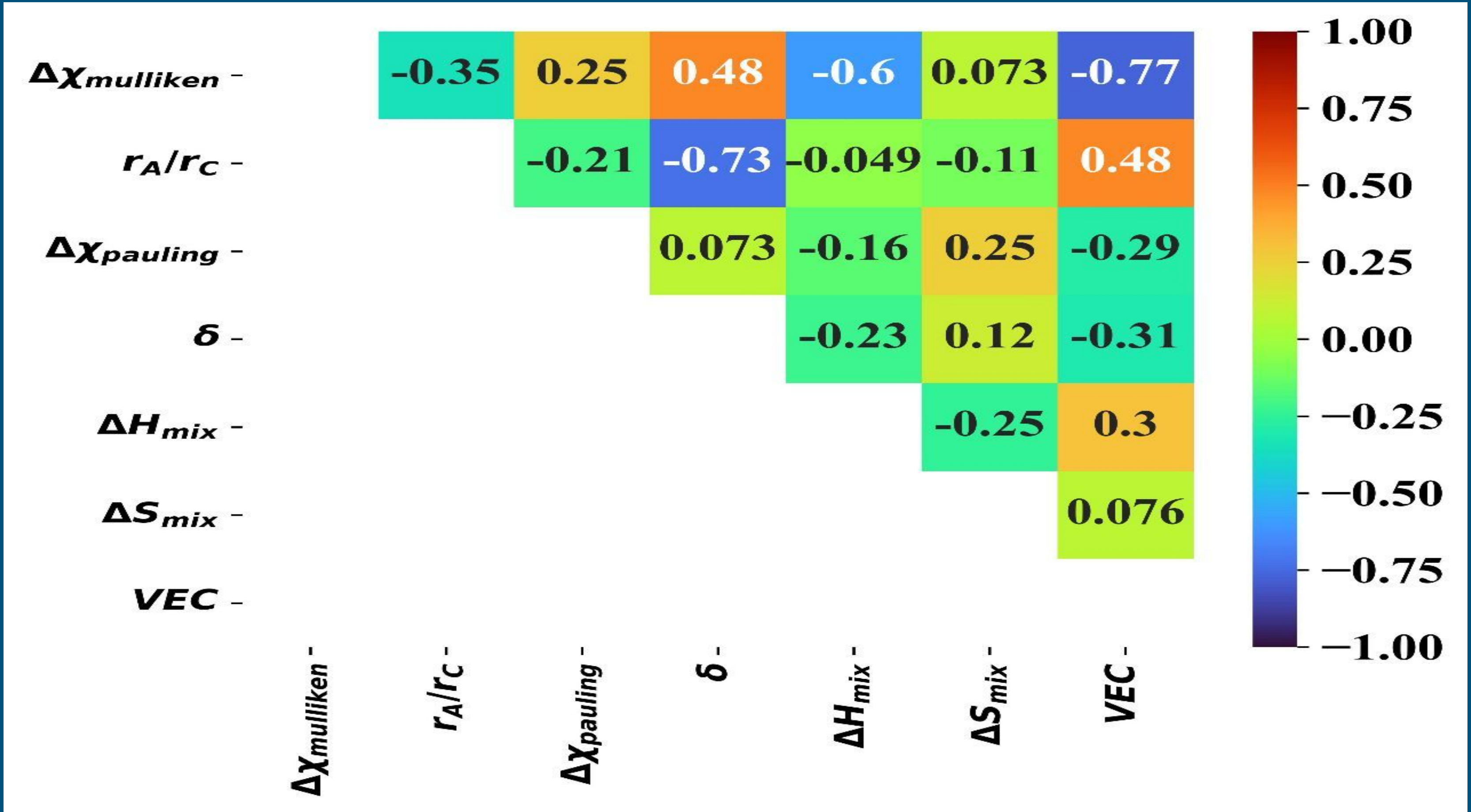
1.89 ≤ Prad ≤ 2.54

-2.72 ≤ ΔHmix ≤ 0.76

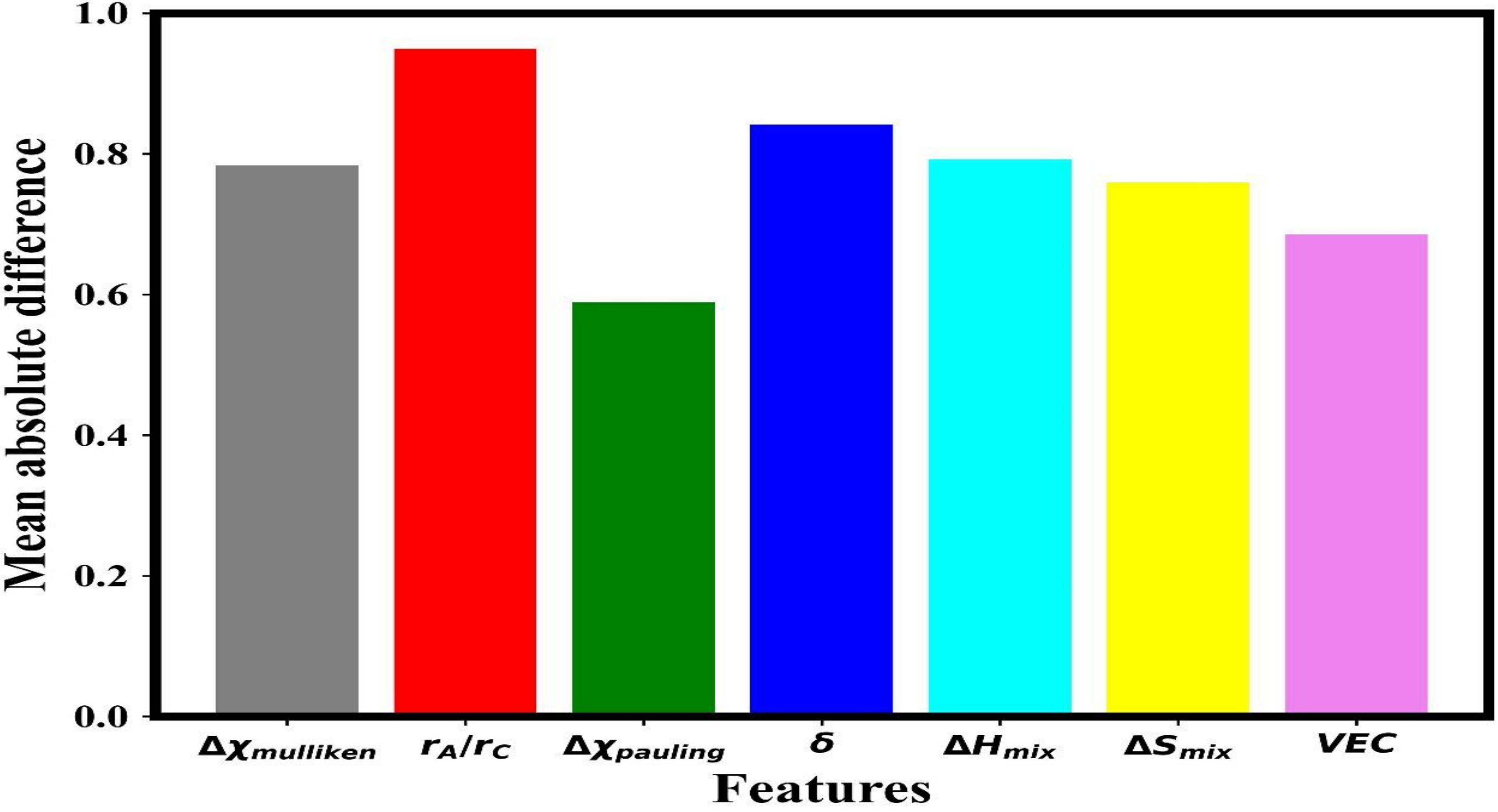
Approach:

1. *Generated semi-synthetic dataset using an atomic environment mapping based structural plot with the help of existing dataset .*
2. *Developed two distinct sorting criteria for quinary and quaternary compositions to produce candidates for single- phase class.*
3. *Implemented ADASYN (***Adaptive Synthetic Sampling Approach for Imbalanced Learning***)technique to generate synthetic dataset from structural modeling and literature dataset to oversample the minority class.*
4. *Implemented 4 ML algorithms KNN,RF,SVM,GNB and trained it on balance and imbalance dataset.*
5. *Designed the feature pool of model and ideated through structural and thermodynamic parameters.*

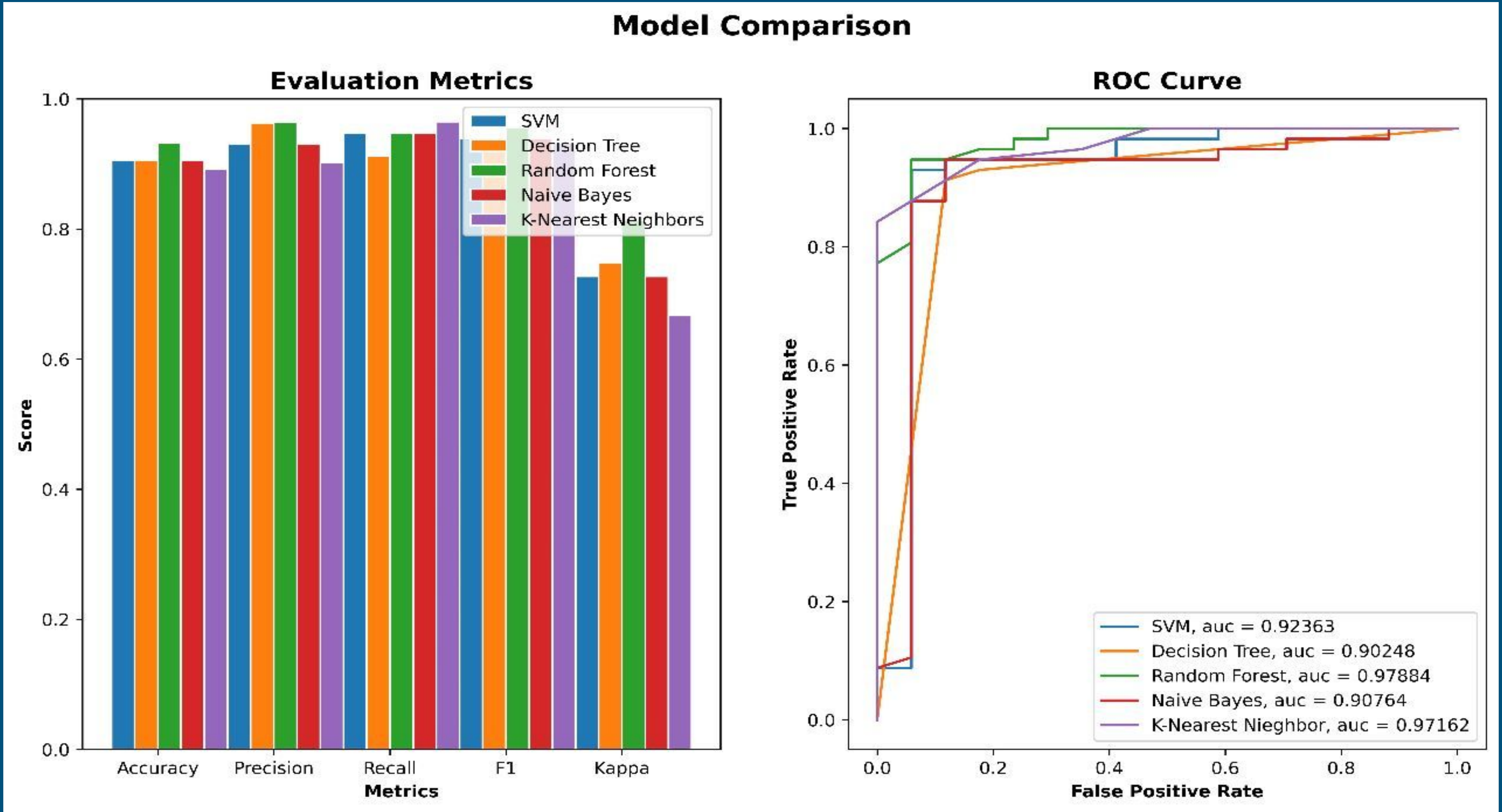
X Mulliken
rA/rC
X Pauling
Delta
H mixing
S mixing
VEC
Formation
Single Phase
Multi Phase



Correlation Matrix

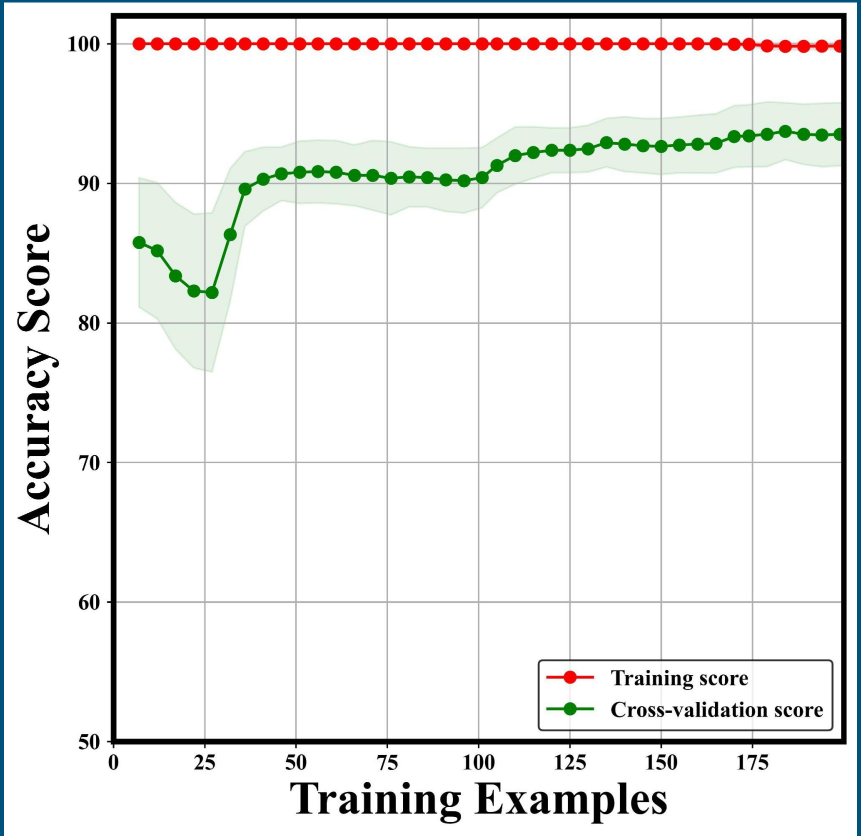


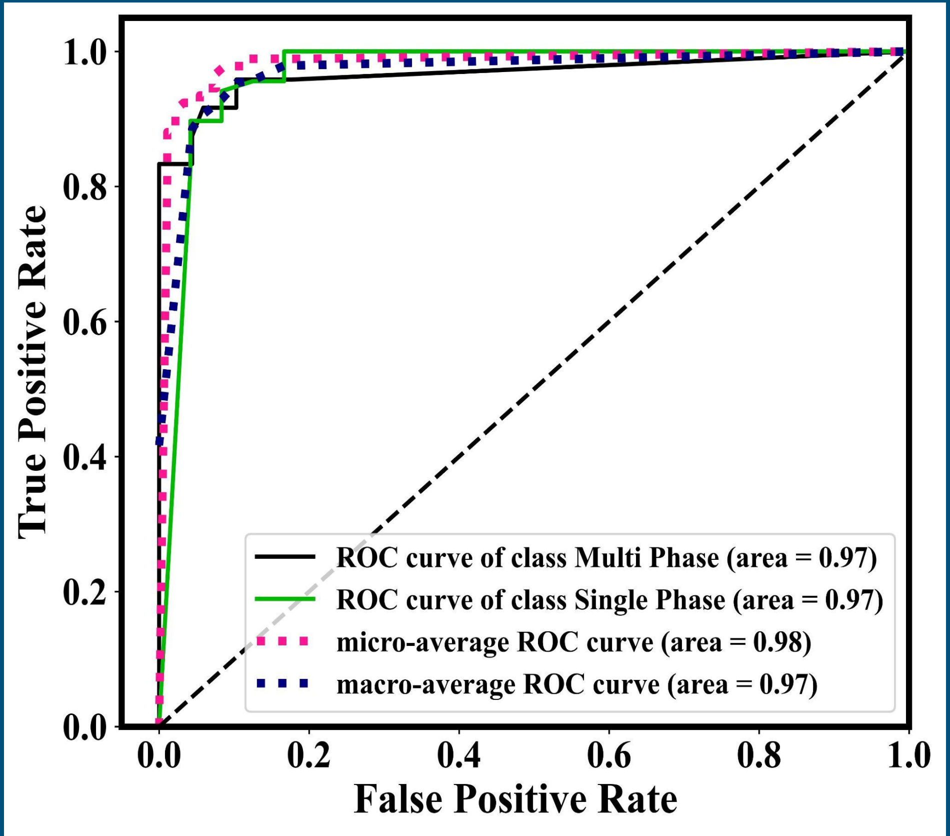
1.0
0.8
0.6
0.4
0.2
0.0
Mean absolute difference
$\Delta\chi_{mulliken}$
r_A/r_C
$\Delta\chi_{pauling}$
δ
ΔH_{mix}
ΔS_{mix}
VEC
Features



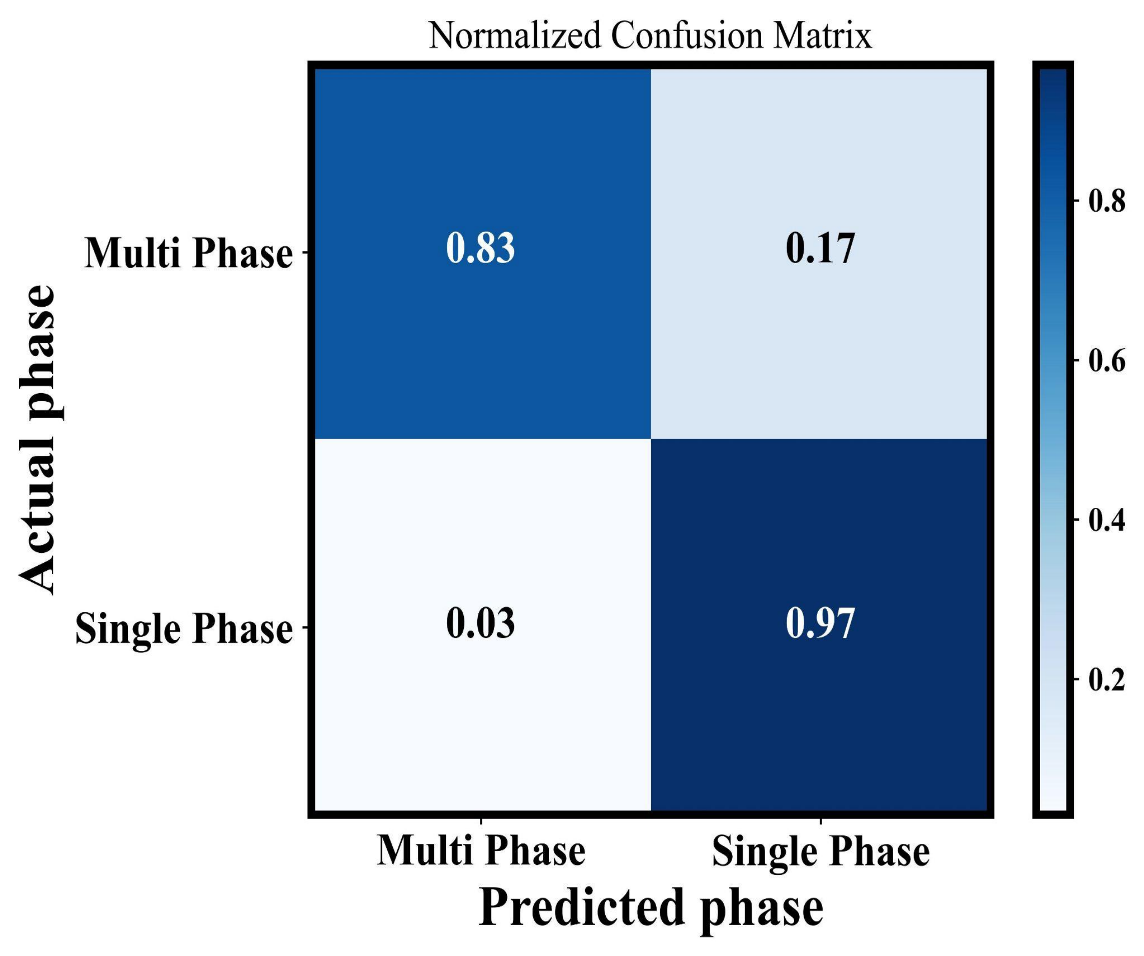
Model Comparison
Evaluation Metrics
ROC Curve
SVM
Decision Tree
Random Forest
Naive Bayes
K-Nearest Neighbors
Score
Metrics
Accuracy
Precision
Recall
F1
Kappa
1.0
0.8
0.6
0.4
0.2
0.0
True Positive Rate
False Positive Rate
SVM, auc = 0.92363
Decision Tree, auc = 0.90248
Random Forest, auc = 0.97884
Naive Bayes, auc = 0.90764
K-Nearest Nieghbor, auc = 0.97162

Learning curve and Accuracy metrices





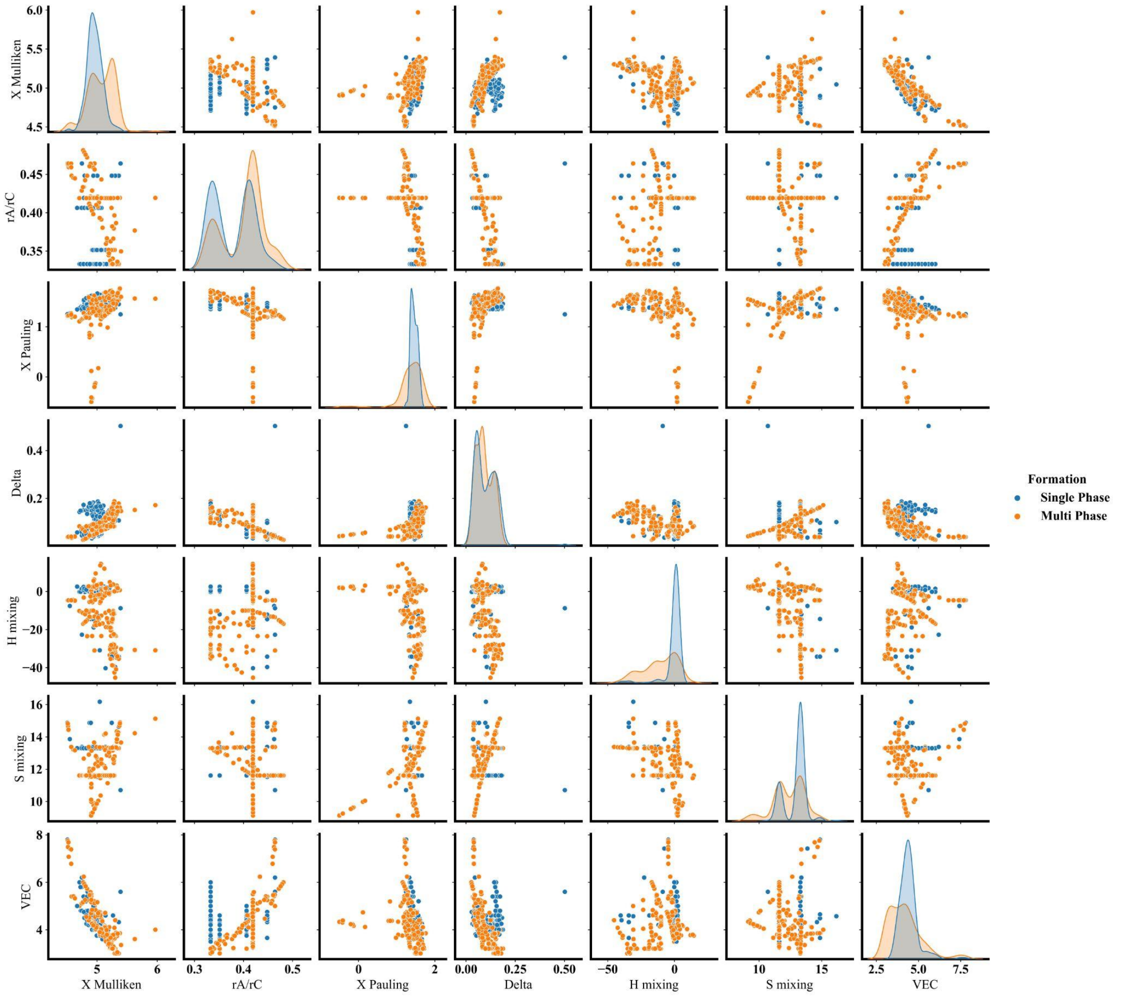
Confusion Matrix



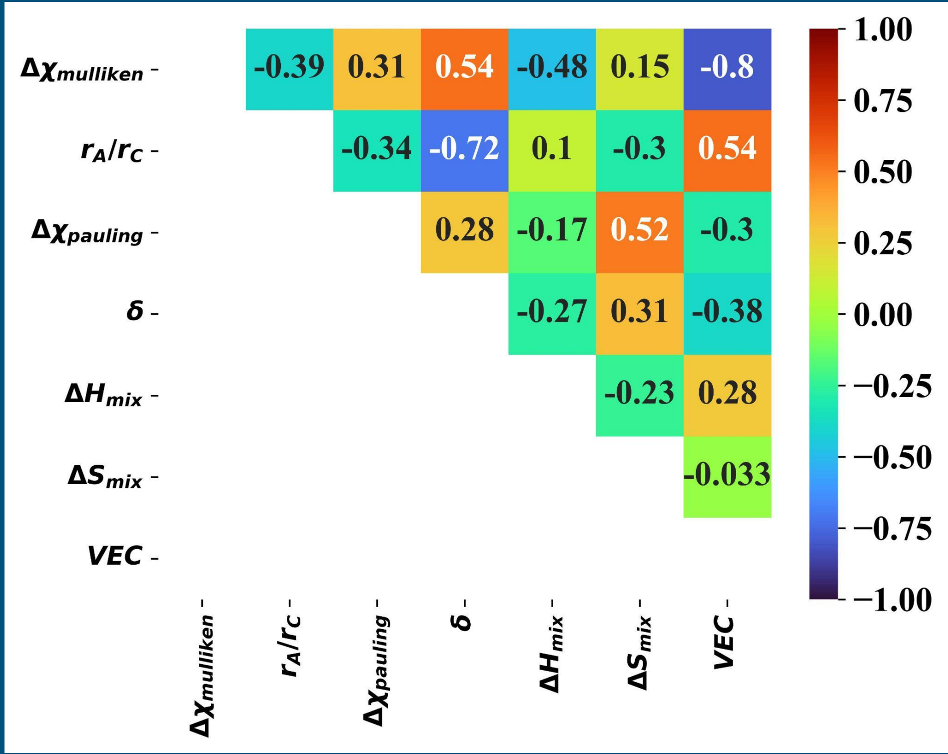
Impact :

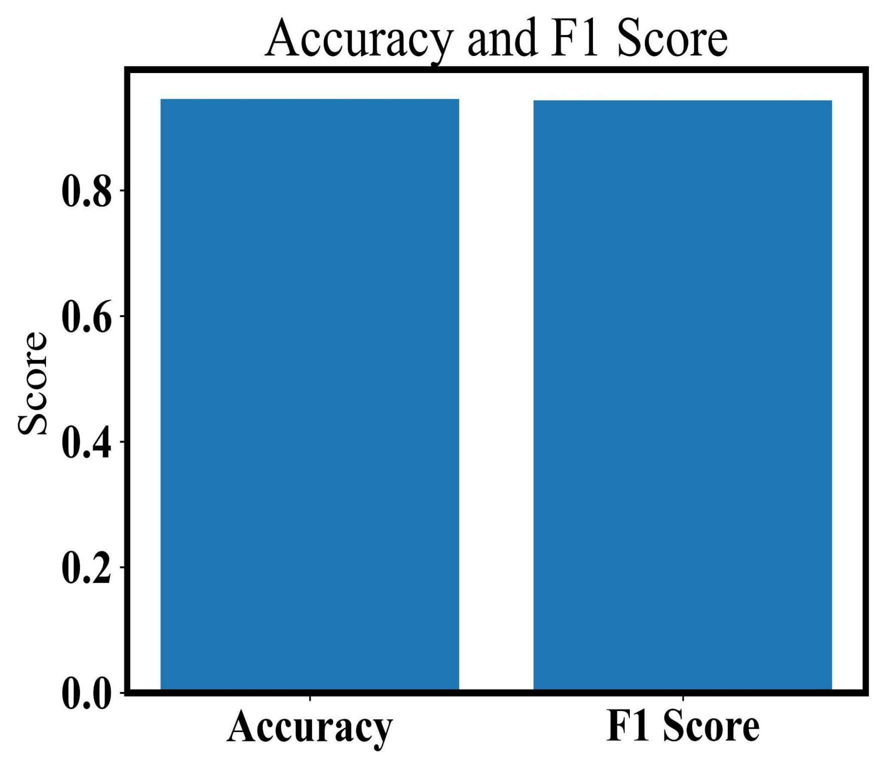
-KNN model achieved an accuracy of 99.63% and 93.4 % on train and test data after Cross validation.

Scatterplot after ADASYN

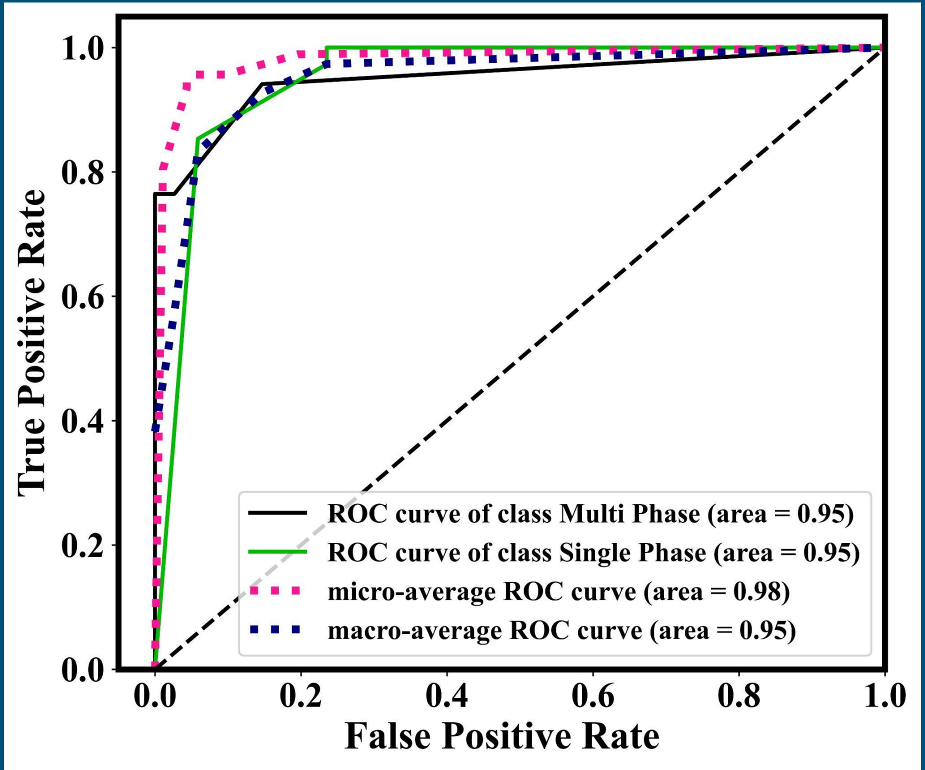


Accuracy and F1 score after ADAYSYN

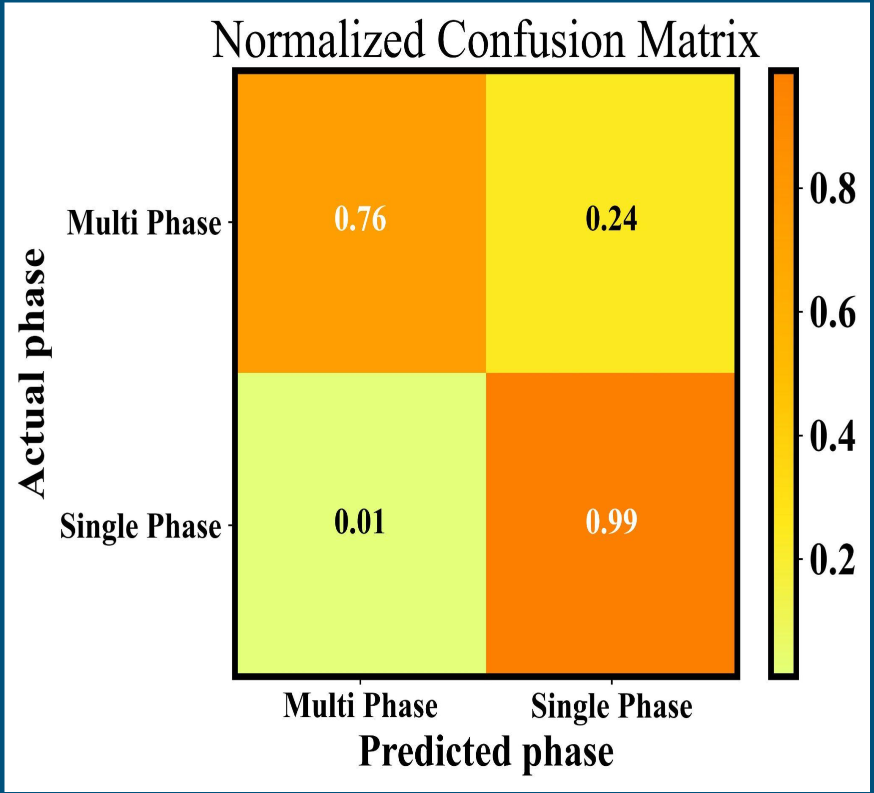




ROC-AUC curve after ADASYN



Confusion Matrix after ADAYSYN



Impact :

-KNN model achieved an accuracy of 96.7% and 94.2 % on train and test data after Cross validation.

THANK YOU!!