**Parametric Optimization of Transition Metal based Nanocomposite Electrocatalysts for Oxygen Evolution Reaction in Alkaline Media**

***Vedasri Bai Khavalaa, Abhijai Velluvaa, Adhithyan Kathiravana, Harish Kuruvaa, Chandan Singh Khavalab, ,B.S. Murtya,c\**,*Tiju Thomasa\****

*a*Department of Metallurgical and Materials Engineering, Indian Institute of Technology Madras, Chennai-600036, Tamil Nadu, India

*b*Department of Metallurgical and Materials Engineering, *Visvesvaraya National Institute of Technology, Nagpur – 440010, Maharastra, India*

*c* Indian Institute of Technology Hyderabad, Kandi -502285, Telangana, India

Email: [bsm@iith.ac.in](mailto:bsm@iith.ac.in) (B.S.Murty), [tijuthomas@iitm.ac.in](mailto:tijuthomas@iitm.ac.in) (Tiju Thomas) and  
[vkhavala97@gmail.com](mailto:vkhavala97@gmail.com) (Vedasri Bai Khavala)

**DATA AND CODE**

To make our work reproducible, we have uploaded the following files in our GitHub repository located at <https://github.com/adhi1910/TMN_OER>,

1. TMON\_dataset.csv – The whole dataset, consisting of 525 datapoints corresponding to 26 features, is contained in this file and was used to train the machine learning algorithms.
2. TMON\_model\_classifier. ipynb – This Jupyter notebook file contains the code for classification of the TMON dataset, containing all the steps of our work from the beginning to the end.
3. TMON\_model\_regressor. ipynb - This Jupyter notebook file contains the code for performing the regression for the TMON dataset.
4. TMO\_data.csv - The whole dataset, consisting of 295 TMO datapoints corresponding to 25 features, is contained in this file and was used to train the machine learning algorithms
5. Literature Sources.xlsx – This file contains sources of data points which include literature details.
6. Dataset\_with\_iR and Stability data – This file contains the original data collected and also shows the iR correction if performed or not for every datapoint.

**S1. LIST OF FEATURES**

The categorical (11) and continuous (15) features in TMON dataset. The TMO dataset contains same features excluding the oxided or nitrided categorical feature i.e., TMO dataset contains 10 categorical and 15 continuous features. The following features being used in the model highlight both materials and device perspective and is indicated in brackets following the feature:

* **Categorical features**

The encoding used for the corresponding feature are indicated using single quotes ‘’:

1. Compound Class (materials perspective):

(ⅰ) cobalt class (‘Co’), (ⅱ) nickel class (‘Ni’), (ⅲ) iron class (‘Fe’), and (ⅳ) other transition metals class (molybdenum, vanadium, copper, manganese, chromium, zinc & titanium) (‘Other TMs’).

1. Composite (Y or N) (materials perspective):

Composite= ‘1’, non-composite = ‘0’

1. Oxided or nitrided (materials perspective):

'1' if nitrided; otherwise, '0'

1. Presence of Carbon (materials perspective):

Yes=’1’, No =’0’

1. Carbon Structures (materials perspective):

(i) no carbon ‘0’, (ⅱ) 1D carbon structures (carbon nanotube, carbon nanowires, carbon fibers, etc.) ‘1’, (ⅲ) 2D carbon structures (graphene, carbon nanosheets, etc.) ‘2’, (ⅳ) 3D carbon structures (carbon cloth, carbon nanocube, crumbled graphene, etc.) ‘3’ and (ⅴ) reduced graphene oxide ‘4’

1. N-doped Carbon Structure (0 or 1) (materials perspective)

‘0’ is absence of nitrogen, ‘1’ presence of nitrogen

1. Doping (other than TMs) (materials perspective):

‘0’ means undoped, ‘1’ means doped

1. Morphology Category (materials perspective):

(ⅰ) bulk ‘1’ (ⅱ) nanorod, nanotubes, nanowires, nanofibers, nanoneedles, etc. ‘2’ (ⅲ) nanoparticles, nanospheres & microspheres ‘3’ (ⅳ) combinatorial structures (such as core-shell, a structure with multiple morphologies) ‘4’ (ⅴ) thin films ‘5’ (ⅵ) 3D nanomaterials with unique morphology (flower-like, urchins, nanocrystals, nano boxes, nanocubes, etc.) ‘6’ (ⅶ) nanosheets, nanomeshes, nanoflakes, nanoplates, etc. ‘7’ (ⅷ) powders ‘8’.

1. Grown on (materials perspective):

This feature indicates the substrate on which the catalyst is synthesized on.

(ⅰ) free-standing/ not grown on any substrate (‘0’), (ⅱ) nickel foam or nickel foil (‘1’), (ⅲ) carbon matrixes (‘2’), and (ⅳ) Other metallic foams, ligaments, nanosheets and foils (‘3’).

1. Porous - 1, Non-Porous – 0 (materials perspective)
2. Substrates (device perspective):

This feature indicates the substrate on which electrochemical measurements are performed for the catalyst.

nickel foam (‘NF’), glassy carbon (‘GC’), carbon-based (other than GC, like carbon cloth (CC), carbon paper (CP), carbon fiber paper (CFP), graphite disk, etc.) (‘carbon\_based’), and TM based (iron foam, copper foam, titanium foil, titanium mesh, etc.) (‘TM\_based’).

* **Continuous features**

1. Presence of atoms (other than TMs) (materials perspective):

filled by the atomic number of respective atoms

1. Cobalt atoms (materials perspective):

Filled with number of atoms

1. Nickel atoms (materials perspective):

Filled with number of atoms

1. Iron atoms (materials perspective):

Filled with number of atoms

1. Molybdenum atoms (materials perspective):

Filled with number of atoms

1. Manganese atoms (materials perspective):

Filled with number of atoms

1. Copper atoms (materials perspective)

Filled with number of atoms

1. Zinc atoms (materials perspective):

Filled with number of atoms

1. Chromium Atoms (materials perspective):

Filled with number of atoms

1. Vanadium Atoms (materials perspective):

Filled with number of atoms

1. Tungsten atoms (materials perspective):

Filled with number of atoms

1. Titanium atoms (materials perspective):

Filled with number of atoms

1. Average d electrons (materials perspective):

Filled by taking average of total number of d-electrons available in the electrocatalyst.

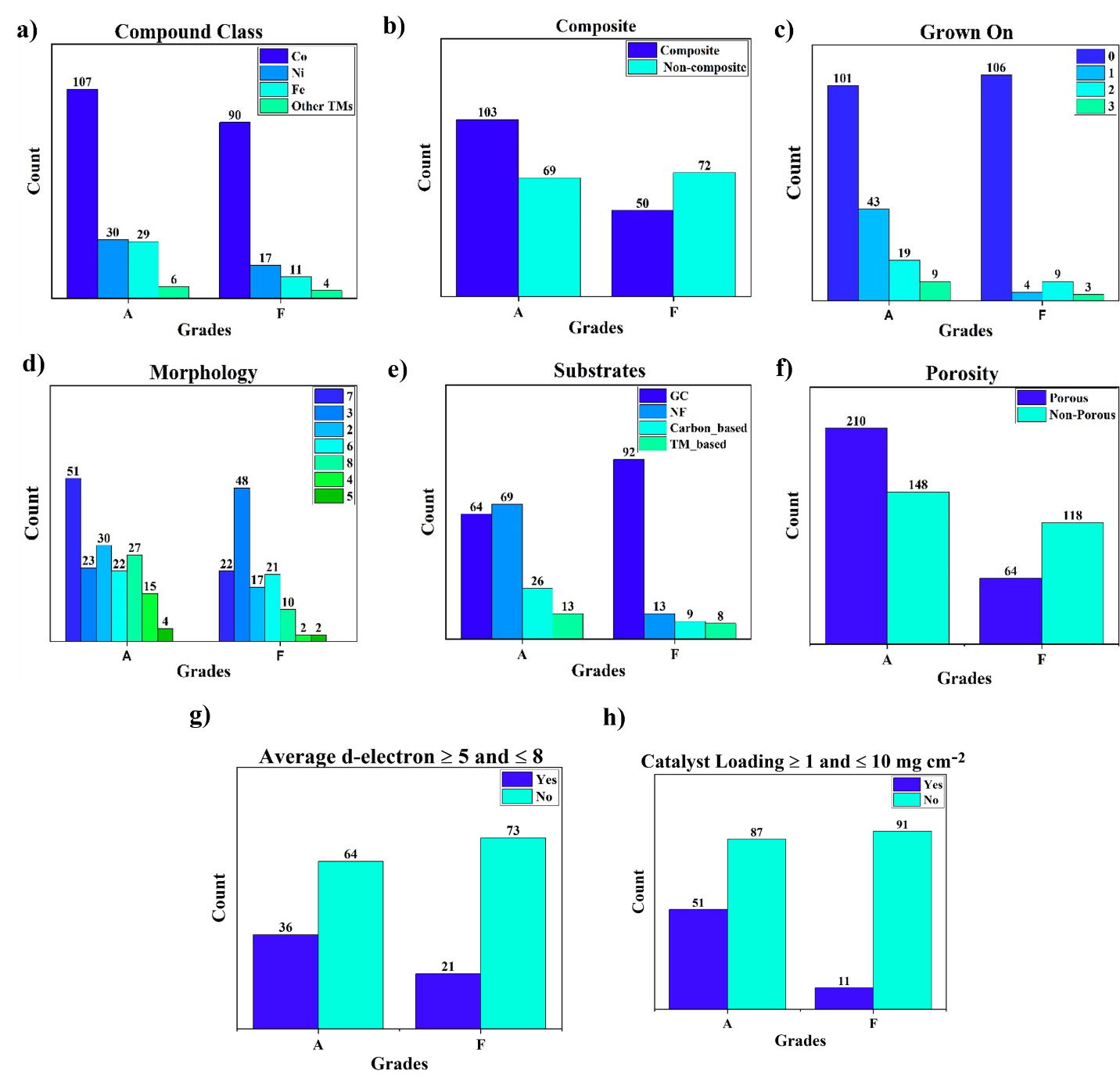
, for non-composite

, for composite

1. pH (device perspective):

pH values of the respective electrolyte used

1. Catalyst Loading (mg cm-2) (device perspective)

****

**Figure S1:** TMO dataset representation (a) Compound Class, (b) Composite, (c) Grown on, (d) Morphology, (e) Substrates, (f) Porosity, (g) Average d electrons ≥ 5 and ≤8, and (h) Catalyst Loading ≥ 1 and ≤ 10 mg cm-2

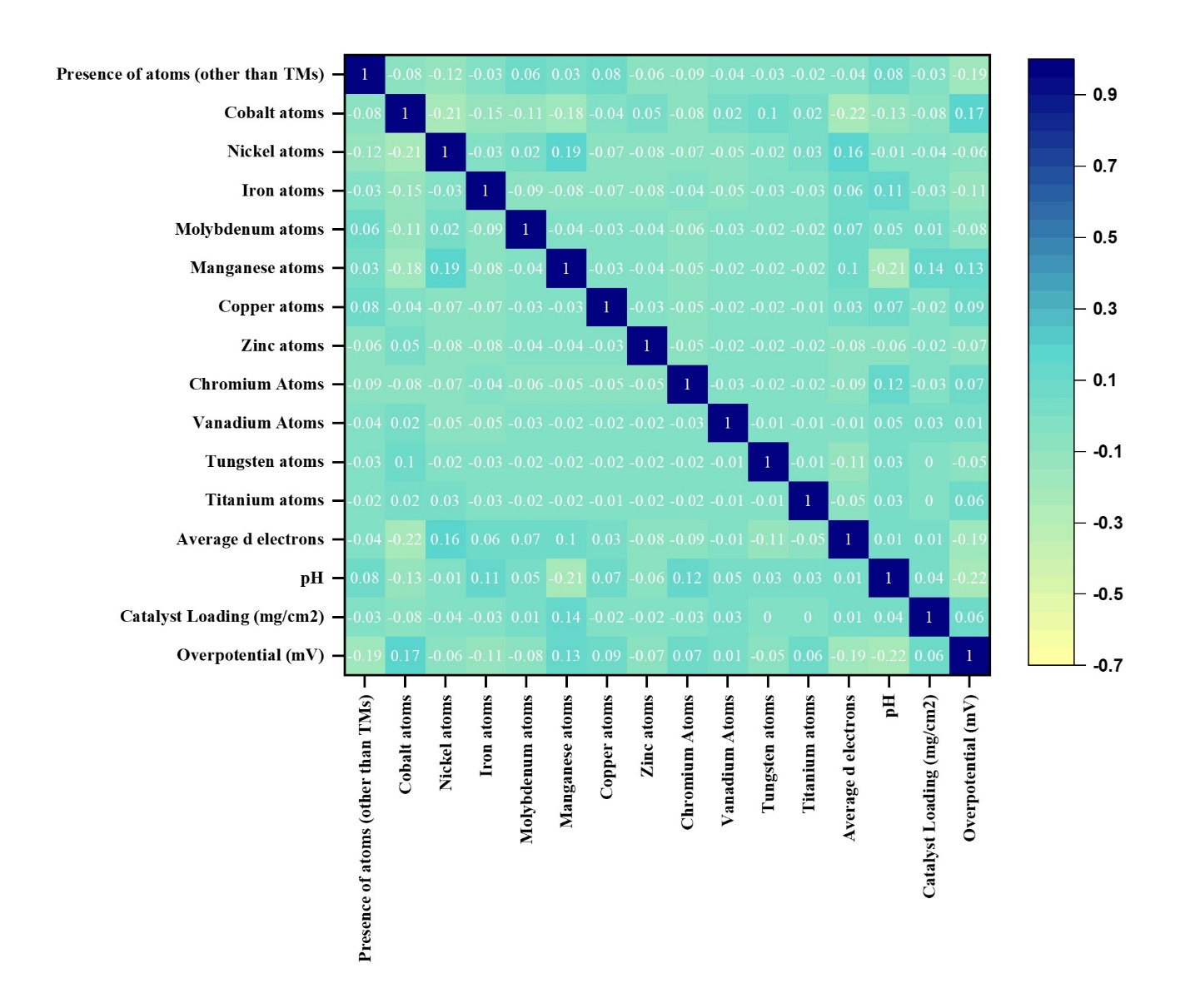


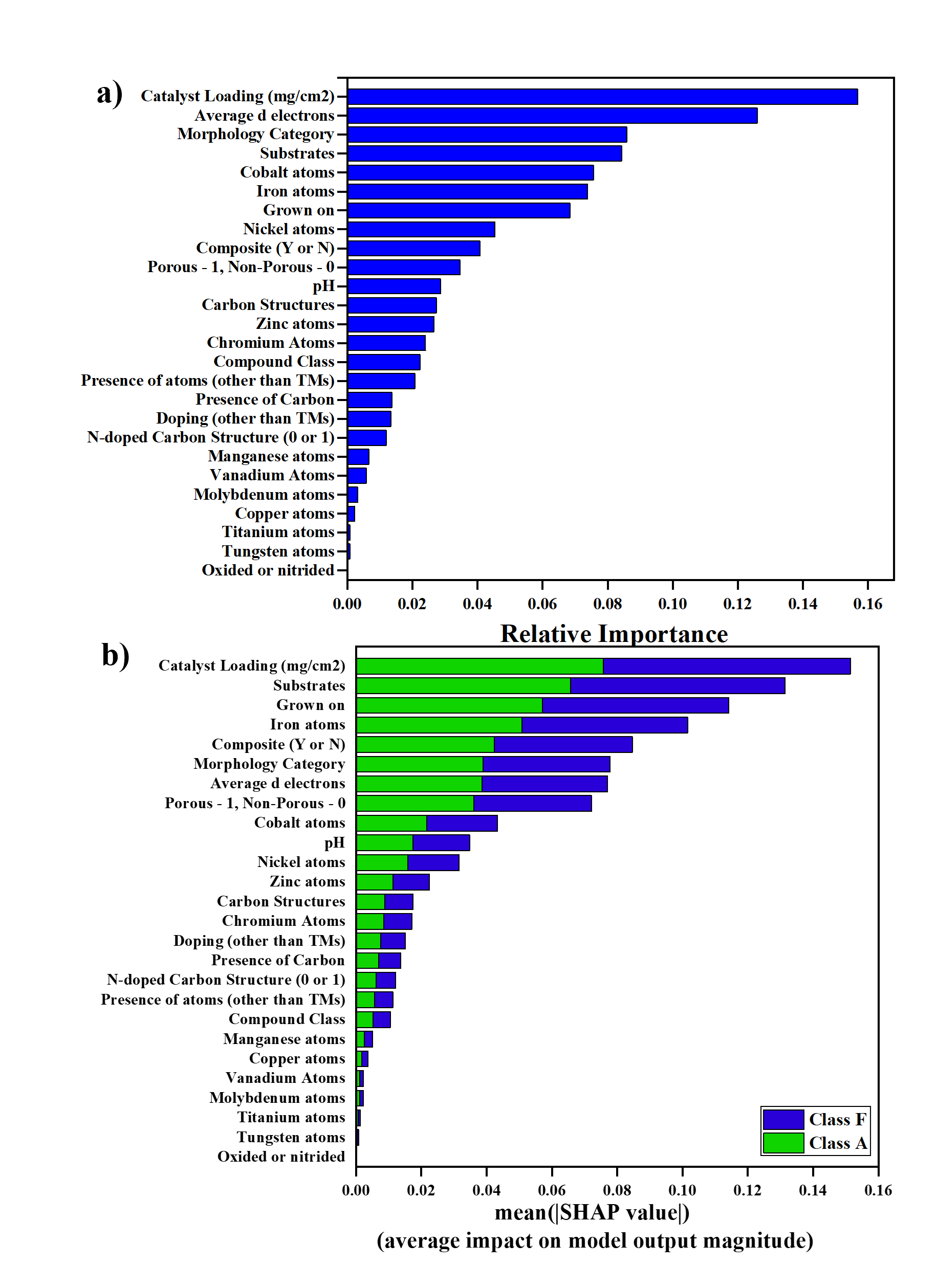
**Figure S2:** Distribution of the target variable η10 in TMON dataset, which has a range of 180 to 460 mV and a mean of 341.85 mV.

**Figure**  **S3:** Distribution of the target variable η10 in TMO dataset. The distribution of is in the range of 180 to 460 mV with a mean of 356.9 mV.

**S2. FEATURE ENGINEERING AND DATA PRE-PROCESSING**

Each compound is introduced to the model by multiplying the number of transition metal atoms present in the compound by their respective atomic masses. The "average d-electrons" feature is calculated for all possible data points, which is an important feature that tunes the catalytic properties. The missing values in "catalyst loading" and "average d-electrons" are filled with the mean of the distribution. The nominal categorical features such as "compound class" and "substrates" are label encoded. Figure S4 for TMOs represent the heat map correlation matrix which gives Pearson correlation (-1: perfect negative correlation, 0: no correlation, and +1: perfect positive correlation) between the various features used in the model and found that no correlation between the continuous variables in the final dataset was higher than 0.75).

**Figure S4:** Heat map showing the Pearson correlation between features. where a value of -1 denotes a perfect negative correlation, a value of 0 denotes no correlation at all, and a value of +1 denotes a perfect positive correlation.



**Figure S5:** (a) Feature Importance Plots of trained features of RF model trained on the collected TMO dataset, with the catalyst loading and average d-electrons as the highest priority. (b) SHAP analysis of trained features of RF model trained on TMO dataset showing the most important features are catalyst loading and substrates.

**Table S1**: Probability of Compound Class, Composite, Morphology types, Grown On, Porous, substrates, Catalyst loading ≥ 1 and ≤ 10 mg cm-2and average d-electrons ≥ 5 and ≤ 8 sub-classes being grade ‘A’ for TMON dataset. The category which is highlighted in green will have more probability of being grade ‘A’, compared to other categories in respective feature.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Categories** | **Probability of Being Grade ‘A’ (in %)** |
| **Compound Class** | Co | 62.67 |
| **Ni** | **73.68** |
| **Fe** | **69.12** |
| Other TMs | 53.33 |
| **Composite** | **Yes** | **74.19** |
| No | 53.92 |
| **Morphology Category** | 1 | 54.54 |
| 2 | 75.47 |
| 3 | 45.97 |
| **4** | **75.75** |
| 5 | 75 |
| 6 | 66.67 |
| **7** | **72.65** |
| **8** | **70.45** |
| **Grown On** | 0 | 56.85 |
| 1 | 91.67 |
| 2 | 76.74 |
| 3 | 77.27 |
| **Porous** | 0 | 54.68 |
| **1** | **76.4** |
| **Substrates** | GC | 53.70 |
| NF | 85.95 |
| Carbon\_based | 75.25 |
| TM\_Based | 61.29 |
| **Catalyst loading ≥ 1 and ≤ 10 mg cm-2** | **Yes** | **85.32** |
| No | 57.60 |
| **Average d-electrons ≥ 5 and ≤ 8** | **Yes** | **74.81** |
| No | 56.22 |
| **Oxided/Nitrided** | **Nitrided** | **75.11** |
| Oxided | 58.5 |

**Table S2**: Probability of Compound Class, Composite, Morphology types, Grown On, Porous, substrates, Catalyst loading ≥ 1 and ≤ 10 mg cm-2and average d-electrons ≥ 5 and ≤ 8 sub-classes being grade ‘A’ for TMO dataset. The category which is highlighted in green will have more probability of being grade ‘A’, compared to other categories in respective feature.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Categories** | **Probability of**  **Being Grade ‘A’ (in %)** |
| **Compound Class** | Co | 54.31 |
| **Ni** | **63.83** |
| **Fe** | **72.5** |
| Other TMs | 60 |
| **Composite** | **Yes** | **67.32** |
| No | 48.93 |
| **Morphology Category** | 2 | 63.82 |
| 3 | 32.39 |
| **4** | **88.24** |
| 5 | 66.67 |
| 6 | 51.16 |
| **7** | **69.86** |
| **8** | **72.97** |
| **Grown On** | 0 | 48.79 |
| 1 | 91.48 |
| 2 | 67.85 |
| 3 | 75 |
| **Porous** | 0 | 49.38 |
| **1** | **69.4** |
| **Substrates** | GC | 41.02 |
| NF | 84.15 |
| Carbon\_based | 74.28 |
| TM\_Based | 61.9 |
| **Catalyst loading ≥ 1 and ≤ 10 mg cm-2** | **Yes** | **85.71** |
| No | 48.91 |
| **Average d-electrons ≥ 5 and ≤ 8** | **Yes** | **63.15** |
| No | 46.72 |

**S3. RF regressor with TMON dataset and TMO dataset**

**Table S3**: Values of optimized parameters for RF Regression on TMON dataset

|  |  |
| --- | --- |
| **Parameter** | **Value** |
| 'bootstrap' | ‘True’ |
| 'ccp\_alpha' | 0.0 |
| 'criterion' | 'squared\_error' |
| 'max\_depth' | ‘None’ |
| 'max\_features' | 'sqrt' |
| 'max\_leaf\_nodes' | ‘None’ |
| 'max\_samples' | ‘None’ |
| 'min\_impurity\_decrease' | 0.0 |
| 'min\_samples\_leaf' | 1 |
| 'min\_samples\_split' | 2 |
| 'min\_weight\_fraction\_leaf' | 0.0 |
| 'n\_estimators' | 100 |
| 'n\_jobs' | ‘None’ |
| 'oob\_score' | ‘False’ |
| 'random\_state' | 0 |
| 'verbose' | 0 |
| 'warm\_start' | ‘False’ |

**Errors**:

**Table S4**: Values of the error metrics for RF Regression on TMON dataset

|  |  |
| --- | --- |
| **Errors** | **Error Value (mV)** |
| MSE | 2043.293 |
| RMSE | 45.203 |
| MAE | 36.468 |
| MAPE | 0.115 |

**Table S5**: Values of optimized parameters for RF Regression on TMO dataset

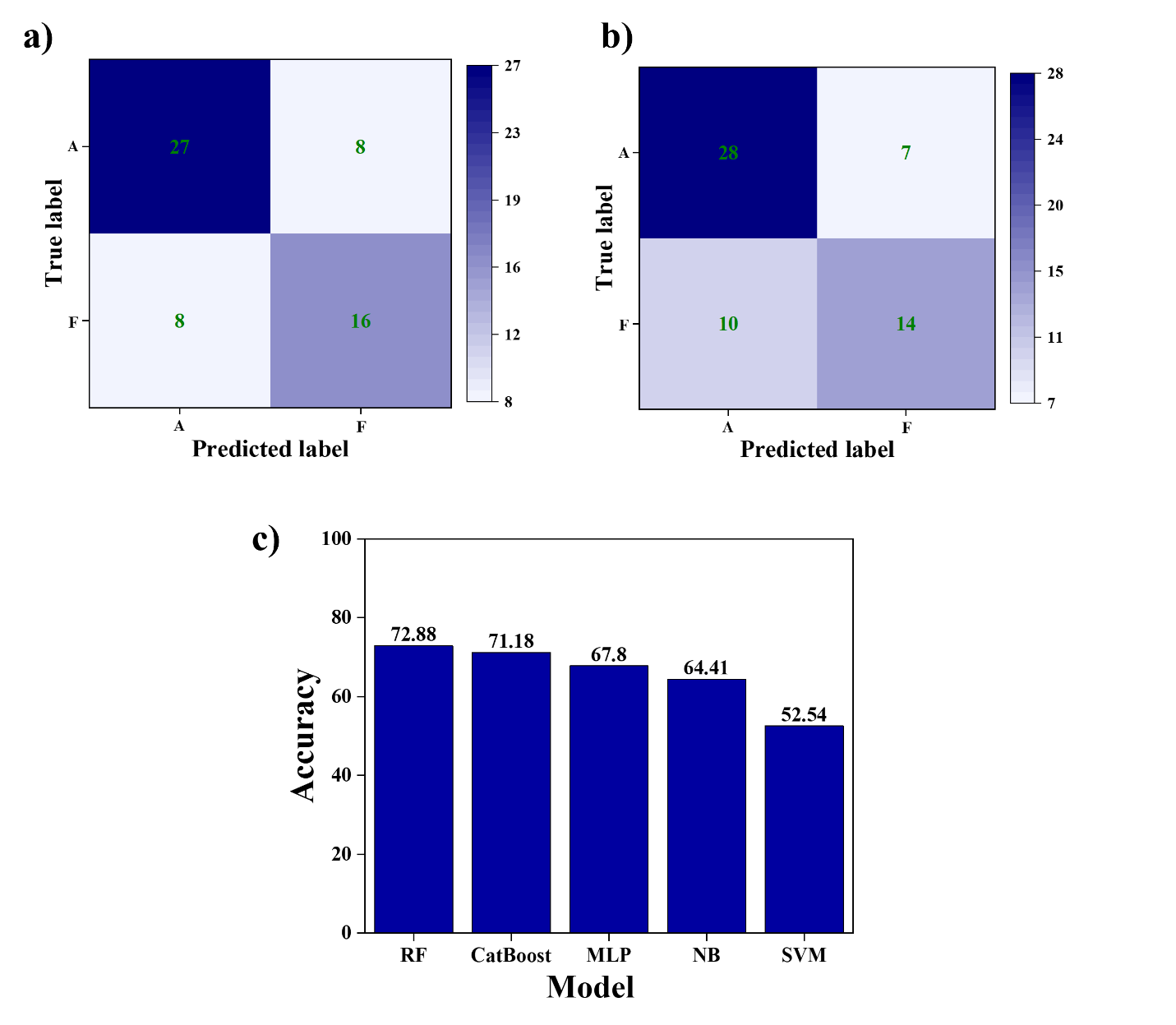
|  |  |
| --- | --- |
| **Parameter** | **Value** |
| 'bootstrap' | ‘True’ |
| 'ccp\_alpha' | 0.0 |
| 'criterion' | 'squared\_error' |
| 'max\_depth' | ‘None’ |
| 'max\_features' | 'sqrt' |
| 'max\_leaf\_nodes' | ‘None’ |
| 'max\_samples' | ‘None’ |
| 'min\_impurity\_decrease' | 0.0 |
| 'min\_samples\_leaf' | 1 |
| 'min\_samples\_split' | 2 |
| 'min\_weight\_fraction\_leaf' | 0.0 |
| 'n\_estimators' | 100 |
| 'n\_jobs' | ‘None’ |
| 'oob\_score' | ‘False’ |
| 'random\_state' | 0 |
| 'verbose' | 0 |
| 'warm\_start' | ‘False’ |

**Errors**:

**Table S6**: Values of the error metrics for RF Regression on TMO dataset

|  |  |
| --- | --- |
| **Errors** | **Error Value (mV)** |
| MSE | 1803.069 |
| RMSE | 42.463 |
| MAE | 32.691 |
| MAPE | 0.095 |

**S4. Classification models for TMO dataset**



**Figure S6.** Confusion matrix of TMOs for (a) RF model (b) CatBoost model (c) Accuracy comparison of various ML models for TMOs. Diagonal dominance can be observed in both the confusion matrix when TMOs are classified using random forest and TMOs classified using Catboost model. Random Forest is the most accurate ML Model for TMOs compared to all tested models. Using TMO dataset the most accurate RF model shows F1 scores of 0.77 and 0.67 for grades 'A' and' F,' respectively

**Table S7:** Comparison of different ML models during the testing phase of overpotential grade prediction for the TMO dataset.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **RF** | **CatBoost** | **SVM** | **MLP** | **NB** |
| **Accuracy (%)** | 72.88 | 71.18 | 52.54 | 67.80 | 64.41 |
| **F1 score [A, F]** | [0.77, 0.67] | [0.77, 0.62] | [0.63, 0.33] | [0.76, 0.51] | [0.72, 0.51] |
| **Precision [A, F]** | [0.77, 0.67] | [0.74, 0.67] | [0.59, 0.39] | [0.68, 0.67] | [0.68, 0.58] |
| **Recall [A, F]** | [0.77, 0.67] | [0.80, 0.58] | [0.69, 0.29] | [0.86, 0.42] | [0.77, 0.46] |