**Data-Driven Prediction of Transition Metal (Oxy)Nitrides as Electrocatalysts for Enhanced Oxygen Evolution Reaction**

***Vedasri Bai Khavalaa, Abhijai Velluvaa, Adhithyan Kathiravana, Chandan Singh Khavalab, Harish Kuruvaa, 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**

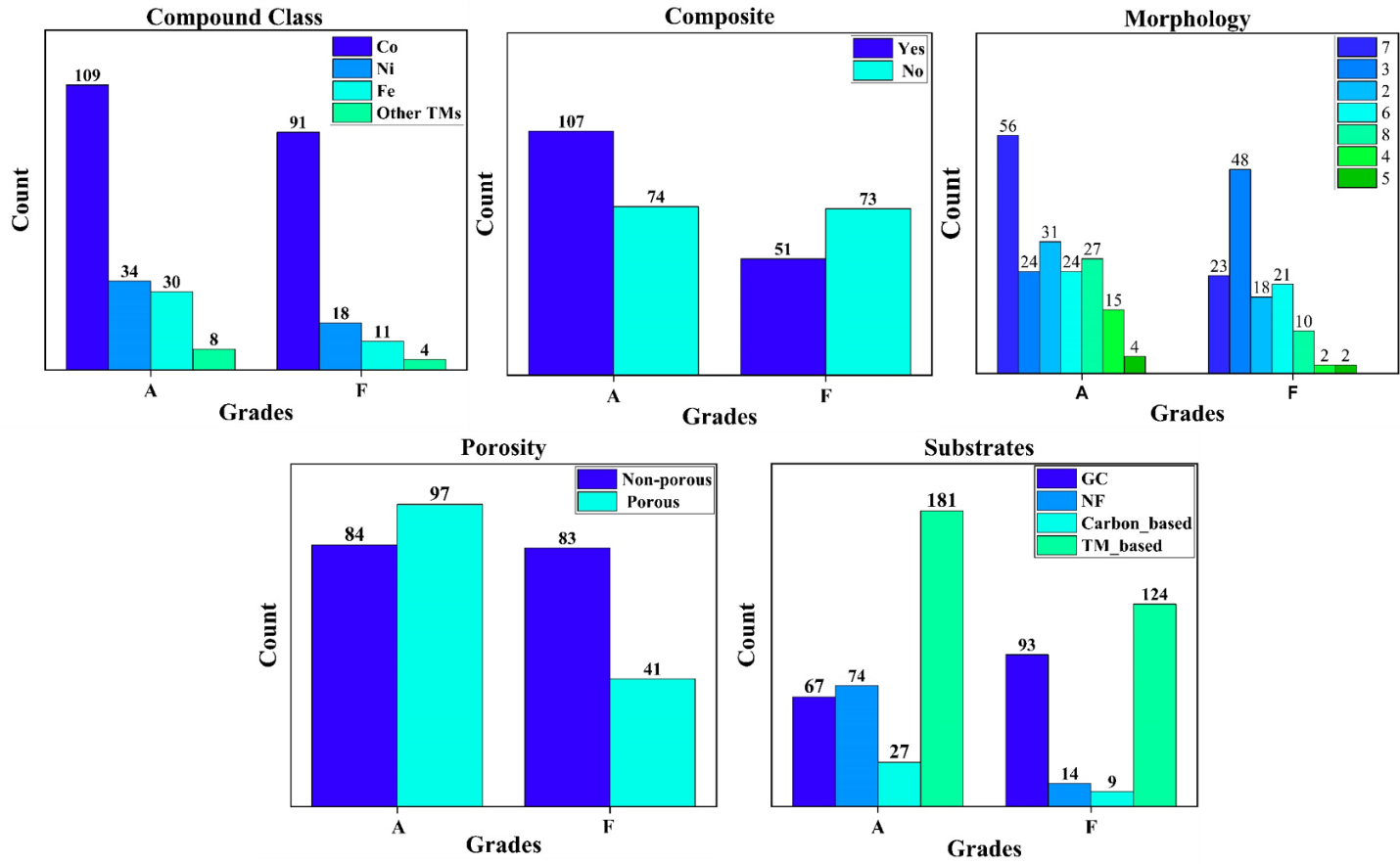
To make our work reproducible, we have uploaded the following files in our GitHub repository located at <https://github.com/adhi1910/TMN_OER>,

1. TM(O)N\_dataset.csv – The whole dataset, consisting of 540 datapoints corresponding to 26 features, is contained in this file and was used to train the machine learning algorithms.
2. TM(O)N\_model\_classifier.ipynb – This Jupyter notebook file contains the code for classification of the TMON dataset, contaning all the steps of our work from the beginning to the end.
3. TM(O)N\_model\_regressor.ipynb - This Jupyter notebook file contains the code for performing the regression for the TM(O)N dataset.
4. TMO\_data.csv - The whole dataset, consisting of 305 TMO datapoints corresponding to 25 features, is contained in this file and was used to train the mac hine learning algorithms

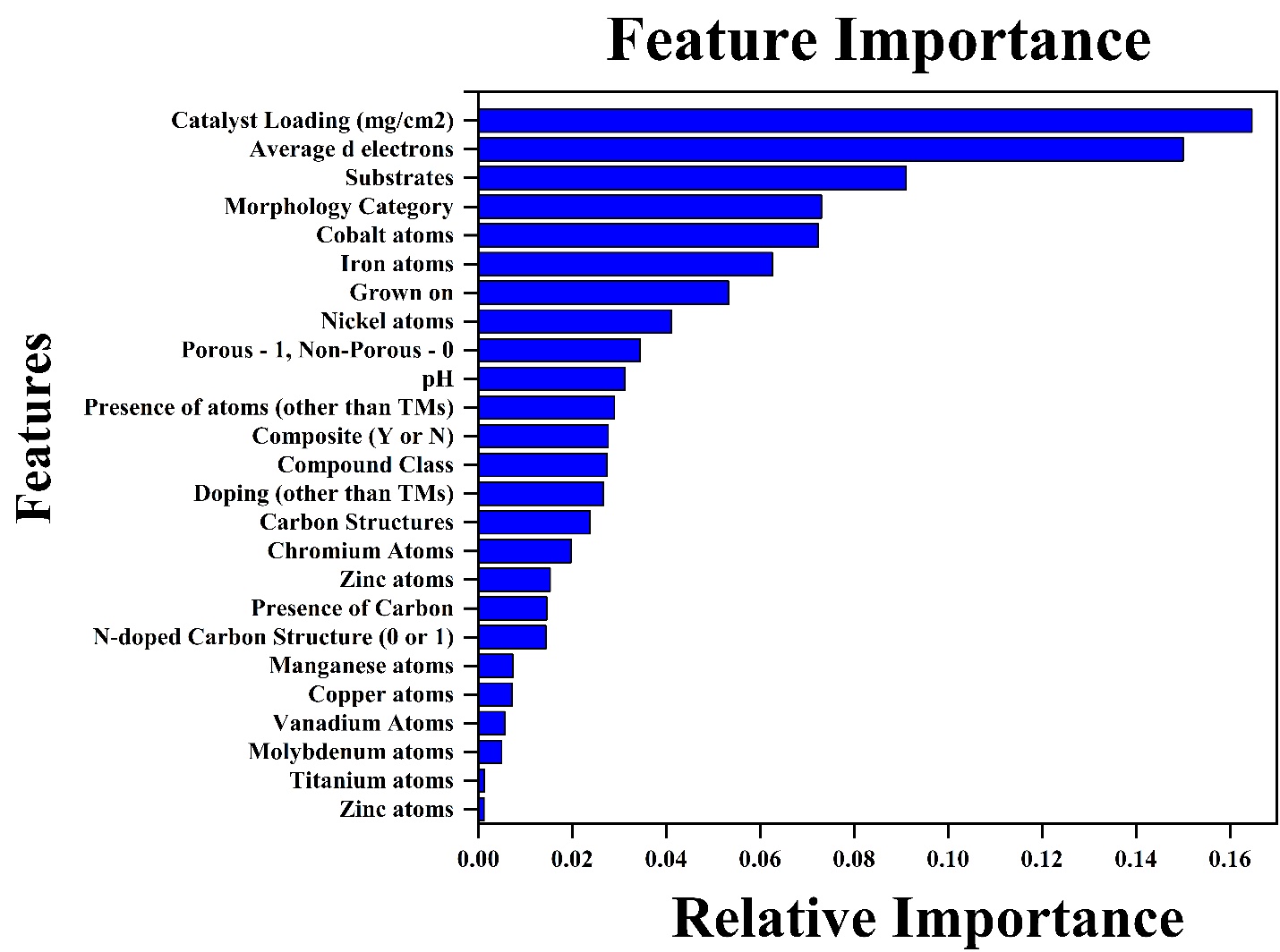
**LIST OF FEATURES**

1. Compound Class
2. Composite (Y or N)
3. Oxided or nitrided
4. Presence of Carbon
5. Carbon Structures
6. N-doped Carbon Structure (0 or 1)
7. Doping (other than TMs)
8. Morphology Category
9. Grown on
10. Porous - 1, Non-Porous - 0
11. Substrates
12. Presence of atoms (other than TMs)
13. Cobalt atoms
14. Nickel atoms
15. Iron atoms
16. Molybdenum atoms
17. Manganese atoms
18. Copper atoms
19. Zinc atoms
20. Chromium Atoms
21. Vanadium Atoms
22. Tungsten atoms
23. Titanium atoms
24. Average d electrons
25. pH
26. Catalyst Loading (mg/cm2)

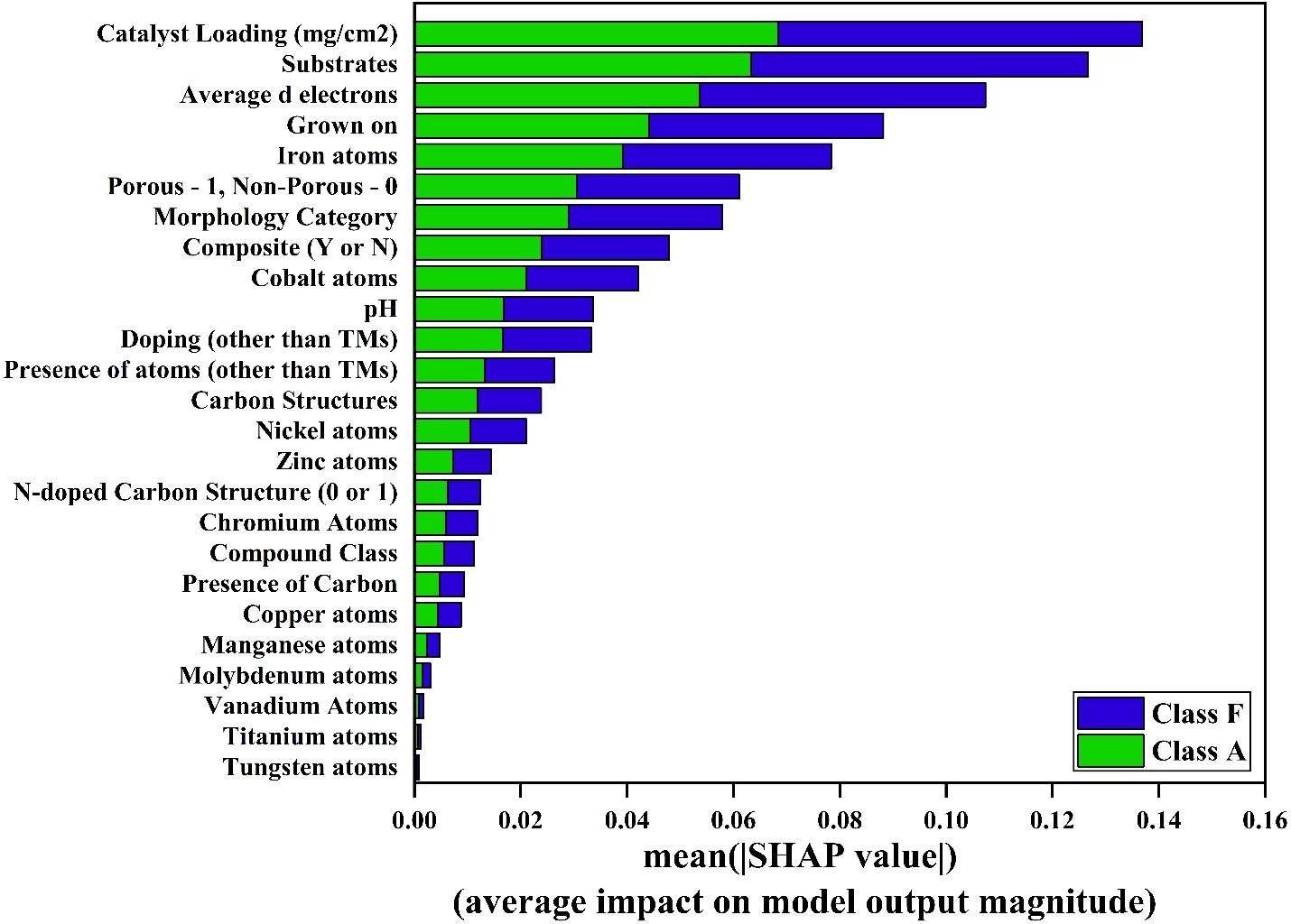
**TMO Dataset Results**

****

**Figure S1:** TMO dataset representation (a) Compound class, (b) Composite, (c) Morphology, (d) Substrates and (e) Porosity with respect to grade A and F. Grade A requires composite electrocatalysts with a morphology of nanosheets, high porosity and TM based substrates.



**Figure S2:** 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.



**Figure S3.** SHAP analysis of trained features of RF model trained on TMO dataset showing the most important features are catalyst loading, substrates, average d electrons and grown on.