Optimization-free Inverse Design of High-Dimensional Nanoparticle Electrocatalysts using Multi-Target Machine Learning

Sichao Li[1], Jonathan Y. C. Ting[1], and Amanda S. Barnard[1][0000−0002−4784−2382]

Australian National University, Acton ACT 2601, Australia
amanda.s.barnard@anu.edu.au

Abstract. Inverse design that directly predicts multiple structural characteristics of nanomaterials based on a set of desirable properties is essential for translating computational predictions into laboratory experiments, and eventually into products. This is challenging due to the high-dimensionality of nanomaterials data which causes an imbalance in the mapping problem, where too few properties are available to predict too many features. In this paper we use multi-target machine learning to directly map the structural features and property labels, without the need for exhaustive data sets or external optimization, and explore the impact of more aggressive feature selection to manage the mapping function. We find that systematically reducing the dimensionality of the feature set improves the accuracy and generalizability of inverse models when interpretable importance profiles from the corresponding forward predictions are used to prioritise inclusion. This allows for a balance between accuracy and efficiency to be established on a case-by-case basis, but raises new questions about the role of domain knowledge and pragmatic preferences in feature prioritization strategies.

Keywords: inverse design · machine learning · catalysis

1 Introduction

Chemical reactions are essential to maintaining sustainable production of fuels, medicines and materials, and the majority of industrially important reactions are facilitated by catalysts, which are additives that speed up chemical reactions without being consumed in the process [1]. Catalysts enable a shorter reaction time without affecting the yield, and considerable research has been directed toward finding the right catalyst for a given reaction and industrial objective. Applications include energy generation [2–6] using metallic nanoparticles (typically 1-100 nm), which include single crystals, polycrystals and clusters in a variety of sizes and shapes [7, 8]. Metallic nanoparticles have high surface-area-to-volume ratio, enabling high catalytic activity to be achieved since more active sites are available for reactant adsorption compared to larger catalysts, and offer a large number of engineering degrees of freedom [9]. This presents both an opportunity and a challenge.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

Nanoinformatics [10] has recently emerged as an complementary approach to typical trial-and-error nanomaterials design, ideal for situations where the design space is large; employing both computational simulations and data-driven methods such as machine learning in conjunction with standard experimental and computational tools [11] to infer relationships between the structural characteristics (features, that can be control) and the properties (target labels, that are required) [12, 13]. Past studies have focused on forward prediction [14–17], but an alternative approach is to use inverse predictions, that provide a 'recipe' for experimentalists to follow [18–20]. This is difficult however, due to the highly imbalanced mapping functions; we have many more structural features we need to predict than properties we have to base them on.

Recently an entirely machine learning approach to inverse design has been reported which directly maps multiple structural features to multiple property labels, simultaneously, without the need for additional global searching or optimization [21]. This method has considerable advantages over conventional inverse design that requires exhaustive data sets, external optimization algorithms, and can potentially predict multiple candidates with no means for discrimination. In this method the mapping imbalance problem is alleviated by using a matching forward model for features selection, but has so far only been demonstrated for low-dimensional cases or where the mapping can be reduced to 1-to-1, which is unlikely to be true for complex materials such as electrocatalysts. In this study we explore the use of the optimization-free inverse design method for a high-dimensional set of platinum nanoparticle catalysts and show that, while restricting the models to features identified using importance profiles and recursive feature elimination, a systematic improvement in model performance can be achieved by balancing the mapping function, even if it is imbalanced in the corresponding forward model.

2 Methods

2.1 Data set

The raw data used in this work are generated through molecular dynamics simulations of the sintering and coarsening of platinum nanoparticles under different temperatures and atomic deposition rates, which exist in the form of three dimensional spatial coordinates [22], and can be obtained from Reference [23]. The entire data set contains 1300 instances, over a range of temperatures, sizes (54 to 15837 atoms), growth rates and shapes. A complete list of the manually extracted features and labels is provided with the meta data in the online repository [23]. The data set is labelled by the proportion of surface atoms with particular ranges of coordination number, or surface coordination numbers as an indicator of catalytic activity of face-centred cubic (fcc) metallic nanoparticles for carbon monoxide oxidation reactions, oxygen reduction reactions, and hydrogen oxidation and evolution reactions [24–26].

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

2.2 Multi-target random forest regression

Previous works have demonstrated that random forest (RF) methods can be used to fit high-dimensional target labels simultaneously with high accuracy [21]. Random forest is an ensemble technique based on decision trees with bootstrap aggregation (bagging), that has been shown to perform well for predicting the properties of metallic nanoparticle compared with other regressors [27]. A decision tree predicts the value of a label following decision rules inferred from the features and bagging randomly sampled from the training set with replacement, reducing the variance.

For each tree in the forest a bootstrap sample, i is selected from S, giving $S^{(i)}$. A decision-tree is learnt such that instead of examining all possible feature-splits at each node of the tree, a randomly subset of the features $f \subseteq F$ is used, where F is a set containing all features. Each node splits on the best feature in f, which is smaller than F, making it more efficient, and allows for the accumulation of importance profiles based on how many times a decision is split on a given feature. The pseudocode is illustrated in Alg. 1.

Algorithm 1 Random forest

Require: A training set $S := (x_1, y_1), \ldots, (x_n, y_n)$, features F, and number of trees in forest B.
  function RandomForest(S, F)
    $H \leftarrow 0$
    for $i \in 1, ..., B$ **do**
      $S^{(i)} \leftarrow$ A bootstrap sample from S
      $h_i \leftarrow$ RandomizedTreeLearn($S^{(i)}, F$)
      $H \leftarrow H \cup \{h_i\}$
    end for
    return H
  end function
  function RandomizedTreeLearn(S, F)
    for Each node: **do**
      $f \leftarrow$ very small subset of F
      Split on best feature in f
    end for
    return The learned tree
  end function

Constructing a large number of decision trees with random feature selection grows into a random forest, in which the decision of the individual tree is counted to output the mode of the classes for classification or average prediction for regression [28, 29]. Features are ranked during training based on variance reduction during regression tasks, increasing the diversity and avoiding over-fitting. Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

total number of samples. The feature selection criterion for internal nodes is the Gini impurity or information gain for classification and variance reduction for regression. This machine learning method can intrinsically handle multi-task problems since the leaf nodes can refer to any collection of relevant labels. To extend the traditional single-target ensemble predictor to solve multi-target RF learning problems, users can simply substitute the typical univariate trees for multivariate trees, where leaf nodes refer to multiple classes or target labels [30].

2.3 Workflow

The workflow for multi-target machine learning-based forward and inverse design requires data preprocessing (outlier removal and feature engineering) and splitting the data for model training and validation. We used Tukey's method [31] to detect outliers based on the quartiles of the data, and the results were optimal when the threshold was set to 4, leaving 1114 instances. Constant features and strongly correlated features (with $> 95\%$ correlation) were omitted to avoid introducing bias during learning and each label was stratified to reduce the impact of imbalanced distributions, resulting in 37 features retained in the feature space. All data are standardised and normalised (both features and labels). Essential hyper-parameter optimization of all models was undertaken here using a random grid search with 5-fold cross validation and evaluated using the mean squared error (MSE) and the mean absolute error (MAE). 10-fold cross validation was used during training, and all models used stratified 25/75 test/train splits with the same random seed (both forward and inverse, during optimization and training).

When applying this approach to inverse design, a preliminary multi-target forward model is often necessary to reduce the number of features and focus the inverse model on attributes that contribute to a strong relationship. Robust feature selection in a forward ML workflow is typically data-driven, such that final feature set used to describe the raw data is selected by computational algorithms [32]. This reduces model complexity, improves performance, and indicates which features are likely to be most influential. When using RF regression, this involves using the feature importance profile and recursive feature elimination (RFE) to extract the subset of important features that are sufficient to simultaneously predict the target labels without significant loss of performance. A multi-target inverse model can be re-optimized and then re-trained to simultaneously predict the reduced (important) set of structural features (referred to as "meta-labels") using the set of labels (referred to as "meta-features"), as reported elsewhere [21].

The entire workflow is provided in Figure 1.

3 Discussion of Results

After data cleaning 37 features were retained to predict the 4 target labels. This was done both separately (single target RF regression) and simultaneously

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

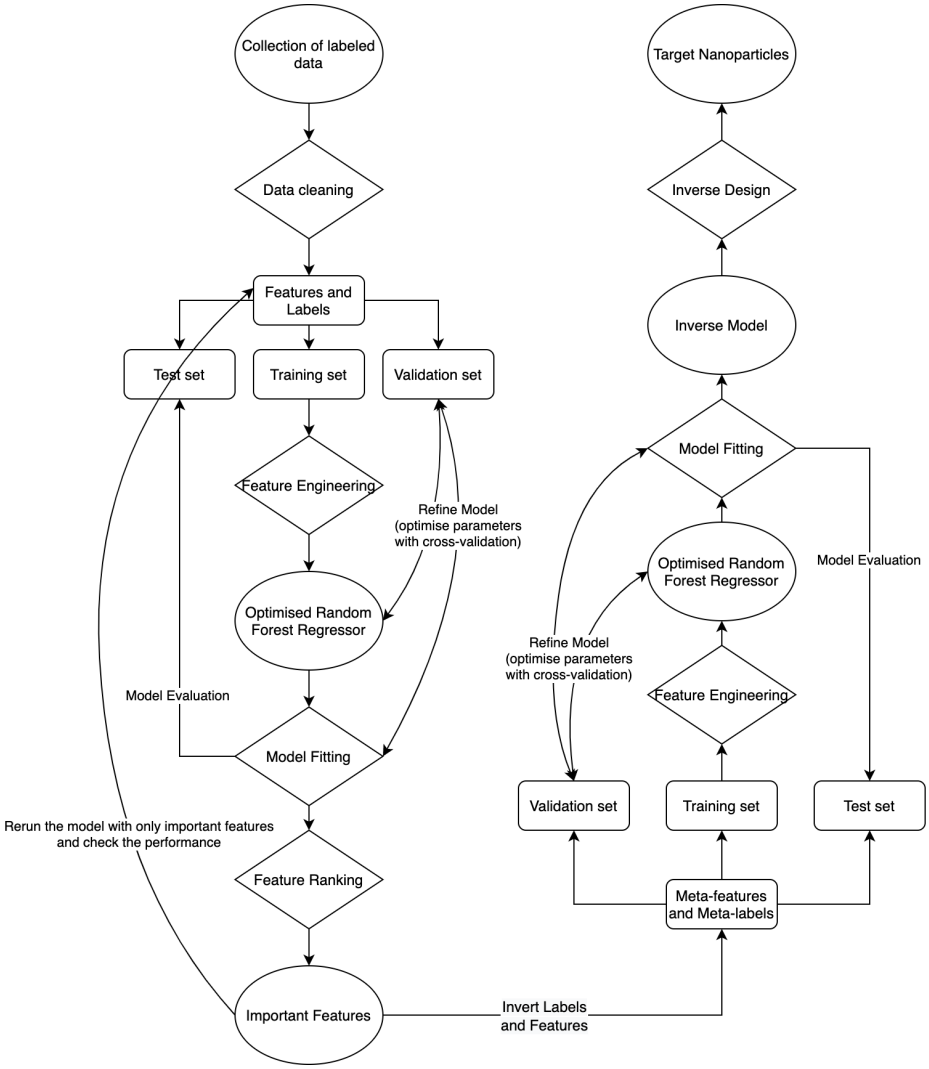

Fig. 1. Workflow for the inverse design methodology, based on multi-target regression and feature selection using a forward model to reduce the mapping function. Reproduced with permission from Reference [21].

(multi-target RF regression) as summarised in Table 1, providing a baseline. The results for each of the four single-target models (which provide the baselines) are compared with the multi-target prediction in Table 1 and presented in the Figure 2. We can see from Figure 2 that there is no bias error (under-fitting) and less than 0.5% variance error (over-fitting) when we use all of the 37 features retained after cleaning, indicating the forward model can achieve high accuracy and generalizability. The relative importance (magnitude) rapidly decays after

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

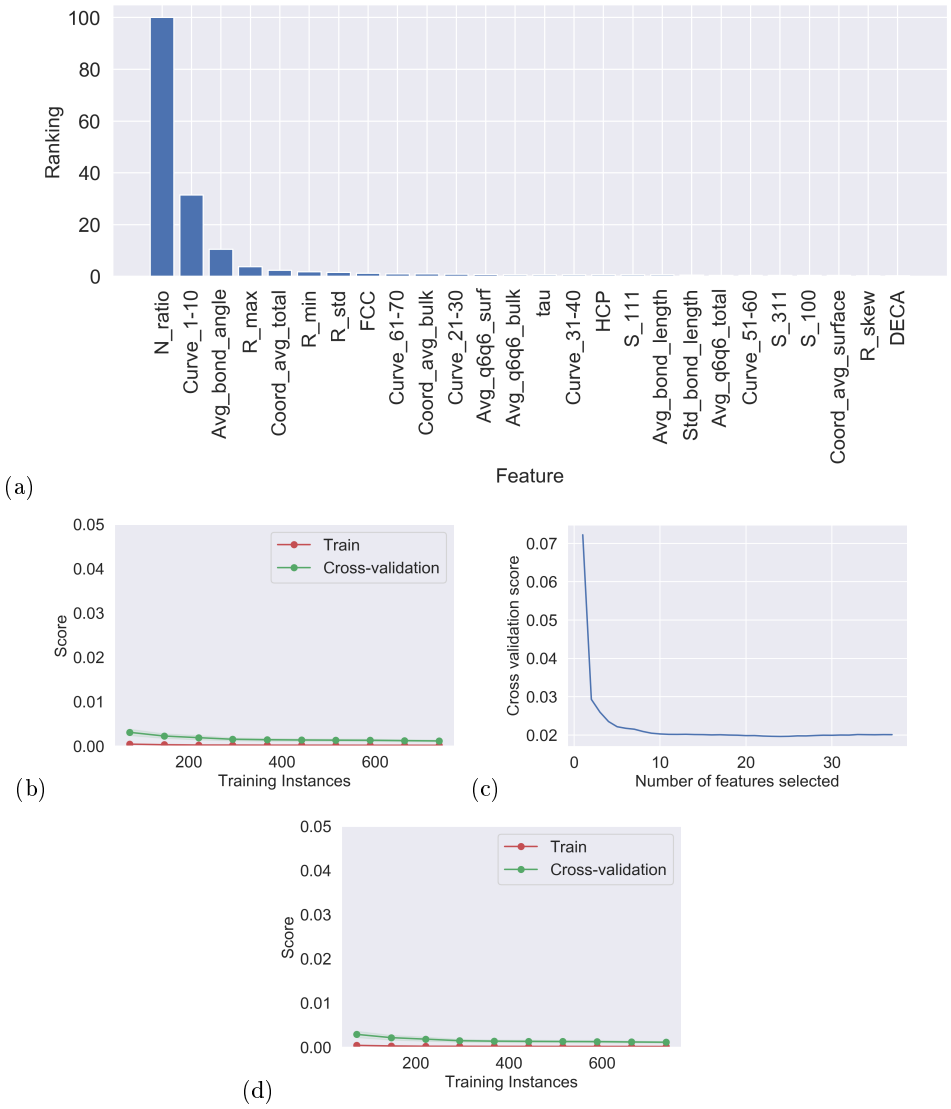

Fig. 2. Results for forward multi-target regression simultaneously predicting all four target properties of the Pt catalysts trained on 37 features retained after data cleaning, including (a) the feature importance profile rankings, (b) the learning curve trained on the 37 retained features, (c) the recursive feature elimination selecting the top 24 features, and (d) the learning curve trained on the 24 most important features, showing the accuracy and generalizability. Scores are the mean squared error.

the top three features in the feature importance profile. The forward features selection using RFE indicates that the model may be reduced to the top 24 fea-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

Table 1. Forward predictions using the multi-target regression model of Pt catalyst nanoparticles, with all of the features retained after data cleaning, the top 24 important features identified using recursive feature elimination to predict single property individually, or all 4 properties simultaneously; and inverse predictions using the 24 most important features or the 16, 8 or 4 top features as ranked in the forward feature importance profile. The results are evaluated using the mean absolute error (MAE) and mean squared error (MSE).

Prediction	Feature Set (Number)	Target (Number)	MAE	MSE
Forward	Retained (37)	`Surface_Defects` (1)	0.0205	0.0016
Forward	Retained (37)	`Surface_Microstructures` (1)	0.0155	0.0007
Forward	Retained (37)	`Surface_Facets` (1)	0.0181	0.0007
Forward	Retained (37)	`Formation_E` (1)	0.0129	0.0003
Forward	Retained (37)	All Properties (4)	0.0211	0.0012
Forward	Important (24)	All Properties (4)	0.0208	0.0012
Inverse	All Properties (4)	Important (24)	0.0540	0.0085
Inverse	All Properties (4)	Top (16)	0.0499	0.0071
Inverse	All Properties (4)	Top (8)	0.0410	0.0048
Inverse	All Properties (4)	Top (4)	0.0284	0.0022

tures, which exhibits an increase in performance due to the significant reduction in model complexity and the removal of nuisance variables, as measured by the MAE and MSE in Table 1. The training process for the final multi-target forward model is visualised in Figure 2(d). Overall the multi-target model exhibited performance comparable to the single-target models, and all models converge well with almost no loss of accuracy or generalizability.

To develop the inverse model the RF regressor was re-optimized and retrained on the same data set, but the properties become "meta-features" and the structural characteristics become "meta-labels". The models are evaluated using the same strategy, and the results are presented in Table 1. All training processes are visualised using learning curves in Figure 3(a), and the comparison of the predicted values of an instance from the testing set with its corresponding ground truth in Figure 3(b), for the set containing the top 24 meta-labels. The MAE for the inverse model is slightly more than double the MAE for the forward model, due to the significant imbalance between the number of meta-features and meta-labels, but is still less than 6%. As we reduce the feature set below the optimal list identified using RFE in the forward model, a systematic improvement in inverse model performance can be achieved by balancing the mapping function, even if it is imbalanced in the corresponding forward model. This is shown in the decrease in the MAE and MSE in Table 1 and in Figure 3(c,d) for 16 meta-labels, Figure 3(e,f) for 8 meta-labels, and Figure 3(g,h) for 4 meta-labels (representing a balanced function).

Conventional inverse design approaches are based on screening of a set of (forward) structure/property relationships [33–38] to determine the structures with the right properties, which eventually casts it as an optimization problem. Several approaches have been proposed [39, 40]. Tominaga *et al.* designed

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

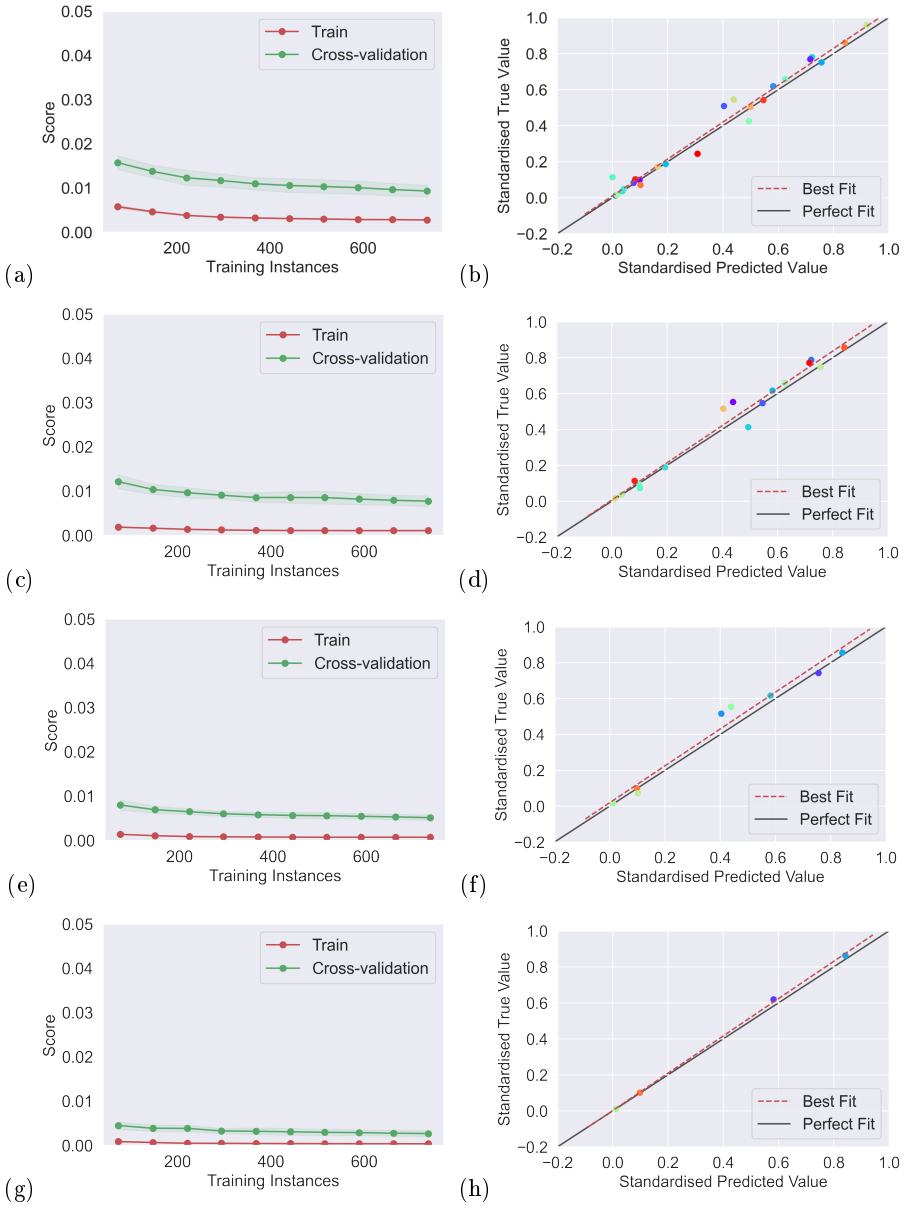

Fig. 3. Results for inverse multi-target regression simultaneously predicting (a,b) 24 meta-labels (structural characteristics), (c,d) 16 meta-labels, (e,f) 8 meta-labels, and (g,h) 4 meta-labels, for the Pt catalysts trained on all 4 target meta-features (nanoparticle stability and catalytic property indicators). The learning curves are shown to the left (a,c,f,g) and the 45 degree plot comparing the prediction for an instance in the testing set with the ground truth to the right (b,d,f,h). Scores are the mean squared error, and the colours in (b,d,f,h) are the features in the profile shown in Figure 2(a).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

a procedure that used genetic algorithms and Yang *et al.* developed IM^{2}ODE (Inverse design of Materials by Multi-Objective Differential Evolution) based on multi-objective differential evolution, both capable of global searching of large databases. Zunger *et al.* also identified materials with specific functionalities using an inverse design framework based on a global searching task combined with high-throughput density functional theory [41–43]. These approaches suffer from introducing more overhead (beyond the model training), more uncertainty (as optimization is imperfect), and can only identify candidate materials that are already in the set. They also become computationally intensive and impractical in nanomaterials design, where the design space is larger, and the probability of predicting implausible configurations increases [44]. By eliminating the need for the optimization step the workflow is accelerated, with the added advance of increasing specificity and relaxing the need for big data.

Given a profile of desired properties for a catalyst, the well-trained inverse models can output the most appropriate values of each structural feature for a given set of industrial requirements; property indicators for the oxygen reduction reaction could be suppressed in favour of indicators for hydrogen evolution reactions, for example, or a particle could be designed that can be simultaneously be used for both reactions. A combination of structural feature could be predicted for the specified properties without requiring additional optimization or global searching. A question that arises, however, is how flexible the feature (meta-label) selection process can be? Some of the top 24, 16, 8 or even top 4 features cannot be directly controlled in the lab. Including domain knowledge and selecting features based on practical considerations could accelerate translations of predictions in to the lab, and eventually the factory, but would represent a departure from an entirely data-driven, evidence based approach. The impact of combining, or prioritising, user preference over data-driven feature selection is the topic of future work, to explore how tolerant inverse models (such as these) will be to human intervention and whether inverse models can be tailored to experimental or industrial needs.

4 Conclusions

In the present study we have demonstrated the use of forward and inverse design models to predict the structure/property and property/structure relationship of Pt nanoparticle catalysts, respectively. These models included performance indicators relevant to the nanoparticle stability, carbon monoxide oxidation reactions, oxygen reduction reactions, and hydrogen oxidation and evolution reactions. We have used interpretable multi-target random forest regression to predict multiple property indicators and structural characteristics simultaneously, which better accounts for the fact that material properties can be correlated and that certain structural features can drive more than one type of reaction. The models generally show low errors, less than 2%, with no bias error and minimal variance error. The absolute prediction error for the inverse model is more than twice the error in the corresponding forward model, likely due (at least in part)

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

to the significant imbalance in the mapping function; predicting 24 meta-labels with only 4 meta-features.

To improve the performance of the inverse model, we explored the impact of artificially reducing the set of meta-labels to better balance the mapping function, while retaining the most important features in the forward model at each stage. We find that systematically reducing the number of meta-labels lowers the model error, and when the number of meta-features and meta-labels are balanced a similar performance to the forward model can be achieved. Reducing the meta-labels balance also simplifies the inverse model and can offer some advantages when using it in practice. This raises questions as to whether human intervention in the feature selection process to accommodate practical consideration, will be detrimental to performance once a fully data-driven approach is abandoned. This is an interesting topic of ongoing work and will be reported elsewhere.

Acknowledgements Computational resources for this project were supplied by the National Computing Infrastructure national facility under grant p00.

References

1. Laidler, K.J.: A glossary of terms used in chemical kinetics, including reaction dynamics (IUPAC Recommendations 1996). Pure and Applied Chemistry **68**(1), 149–192 (1996).
2. Shokrlu, Y.H., Babadagli, T.: Viscosity reduction of heavy oil/bitumen using micro- and nano-metal particles during aqueous and non-aqueous thermal applications. Journal of Petroleum Science and Engineering **119**, 210–220 (2014).
3. González-Gil, R., Herrera, C., Larrubia, M.A., Mariño, F., Laborde, M., Alemany, L.J.: Hydrogen production by ethanol steam reforming over multimetallic RhCeNi/Al2O3 structured catalyst. Pilot-scale study. International Journal of Hydrogen Energy **41**(38), 16786–16796 (2016).
4. Guo, W., Vlachos, D.G.: Patched bimetallic surfaces are active catalysts for ammonia decomposition. Nature Communications **6**(1), 8619 (2015).
5. Wu, Z.P., Caracciolo, D.T., Yazan, M., Jianguo, W., Zhijie, K., Shiyao, S., Vargas, J.A., Yan, S., Hopkins, E., Keonwoo, P., Sharma, A., Yang, R., Valeri, P., Wang, L., Chuan-Jian, Z.: Alloying–realloying enabled high durability for Pt–Pd-3d-transition metal nanoparticle fuel cell catalysts. Nature Communications **12**(1) (2021).
6. Sheng, W., Bivens, A.P., Myint, M., Zhuang, Z., Forest, R.V., Fang, Q., Chen, J.G., Yan, Y.: Non-precious metal electrocatalysts with high activity for hydrogen oxidation reaction in alkaline electrolytes. Energy & Environmental Science **7**(5), 1719–1724 (2014).
7. Wu, Z., Yang, S., Wu, W.: Shape control of inorganic nanoparticles from solution. Nanoscale **8**(3), 1237–1259 (2016).
8. Baig, N., Kammakakam, I., Falath, W.: Nanomaterials: a review of synthesis methods, properties, recent progress, and challenges. Materials Advances **2**(6), 1821–1871 (2021).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

9. Rodrigues, T.S., da Silva, A.G.M., Camargo, P.H.C.: Nanocatalysis by noble metal nanoparticles: controlled synthesis for the optimization and understanding of activities. Journal of Materials Chemistry A **7**(11), 5857–5874 (2019).
10. Barnard, A.S., Motevalli, B., Parker, A.J., Fischer, J.M., Feigl, C.A., Opletal, G.: Nanoinformatics, and the big challenges for the science of small things. Nanoscale **11**(41), 19190–19201 (2019).
11. Rajan, K.: Materials informatics. Materials Today **8**(10), 38–45 (2005).
12. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering **160**(1), 3–24 (2007).
13. Sammut, C., Webb, G.I. (eds.): Supervised Learning, p. 941. Springer US, Boston, MA (2010).
14. Parker, A.J., Opletal, G., Barnard, A.S.: Classification of platinum nanoparticle catalysts using machine learning. Journal of Applied Physics **128**(1), 1–11 (2020).
15. Sun, B., Fernandez, M., Barnard, A.S.: Machine Learning for Silver Nanoparticle Electron Transfer Property Prediction. Journal of chemical information and modeling **57**(10), 2413–2423 (2017).
16. Janet, J.P., Kulik, H.J.: Predicting electronic structure properties of transition metal complexes with neural networks. Chemical Science **8**(7), 5137–5152 (2017).
17. Takigawa, I., Shimizu, K.i., Tsuda, K., Takakusagi, S.: Machine-learning prediction of the d-band center for metals and bimetals. RSC Advances **6**(58), 52587–52595 (2016).
18. Christiansen, R.E., Michon, J., Benzaouia, M., Sigmund, O., Johnson, S.G.: Inverse design of nanoparticles for enhanced Raman scattering. Optics Express **28**(4), 4444–4462 (2020).
19. Lee, J.W., Park, W.B., Do Lee, B., Kim, S., Goo, N.H., Sohn, K.S.: Dirty engineering data-driven inverse prediction machine learning model. Scientific Reports **10**(1), 20443 (2020).
20. Hassan, S.A.: Artificial neural networks for the inverse design of nanoparticles with preferential nano-bio behaviors. The Journal of Chemical Physics **153**(5), 54102 (2020).
21. Li, S., Barnard, A.S.: Inverse design of nanoparticles using multi-target machine learning. Advanced Theory and Simulation, **5**(2), 2100414 (2022).
22. Barron, H., Opletal, G., Tilley, R.D., Barnard, A.S.: Dynamic evolution of specific catalytic sites on Pt nanoparticles. Catalysis Science and Technology **6**(1), 144–151 (2016).
23. Barnard, A., Opletal, G.: Platinum nanoparticle data set, v1. CSIRO Data Collection (2019), https://doi.org/10.25919/5d3958d9bf5f7
24. Zhao, Z., Chen, Z., Zhang, X., Lu, G.: Generalized Surface Coordination Number as an Activity Descriptor for CO2 Reduction on Cu Surfaces. The Journal of Physical Chemistry C **120**(49), 28125–28130 (2016).
25. Sun, B., Barron, H., Wells, B., Opletal, G., Barnard, A.S.: Correlating anisotropy and disorder with the surface structure of platinum nanoparticles. Nanoscale **10**(43), 20393–20404 (2018).
26. Parker, A.J., Barnard, A.S.: Machine learning reveals multiple classes of diamond nanoparticles. Nanoscale Horizons **5**(10), 1394–1399 (2020).
27. Barnard, A.S., Opletal, G.: Selecting machine learning models for metallic nanoparticles. Nano Futures **4**(3), 035003 (2020).
28. Breiman, L.: Random forests. Machine Learning **45**(1), 5–32 (2001).
29. Breiman, L.: Bagging predictors. Machine Learning **24**, 123–140 (1996).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39

30. Kamiński, B., Jakubczyk, M., Szufel, P.: A framework for sensitivity analysis of decision trees. Central European Journal of Operations Research **26**(1), 135–159 (2018).
31. Hoaglin, D.C., Iglewicz, B., Tukey, J.W.: Performance of some resistant rules for outlier labeling. Journal of the American Statistical Association **81**(396), 991–999 (1986).
32. Liu, T., Barnard, A.S.: Fast derivation of shapley based feature importances through feature extraction methods for nanoinformatics. Machine Learning: Science and Technology **2**(3), 035034 (2021).
33. Zunger, A.: Inverse design in search of materials with target functionalities. Nature Reviews Chemistry **2**, 0121 (2018).
34. Sanchez-Lengeling, B., Aspuru-Guzik, A.: Inverse molecular design using machine learning: Generative models for matter engineering. Science **361**, 360 (2018).
35. Jørgensen, P. B., Schmidt, M. N., Winther O.: Deep generative models for molecular science. Molecular Informatics **37**, 1700133 (2018).
36. Hanakata, P. Z., Cubuk, E. D.,. Campbell, D. K, Park, H. S.: Accelerated search and design of stretchable graphene kirigami using machine learning. Physical Review Letters **121**, 255304 (2018).
37. Wan, J., Jiang, J.-W., Park, H. S.: Thermal conductivity versus the density of holes for porous graphene at room temperature. Carbon **157**, 262 (2020).
38. Ma, C., Zhang, Z., Luce, B., Pusateri, S., Xie, B., Rafiei, M., Hu, N.: Accelerated design and characterization of non-uniform cellular materials via a machine-learning based framework. npj Computational Materials **6**, 40 (2020).
39. Tominaga, D., Koga, N., Okamoto M.: in, Proceedings of the 2nd Annual Conference on Genetic and Evolutionary Computation, ACMPress, New York pp. 251–258 (2000)
40. Zhang ,Y.-Y., Gao, W., Chen, Xiang, S., H., Gong, X.-G.: Inverse design of materials by multi-objective differential evolution. Computational Materials Science **98**, 51 (2015).
41. Dudiy, S., Zunger, A.: Searching for alloy configurations with target physical properties: Impurity design via a genetic algorithm inverse band structure approach. Physical Review Letters **97**, 046401 (2006).
42. Yu, L., Kokenyesi, R. S., Keszler, D. A., Zunger, A.: Inverse design of high absorption thin-film photovoltaic materials. Advanced Energy Materials **3**, 43 (2013).
43. Yang, D., Lv, J., Zhao, X., Xu, Q., Fu, Y., Zhan, Y., Zunger, A., Zhang, L.: Functionality-directed screening of Pb-free hybrid organic-inorganic perovskites with desired ontrinsic photovoltaic functionalities. Chem. Mater. **29**, 524 (2017).
44. Zunger, A.: Beware of plausible predictions of fantasy materials. Nature **566**, 447 (2019).

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08754-7_39