

Supporting Information

Machine learning-assisted optimization of multi-metal hydroxide electrocatalysts for overall water splitting

Carina Yi Jing Lim^a, Riko I Made^a, Zi Hui Jonathan Khoo^{a,b}, Chee Koon Ng^a, Yang Bai^{a,c}, Jianbiao Wang^a, Gaoliang Yang^a, Albertus D. Handoko^{a,b}, Yee-Fun Lim^{a,b*}*

^a*Institute of Materials Research and Engineering (IMRE), Agency for Science, Technology and Research (A*STAR), 2 Fusionopolis Way, Innovis, Singapore 138634, Singapore*

^b*Institute of Sustainability for Chemical, Energy and Environment (ISCE²), Agency of Science, Technology and Research (A*STAR), 1 Pesek Road, Singapore 627833, Singapore.*

^c*Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 3G4, Canada*

*Email: Handoko.Albertus@isce2.a-star.edu.sg

*Email: limyf@isce2.a-star.edu.sg

S1. Experimental

S1.1 Materials

Cobalt(II) nitrate hexahydrate, nickel(II) nitrate hexahydrate (99.999%, trace metal basis), iron(III) nitrate nonahydrate (99.99+%), urea and triethanolamine was obtained from Sigma Aldrich while ammonium molybdate (para) tetrahydrate and ammonium tungsten oxide hydrate (lot no. B23W009, hydration level, $x = 4.8$) were purchased from Alfa Aesar. Perfluorinated resin solution containing Nafion of 5 wt.% Nafion in lower aliphatic alcohols and water purchased from Sigma Aldrich was used in the catalyst ink. Potassium hydroxide, 99.98% (metal basis) purchased from Alfa Aesar was used for the electrolyte.

All chemicals and materials were used as received.

S1.2 Synthesis

Synthesis on Automated Workstation. Aqueous solutions of 0.2 M triethanolamine (TEA) and 0.4 M urea and 0.2 M for each of the five metal precursors (normalised to the molar ratio of the metal in each salt, i.e., 0.2 M of $(\text{NH}_4)_{0.5}\text{WO}_{3.25}\cdot 0.4\text{H}_2\text{O}$ and $(\text{NH}_4)_{0.86}\text{MoO}_{3.43}\cdot 0.57\text{H}_2\text{O}$ for the W and Mo precursor) were prepared. A high-throughput automated workstation (Lissy Chemical Workstation from Zinsser Analytics) was initially used to synthesise the catalysts. The workstation allows for the auto-pipetting of pre-determined volumes of liquids as well as the shaking and heating of solution vials. For the workstation's setup, glass vials containing the 7 precursor solutions and de-ionised water was loaded into the workstation blocks. 8 mL glass vials, each with two 5 mm zirconia balls were used as the reaction vessels. Two zirconia grinding balls were included in each vial to facilitate homogenisation during shaking steps, and all shaking steps were conducted with shaking speed of 1200 rpm.

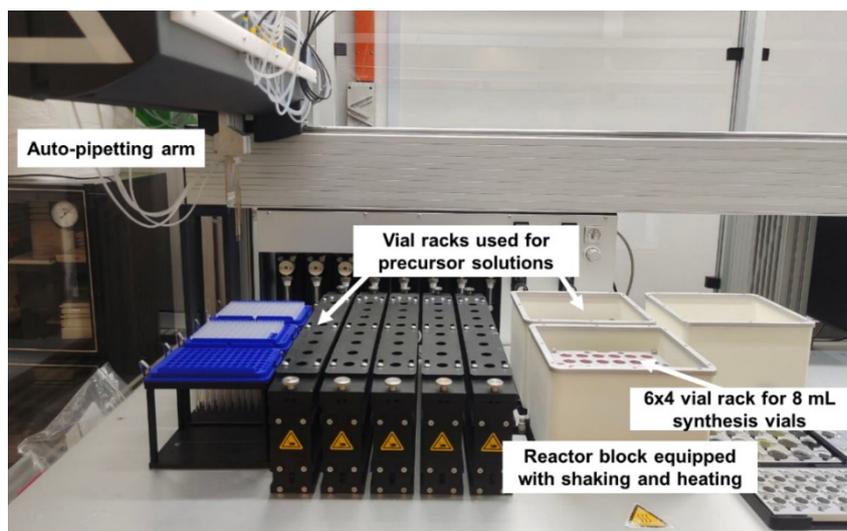


Figure S1.1 Image of high-throughput automated workstation set-up.

The synthesis workflow is summarised in **Figure S1**. Into the 8 mL glass vials, 2.25 mL of deionised water was first pipetted, followed by the addition of the metal salt solutions. Varying volumes of each of the five metal solutions were added, totalling to 1.25 mL. After the addition of the metal salt solutions, 1.25 mL of the urea solution was also added before the vials were shaken for 5 min at room temperature. Subsequently, 5 additions of 250 μ L aqueous TEA was introduced to the vials, interspaced by 1 min shaking at room temperature. After 5 additions of TEA, the vials were first shaken for 24 h at room temperature, following by shaking for an additional 48 h at 95 $^{\circ}$ C.

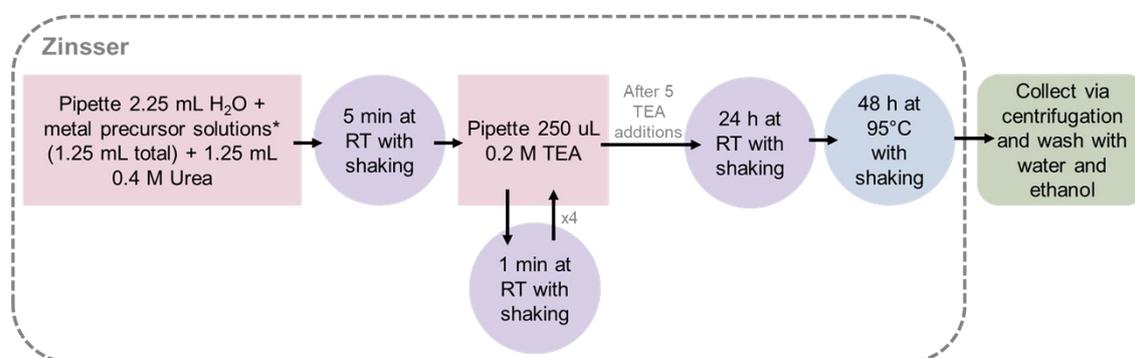


Figure S1.2 Summary of synthesis workflow on the high-throughput automated workstation.

Scaled-up Synthesis Procedure. To scale up the synthesis using the optimized ratio of the precursors, a round-bottomed flask with magnetic stirring at 400 rpm was used. 287.5 and 280.9 mg of the Co and Fe salts were first dissolved in 30 mL of DI water. The Mo solution, consisting of 144.1 mg of the Mo precursor in 6 mL of water, was then added dropwise to the reaction mixture. Separately, 300.3 mg of urea and 373.0 mg of TEA were then dissolved in 12 mL of DI H₂O. The urea solution was gradually added to the reaction flask, followed by the dropwise addition of the TEA solution. The reaction mixture was left to stir for 24 h at room temperature followed by an additional 48 h stirring at 95 $^{\circ}$ C.

For all synthesised catalysts, the solid products were collected and washed twice with ethanol and water with centrifugation at 6000 rpm (4830 \times g) for 4 mins, before drying in a 60 $^{\circ}$ C oven overnight.

Synthesis of Additional Samples. Additional samples include the NiFe benchmark (Ni₃Fe₁) as well as additional samples were synthesised to investigate the influence of the metal precursors (discussed in **Section S4.2** below). NiFe (Ni₃Fe₁) was synthesised according to literature,¹ where in 80 mL of water, Ni, Fe and urea precursors were dissolved to achieve concentrations of 7.5, 2.5, and 17.5 mM respectively. This is followed by the dropwise addition of 0.8 mmol of TEA under magnetic stirring. The solution was left to stir at room temperature for 24 h followed by an additional 48 h at 100 °C. The Ni₄Fe₃ and Co₄Fe₃ was synthesised similarly, altering the metal precursor ratios whilst keeping the total molar amount of the metal salts consistent. CoFe-W was synthesised by replacing the Mo precursor with the W precursor in the procedure used to prepare NN-12, as detailed in the *Scaled-up Synthesis Procedure* above.

S1.3 Materials Characterisation

X-ray diffraction (XRD) was performed on a Bruker Advance D8 diffractometer (Cu K α λ = 1.541058 Å) equipped with Ni filter and LynxEye XE energy dispersive 1-D detector over a range of 8 – 90° 2 θ with a step size of 0.02° and collection time of 0.3 s per step. A variable divergence slit was programmed to allow a constant irradiated sample length of 10 mm to enhance the diffractions at higher 2 θ range. A beam knife with automatic height adjuster was placed above the samples to minimize air scattering at low 2 θ range. Approximately 50-100 mg of powders were filled into 10 mm diameter cavity of an off-cut zero background single crystal Si XRD holders (MTI Crystals Ltd.) to minimize background contribution. The sample holder was spun at 2 rpm to minimize texturing effect. Scanning electron microscopy (SEM) and energy dispersive X-ray (EDX) spectroscopy data were collected on a Hitachi SU8200 with 5 kV accelerating voltage. X-ray photoelectron spectroscopy (XPS) was performed on a Thermo-Scientific Theta Probe spectrometer fitted with an Al K α X-ray source and all spectra were calibrated to the graphitic carbon peak. Raman spectra were collected of the powdered samples on a glass slide using a Renishaw inVia confocal Raman microscope spectrometer with 785 nm red laser as the excitation source. Data collection was performed at 0.5% attenuated laser power with 50 \times objective lens, 1 s exposure time over 10 accumulations. Fourier transform infrared spectroscopy – attenuated total reflectance (FTIR-ATR) was collected on a Perkin Elmer Frontier with the Universal ATR attachment for 16 scans from 4000 cm⁻¹ to 600 cm⁻¹.

S1.4 Electrochemical Testing

All electrochemical measurements were conducted at room temperature (~25 °C) using a calibrated Gamry 600+ potentiostat. Only cyclic and linear sweep measurements are iR-compensated using Gamry's Current Interrupt function. With exception to the H₂ gas quantification experiment, both cathode and anodes were in the same compartment, ~1.5 cm apart, with no membrane/separators/frit used. A Hg/HgO reference electrode (CH 152, CH Instruments) was used, which was calibrated to a reversible hydrogen electrode (RHE, Hydroflex, Gaskatel GmbH) in 1 M KOH. All potentials reported are converted to RHE via the Nernst equation (pH = 14 was taken for 1 M KOH).

For the screening of catalysts, 0.196 cm² (0.5 cm diameter) vitreous carbon discs polished to a mirror finish with 0.3 μ m alumina slurry were used as the substrate. A 5 mg mL⁻¹ dispersion of each catalyst in 80:20:2 ethanol:water:Nafion was prepared via ultrasonication in an ice bath for 30 mins. 19.6 μ L of the dispersion was deposited onto the polished glassy carbon substrates and dried under a 100 W infrared lamp to achieve a mass loading of 0.5 mg cm⁻². The electrodes were then tested using a rotating disc electrode set-up (**Figure S1.3a**) (MSR Rotator, Pine Research Instrumentation) at 1600 rpm in de-gassed and Ar-purged 1 M high purity KOH (Alfa Aesar, 99.8% metals basis), with a carbon rod counter electrode (Grade 1 Spec-Pure, Ted Pella). To stabilise the catalysts, 5 cyclic voltammetry scans at 10

mV s⁻¹ were conducted at the voltage range of electrochemical testing, followed by a linear sweep voltammetry (LSV) with 1 mV step size to measure the electrochemical performance.

Ni foam electrodes were loaded on both sides via air-brush spraying with nitrogen gas, with 5 mg mL⁻¹ catalyst dispersion in an ethanol-Nafion solution to achieve average mass loadings on the foams achieved was 1.55-1.65 mg cm⁻². Electrical wires were soldered on to the edge of the foams and epoxy resin was applied over the solder, wire, and excess foam areas, such that only 2 cm² (1 cm² for each side) of the active area would be exposed to the electrolyte (**Figure S1.3c**). Electrochemically active surface areas of the foam-based electrodes are reflected in **Figure S5.1**. Half-cell electrochemical tests were also conducted with similar set-up as the RDE – with the working electrode as the foam electrode instead, whilst overall water splitting was conducted with both foam electrodes as in **Figure S1.3b**.

All electrode preparation and testing procedures for any benchmarks are unchanged with procedures used for studied catalysts.

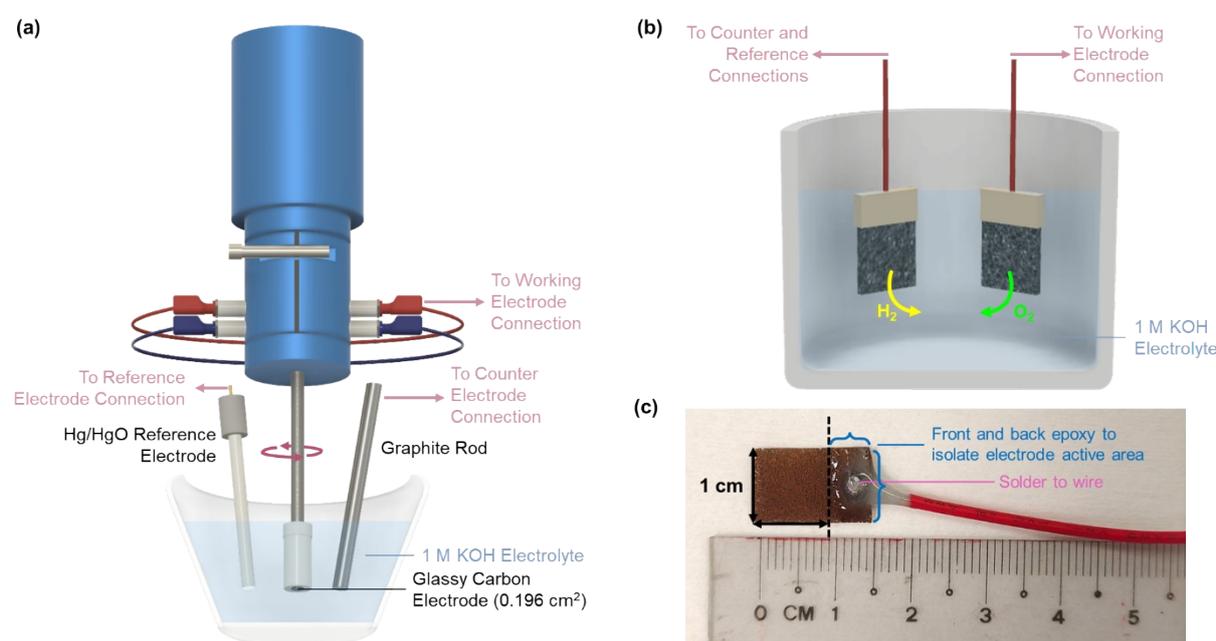


Figure S1.3 Schematics of the (a) 3 electrode rotating disc electrode set-up and (b) overall water splitting set-up. (c) Image of catalyst deposited on Ni foam and electrode set-up.

Quantification of hydrogen product was conducted with a two-electrode set-up in a PTFE gas-tight H-cell with electrolyte compartments separated by a Nafion 117 membrane. 15 mm diameter glassy carbon electrodes were used, with a 0.5 mg cm⁻² catalyst loading achieved via similar ink preparation and drop-cast deposition methods as for the RDE electrodes. The exposed geometric area of the working electrode in the H-cell was 1.13 cm² and Ar was bubbled into the electrolyte using a calibrated mass flow controller (Alicat Scientific MC) at 20 sccm. An automated valve gas chromatograph (GC) based on Agilent 7890B mainframe (Wasson ECE Instrumentation) was used with a thermal conductivity detector (N₂ carrier gas, 99.9995%, Leeden National Oxygen) was used for detecting H₂. The electrode was subjected to a chronopotentiometric measurement at 10 mA cm⁻² with periodic sampling of the gaseous products every 30 mins. The H₂ Faradaic efficiency (FE) was calculated as per the equation:

$$FE = \frac{[H_2] * n * F * V * 101325}{R * T * 10^{12} * 60 * I}$$

where $[H_2]$ is the recorded concentration of H_2 from the GC in ppm mol, $n = 2$ electrons required to form one molecule of H_2 , F is the Faraday constant (96485 C mol^{-1}), \dot{V} is the flow rate (mL min^{-1}), R is the gas constant ($8.314 \text{ m}^3 \text{ Pa K}^{-1} \text{ mol}^{-1}$), T is the temperature (K), and $I=11.33 \text{ mA}$, corresponding to a current density of 10 mA cm^{-2} .

Electrochemical impedance spectroscopy (EIS) data was collected with glassy carbon electrodes with 0.5 mg cm^{-2} loading, prepared with the same procedure as the electrodes used in the hydrogen quantification experiment as described above. All measurements were collected in Ar purged 1 M KOH , and the catalysts were pre-reduced with 5 cyclic voltammetry scans at 10 mV s^{-1} . Potentiostatic EIS runs were conducted at 1.65 and -0.35 V vs. RHE , for OER and HER respectively. The spectra were collected from 0.2 Hz to 100 kHz at 10 points per decade and were then fitted using the Simplex method on Gamry Echem Analyst (v7.9.0).

S2. Machine Learning

S2.1 Methods and Details of Model Selection and Optimisation of Catalyst Composition

Herein, machine learning (ML) is used as an unbiased guide for optimization of synthesis parameters, which encompasses constructing the objective surrogate model, generating suggestions for the next iteration of experiments, experimental validations, and model retraining. Plots supporting the experimental validation may be found in **Section 3** below.

As described in the main text, three surrogate models separately optimizing for η_{OER} , η_{HER} and $Y_{bifunc.}$ were built from the initial dataset to assist the optimization of the catalysts as follows:

$$\eta_{OER} = f_{OER}(X) \quad (1)$$

$$\eta_{HER} = f_{HER}(X) \quad (2)$$

$$Y_{bifunc.} = f_{bifunc.}(X) = -(\eta_{OER}^2 + \eta_{HER}^2)^{-1} \quad (3)$$

where X represents the molar ratio of the metal precursors; f_{OER} , f_{HER} and $f_{bifunc.}$ are the surrogate functions for η_{OER} , η_{HER} and bifunctional activity ($Y_{bifunc.}$), respectively. For GB, $Y_{bifunc.}$ was derived from f_{HER} and f_{OER} . In constructing the surrogate models, we explored three commonly used machine learning algorithms, namely Gradient Boosting (GB), Neural Network (NN) and Gaussian Process Regression (GPR) to map the inputs – which were the ratio of the precursor solutions, to the outputs – which were η_{OER} , η_{HER} and $Y_{bifunc.}$. To increase the likelihood of achieving a good generalised model, we partitioned the dataset into training and test with proportions of 70% and 30%, respectively (**Figure S2.1**). Of the test sets, a third of the test set comprises of the best 10% of the datapoints for the given objective to evaluate the models' ability to extrapolate beyond training dataset.

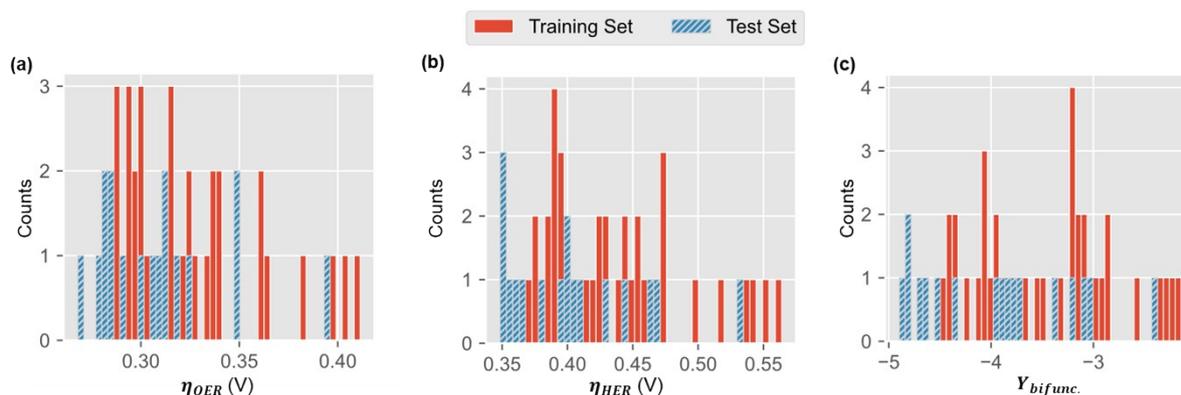


Figure S2.1. Data partitioning for the train and test set based on the initial sampling points for (a) η_{OER} , (b) η_{HER} and (c) $Y_{bifunc.}$. Note that we intentionally select some of the *test set* to be out of distribution from the *training set* to increase our chances to get extrapolative model.

To further increase the chances of obtaining a generalised model, the hyperparameters were also fully tuned. The training set was used to train the model parameters for a given model and set of hyperparameters. The first pass of model generalisation was evaluated with a third of the test set, where the model performance was quantified from the mean squared error (MSE) and the coefficient of correlation (R^2). The hyperparameters were then adjusted based on the results of the model evaluation with that third of test dataset. Bayesian optimization with Gaussian Process surrogate (BO-GP) was utilised to search for the optimum hyperparameters set from the hyperparameters space.

For the GPR-flavoured catalyst surrogate model, the hyperparameters space includes the choice of kernel and the associated kernel's parameters, such as the length scale. Meanwhile, for the GB-flavoured catalyst model, the hyperparameters to be optimized included the learning rate, number of

estimators, number, and weight of leaf, and the depth of the tree. Lastly, for NN, the search space included the number of layers, number of nodes for each layer, type of activation function, single dropout layer, the position of the dropout layer, and type of optimizer. After hyperparameter optimization, 10 models were generated for each objective using each algorithm, with an individual set of hyperparameters for each of the models.

To better illustrate, plots of the predicted against test values for η_{OER} (**Figure S2.2**), η_{HER} (**Figure S2.3**), and Y_{bifunc} (**Figure S2.4**), derived from the 10 best performing models of the NN algorithm are presented below (not sorted in any order of performance). In general, NN modelling performed better for the η_{HER} , that different set of hyperparameters able to reasonably map the input composition to the η_{HER} , i.e model 4, 5, 6, and 9 (**Figure S2.3, model**). Model-9, which is the best performing has 21 hidden layers with 328 hidden nodes, while Model-6 has 26 hidden layers with 462 hidden nodes.

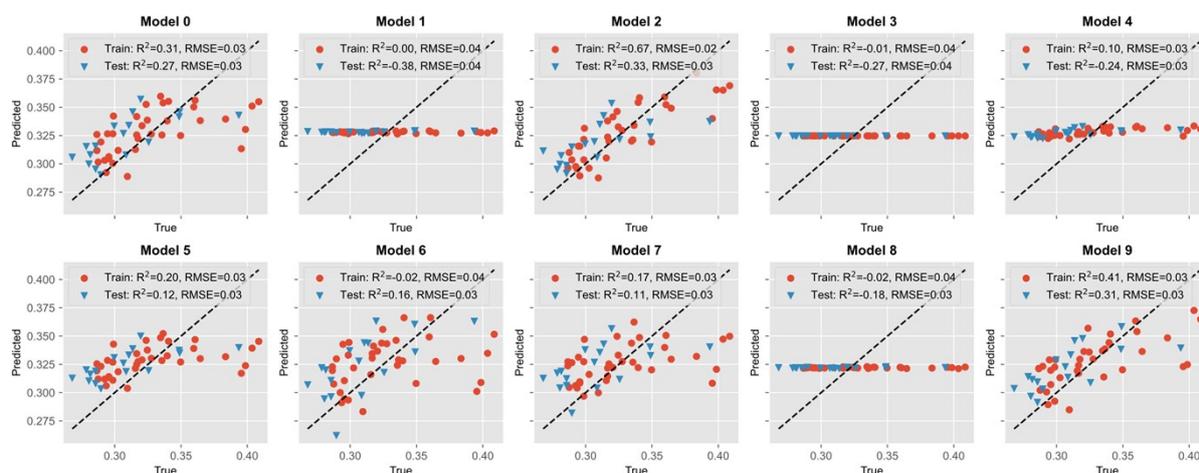


Figure S2.2. Predicted vs. true values of η_{OER} (V) for the 10 best performing models from the NN hyperparameters optimization.

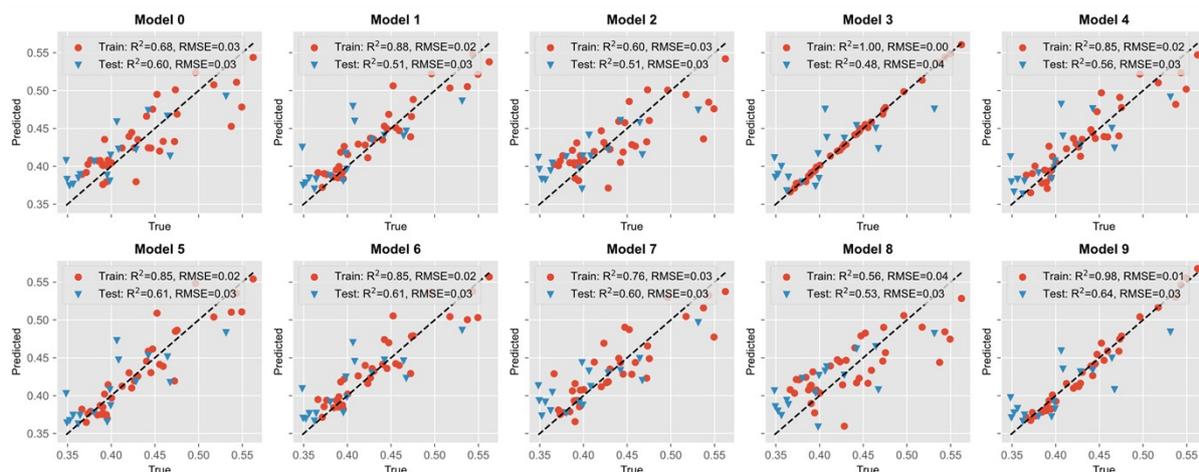


Figure S2.3. Predicted vs. true values of η_{HER} (V) for the 10 best performing models from the NN hyperparameters optimization.

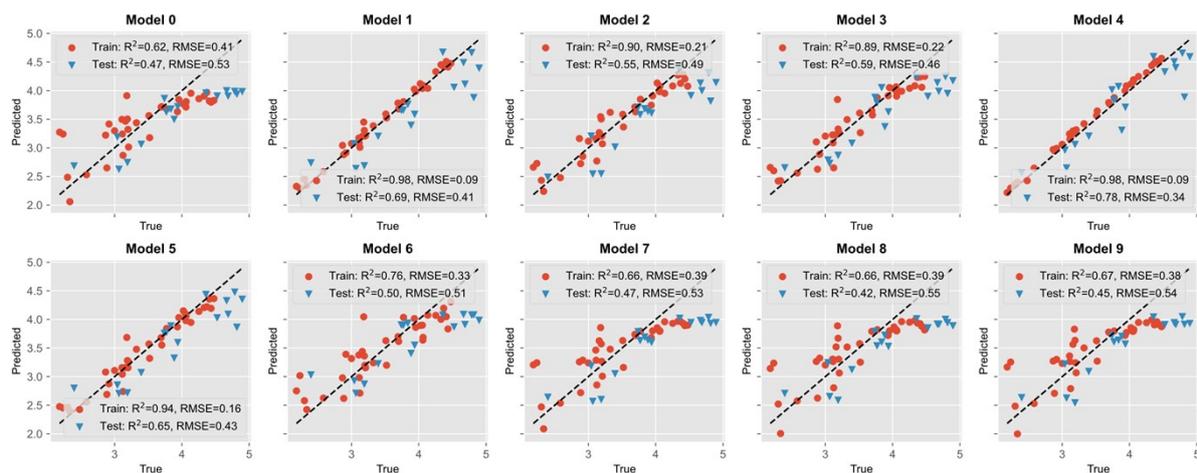


Figure S2.4. Predicted vs. true values of $-Y_{bifunc}$ for the 10 best performing models from the NN hyperparameters optimization.

Given the use of NN as the main algorithm to model and optimise the catalyst for the text, details of the models and hyperparameters are compiled in **Table S1-3** for the η_{OER} , η_{HER} and Y_{bifunc} models.

Table S1. Hyperparameters of neural network models for η_{OER} .

Model #	# Input Nodes	# Hidden Layers	# Hyperparameters	# Output Nodes	Activation Function	Dropout	Dropout Position	Training Optimizer	Learning Rate	L2
0	5	18	266	1	elu	0	-	Adagrad	0.000265	2.93E-10
1	5	28	156	1	selu	0	-	SGD	0.009463	1.78E-05
2	5	4	204	1	selu	0	-	Adadelta	0.00384	0.001575
3	5	45	217	1	leaky_relu	0.325722	80	Adagrad	0.0006	1.58E-10
4	5	23	450	1	selu	0	-	Adadelta	0.000527	9.62E-10
5	5	11	119	1	elu	0	-	Adagrad	0.000287	9.46E-05
6	5	2	500	1	tanh	0	-	Adadelta	0.001223	1E-10
7	5	7	111	1	selu	0	-	Adadelta	0.001852	0.019206
8	5	12	304	1	selu	0	-	SGD	0.004808	0.061234
9	5	2	500	1	elu	0	-	Adadelta	0.000421	1.12E-08

Table S2. Hyperparameters of neural network models for η_{HER} .

Model #	# Input Nodes	# Hidden Layers	# Hyperparameters	# Output Nodes	Activation Function	Dropout	Dropout Position	Training Optimizer	Learning Rate	L2
0	5	11	395	1	elu	0	-	Adagrad	0.004955	1.79E-05
1	5	16	500	1	elu	0.203814	30	Adagrad	0.001912	4.32E-08
2	5	6	110	1	selu	0	-	Adadelta	0.001738	0.036513
3	5	16	500	1	selu	0	-	Adagrad	0.0001	4.66E-07
4	5	22	285	1	selu	0	-	Adagrad	0.0001	3.16E-07
5	5	40	338	1	elu	0.367926	36	Adagrad	0.003464	1.21E-10
6	5	26	462	1	elu	0	-	Adagrad	0.000208	0.000429
7	5	13	491	1	elu	0	-	Adagrad	0.004409	9.78E-10
8	5	5	27	1	elu	0	-	Adagrad	0.000568	2.53E-09
9	5	21	328	1	selu	0.360234	26	RMSprop	0.000224	7.06E-07

Table S3. Hyperparameters of neural network models for $Y_{bifunc.}$

Model #	# Input Nodes	# Hidden Layers	# Hyperparameters	# Output Nodes	Activation Function	Dropout	Dropout Position	Training Optimizer	Learning Rate	L2
0	5	2	37	1	selu	0	-	Adadelata	0.002668142	0.1
1	5	6	217	1	relu	0.5	6	SGD	0.007723899	9.55222E-06
2	5	5	92	1	selu	0	-	Adadelata	0.001851168	0.011536374
3	5	2	286	1	relu	0	-	Adadelata	0.001349933	0.00034116
4	5	2	410	1	relu	0	-	NAdam	0.00276899	8.45286E-07
5	5	2	500	1	relu	0	-	SGD	0.00083601	1E-10
6	5	10	147	1	relu	0	-	Adadelata	0.001277387	0.009563851
7	5	2	152	1	selu	0	-	Adadelata	0.001005118	0.003136986
8	5	2	184	1	selu	0	-	Adadelata	0.000746804	0.000466103
9	5	2	174	1	selu	0	-	Adadelata	0.000840144	0.000607635

Following which, the best model from each algorithm, for each objective, (3 algorithms \times 3 objectives = 9 models total) was then chosen to generate the suggested catalysts to minimise either η_{OER} , η_{HER} or $Y_{bifunc.}$. To generate suggestions for ratio of the metal precursors to be used, the objectives were set to minimise equations (1), (2) and (3) separately. For each objective, the best model was wrapped into objective functions to be minimised, and the *gp_minimize* function implemented in *scikit-optimize* was used. The precursor composition was allowed to be freely varied from 0 to 1, with the constraint that the fractions of the five precursors must sum to 1. Sets of various metal precursor ratios were suggested, with the suggestions denoted as X^* . As an example, the sets of X^* generated from the NN models minimising η_{OER} , η_{HER} and $Y_{bifunc.}$ are shown in **Figure S2.5**, sorted in descending order of their model-predicted performance.

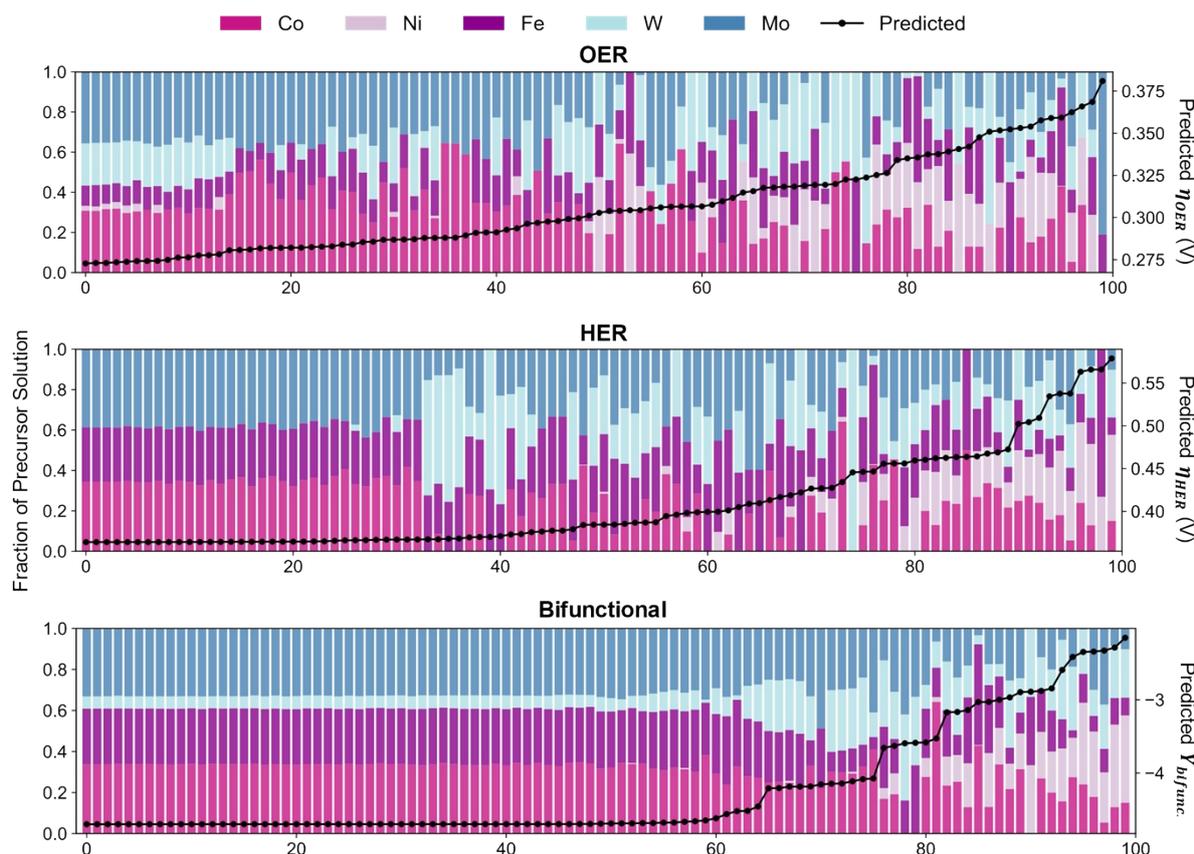


Figure S2.5. Suggested ratios of precursor solutions to minimise η_{OER} , η_{HER} and $Y_{bifunc.}$ from the Neural Network models, accompanied by their predicted values.

Two X^* for each of the f_{OER} and f_{HER} as well as four X^* for bifunctional catalysts $f_{bifunc.}$ were then synthesised and tested (**Figure S3.1**), totalling to an additional 24 data points ((2+2+4) for 3 algorithms). The selection of X^* was mainly based on the best predicted performance, though X^* that are too similar to the other suggestions or existing datapoints were skipped to ensure good variation in the datapoints. **Figure S2.6** reflects the true vs. predicted values for the GB and GPR models for both the initial data set as well as the model-suggested compositions. The plot for NN can be found in the main text, **Figure 2b**.

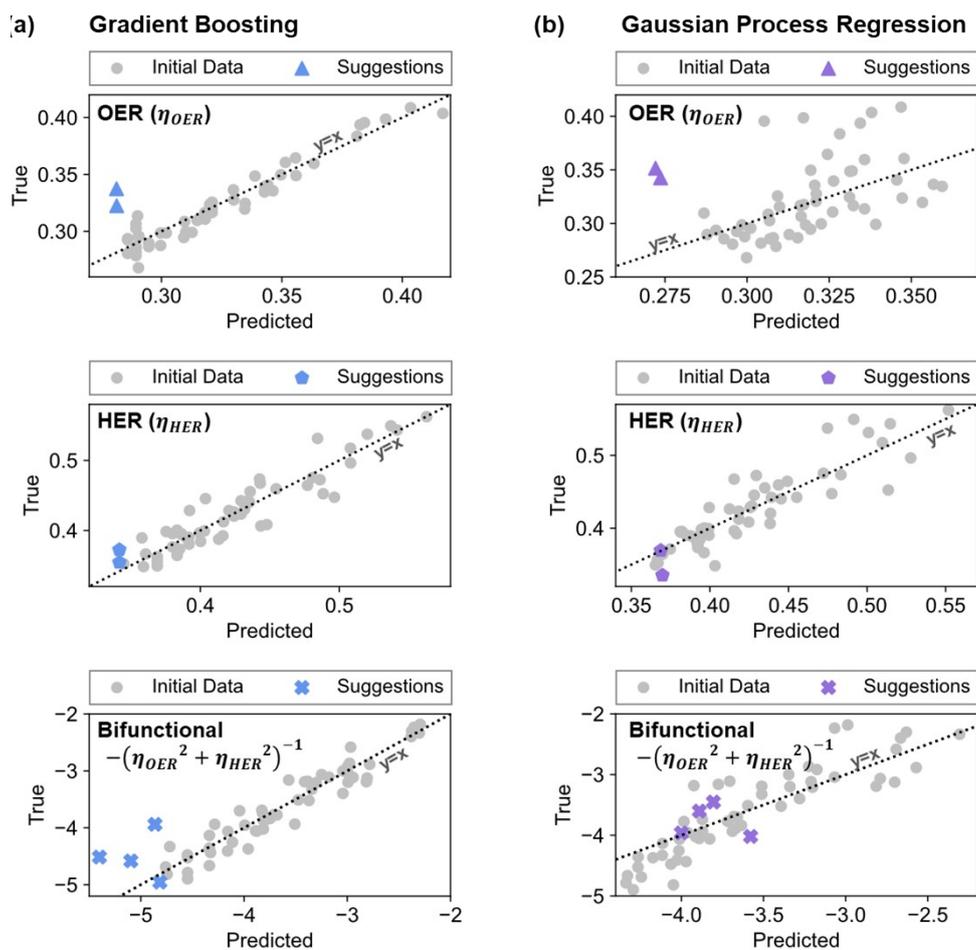


Figure S2.6. True against predicted values of the (a) Gradient Boosting and (b) Gaussian Process Regression models used for η_{OER} , η_{HER} and $Y_{bifunc.}$, plotted in grey. The suggestions from each of the models (2 each for the η_{OER} and η_{HER} models, alongside 4 for the $Y_{bifunc.}$ models) were synthesised and tested. Their corresponding true vs. predicted values are reflected in the plots with the coloured points.

The 24 additional data points acquired from the three different algorithms were then added back to the dataset pool and the process of modelling, hyperparameter tuning, and generating suggested compositions were then repeated for subsequent iterations using NN, since NN demonstrated the best performance of the three algorithms, as covered in the main text.

S2.2 Uniqueness and Feature Importance

The uniqueness of the suggested composition with respect to the initial 53 datapoints may be visualised by projecting the five-input composition to two transformed axis using Uniform Manifold Approximation and Projection (UMAP).² The initial sampling dataset (represented by the round symbols in **Fig S2.6**) are well-distributed upon projection to the UMAP's hyperspace. A general trend is observed where η_{OER} , η_{HER} and $Y_{bifunc.}$ decreases in the direction of increasing UMAP-0 axis and decreasing UMAP-1 axis (i.e., towards the 4th quadrant, along the arrows). The ML suggestions lie *beyond* the boundaries of the initial dataset (as represented by the diamond symbols), with the higher-performing catalysts similarly clustered around the 4th quadrant.

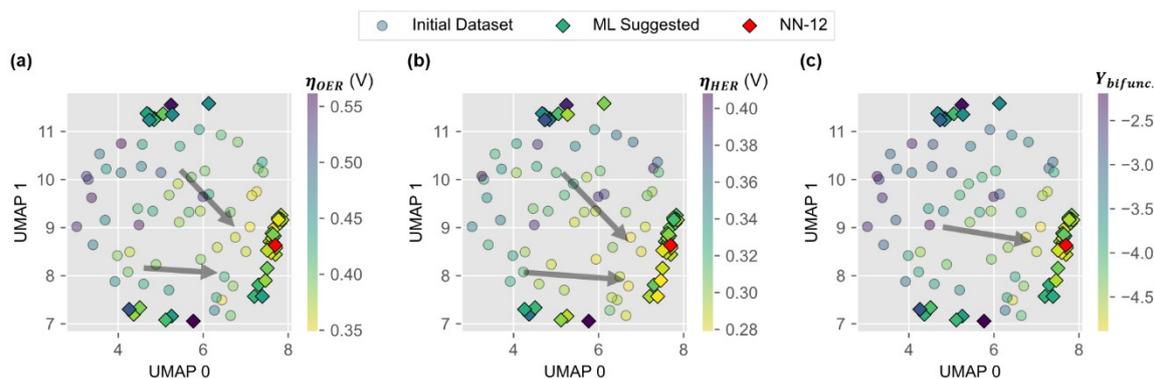


Figure S2.7. UMAP comparing the initial 53 datapoints, plotted with circular markers, with the ML suggested points, plotted with diamond markers for (a) η_{OER} , (b) η_{HER} and (c) $Y_{bifunc.}$. The best performing catalyst, NN-12 is highlighted in red. The ML suggestions mainly lie beyond the UMAP boundaries of the initial dataset, and the performance of the catalysts seem to follow a general trend in the UMAP (generally improving towards the 4th quadrant, i.e., along the direction of the grey arrow).

We further delve into the influence of the five metal precursors is the most determining in the electrocatalytic activity by extracting the NN models' SHapley Additive exPlanations (SHAP) values,³ which have been plotted in bee-swarm style figures (**Figure S2.7**). SHAP deconstructs the machine learning predictions into a sum of contributions from each of the models' input variables. The x-axis represents the impact of the metal precursor on η_{OER} , η_{HER} and $Y_{bifunc.}$, with more negative values reflecting higher impact on minimising the objectives' values; the five compositions are represented categorically on the y-axis, ranked in order of feature importance; and each point is a single observation with the relative value of the metal ratio in the composition reflected by its colour.

Consistently across η_{OER} , η_{HER} and $Y_{bifunc.}$, lower values of Ni have negative SHAP values (i.e., points extending towards the negative x-axis are increasingly blue). This influence is exceptionally strong for $Y_{bifunc.}$, where Ni is also ranked as the most important feature. The opposite is true for Mo, where higher Mo values reflected negative SHAP values for all three outputs, alongside being the most important feature for η_{HER} . For η_{OER} , Co content was the most important feature, with higher Co content correlated to lower SHAP values. The observations from these SHAP values are consistent with our hypotheses drawn from experimental data. Specifically, the exclusion of Ni and higher amounts of Co and Mo is preferred to yield catalysts with poorer crystallinity and stacking, which is correlated with higher electrochemical performance herein, as discussed in **Section 4.2** below.

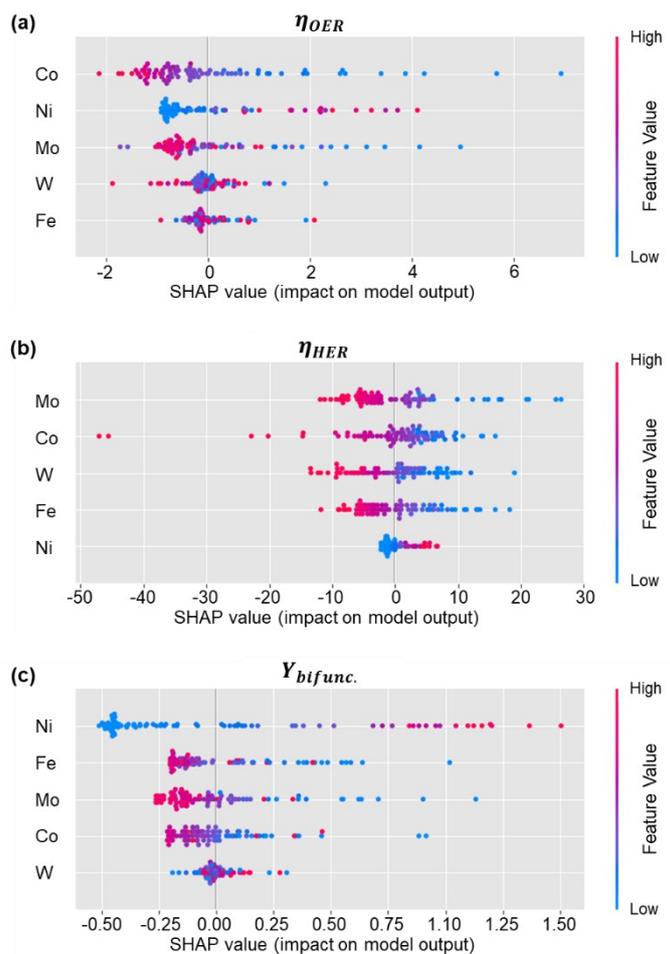


Figure S2.8. SHapley Additive exPlanations (SHAP) values of the models predicting for (a) η_{OER} , (b) η_{HER} and (c) $Y_{bifunc.}$.

We would like to highlight a caveat in this analysis. In calculating the SHAP values, KernelSHAP treats the input as independent variables by replacing a feature with a value from a random instance. This treatment violates the normalization constraint on the actual experiment, where the sum of the five metal precursors ratio is normalised to 1. Thus, we strongly recommend that the SHAP analysis presented here be interpreted only qualitatively.

Despite the lingering caveat, SHAP remains to be one of the most established and accepted methods for evaluating feature importance. The observations from these SHAP values are consistent with our hypotheses drawn from experimental data. Specifically, the exclusion of Ni and higher amounts of Co and Mo is preferred to yield catalysts with poorer crystallinity and stacking, which is correlated with higher electrochemical performance herein.

S2.3 Modelling Other Machine Learning Methods

Gradient boosting, neural network, and gaussian process regression was chosen as prior literature suggests that these algorithms may be more suited for problems requiring extrapolation beyond the training dataset,⁴ since the current objective of the optimisation experiment in this paper centres around optimising for precursor ratios to yield electrocatalysts which outperforms the initial dataset.

As an extension to the initial three machine learning methods investigated, we also demonstrate the use of other machine learning algorithms, which may be more applicable for other research questions. The additional models included are namely multivariate linear regression, polynomial regression (degree = 2), random forest, as well as K-nearest neighbours (2 nearest neighbours), are used as implemented in python using the *scikit_learn* library,⁵ without hyperparameter optimisation. Identical datasets as the main models covered in the manuscript, including the training and test set splits were adopted.

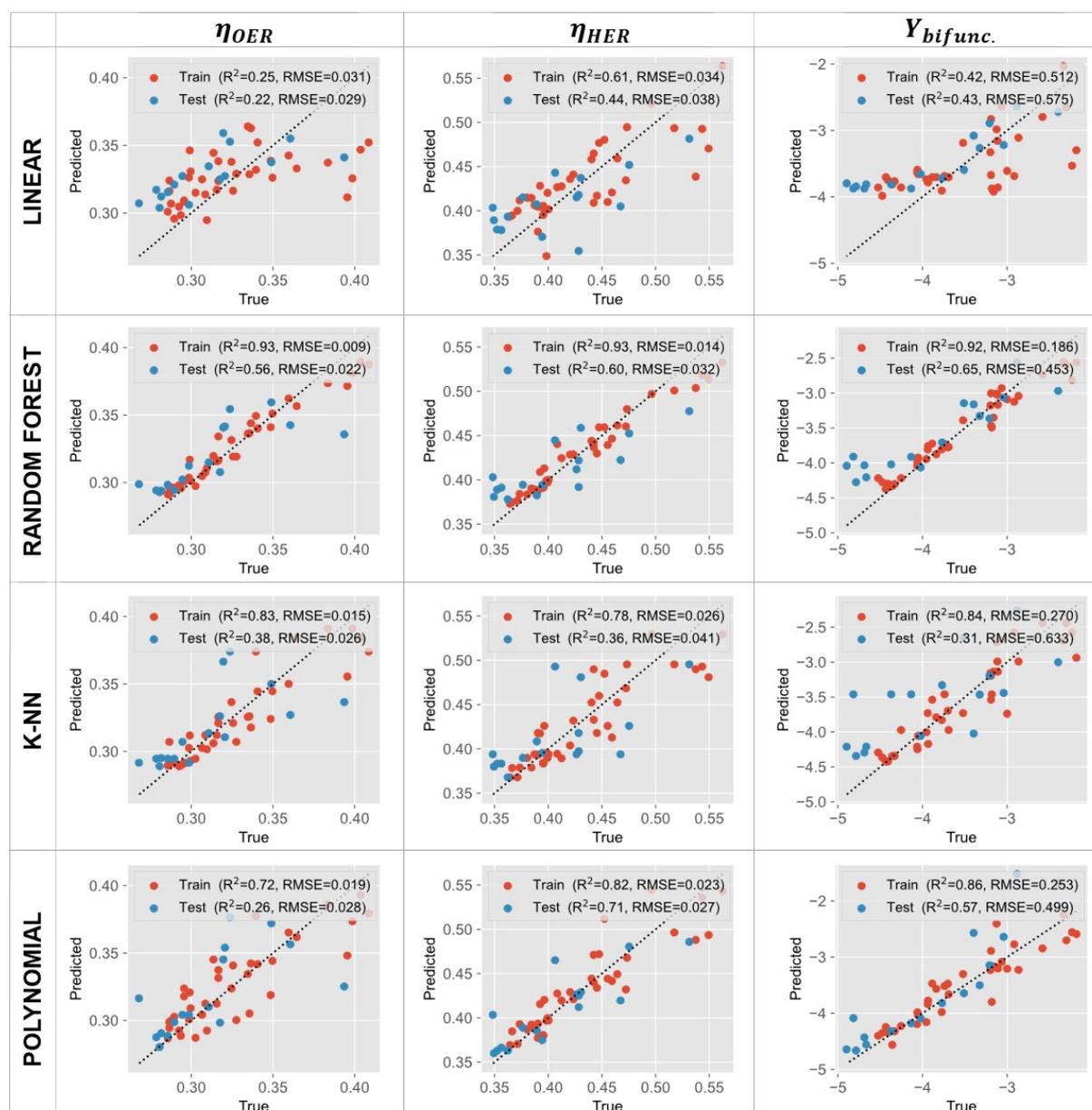


Figure S2.9. Additional multivariate linear regression, random forest, K-nearest neighbours, and polynomial (quadratic) models fitting for η_{OER} , η_{HER} , and $Y_{bifunc.}$

The models for the four additional algorithms for η_{OER} , η_{HER} , and Y_{bifunc} are plotted in **Figure S2.9**, alongside their R^2 and RMSE values. As mentioned in **Section S2.1**, we employed a 0.7:0.3 train:test split with a third of the test set comprising of the best 10% of the datapoints for the given objective to evaluate the models' ability to extrapolate beyond training dataset. In other words, a third of the test set has true values lower than the minimum true value of the train test.

With the exception of the polynomial regression, the limitations of the other regressors for extrapolation is apparent – none of the predicted values of the test set is less than the minimum true value of the train set. We illustrate this in the **Figure S2.10** below, where the red dotted lines represent the minimum true value of the train set. Excluding the polynomial regression model, none of the predicted values of the test sets falls below the red dotted line. The original neural network model is also included in **Figure S2.10** as a comparison.

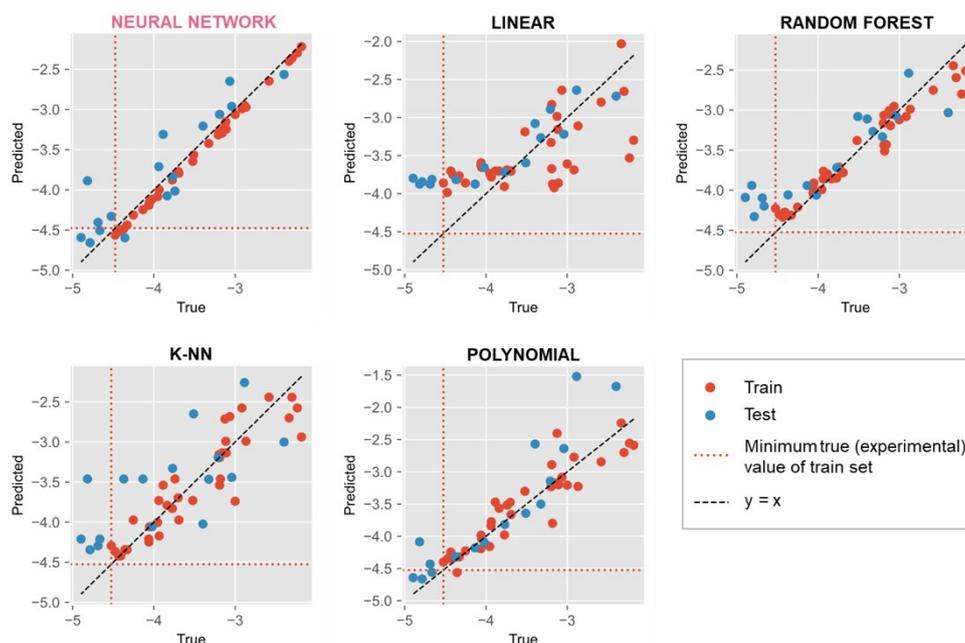


Figure S2.10. Comparison of neural network, linear regression, random forest regression, K-nearest neighbours, and polynomial (quadratic) regression.

Overall, a more comprehensive evaluation of the suitability of each algorithm for various types of datasets and experiment objectives is necessary and would be a valuable area for future investigation. The development of better extrapolative methods and techniques would also greatly facilitate more efficient and effective implementation of machine learning in experiments.

S3. Optimization Experiments

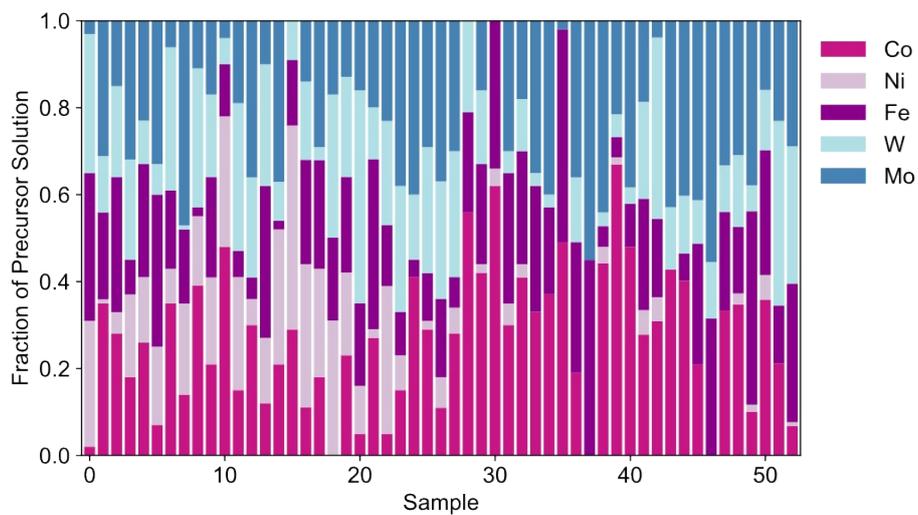


Figure S3.1. Metal precursor ratios of the initial dataset.

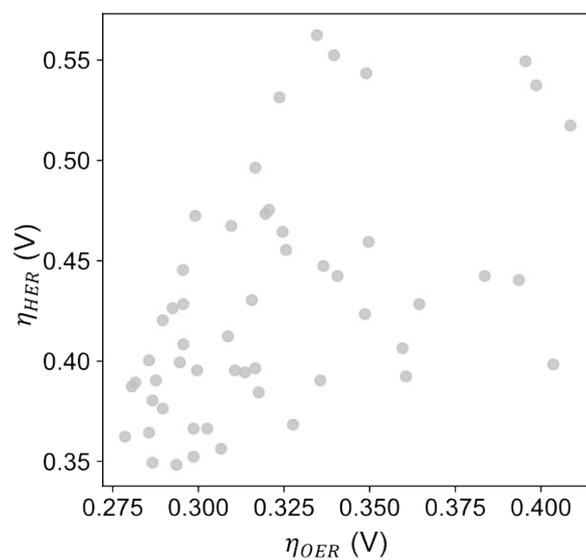


Figure S3.2. Scatter plot of η_{HER} against η_{OER} of the initial dataset.

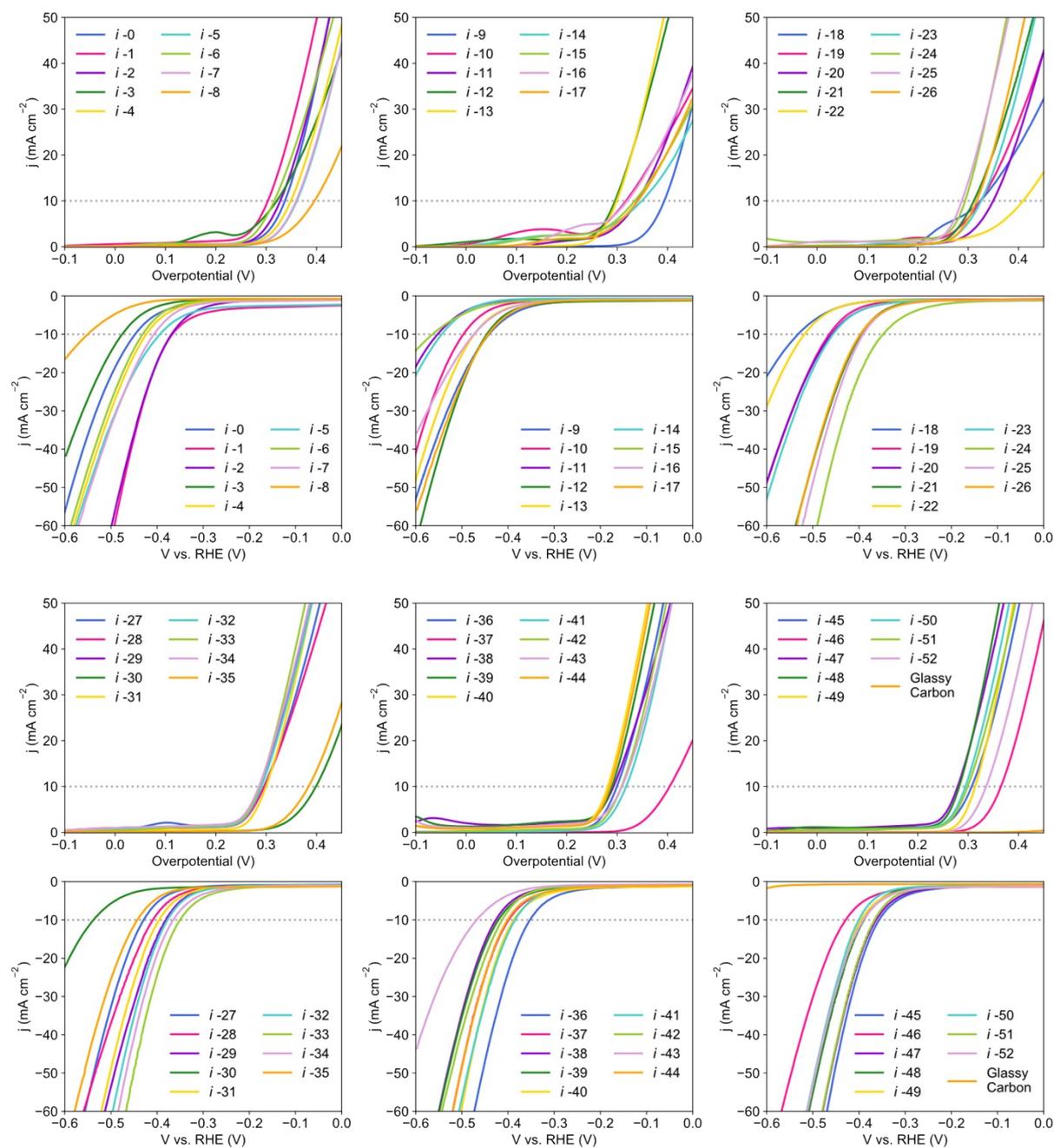


Figure S3.3. Linear sweep voltammograms of the initial dataset and blank glassy carbon substrate.

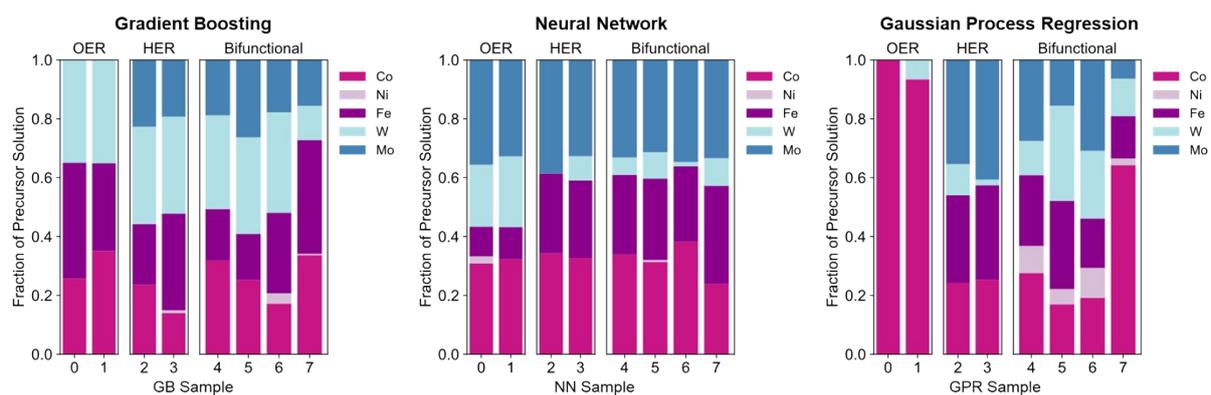


Figure S3.4. Metal precursor ratios suggested by each ML method, with three separate models for each method used to optimize OER, HER and catalysts for bifunctional water splitting.

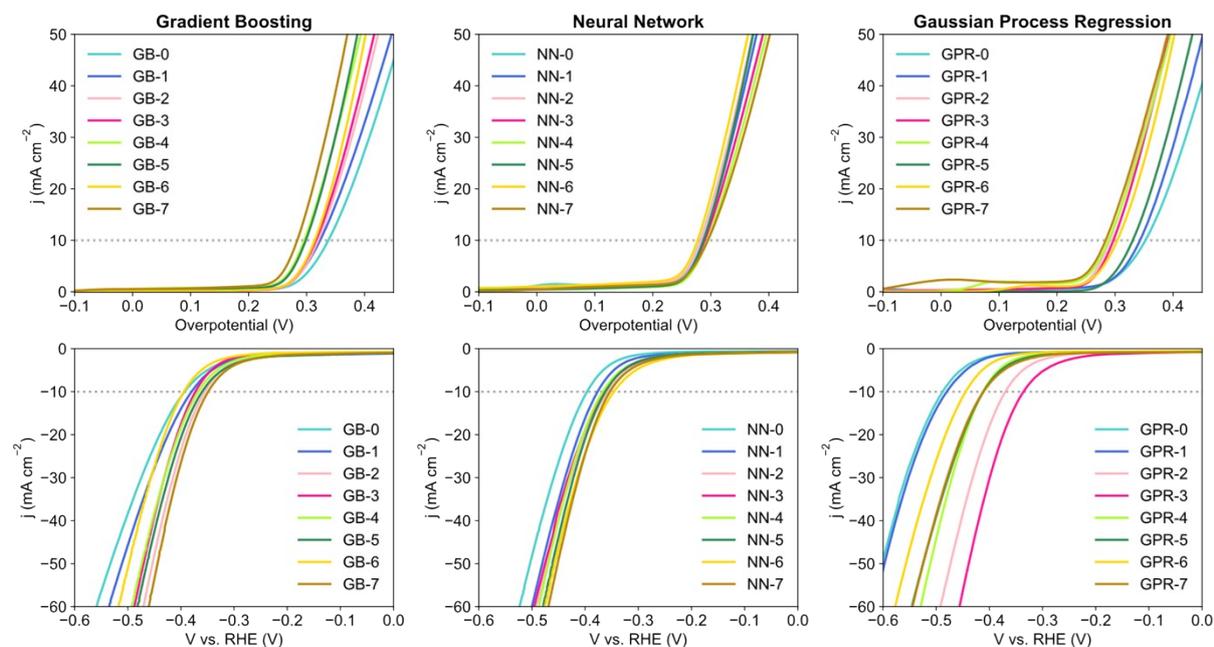


Figure S3.5. Linear sweep voltammograms of the compositions suggested by the three ML methods.

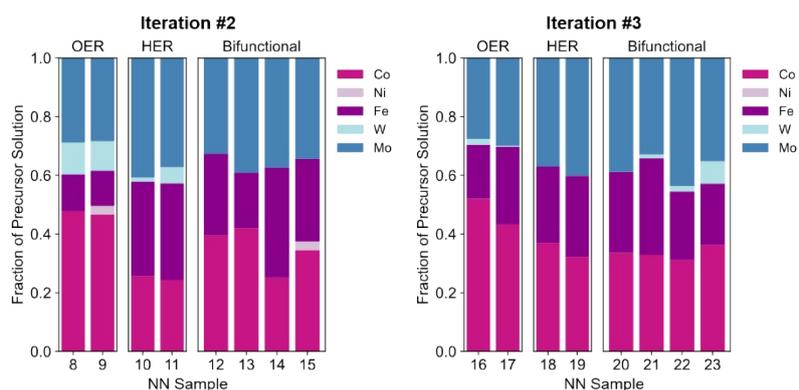


Figure S3.6. Metal precursor ratios suggested by the models from the 2nd and 3rd Neural Network iterations.

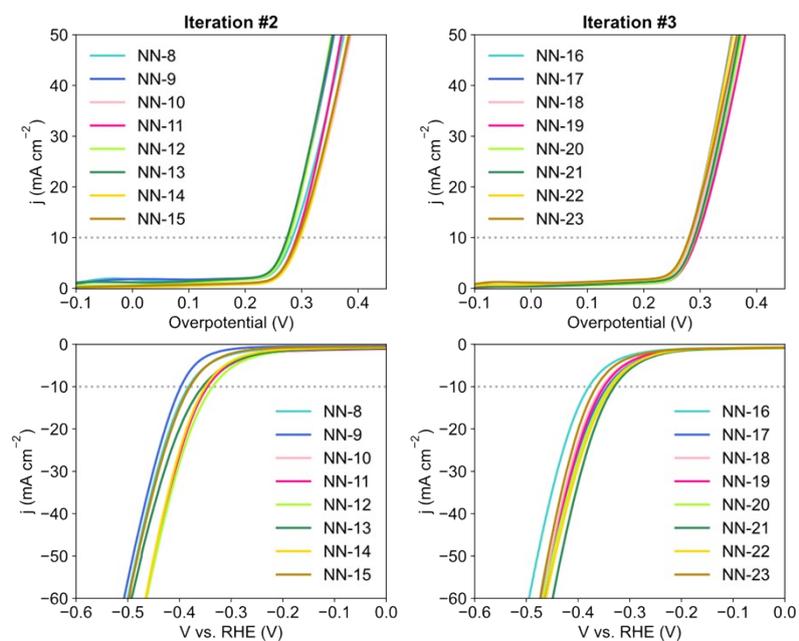


Figure S3.7. Linear sweep voltammograms of the samples from the 2nd and 3rd Neural Network Iterations.

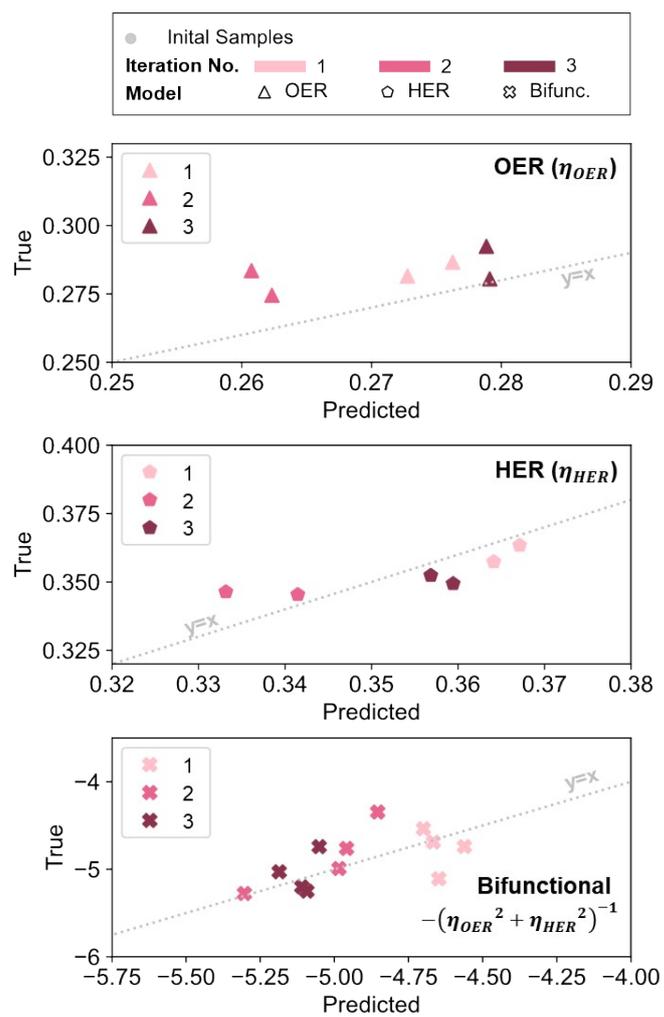


Figure S3.8. Parity plot of experimental values against model-predicted values for the three Neural Network iterations.

S4. Material Characterisations

S4.1 Characterisation Data

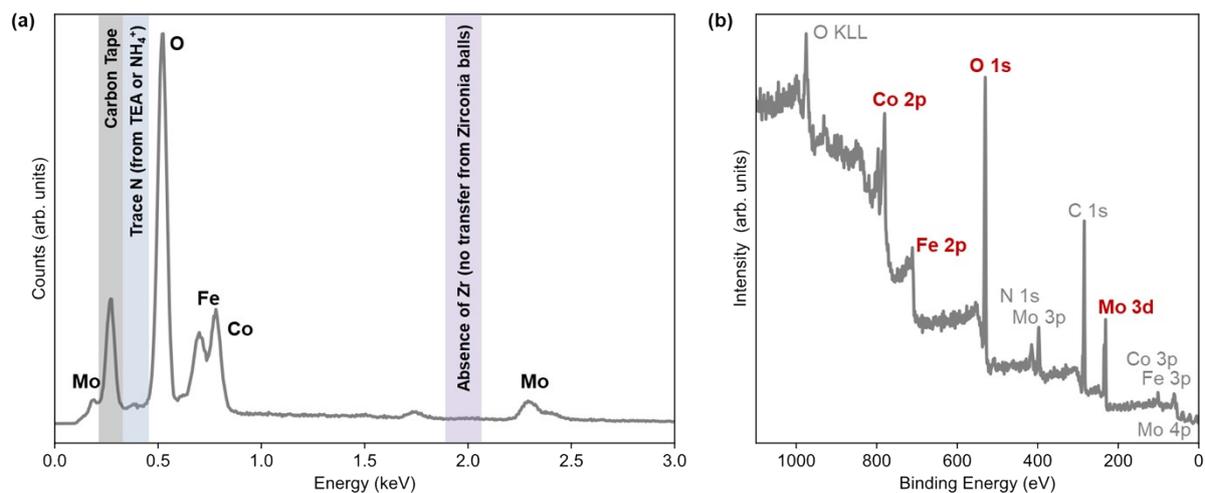


Figure S4.1. (a) EDX spectrum and (b) XPS survey spectrum reflecting the elemental composition of the best performing bifunctional catalyst, NN-12.

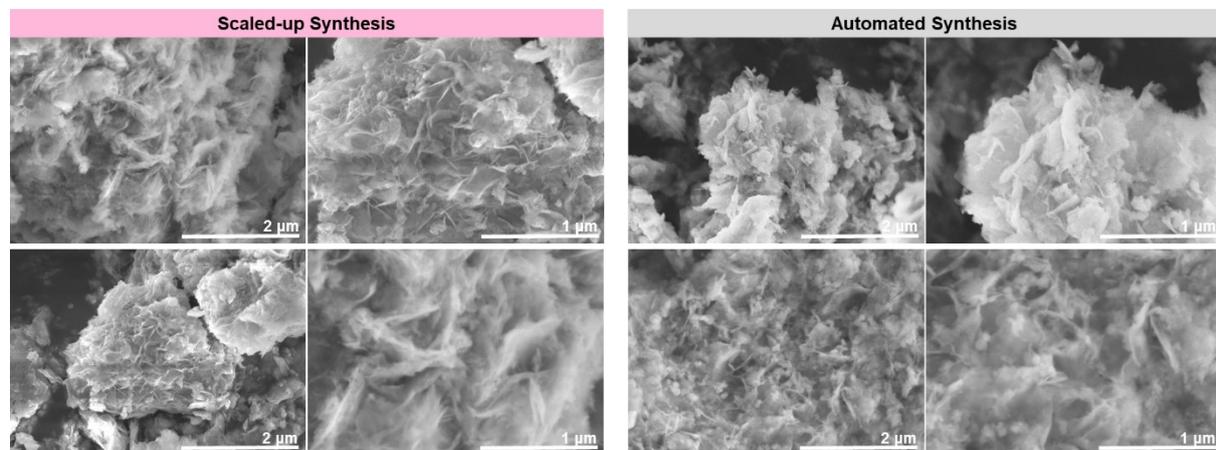


Figure S4.2. Additional SEM images reflecting the morphologies of the NN-12 catalysts.

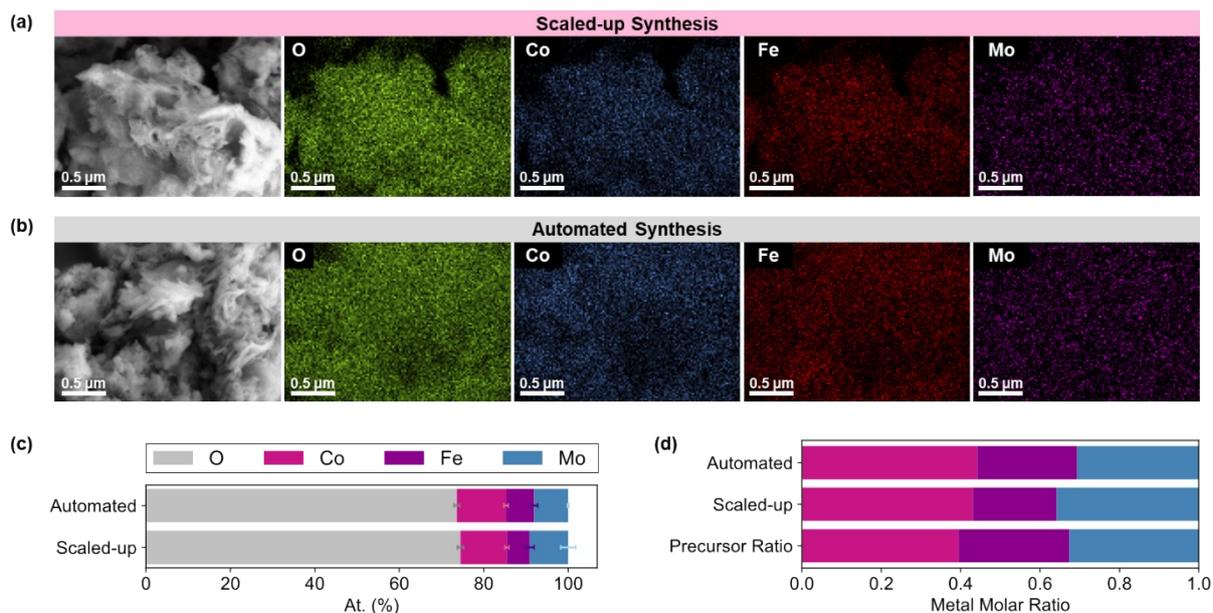


Figure S4.3. SEM images and corresponding EDX maps of NN-12 synthesised by the (a) automated workstation and (b) the scaled-up synthesis. (c) Comparison of the compositions of the two samples, averaged across the EDX data of three areas per sample and (d) comparison between the metal ratios of the two samples and the ratio of the synthesis precursors.

Table S4. Elemental compositions of NN-12 from EDX spectroscopy, averaged across three maps.

Synthesis Method	Elemental Composition (at. %)			
	Co	Fe	Mo	O
Automated	11.65	6.63	8.09	73.63
Scaled-up	10.98	5.38	9.13	74.51

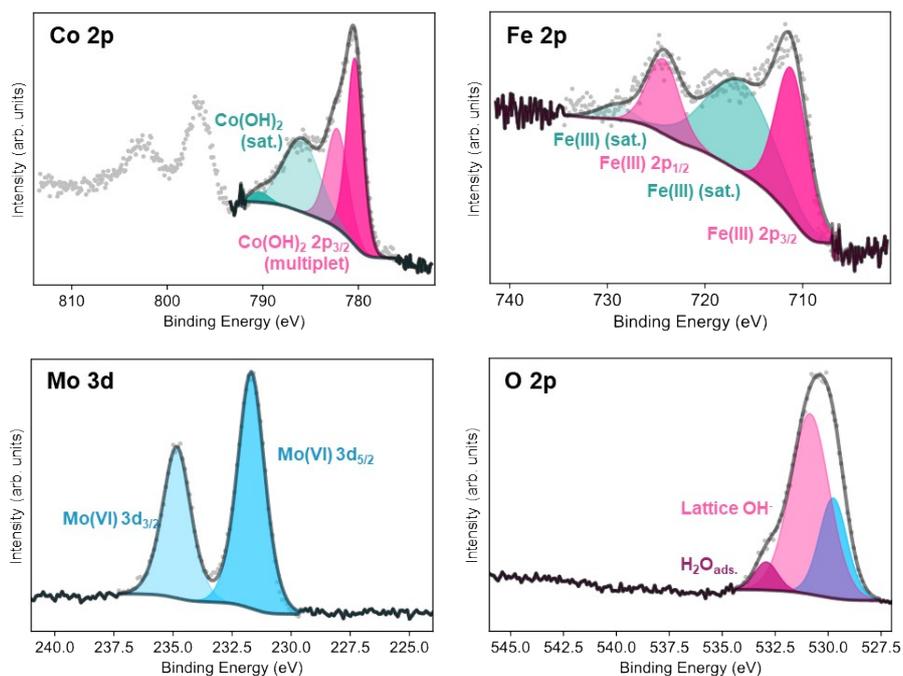


Figure S4.4. Deconvoluted high resolution XPS spectra of cobalt (Co) 2p, Iron (Fe) 2p, Molybdenum (Mo) 3d and Oxygen (O) 1s from the scaled-up synthesis of NN-12.

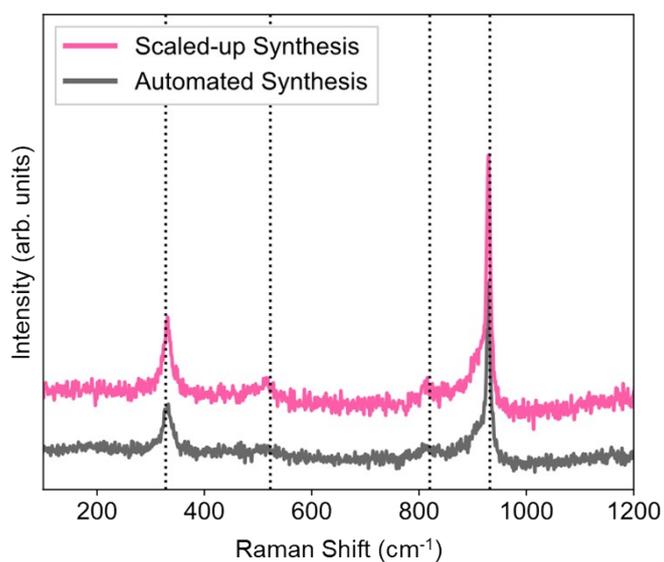


Figure S4.5. Raman spectra of NN-12 from automated and scaled up synthesis, collected with 785 nm laser excitation.

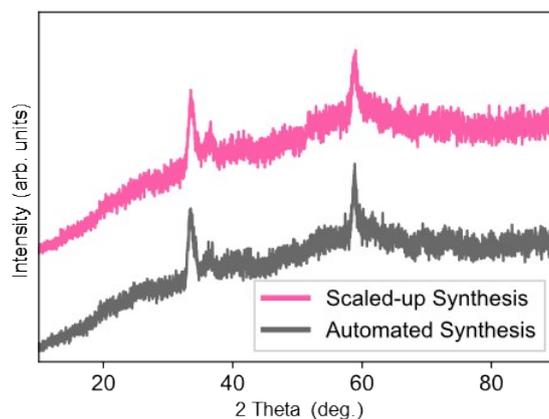


Figure S4.6. X-ray diffraction patterns of both NN-12 samples from the automated and scaled-up synthesis methods.

S4.2 Additional Discussion on the Effects of the Metal Precursors

To better rationalise the composition selected by the ML models as well as to propose physical reasons for the catalysts' observed performance, additional samples with systematic change in precursor composition were synthesised and characterised. Details of these synthesis methods may be found in **Section S1.2** above.

The first set of samples involve four LDH compositions, namely NiFe 3:1 as per the original literature from which the synthesis procedure was adapted,¹ NiFe 4:3, as well as CoFe 4:3. The 4:3 ratio of Co:Fe is close to that of our best performing catalyst in the manuscript (labelled NN-12 in the main text). The X-ray diffraction patterns of the samples are presented in **Figure S4.7a**. Here, the relative intensity of the (00 l) reflexion is indicative of the long-range order along the c -axis, reflecting the interlayer stacking behaviour of the catalysts. Comparing Ni₃Fe₁ and Ni₄Fe₃, it is observed that increasing the ratio of Fe decreases the crystallinity and stacking order of the LDH structure, arising from the increased substitution of Fe³⁺ into the Ni(OH)₂ layers. Following which, replacing Ni with Co during the synthesis further decreases the stacking order in Co₄Fe₃.

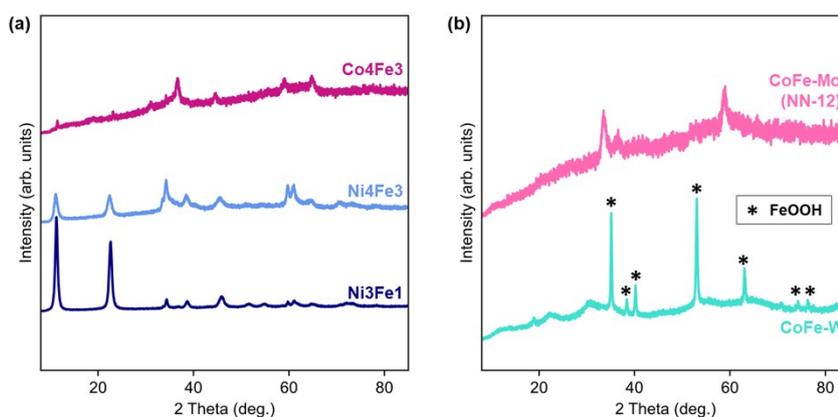


Figure S4.7. XRD patterns of synthesised samples (a) with varying ratios of Ni:Co:Fe – namely NiFe 3:1, NiFe 4:3, NiCoFe 2:1:1 and CoFe 4:3, and (b) with either molybdate or tungstate precursor for CoFe.

The second set of LDH to be compared would be our best performing catalyst, NN-12, and an additional sample synthesised with W precursor instead of Mo (**Figure S4.7b**). The incorporation of the tungstate precursor led to the precipitation of ferrihydrite ($\delta\text{-Fe}^{3+}\text{O}(\text{OH})$), which is not observed when using the molybdate precursor. In addition, the molybdate-incorporated catalyst also possesses poorer crystallinity than with tungstate.

Drawing from the observations above, we attribute the trend of the machine learning models to favour the exclusion of Ni and W to a few reasons. First, the Co-based catalysts are poorly stacked relative to the Ni-based catalysts. The poor stacking order is further exacerbated by the rather high Fe content as well as the incorporation of molybdate. Relative to well stacked layers, structures with poorer crystallinity and stacking has increased surface area, exposure of rich active sites as well as abundant defects, which may all contribute to enhanced electrochemical activity.^{6,7} This is consistent with prior reports of Co-modulated NiFe crystallinity where the sample with the poorest crystallinity demonstrated the highest OER activity.⁸ Second, the addition of W induced the phase separation of FeOOH, which may degrade the catalyst's electrochemical activity when present in high ratios as it has been found to be an electrical insulator whilst also being chemically unstable to dissolution under OER conditions in alkaline solutions.⁹

Ergo, whilst NiFe LDH is often reported to have slightly higher OER activity than CoFe LDH,¹⁰ we posit that with our current synthesis method, the actual catalytic activity may be affected more strongly by the catalysts' nanostructures. We acknowledge that the observations relating to Ni and Co, as well as Mo and W may be unique to the synthesis methods and other parameters adopted herein – including the amount of the triethanolamine and urea precursors as well as reaction temperature and pressure. Different synthesis methods and parameters may alter the crystallinity, stacking behaviour, and phases of the oxyhydroxides.^{8, 11, 12}

S5. Additional Electrochemical Testing

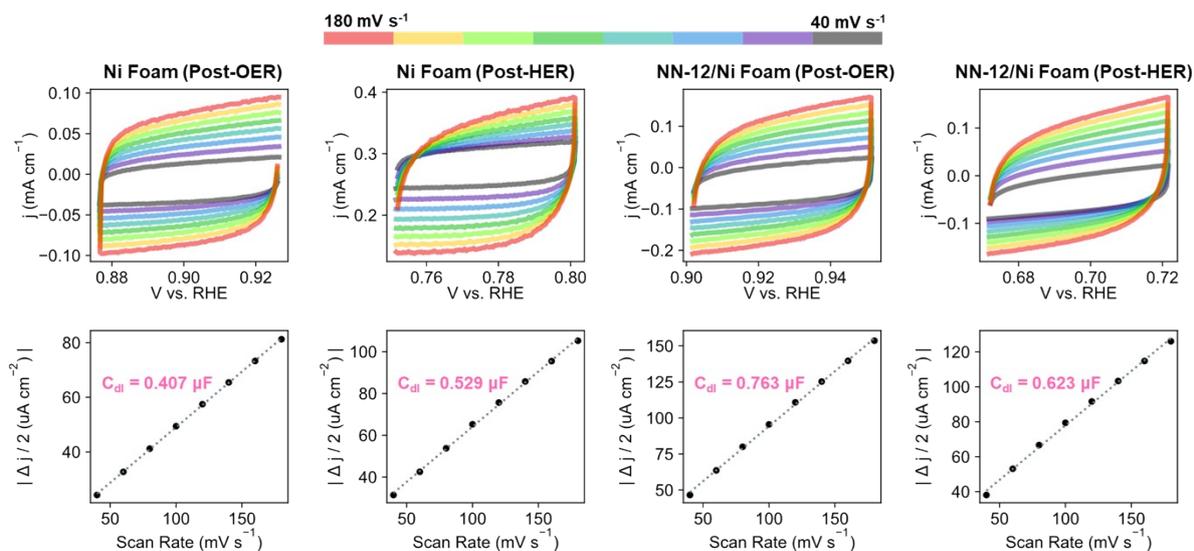


Figure S5.1. Electrochemically active surface area measurements, derived from variable scan rate cyclic voltammetry from 40 mV sec^{-1} to 180 mV sec^{-1} , measured in the capacitive regions of the electrodes.

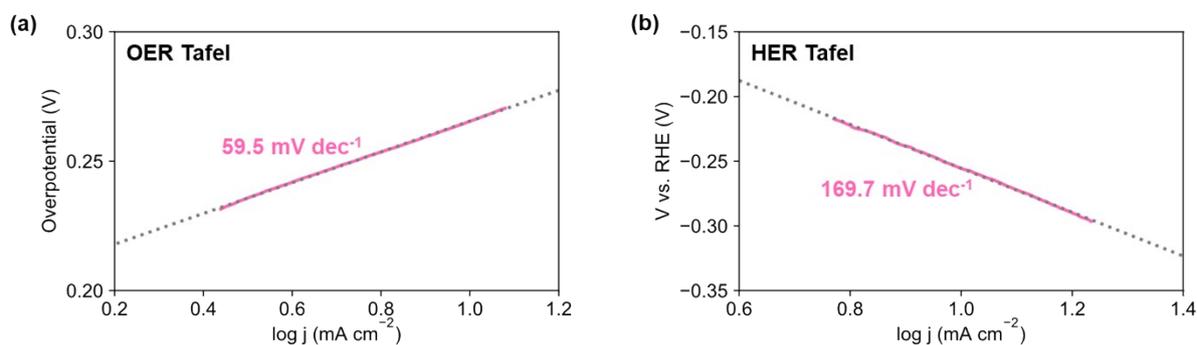


Figure S5.2. Tafel plots of NN-12 from the scaled-up synthesis deposited on Ni foam, derived from linear sweep voltammograms collected at 5 mV s^{-1} (main text **Figure 5b-d**) for (a) OER and (b) HER.

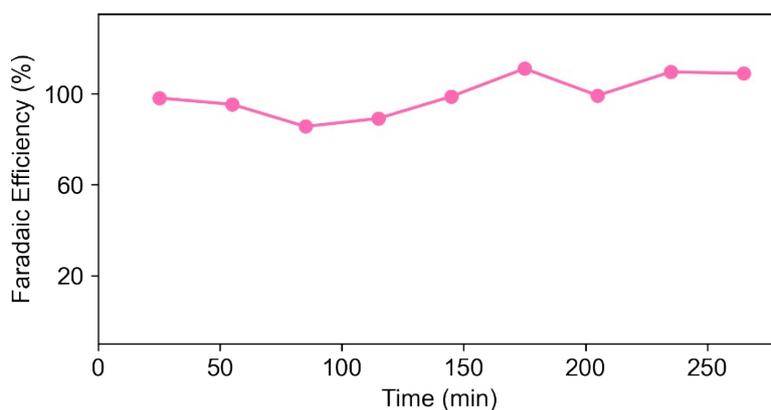


Figure S5.3. Plot of hydrogen Faradaic efficiencies against time, quantified by gas chromatograph (Section S1.4) at 10 mA cm^{-2} . Variations in the measured Faradaic efficiency values are expected due to the dynamic nature of gaseous product release from the electrode causing fluctuation in pressure/flow rate.

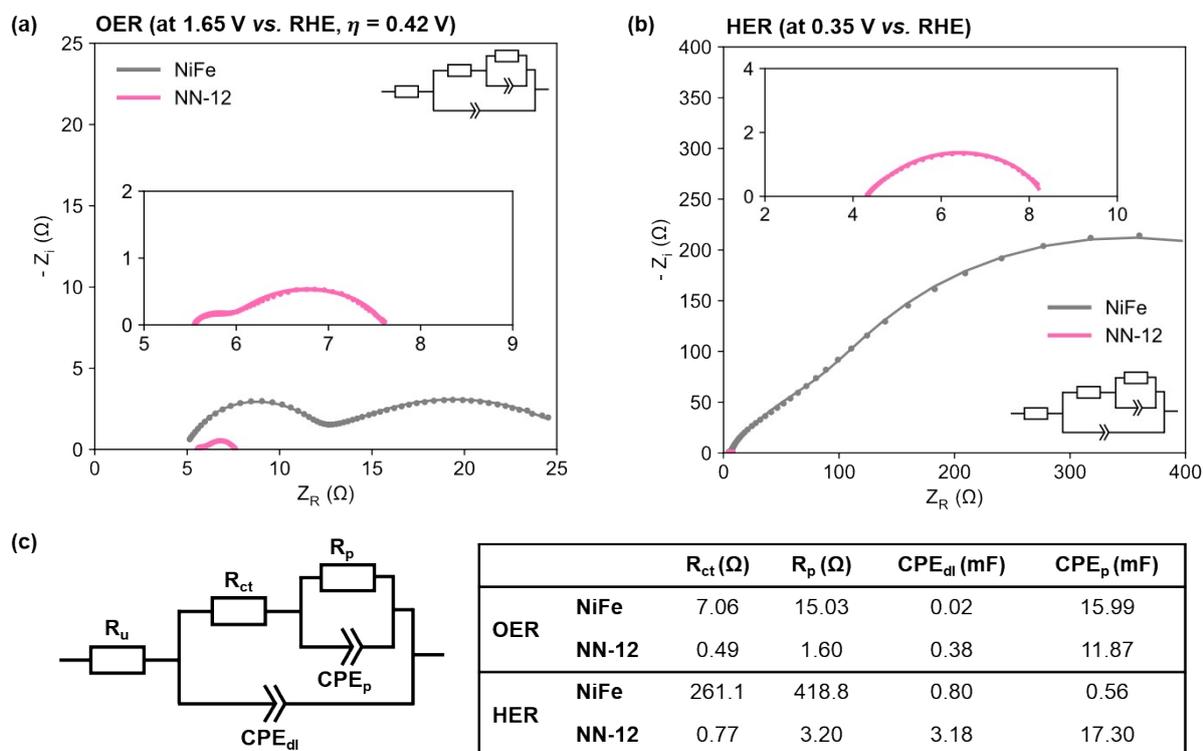


Figure S5.4. Nyquist plots from electrochemical impedance spectroscopy for (a) OER and (b) HER for NN-12 benchmarked against NiFe LDH synthesised as per literature.¹ (c) Armstrong-Henderson equivalent circuit used to fit the EIS spectra, alongside the fitted values of the circuit components.

Potentiostatic electrochemical impedance spectroscopy (EIS) was conducted at low overpotentials to minimise disruptions to the electrode interface arising from vigorous bubbling from gas evolution. Constant phase elements (CPE) were used in place of ideal capacitors to reflect the frequency dispersion of the capacitance at the electrode/solution interface. Armstrong-Henderson equivalent circuit was used to fit the obtained spectra, with R_u representing the uncompensated solution resistance; R_{ct} representing

the charge transfer resistance; CPE_{dl} representing the double-layer capacitance; while R_p and CPE_p represents the parameters associated with the pseudo-resistance and pseudo-capacitance related to the adsorption, desorption, and mass transfer limitations of the adsorbed species.^{13,14} Relative to NiFe LDH, the NN-12 catalyst demonstrates decreased charge-transfer resistance for both OER and HER at the low overpotentials applied. Given that R_p is related to the rate of surface intermediates formed,¹⁴ the lower R_p values reflect greater ease of intermediate formation for both OER and HER on NN-12 relative to NiFe LDH. Overall, these findings are consistent with the linear sweep voltammograms of the two materials (main text **Figure 3b-c**), with the significantly higher catalytic activity of NN-12, particularly for HER.

Table S5. Catalyst performance of NN-12 in comparison to prior literature tested in 1 M KOH.

Catalyst	η_{OER} (at j mA cm ⁻²) (mV)	η_{HER} (at j mA cm ⁻²) (mV)	Overall Water Splitting Potential (V)	Stability (h)	Ref.
OER Catalysts					
Ultrathin NiFe LDH nanosheets with MoO ₄ ²⁻ intercalation	280 (10 mA cm ⁻²)	-	-	20	15
NiFe-WO ₄ -LDH	290 (10 mA cm ⁻²)	-	-	20	16
Nitrogen-incorporated CoNiFe LDH	318 (10 mA cm ⁻²)	-	-	12	17
Bifunctional Catalysts					
Pd ²⁺ -doped NiFe-LDH	254 (30 mA cm ⁻²)	316 (10 mA cm ⁻²)	1.74 (10 mA cm ⁻²)	24	18
CoFe-LDH@NiFe-LDH core-shell structures	190 (10 mA cm ⁻²)	240 (10 mA cm ⁻²)	1.59 (10 mA cm ⁻²)	30	19
Alkali-treated NiCo-LDH	317 (10 mA cm ⁻²)	180 (10 mA cm ⁻²)	1.89 (10 mA cm ⁻²)	20	20
Hierarchical CoFe LDH@g-C ₃ N ₄	275 (10 mA cm ⁻²)	417 (10 mA cm ⁻²)	1.55 (10 mA cm ⁻²)	12	21
ML-optimized molybdate-intercalated CoFe LDH	266 (10 mA cm ⁻²)	272 (10 mA cm ⁻²)	1.78 (10 mA cm ⁻²) 2.04 (50 mA cm ⁻²)	50	This work

References

- 1 S. Jaśkanić, C. Hobbs, A. Seral-Ascaso, J. Coelho, M. P. Browne, D. Tyndall, T. Sasaki and V. Nicolosi, *Sci. Rep.*, 2018, **8**, 4179.
- 2 L. McInnes, J. Healy and J. Melville, *ArXiv e-prints*, 2020, arXiv:1802.03426v03423 [stat.ML].
- 3 S. M. Lundberg and S.-I. Lee, in *31st Conference on Neural Information Processing Systems (NIPS)*, eds. U. v. Luxburg, I. Guyon, S. Bengio, H. Wallach and R. Fergus, Curran Associates Inc., Long Beach, CA, USA, 2017, pp. 4768–4777.
- 4 Y.-F. Lim, C. K. Ng, U. S. Vaiteswar and K. Hippalgaonkar, *Adv. Intell. Syst.*, 2021, **3**, 2100101.
- 5 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 6 P. M. Bodhankar, P. B. Sarawade, G. Singh, A. Vinu and D. S. Dhawale, *J. Mater. Chem. A*, 2021, **9**, 3180-3208.
- 7 F. Song and X. Hu, *Nat. Commun.*, 2014, **5**, 4477.
- 8 Y. Yang, S. Wei, Y. Li, D. Guo, H. Liu and L. Liu, *Appl. Catal., B.*, 2022, **314**, 121491.
- 9 M. S. Burke, M. G. Kast, L. Trotochaud, A. M. Smith and S. W. Boettcher, *J. Am. Chem. Soc.*, 2015, **137**, 3638-3648.
- 10 F. Dionigi, Z. Zeng, I. Sinev, T. Merzdorf, S. Deshpande, M. B. Lopez, S. Kunze, I. Zegkinoglou, H. Sarodnik, D. Fan, A. Bergmann, J. Drnec, J. F. D. Araujo, M. Gliech, D. Teschner, J. Zhu, W.-X. Li, J. Greeley, B. R. Cuenya and P. Strasser, *Nat. Commun.*, 2020, **11**, 2522.
- 11 L. Trotochaud, J. K. Ranney, K. N. Williams and S. W. Boettcher, *J. Am. Chem. Soc.*, 2012, **134**, 17253-17261.
- 12 J. Wu, T. Yang, R. Fu, M. Zhou, L. Xia, Z. Wang and Y. Zhao, *Adv. Funct. Mater.*, 2023.
- 13 A. Alobaid, C. Wang and R. A. Adomaitis, *J. Electrochem. Soc.*, 2018, **165**, J3395.
- 14 L. Negahdar, F. Zeng, S. Palkovits, C. Broicher and R. Palkovits, *ChemElectroChem*, 2019, **6**, 5588-5595.
- 15 N. Han, F. Zhao and Y. Li, *J. Mater. Chem. A*, 2015, **3**, 16348-16353.
- 16 X. Xue, F. Yu, B. Peng, G. Wang, Y. Lv, L. Chen, Y. Yao, B. Dai, Y. Shi and X. Guo, *Sustain. Energy Fuels*, 2019, **3**, 237-244.
- 17 A. T. N. Nguyen, M. Kim and J. H. Shim, *RSC Adv.*, 2022, **12**, 12891-12901.
- 18 D. Liu, J. Liu, B. Xue, J. Zhang, Z. Xu, L. Wang, X. Gao, F. Luo and F. Li, *ChemElectroChem*, 2023, **10**, e202201025.
- 19 R. Yang, Y. Zhou, Y. Xing, D. Li, D. Jiang, M. Chen, W. Shi and S. Yuan, *Appl. Catal., B.*, 2019, **253**, 131-139.
- 20 H. Yang, Z. Zhou, H. Yu, H. Wen, R. Yang, S. Peng, M. Sun and L. Yu, *J. Colloid Interface Sci.*, 2023, **636**, 11-20.
- 21 M. Arif, G. Yasin, M. Shakeel, M. A. Mushtaq, W. Ye, X. Fang, S. Ji and D. Yan, *Mater. Chem. Front.*, 2019, **3**, 520-531.