Supporting Information

Machine Learning-assisted Optimization Design for Enhanced Oxygen Evolution Reaction Based on Vanadium-doped Nickel-Cobalt Layered Double Hydroxides

Chandrasekaran Pitchai,[a#] Ting-Yu Lo,[a#] Hou-Chien Chang,[a] Hung-Chung Li,[be*] Ming-Der Yang [cde*] and Chih-Ming Chen [ade*]

[a] Department of Chemical Engineering, National Chung Hsing University, Taichung 402202, Taiwan

[b] Undergraduate Program of Intellectual Creativity Engineering, National Chung Hsing University, Taichung 402202, Taiwan

[c] Department of Civil Engineering, National Chung Hsing University, Taichung 402202, Taiwan

[d] Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, Taichung 402202, Taiwan

[e] Smart Sustainable New Agriculture Research Center (SMARTer), National Chung Hsing University, Taichung 402202, Taiwan

* Correspondence:

Hung-Chung Li, E-mail: hcli01@nchu.edu.tw, Tel: +886-4-22840430 ext.702

ORCID: https://orcid.org/0000-0003-4143-0050

Ming-Der Yang, E-mail: mdyang@nchu.edu.tw

ORCID: https://orcid.org/0000-0003-2904-5838

Chih-Ming Chen, E-mail: chencm@nchu.edu.tw, Tel: +886-4-22840510 ext.511

ORCID: https://orcid.org/0000-0003-4481-1094

[#] Equal contribution

S1. Machine Learning

The catalyst materials were prepared based on 35 randomly selected combinations of feature parameters within the ranges specified in Table S1, and the oxygen evolution reaction (OER) process was conducted to obtain the overpotential data for each combination as shown in Fig. S1 and Fig. S2(a). The ranges specified for the input features were determined based on the process parameters commonly used in the literature. For the electrolyte (KOH) concentration, 1 M was used commonly [1-3] while some reports used the values less than 1 M [4-6]. Therefore, the range of the electrolyte concentration was set from 0.1 M to 1 M to cover a wide range. The input features set for the catalyst synthesis include elemental composition, temperature, and urea concentration, all of which have specific ranges in the literature [7-10]. To explore the effect of the V doping, the composition ratio of Ni and Co was set as identical (0.35 – 0.45) to eliminate the interference caused by excessive addition of each transition metal. The function of urea is to offer the OH^- species through the hydrolysis reaction for the interaction with the metal cations (Ni^{2+} and Co^{2+}) to form the layered double hydroxides (LDHs) [11]. Therefore, an abundant amount of urea was used as listed in Table S1. The synthesis temperature was set from 150 ^oC to 200 ^oC based on the published reports [7-9].

The 35 experimental data were used as the dataset for machine learning (ML) model training. In Fig. S2(b), an analysis of the distribution of the input features revealed that the variation range of temperature was significantly larger than the other features, and the peak of the overall kernel density estimation (KDE) curve was skewed to the left, posing a challenge for model training. Therefore, data preprocessing was applied to eliminate the issue of data skewness by using a natural logarithm as equation (S1) to transform the temperature feature. Due to zero values inside the logarithm function, a value of 1 was specially added to prevent errors. After logarithmic transformation, the temperature range was adjusted from 150 °C – 200 °C to 5.017 – 5.303, which helped mitigate the skewness problem. Similarly, the concentration distribution of urea exhibited skewness, as shown in Fig. S2(c), with most data points concentrated at higher concentrations. Thus, a square root transformation as equation (S2) was applied to flatten the distribution.

Table S1. Feature variables used for catalyst preparation and OER process.

Feature variables	Unit	Range
Catalyst composition	Molar fraction	Ni: 0.35 – 0.45 Co: 0.35 – 0.45 V: 0.1 – 0.3
Urea	mM	5.83 – 7.43
Electrolyte (KOH)	M	0.1 – 1
Temperature	^{o}C	150 – 200

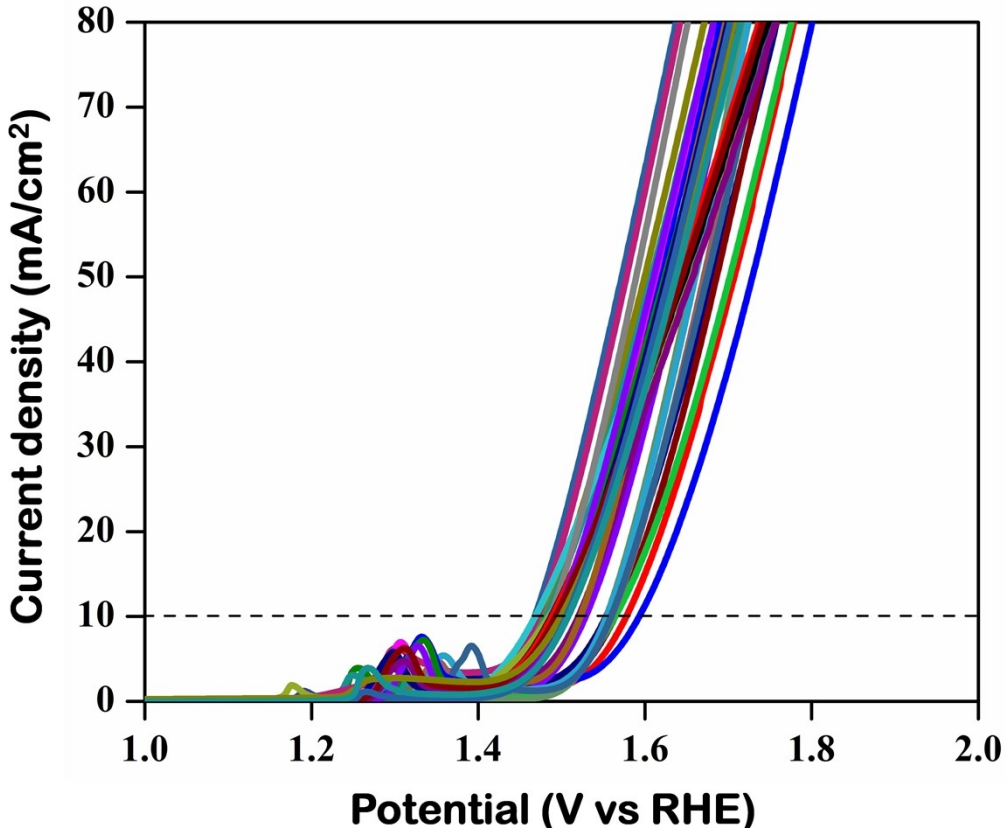

Fig. S1. LSV curves of 35 different NiCoV LDHs.

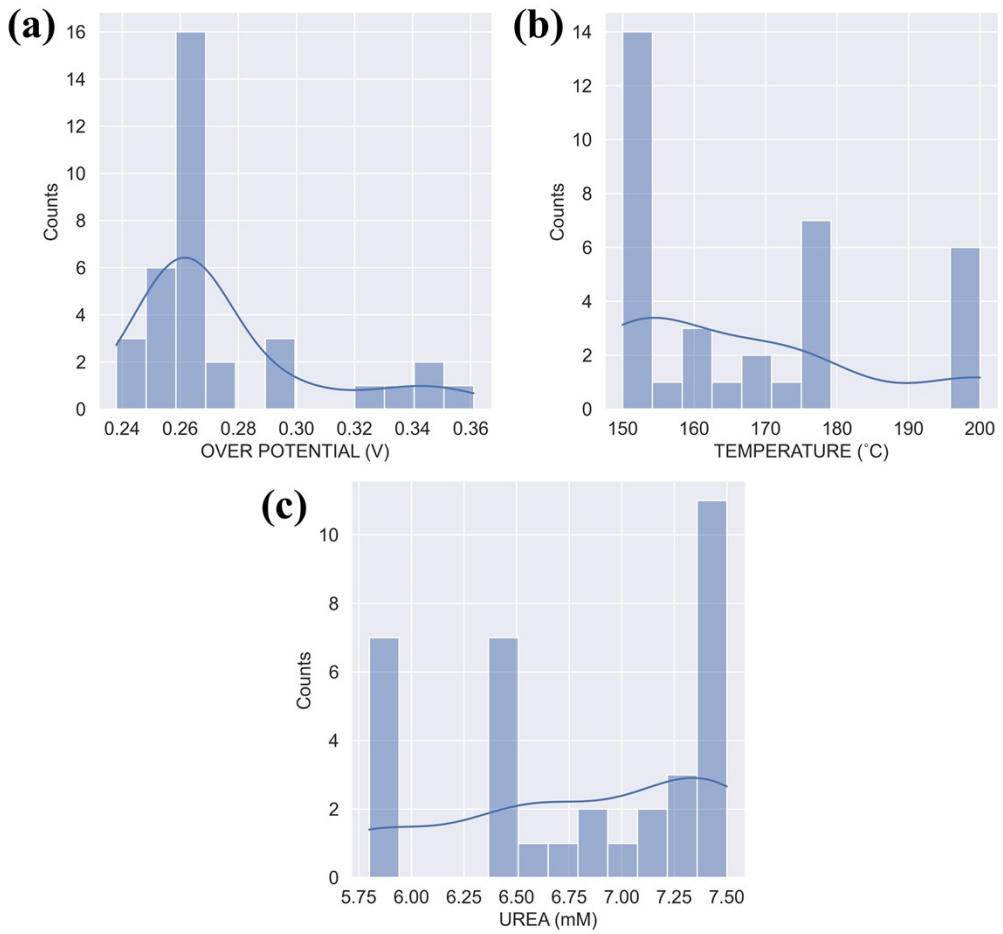

Fig. S2. The original distribution of the 35 data points: (a) overpotential, (b) synthesis temperature, (c) urea concentration.

$$X_{log} = \log_e (X_{original} + 1) \quad \text{(S1)}$$

$$X_{sqrt} = (X_{original})^{0.5} \quad \text{(S2)}$$

This study utilized the Pipeline tool from the *Scikit-learn* module to integrate data transformation and model construction steps, enabling a more efficient machine learning workflow. Within the Pipeline, the Yeo-Johnson transformation was first applied to make the data more normally distributed. Next, *PolynomialFeatures* was used to generate interaction terms between features, followed by min-max normalization, which compressed the data within the 0 and 1 range while preserving relative proportions and accelerating the training convergence. Finally, the transformed data was processed by a random forest regressor for feature selection, retaining the most influential interaction terms to train the polynomial regression model. A grid search method of hyperparameter tuning was employed to identify

the optimal hyperparameter combination from a predefined set. For polynomial features, the hyperparameter corresponded to the polynomial degree. Also, for feature selection using a random forest regressor, the hyperparameters to be optimized included the feature importance threshold, the number of trees in the random forest, and the maximum tree depth. Leave-one-out cross-validation (LOOCV) was then performed to determine the hyperparameter set that minimized the root mean square error (RMSE), establishing the model. Since this study had only 35 training samples, LOOCV was adopted as it provided a more reliable performance evaluation for small datasets.

Fig. S3 presents the results of the Yeo-Johnson transformation applied to various parameters. The results indicated that all data tended to form a normal distribution within ±3 after the transformation. This was because the transformation adjusted the mean to 0 and the standard deviation to 1, ensuring that most data points fell within this range. The transformed logarithmic temperature data did not exhibit a normal distribution due to the high redundancy of identical temperature values in the dataset of 35 training samples, which prevented the Yeo-Johnson transformation from effectively reshaping the distribution into a normal form.

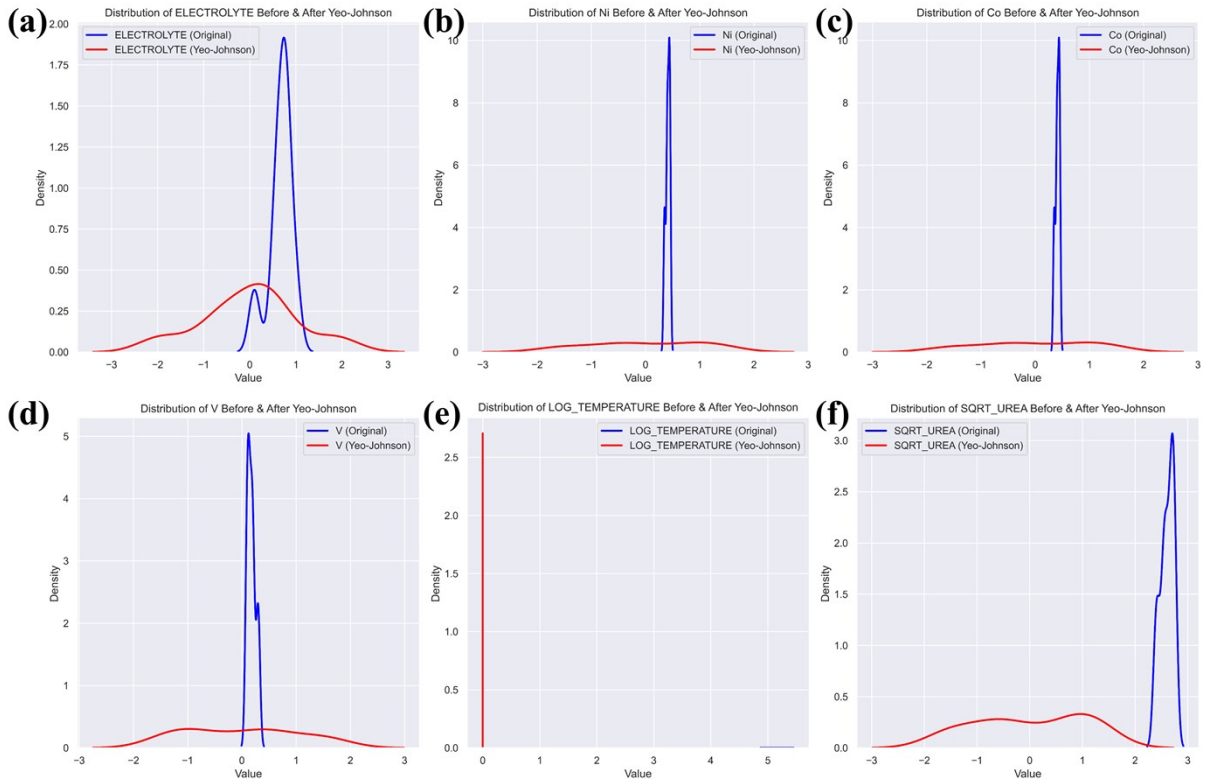

Fig. S3. KDE curves before and after the Yeo-Johnson transformation: (a) electrolyte concentration, (b) Ni molar fraction, (c) Co molar fraction, (d) V molar fraction,

(e) logarithmic temperature, (f) SQRT_Urea. (Blue and red represent the original and transformed data, respectively.)

After polynomial feature generation, different degrees of polynomial features were analyzed, as shown in Fig. S4(a), to evaluate the appropriate selection of the polynomial degree. According to the analysis, when the polynomial degree reached 3, multiple outliers (denoted by "o") appeared because the interaction terms in the third-degree polynomial created a more complex model, leading to overfitting. As a result, the model captured noise and anomalies from the training data, reducing its generalizability to new datasets and increasing the prediction error, causing extreme Mean Square Error (MSE) values. Therefore, this study focused on a first-degree polynomial or a second-degree polynomial. As shown in Fig. S4(b), the second-degree polynomial exhibited the lowest MSE, indicating the best model performance. Consequently, the second-degree polynomial was selected for further model training.

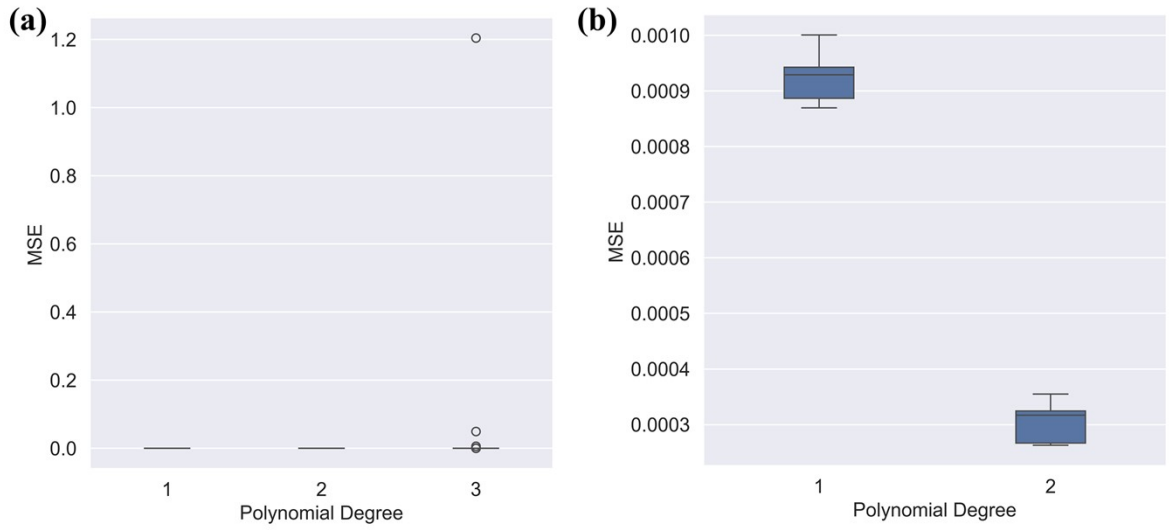

Fig. S4. Box plot analysis of different degrees of polynomial features: (a) a comparison of the 1st to 3rd-degree polynomials, (b) a magnified analysis of the 1st and 2nd-degree polynomials.

According to Fig. S5, during feature selection with the random forest regressor, whether using a first-degree polynomial or a second-degree polynomial, the values were concentrated around 7.0, which indicated that the number of features the model selected was the same for

both polynomial degrees, with seven features being chosen. Additionally, when using a second-degree polynomial (poly_degree = 2), the MSE values of the data points were lower, suggesting that the second-degree polynomial was more suitable than the first-degree polynomial. Therefore, the difference in model performance in this study primarily stemmed from the degree of the polynomial rather than variations in the number of features selected.

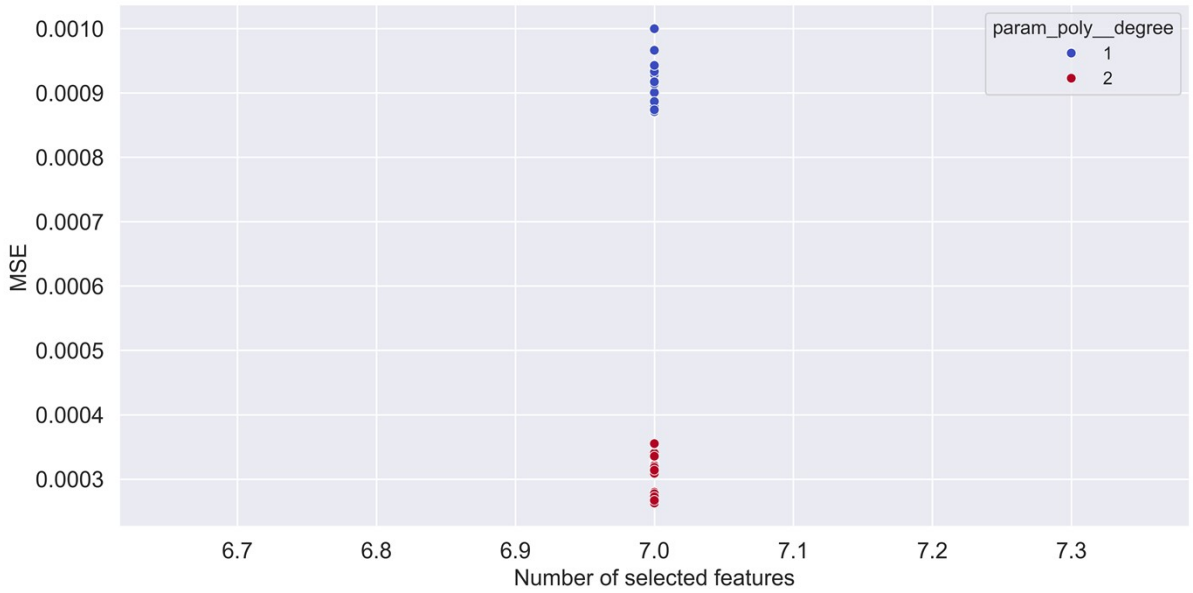

Fig. S5. The relationship between the number of selected features and MSE.

After performing LOOCV, the results calculated through KDE are shown in Fig. S6. The most frequent MSE value among the 35 data points was 0.00016. The lowest MSE was 0.00026 after averaging, with an RMSE of 0.016. The hyperparameters of the polynomial regression model are listed in Table S2.

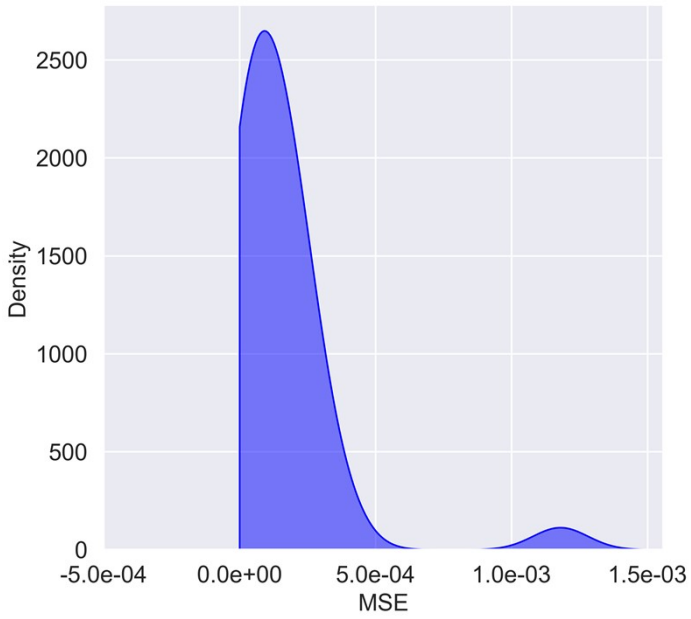

Fig. S6. KDE distribution plot of LOOCV.

Table S2. Hyperparameters of the polynomial regression model.

# Poly__degree	# Estimator_max_depth	# Estimator__n_estimators	# Threshold
2	None	300	0.01

S2. Experimental

S2.1. Material synthesis

Initially, nickel nitrate hexahydrate ($Ni(NO_3)_2{\cdot}6H_2O$), cobalt nitrate hexahydrate ($Co(NO_3)_2{\cdot}6H_2O$), vanadium chloride (VCl_3), and urea were dissolved in deionized water. The mixture was stirred continuously until a clear solution was achieved. Subsequently, 1 M solutions of sodium carbonate and sodium hydroxide were added dropwise to the stirring solution until the pH adjusted to 10. This prepared solution was then transferred into a 100 mL Teflon-lined stainless steel autoclave and subjected to hydrothermal treatment at 150 °C for 12 hours. Once the reaction was complete, the product was isolated by centrifugation and thoroughly washed several times with water and ethanol to eliminate residual impurities. The resulting composite was designated as NiCoV LDHs.

S2.2. Characterization of NiCoV LDHs

The morphology and structural arrangement of the synthesized NiCoV LDHs were characterized using field emission transmission electron microscopy (FE-TEM) and field emission scanning electron microscopy (FE-SEM), complemented by energy-dispersive X-ray spectroscopy (EDX) from the JEOL JEM-2010 (operating at 200 kV) and ZEISS ULTRA PLUS, respectively. The crystalline structure of the NiCoV LDHs was confirmed through X-ray diffraction (XRD) using a D8 SSS diffractometer (Bruker, Germany) with Cu Kα radiation (λ = 1.54 Å). The electronic states were further investigated via X-ray photoelectron spectroscopy (XPS) employing the ULVAC-PHI, PHI 5000 VersaProbe. Finally, the Brunauer–Emmett–Teller (BET) surface area was determined by nitrogen adsorption–desorption measurements using the ASAP 2020 instrument (Micromeritics, Norcross, GA) at 77.4 K.

S2.3. Electrochemical studies

Electrochemical evaluations were performed using a CHI6279E electrochemical workstation to assess the OER performance of the NiCoV LDHs catalyst. Techniques such as linear sweep voltammetry (LSV), Tafel analysis, chronopotentiometry, electrochemical impedance spectroscopy (EIS), and electrochemical active surface area (ECSA) measurements were systematically conducted. A conventional three-electrode configuration was employed, where the NiCoV LDHs, in its as-synthesized state, was uniformly coated onto a nickel foam (NF) substrate to serve as the working electrode. A Hg/HgO electrode functioned as the reference, and a platinum wire acted as the counter electrode.

To prepare the working electrode, a 1 × 1 cm² piece of nickel foam was initially cleaned by sonication in acetone followed by treatment in 3 M HCl for approximately 20 minutes each, ensuring the removal of impurities. The foam was then thoroughly rinsed with deionized water and ethanol and dried completely. Next, 10 mg of the synthesized NiCoV LDHs was mixed with 20 µL of a 5% Nafion solution in an ethanol/water mixture, and the resulting slurry was evenly drop-cast onto the nickel foam surface. The modified electrode was allowed to dry completely before testing. All electrochemical measurements were carried out in a 1 M KOH solution, providing a consistent alkaline environment for the OER experiments.

S2.4. Linear sweep voltammetry and Tafel slope measurements

For evaluating the OER performance of the NiCoV LDHs catalyst, LSV measurements were performed over a potential range of 0 V to 1 V at a scan rate of 10 mV/s. All measured potentials were converted to the reversible hydrogen electrode (RHE) scale using the following conversion:

$$\mathbf{E\ (vs.\ RHE) = E\ (vs.\ Hg/HgO) + 0.098\ V + 0.0591\ V \times pH} \qquad \mathbf{(S3)}$$

where E (vs. RHE) is the converted potential versus the reversible hydrogen electrode, E (vs. Hg/HgO) is the experimentally measured potential vs. the Hg/HgO reference electrode, 0.098 V is the standard potential of the Hg/HgO reference electrode vs. RHE at 25 °C in 1 M KOH, pH is the hydrogen ion concentration of the electrolyte, and 0.0591 V × pH is the Nernstian correction term for converting to the RHE scale.

The Tafel slope value was calculated from the equation (S4) where η is the overpotential, b and j are the Tafel slope and the current density, respectively.

$$\eta = a + b \log |j| \qquad \text{(S4)}$$

S2.5. Electrochemical impedance spectroscopy measurements

To probe the interfacial charge transfer resistance and capacitive behavior of the NiCoV LDHs catalyst, EIS measurements were carried out over a frequency range from 10^5 Hz to 0.1 Hz with an amplitude of 5 mV, all at open circuit potential (OCP). The resulting impedance spectra were analyzed by fitting them to an appropriate equivalent circuit model, thereby extracting critical parameters associated with the catalytic interface.

S2.6. Chronopotentiometry

The durability and long-term stability of the NiCoV LDHs catalyst were assessed through chronopotentiometry. A constant current density of 10 mA/cm² was applied continuously for 60 hours, ensuring that the catalyst's performance under prolonged OER conditions could be thoroughly evaluated.

S2.7. Calculation of electrochemical active surface area

Cyclic voltammetry (CV) was employed to calculate the electrochemical active surface area by measuring the double-layer capacitance (C_{dl}). CV curves were obtained from the non-faradic region at various scan rates ranging from 10 to 100 mV/s. The double-layer capacitance was determined by analyzing the slope of the average current density versus scan rate through linear fit values, and it was directly linked to the ECSA. Equation (S5) outlines the relationship between double-layer charging current density (j_{dl}), scan rate (ν), and C_{dl}.

$$j_{dl} = \nu C_{dl} \qquad \text{(S5)}$$

From equation (S6), the ECSA was calculated using C_{dl} and C_s, where C_s is the specific capacitance of the sample in identical electrolyte solution. ECSA normalization can be further calculated by equation (S7).

$$\text{ECSA} = C_{dl} / C_s \qquad \text{(S6)}$$

$$\text{Specific activity} = J_{ECSA} = J_{geo} / \text{ECSA} \qquad \text{(S7)}$$

S2.8. Calculation of turn over frequency (TOF)

TOF is simply defined as the number of moles of O_2 formed per unit time from a catalytic active site. The equation used to calculate the TOF value of a catalyst is provided below:

$$\mathbf{TOF = j \times N_A / F \times n \times \tau} \qquad \mathbf{(S8)}$$

Here, j is the current density at a certain overpotential, N_A is the Avogadro number, n is the number of electron transfer (n = 4), F is faraday constant and τ is the number active sites over the working electrode surface. The τ value can be calculated from the redox area curve, obtained from the CV at the highest possible scan rates.

S3. Results

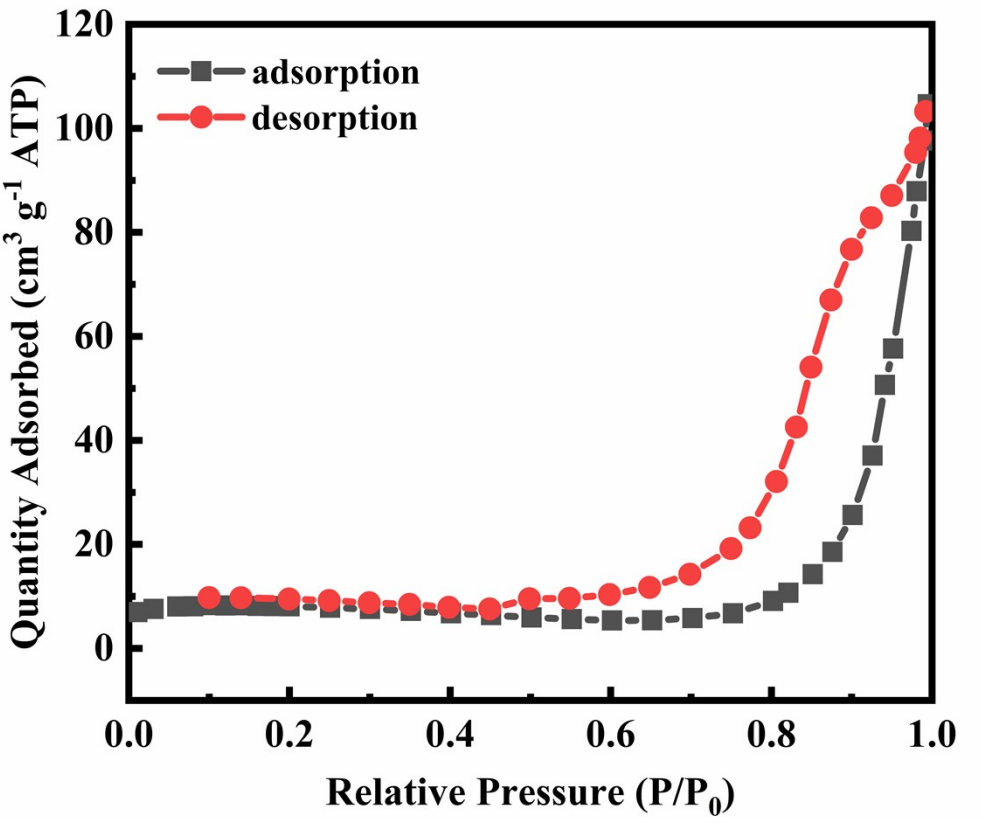

Fig. S7. BET spectrum of NiCoV LDHs.

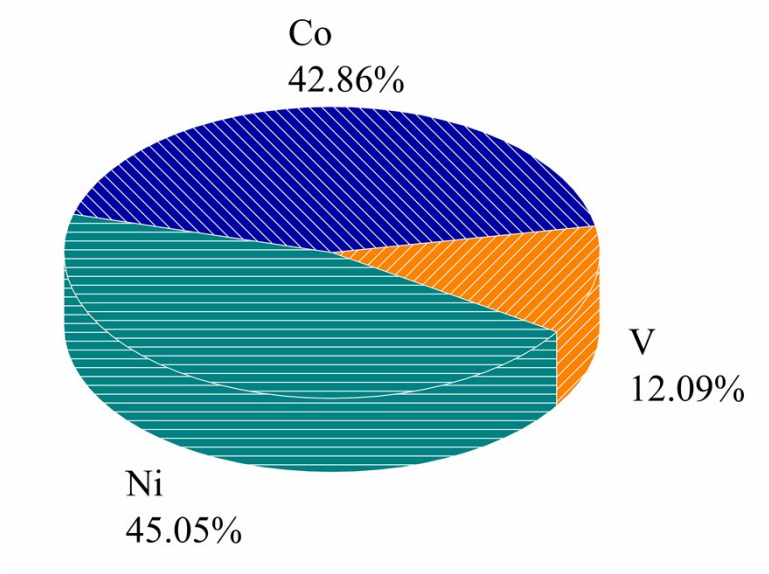

Fig. S8. ICP-OES data for the NiCoV LDHs.

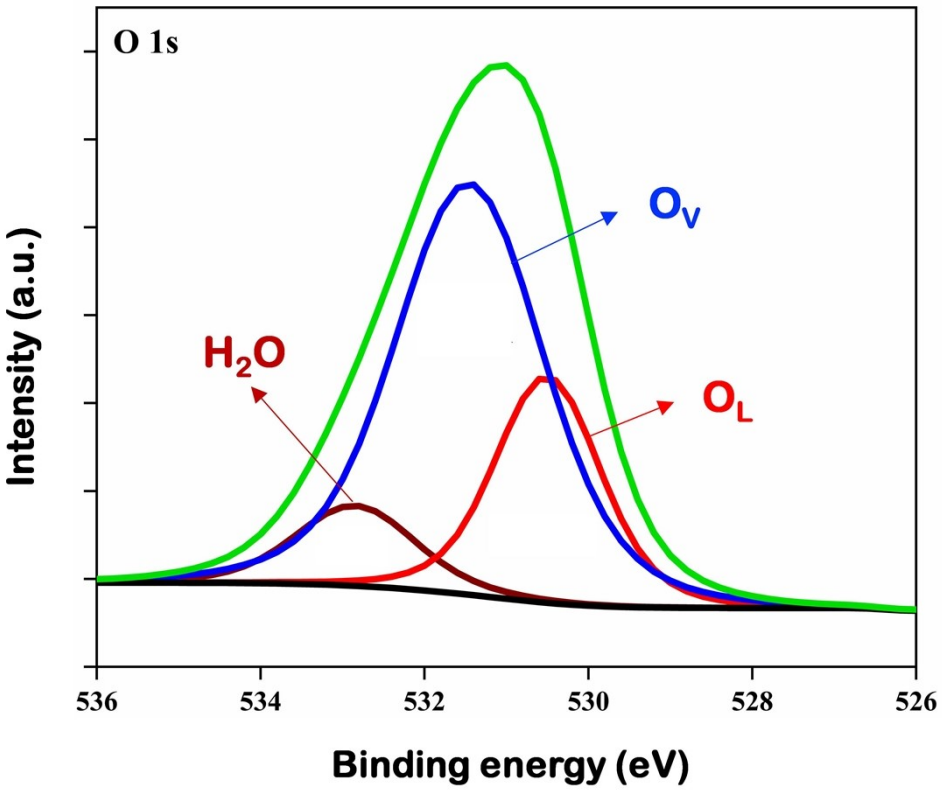

Fig. S9. XPS spectrum of O 1s of NiCo LDHs.

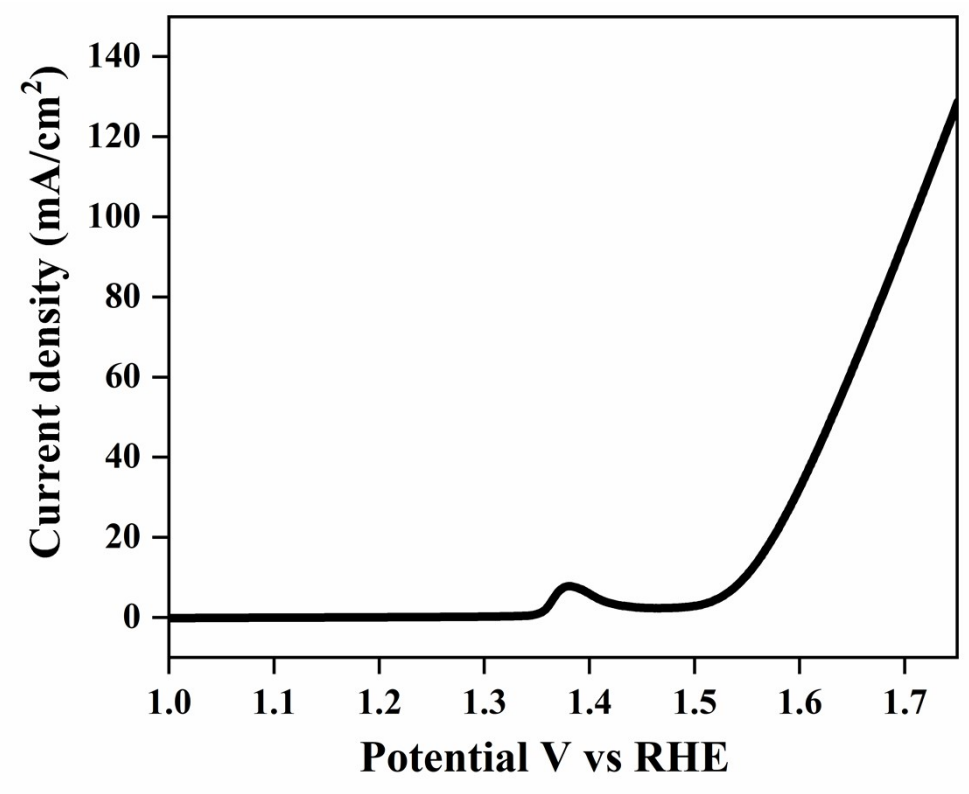

Fig. S10. LSV curve for bare NF in 1 M KOH.

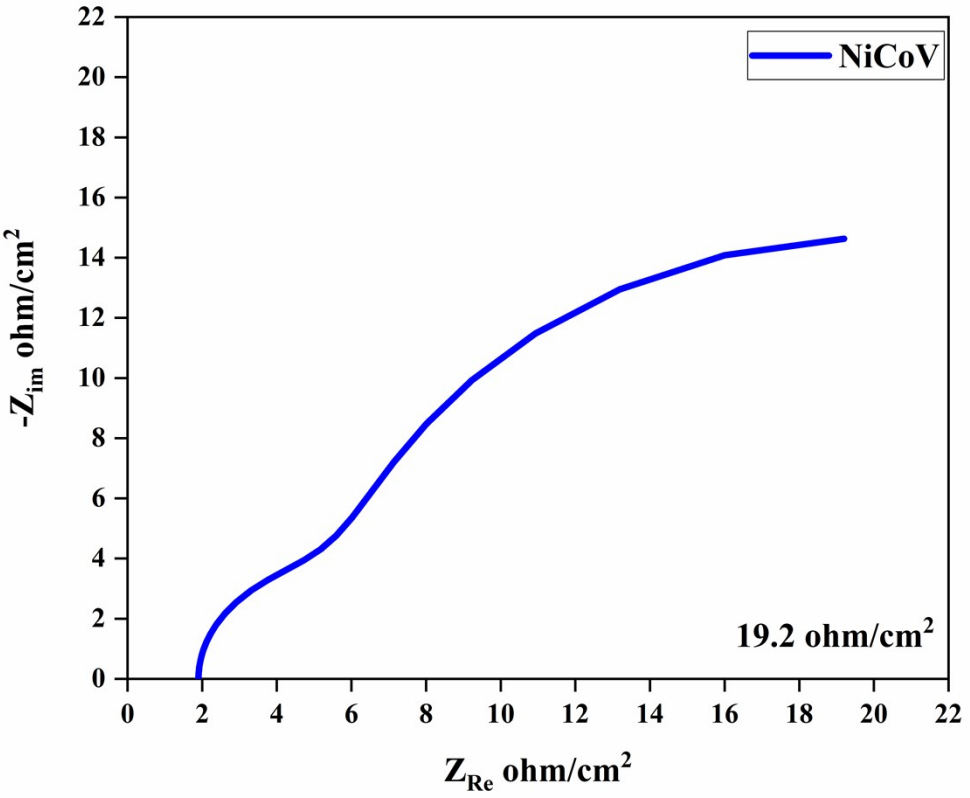

Fig. S11. Nyquist spectrum of NiCoV LDHs.

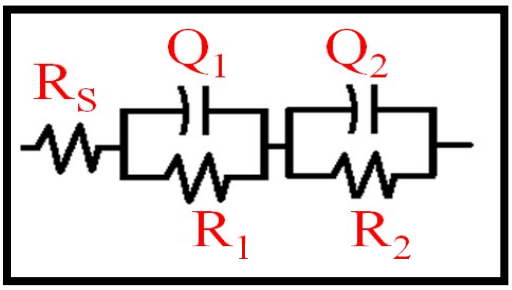

Fig. S12. Equivalent circuit for NiCoV LDHs. (R_2 is the charge transfer resistance, R_{ct}.)

References

1. Y. Wang, S. Tao, H. Lin, G. Wang, K. Zhao, R. Cai, K. Tao, C. Zhang, M. Sun, J. Hu, B. Huang and S. Yang, *Nano Energy*, 2021, **81**, 105606.
2. K. Bera, A. Karmakar, S. Kumaravel, S. S. Sankar, R. Madhu, H. N. Dhandapani, S. Nagappan and S. Kundu, *Inorg. Chem*., 2022, **61**, 4502–4512.
3. Z. Tian, Y. Song, M. Gan, Y. Shen, P. Zhang, P. Liu, C. Liu, B. Xu and J. Guo, *Appl. Surf. Sci.*, 2024, **661**, 160065.
4. X. Shao, A. Zhou, D. Li, L. Zhu, Y. Du, L. Cao and J. Yang, *Energy Fuels*, 2024, **38**, 6938–6949.
5. J. Ahmed, N. Alhokbany, T. Ahamad and S. M. Alshehri, *New J. Chem*., 2022, **46**, 1267–1272.
6. M. Görlin, P. Chernev, J. F. D. Araújo, T. Reier, S. Dresp, B. Pual, R. Krähnert, H. Dau and P. Strasser, *J. Am. Chem. Soc*., 2016**, 138**, 5603–5614.
7. H. N. Dhandapani, D. Mahendiran, A. Karmakar, P. Devi, S. Nagappan, R. Madhu, K. Bera, P. Murugan, B. R. Babu and S. Kundu, *J. Mater. Chem. A*, 2022, **10**, 17488–17500.
8. Y.-G. Fu, Y.-H. Li and Q.-F. Lü, *J. Energy Storage*, 2024, **100**, 113510.
9. X. Yan, Q.-T. Hu, J. Liu, W.-D. Zhang and Z.-G. Gu*, Electrochim. Acta*, 2021, 368, 137648.
10. Y. J. Yang, *J. Appl. Electrochem*., 2020, **50**, 1301–1313.
11. Z. Wu, D. Khalafallah, C. Teng, X. Wang, Q. Zou, J. Chen, M. Zhi and Z. Hong, *J. Alloys Compd*., 2020, **838**, 155604.