Journal of Multiscale Material Informatics
ISSN 3047-5724

Review Article

Machine Learning-Assisted Prediction of Oxygen Evolution Reaction (OER) Activity for Catalyst Discovery: A Review

Wise Herowati[1], Muhamad Akrom[2*], Totok Sutojo[3], Achmad Wahid Kurniawan[4]

[1,2,3,4] Research Group for Quantum Computing and Materials Informatics, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
*Corresponding: m.akrom@dsn.dinus.ac.id

Abstract
The Oxygen Evolution Reaction (OER) is a fundamental process in electrochemical water splitting, playing a crucial role in sustainable hydrogen production. However, its intrinsically sluggish kinetics, involving complex four-electron transfer steps, remain a major bottleneck for efficient energy conversion. In recent years, Machine Learning (ML) has emerged as a powerful approach to accelerate catalyst discovery by enabling data-driven prediction of OER activity and reducing reliance on costly experimental and density functional theory (DFT) calculations. This review systematically summarizes recent advances in ML-assisted OER research, focusing on key aspects including dataset construction, descriptor engineering, model development, and performance evaluation. Various ML techniques, ranging from traditional algorithms such as Random Forest and Support Vector Machines to advanced deep learning approaches, are critically discussed in the context of catalyst screening and activity prediction. Particular attention is given to the role of physicochemical descriptors, including adsorption energies and electronic structure parameters, in governing model performance and interpretability. Furthermore, this review highlights current challenges, such as data scarcity, lack of standardization, and limited model generalization, and discusses emerging trends, including active learning, explainable AI, and integration with high-throughput simulations. By providing a comprehensive overview, this work aims to guide future research toward the development of robust, interpretable, and scalable ML frameworks for accelerating the discovery of efficient OER catalysts.

Keywords: Oxygen evolution reaction; Machine learning; Explainable artificial intelligence; Catalysis.

Received: 27 March 2026 / **Revised:** 29 April 2026 / **Accepted:** 30 April 2026 / **Published:** 1 Mei 2026

1. INTRODUCTION

The global transition toward sustainable and low-carbon energy systems has intensified interest in hydrogen as a clean energy carrier. Among various hydrogen production technologies, electrochemical water splitting has emerged as a promising and environmentally friendly approach. In this process, the Oxygen Evolution Reaction (OER) plays a critical role as the anodic half-reaction. However, OER suffers from intrinsically sluggish kinetics due to its complex four-electron transfer mechanism, inresulting to high overpotentials and reduced overall efficiency. Consequently, the development of highly active, stable, and cost-effective OER catalysts remains a central challenge in energy research [1].

Traditionally, the discovery and optimization of OER catalysts have relied on experimental trial-and-error approaches and first-principles calculations such as density functional theory (DFT). While these methods provide valuable insights into catalytic mechanisms and reaction energetics, they are often time-consuming and computationally expensive, particularly when exploring large chemical spaces. Descriptor-based approaches, including adsorption free energies of reaction intermediates (e.g., OH*, O*, and OOH*), have been widely used to understand catalytic activity; however, scaling relations among these intermediates impose fundamental limitations on catalyst performance [2].

In recent years, Machine Learning (ML) has emerged as a transformative paradigm in materials science, enabling data-driven discovery and accelerated screening of functional materials. By learning complex nonlinear relationships between material descriptors and catalytic performance, ML models can significantly reduce the need for exhaustive experiments and high-throughput DFT calculations. A wide range of ML techniques, including regression models, ensemble learning methods, and deep neural networks, have been successfully applied to predict OER activity, identify promising catalyst candidates, and uncover hidden structure–property relationships [3].

Despite these advances, several challenges remain in applying ML to OER research. These include limited availability of high-quality, standardized datasets, inconsistent descriptor selection, and the lack of

interpretability and generalizability of ML models across different material systems. Furthermore, integrating ML with domain knowledge, such as electrochemical theory and surface science, remains an open research area that warrants further exploration [4].

Given the rapid growth of this interdisciplinary field, a comprehensive and critical review is necessary to consolidate existing knowledge and identify future research directions. Therefore, this paper aims to systematically review recent progress in ML-assisted prediction of OER activity, focusing on key components, including data acquisition, feature engineering, model development, and evaluation strategies. Additionally, emerging trends, including explainable artificial intelligence, active learning, and hybrid physics-informed ML approaches, are discussed to provide insights into next-generation catalyst discovery frameworks.

2. FUNDAMENTAL CONCEPTS

2.1 Mechanism of Oxygen Evolution Reaction (OER)

The Oxygen Evolution Reaction (OER) is a multi-step electrochemical process involving four proton-coupled electron transfer (PCET) steps. It typically occurs on the surface of electrocatalysts in alkaline or acidic environments and can be described through a sequence of adsorbed intermediates: Formation of hydroxyl intermediate (OH*); Conversion to adsorbed oxygen (O*); Formation of hydroperoxyl (OOH*); and Release of molecular oxygen (O_2) [5].

The overall reaction pathway is governed by the adsorption energetics of these intermediates. Due to the complexity of this four-electron process, OER exhibits sluggish kinetics, requiring significant overpotential (η) to proceed at practical rates. The rate-determining step (RDS) is often associated with the formation of OOH*, which has a high energy barrier [6].

2.2 Overpotential and Activity Metrics

A key performance indicator for OER catalysts is the overpotential (η), defined as the additional potential required beyond the thermodynamic equilibrium potential (1.23 V vs. RHE) to drive the reaction. Important evaluation metrics include: Overpotential at a given current density (e.g., η@10 mA/cm²); Tafel slope (reaction kinetics); Turnover frequency (TOF); and Stability and durability. Lower overpotential indicates higher catalytic efficiency, making it the primary target variable in most ML-based OER studies.

2.3 Descriptor-Based Catalyst Design

The rational design of OER catalysts relies heavily on identifying meaningful physicochemical descriptors that correlate with catalytic activity. Among the most widely used descriptors are: Adsorption free energies: ΔG_{OH}, ΔG_{O}, ΔG_{OOH}; d-band center: Reflects the electronic structure of transition metals; Metal–oxygen bond strength; Surface coordination environment; Electronegativity and oxidation states. These descriptors are often derived from Density Functional Theory (DFT) calculations and serve as input features for ML models.

2.4 Scaling Relationships and Limitations

A major challenge in OER catalyst design arises from scaling relationships between adsorption energies of intermediates. For example: ΔG_{OOH} is linearly correlated with ΔG_{OH}. This limits the independent optimization of reaction steps. As a result, even the best catalysts are constrained by a theoretical minimum overpotential (~0.37 V), which gives rise to the well-known volcano plot relationship based on the Sabatier Principle. This intrinsic limitation motivates the use of advanced approaches, including: Breaking scaling relations via novel materials (e.g., heterostructures); Exploring multi-site catalysis; and Leveraging ML to identify non-intuitive descriptor combinations.

2.5 Relevance to Machine Learning

The descriptor-based framework provides a natural bridge to Machine Learning applications. In ML-assisted OER studies: Descriptors act as input features; Overpotential serves as the target variable; and Models learn nonlinear relationships beyond traditional scaling laws. This enables Rapid screening of large material spaces, Identification of hidden patterns, and Prediction of high-performance catalysts without exhaustive simulations [7].

3. MACHINE LEARNING TECHNIQUES

3.1 Conventional Machine Learning Models

The integration of ML into OER research has significantly accelerated catalyst discovery by enabling data-driven prediction of catalytic activity. This section categorizes and critically reviews the major ML approaches applied in OER studies, including conventional models, ensemble methods, deep learning, and emerging paradigms. Conventional ML algorithms have been widely adopted due to their simplicity, interpretability, and relatively low data requirements. Commonly used models include: Linear Regression

(LR), Support Vector Machine (SVM), and k-Nearest Neighbors (k-NN). These models are typically applied to predict OER overpotential based on handcrafted descriptors such as adsorption energies and electronic structure parameters. Strengths: Easy to implement; Interpretable relationships; Effective for small datasets. Limitations: Limited ability to capture complex nonlinear interactions; performance is highly dependent on feature engineering [8].

3.2 Ensemble Learning Approaches

Ensemble methods have demonstrated superior performance in many OER studies by combining multiple weak learners to improve prediction accuracy. Popular techniques: Random Forest (RF); Gradient Boosting (GB, XGBoost, LightGBM). These models are particularly effective in handling nonlinear relationships and feature interactions. Key advantages: High predictive accuracy; Robust to noise and overfitting; Built-in feature importance analysis. In many reported studies, Random Forest consistently achieves high R^2 values (>0.9) in predicting overpotential, making it one of the most reliable baseline models [9].

3.3 Deep Learning Models

Deep learning approaches have gained increasing attention for their ability to automatically learn representations from raw data. Common architectures include Artificial Neural Networks (ANNs), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs), such as CGCNN. Unlike traditional models, deep learning can learn features directly from atomic structures; capture complex spatial and electronic relationships, and reduce reliance on manual descriptor design. Challenges: require large datasets, are computationally intensive, and Have Limited interpretability [10].

3.4 Descriptor Engineering and Feature Representation

Feature engineering remains a critical component in ML-based OER prediction. Types of descriptors: Physics-based descriptors: adsorption energies, d-band center; Composition-based descriptors: elemental properties; and Structure-based descriptors: coordination number, lattice parameters. Recent advances include: Automated feature extraction, Feature selection using SHAP and permutation importance, Dimensionality reduction (PCA, t-SNE). The quality and relevance of descriptors directly influence model performance and generalizability [11].

3.5 Model Evaluation and Validation

To ensure reliability, ML models in OER studies are evaluated using the Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Best practices include: Cross-validation (k-fold)e, External test sets, and Avoidance of data leakage. However, many studies still lack standardized evaluation protocols, making it difficult to benchmark models across datasets [12].

3.6 Emerging Trends in ML for OER

Recent developments show a shift toward more advanced and intelligent frameworks: Explainable AI (XAI) to improve the interpretability of ML predictions; Active Learning to iteratively select the most informative data points; Transfer Learning to leverage knowledge from related systems; and Physics-informed ML to integrate domain knowledge into models. These approaches aim to overcome current limitations and enhance model reliability in real-world applications. Despite significant progress, most ML models for OER remain data-limited and descriptor-dependent, which constrains their extrapolation capability. Future research should prioritize: large, standardized datasets; integration of physics-based constraints; and robust validation across diverse material classes [13].

4. CHALLENGES AND FUTURE PERSPECTIVES

Despite the rapid advancement of ML in OER research, several critical challenges remain that hinder its full potential in catalyst discovery. Addressing these limitations is essential for developing robust, generalizable, and practically applicable ML frameworks.

4.1 Data Scarcity and Quality Issues

One of the most significant bottlenecks in ML-assisted OER research is the limited availability of high-quality and standardized datasets. Most existing datasets are Small in size, heterogeneous in format, and derived from different computational settings (e.g., varying DFT parameters). This lack of consistency introduces bias and reduces model generalizability. Moreover, experimental data are often scarce, further limiting real-world validation. Future direction: Development of large-scale, open, and standardized OER databases; Integration of experimental and computational datasets; Adoption of FAIR (Findable, Accessible, Interoperable, Reusable) data principles [14].

4.2 Descriptor Dependency and Feature Engineering Limitations

Current ML models heavily rely on handcrafted descriptors such as adsorption energies and electronic structure parameters. While these descriptors are physically meaningful, they introduce several limitations: Dependence on prior domain knowledge, Computational cost for descriptor generation (e.g., DFT), and limited transferability across material systems. Future direction: Descriptor-free approaches using deep learning (e.g., GNN); Automated feature learning from raw structural data; Hybrid descriptors combining physics-based and data-driven features [15].

4.3 Model Interpretability and Explainability

Many high-performing ML models, particularly deep learning models, function as "black boxes," making it difficult to interpret their predictions. This lack of transparency limits trust and adoption in scientific research. Future direction: XAI methods such as SHAP and LIME; Development of interpretable ML models aligned with physical laws; and linking ML outputs with catalytic theory (e.g., reaction mechanisms, scaling relations) [16].

4.4 Generalization and Transferability

Most ML models are trained on specific datasets and often fail to generalize to unseen material classes. This issue is particularly critical in OER, where materials span diverse chemical and structural spaces. Future direction: Transfer learning across different catalyst families; Domain adaptation techniques; and Benchmarking across multiple datasets [17]–[19].

4.5 Breaking Scaling Relationships

As discussed earlier, scaling relationships impose fundamental limitations on OER performance, as described by the Sabatier Principle. Most ML models implicitly learn these constraints rather than overcoming them. Future direction: ML-guided discovery of materials that deviate from scaling relations; Exploration of multi-site and bifunctional catalysts; and Integration of ML with quantum simulations to identify unconventional pathways [20]–[24].

4.6 Integration with Advanced Computational Techniques

The next generation of OER research will likely involve tight integration between ML and advanced computational/experimental techniques. Emerging paradigms include: Active Learning: Iterative model improvement with minimal data; High-throughput screening: ML-guided exploration of large chemical spaces; Autonomous laboratories: Closed-loop experimentation with AI control; Quantum-enhanced ML: Leveraging quantum computing for complex systems [25]–[27].

4.7 Toward Real-World Deployment

Despite promising results, most ML models remain at the theoretical or computational level, with limited experimental validation. Future direction: Bridging the gap between ML predictions and experimental synthesis; Developing deployable ML tools for materials scientists; Collaboration between data scientists and experimental researchers. Current ML-assisted OER research is still largely predictive rather than prescriptive. That is, models can identify promising candidates but often fail to provide actionable insights for catalyst design. The future of this field lies in transitioning toward decision-making frameworks that integrate ML predictions with mechanistic understanding and experimental feasibility [28]–[30].

5. CONCLUSION

This review provides a comprehensive overview of the application of Machine Learning to the prediction and discovery of Oxygen Evolution Reaction catalysts. By systematically analyzing recent developments in dataset construction, descriptor engineering, model design, and evaluation strategies, it is evident that ML has significantly accelerated catalyst screening and enabled deeper insights into structure–property relationships. However, key challenges, including data limitations, descriptor dependency, model interpretability, and generalization, continue to restrict broader applicability. Addressing these issues requires integrating advanced ML techniques, standardized datasets, and domain knowledge from electrochemistry and materials science. Looking forward, emerging approaches such as explainable AI, active learning, and physics-informed machine learning are expected to play a pivotal role in overcoming current limitations. Ultimately, the convergence of ML, high-throughput simulations, and experimental validation will pave the way for the rational design of next-generation OER catalysts, advancing sustainable energy technologies.

REFERENCES

[1] Saravanan, A.; Deivayanai, V.; Kumar, P.S.; Rangasamy, G.; Hemavathy, R.; Harshana, T.; Gayathri, N.; Alagumalai, K. A detailed review on advanced oxidation process in treatment of wastewater: Mechanism, challenges and future outlook. Chemosphere 2022, 308, 136524.

[2] Wang, B.; Zhang, H.; Wang, F.F.; Xiong, X.G.Y.; Tian, K.; Sun, Y.B.; Yu, T.T. Application of heterogeneous catalytic ozonation for refractory organics in wastewater. Catalysts 2019, 9, 241.

[3] Ma, D.S.; Yi, H.; Lai, C.; Liu, X.G.; Huo, X.Q.; An, Z.W.; Li, L.; Fu, Y.K.; Li, B.S.; Zhang, M.M.; et al. Critical review of advanced oxidation processes in organic wastewater treatment. Chemosphere 2021, 275, 130104.

[4] Fraiha, O.; Hadoudi, N.; Najlae, Z.A.; Salhi, A.; Amhamdi, H.; Mourabit, F.; Ahari, M.H. Comprehensive review on the adsorption of pharmaceutical products from wastewater by clay materials. Desalination Water Treat. 2024, 317, 100114.

[5] Li, L.X.; Han, J.Z.; Huang, X.H.; Qiu, S.; Liu, X.H.; Liu, L.L.; Zhao, M.J.; Qu, J.W.; Zou, J.L.; Zhang, J. Organic pollutants removal from aqueous solutions using metal-organic frameworks (MOFs) as adsorbents: A review. J. Environ. Chem. Eng. 2023, 6, 111217.

[6] Chen, H.J.; Xiao, R.H.; Huang, D.L.; Deng, R.; Li, R.J.; Chen, Y.S.; Zhou, W. Three kinds of apatite adsorbents prepared by co-precipitation for Pb (II) and Cd (II) removal from wastewater: Performance, competitive effects and mechanisms. J. Mol. Liq. 2024, 400, 124478.

[7] Kavitha, E.; Poonguzhali, E.; Nanditha, D.; Kapoor, A.; Arthanareeswaran, G.; Prabhakar, S. Current status and future prospects of membrane separation processes for value recovery from wastewater. Chemosphere 2022, 291, 132690.

[8] Öcal, Z.B.; Öncel, M.S.; Keskinler, B.; Khataee, A.; Karagündüz, A. Sustainable treatment of boron industry wastewater with precipitation-adsorption hybrid process and recovery of boron species. Process Saf. Environ. Prot. 2024, 182, 719–726.

[9] Bartolomeu, M.; Neves, M.G.; Faustino, M.A.; Almeida, A. Wastewater chemical contaminants: Remediation by advanced oxidation processes. Photochem. Photobiol. Sci. 2018, 17, 1573–1598.

[10] Li, L.X.; He, Z.M.; Liang, T.J.; Sheng, T.; Zhang, F.G.; Wu, D.; Ma, F. Colonization of biofilm in wastewater treatment: A review. Environ. Pollut. 2022, 293, 118514.

[11] Muhamad Akrom, Supriadi Rustad, Hermawan Kresno Dipojono, Ryo Maezono, Hideaki Kasai. Quantum machine learning for ABO3 perovskite structure prediction. Computational Materials Science, Volume 250, Pages 113694, 2025, https://doi.org/10.1016/j.commatsci.2025.113694.

[12] S Rustad, M Akrom, T Sutojo, HK Dipojono. A feature restoration for machine learning on anti-corrosion materials. Case Studies in Chemical and Environmental Engineering 10, 100902, 2024, https://doi.org/10.1016/j.cscee.2024.100902.

[13] Akrom, M., Rustad, S. & Dipojono, H.K. Investigation of Corrosion Inhibition Capability of Pyridazine Compounds via Ensemble Learning. J. of Materi Eng and Perform 34, 14948–14962 (2025). https://doi.org/10.1007/s11665-024-10129-x.

[14] Muhamad Akrom, Wise Herowati, De Rosal Ignatius Moses Setiadi. A quantum circuit learning-based investigation: A case study in iris benchmark dataset binary classification. Journal of Computing Theories and Applications. Volume 2, Issue 3, Pages 355-367, 2024, https://doi.org/10.62411/jcta.11779.

[15] Muhamad Akrom, Supriadi Rustad, Totok Sutojo, De Rosal Ignatius Moses Setiadi, Pulung Nurtantio Andono, Guruh Fajar Shidik, Hermawan Kresno Dipojono, Ryo Maezono. A novel quantum-enhanced model cascading approach based on support vector machine in blood-brain barrier permeability prediction. Materials Today Communications, Volume 45, Pages 112341, 2025, https://doi.org/10.1016/j.mtcomm.2025.112341.

[16] Muhamad Akrom, Usman Sudibyo, Achmad Wahid Kurniawan, Noor Ageng Setiyanto, Ayu Pertiwi, Aprilyani Nur Safitri, Novianto Hidayat, Harun Al Azies, Wise Herawati. Artificial Intelligence Berbasis QSPR Dalam Kajian Inhibitor Korosi. JoMMiT: Jurnal Multi Media dan IT, Volume 7, Issue 1, Pages 015-020, 2023, https://doi.org/10.46961/jommit.v7i1.721.

[17] M. Akrom, T. Sutojo, A. Pertiwi, S. Rustad, H.K. Dipojono, Investigation of Best QSPR-Based Machine Learning Model to Predict Corrosion Inhibition Performance of Pyridine-Quinoline Compounds, J Phys Conf Ser, 2673(1), 012014 (2023), https://doi.org/10.1088/1742-6596/2673/1/012014.

[18] M. Akrom, Green corrosion inhibitors for iron alloys: a comprehensive review of integrating data-driven forecasting, density functional theory simulations, and experimental investigation. J Mult Mater Inf, 1(1), 22–37 (2024), https://doi.org/10. 62411/jimat.v1i1.10495

[19] M. Akrom, S. Rustad, H.K. Dipojono, A machine learning approach to predict the efficiency of corrosion inhibition by natural product-based organic inhibitors, *Phys Scr*, 99(3), 036006 (2024), https://doi.org/10.1088/1402-4896/ad28a9.

[20] M. Akrom, S. Rustad, H.K. Dipojono, Machine learning investigation to predict corrosion inhibition capacity of new amino acid compounds as corrosion inhibitors, Results in Chemistry 6 (2023) 101126, https://doi.org/10.1016/j. rechem.2023.101126.

[21] M. Akrom, S. Rustad, A.G. Saputro, H.K. Dipojono, Data-driven investigation to model the corrosion inhibition efficiency of Pyrimidine-Pyrazole hybrid corrosion inhibitors, Comput. Theor. Chem. 1229 (2023) 114307, https://doi.org/10.1016/ J.COMPTC.2023.114307.

[22] M. Akrom, S. Rustad, H.K. Dipojono, Prediction of Anti-Corrosion performance of new triazole derivatives via Machine learning, Comput. Theor. Chem. 1236 (2024), https://doi.org/10.1016/j.comptc.2024.114599.

[23] M. Akrom, Investigation of natural extracts as green corrosion inhibitors in steel using density functional theory, *Jurnal Teori dan Aplikasi Fisika*, 10(1), 89-102 (2022), https://doi.org/10.23960%2Fjtaf.v10i1.2927.

[24] M. Akrom, S. Rustad, H.K. Dipojono. Development of quantum machine learning to evaluate the corrosion inhibition capability of pyrimidine compounds. Materials Today Communications, 39, 108758 (2024), https://doi.org/10.1016/j.mtcomm.2024.108758.

[25] M. Akrom, S. Rustad, H.K. Dipojono, SMILES-based machine learning enables the prediction of corrosion inhibition capacity, MRS Commun 14 (2024) 379–387, https://doi.org/10.1557/s43579-024-00551-6.

[26] M. Akrom, S. Rustad, A.G. Saputro, A. Ramelan, F. Fathurrahman, H.K. Dipojono, A combination of machine learning model and density functional theory method to predict corrosion inhibition performance of new diazine derivative compounds, Mater. Today Commun. 35 (2023) 106402, https://doi.org/10.1016/J. MTCOMM.2023.106402.

[27] M. Akrom, et al., DFT and microkinetic investigation of oxygen reduction reaction on corrosion inhibition mechanism of iron surface by Syzygium Aromaticum extract, Appl. Surf. Sci. 615 (2023), https://doi.org/10.1016/j. apsusc.2022.156319.

[28] M. Akrom, S. Rustad, H.K. Dipojono. Variational quantum circuit-based quantum machine learning approach for predicting corrosion inhibition efficiency of pyridine-quinoline compounds. Materials Today Quantum, 2, 100007 (2024), https://doi.org/10.1016/j.mtquan.2024.100007.

[29] Muhamad Akrom, Supriadi Rustad, Hermawan Kresno Dipojono, Ryo Maezono, Hideaki Kasai. Enhanced quantum support vector regression with quantum kernels and virtual sampling for ABX3 perovskite formation energy. Expert Systems with Applications, Pages 130817, 2025, https://doi.org/10.1016/j.eswa.2025.130817.

[30] Muhamad Akrom. Quantum Neural Network in Architectures, Learning Mechanisms, and Emerging Applications Across Domains: A Review. Journal of Multiscale Materials Informatics, Volume 2, Issue 2, Pages 30-39, 2025, 10.62411/jimat.v2i2.14929.