



Contents list available at CBIORE journal website

Journal of Emerging Science and Engineering

Journal homepage: <https://journal.cbior.id/index.php/jese/index>



Hydrogen fuel cells for green shipping: A machine learning-based power prediction framework

Thuy Hien Tran^a, Thanh Hieu Chau^b, Duc Chuan Nguyen^{b*}

^aAulac Corporation, Ho Chi Minh city, Vietnam

^bInstitute of Maritime, University of Transport Ho Chi Minh City, Ho Chi Minh City, Viet Nam

Abstract. Hydrogen fuel cells offer a promising route toward low-emission maritime electrification, but reliable prediction of electrical power is essential for their integration into shipboard energy-management systems. This study develops a compact, explainable machine-learning framework using 119 observations from a publicly available stationary hydrogen fuel-cell installation. Ambient, fuel-cell gas, and coolant outlet temperatures; power-generation and consumption setpoints; auxiliary consumption; and operating status were used to predict power supplied to the grid. Linear regression (LR) and Extreme Gradient Boosting (XGBoost) were evaluated using a chronological 80:20 training–testing split, with parity plots, residual diagnostics, and SHapley Additive exPlanations (SHAP) supporting model assessment. LR fitted the training observations closely ($R^2 = 0.9960$) but failed to generalize to testing data ($R^2 = -33.8364$; RMSE = 65.651). In contrast, XGBoost achieved consistent training and testing performance ($R^2 = 0.9740$ and 0.9770 , respectively), with a testing RMSE of 1.687 and MAE of 1.273. SHAP identified auxiliary consumption (mean absolute SHAP = 9.82) and the power-consumption setpoint (2.45) as dominant predictive features. The analysis addresses grid-delivered power rather than propulsion demand or directly measured hydrogen utilization. The findings demonstrate the value of interpretable nonlinear regression for fuel-cell power prediction and offer a methodological foundation for future maritime applications.

Keywords: Hydrogen fuel cells; Green shipping; Maritime electrification; XGBoost; Machine learning; SHAP; Power prediction



@ The author(s). Published by CBIORE. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

Received: 6th August 2026; Revised: 16th Oct 2026; Accepted: 28th Oct 2026; Available online: 1st Nov 2026

1. Introduction

Maritime transportation plays a fundamental role in international trade and economic development. However, its continued dependence on conventional fossil fuels presents significant environmental and sustainability challenges. Emissions of carbon dioxide (CO₂), nitrogen oxides (NO_x), sulfur oxides (SO_x), and particulate matter from marine propulsion systems contribute to climate change and atmospheric pollution (Dzida & Nguyen, 2026). These emissions have become an increasing concern and have led to the development of alternative fuels, renewable energy technologies, and electrified propulsion systems. The International Maritime Organization (IMO) adopted a new greenhouse gas (GHG) strategy in 2023 with a goal of reaching net-zero GHG emissions from international shipping by or around 2050 and intermediate emission reduction targets for 2030 and 2040 (IMO, 2023). Significant technological progress in onboard energy conversion, storage, and management is needed to reach these goals. Fuel-cell systems powered by hydrogen are one of the new options that have been receiving a lot of attention, as they could supply efficient and locally emission-free electrical power for sustainable maritime use (Nguyen et al., 2025; Serbin et al., 2022).

Hydrogen fuel cells are systems that produce electricity via electrochemical reactions, not combustion, and have several benefits: they are very efficient at conversion, they are very quiet, they can be built in modules, and they do not directly emit carbon-containing pollutants (Cavanini et al., 2023; Inal, 2024; Ye et al., 2022). If hydrogen is supplied using renewable energy, these systems can make a significant contribution to the reduction of life cycle GHG emissions. Amongst them, proton exchange membrane fuel cells (PEMFCs) are viewed as particularly promising for maritime electrification due to their relatively low operating temperature, compact design, and fast reaction times (Amat Ventayol et al., 2025; Coraddu et al., 2025; Lee et al., 2020). They can also enhance the flexibility of operation when combined with battery energy storage systems, enabling transient power demand and minimizing fluctuations in fuel-cell load. These are particularly applicable to short sea vessels, ferries, inland waterway transport, and auxiliary shipboard power generation. However, issues related to hydrogen storage, infrastructure, system durability, and dynamic power management persist in the practical implementation of these (Lee et al., 2020; Xie et al., 2024).

Accurate forecasting of the electrical power of fuel cells is crucial to enhance their operation, energy management, and integration with the maritime electrical network. The performance of fuel cells is influenced by a number of interdependent operating parameters such as temperature, electrical loading, auxiliary energy consumption, and operational state. Variations in these parameters can result in nonlinear variations in electrical output, especially at start-up and shut-down and during load changes (Golbabaie et al., 2022; Penga et al., 2024; Yang et al., 2022). The behavior of fuel cells can be modelled with conventional mathematical and electrochemical models, which relate the reaction kinetics, mass transport, thermodynamic properties, and electrical characteristics (Su et al., 2023).

* Corresponding author
Email: chuan.nguyen@ut.edu.vn (D. C. Nguyen)

. While these models offer important physical insights, they often demand a lot of system details, parameter estimation, and computation. Hence, simplified predictive methods that can be directly learned from measured data are an attractive alternative for performance assessment and initial energy-management applications (Klass et al., 2023; Kumar et al., 2025).

New opportunities for modelling complex energy-conversion systems have emerged in recent years due to advances in artificial intelligence and machine learning (ML). Without solving detailed electrochemical equations, regression algorithms can be used to find relationships between measured operating variables and desired responses. Of these methods, linear regression (LR) offers a baseline for detecting approximately linear relationships between independent variables and power output, and is also relatively simple to compute. However, when the relationships include nonlinear interactions or different operating regimes, its predictive power may decrease (Panja & Meharia, 2024; Szemer et al., 2025). On the other hand, Extreme Gradient Boosting (XGBoost) is an ensemble of decision trees built sequentially to model complex nonlinear relationships. It is especially well adapted to structured operational data thanks to its regularization mechanisms, flexible model structure, and capacity to model interactions between predictors. Previous studies have shown the ability of machine learning algorithms such as gradient boosting and XGBoost to model the PEMFC performance under different operating conditions (Shrote et al., 2024; Wang et al., 2020).

However, some practical issues remain concerning the relative performance and interpretability of relatively simple regression models for small datasets of fuel-cell operational data. Prediction results may be affected by limited observations, changing operating states, and correlated process variables, which can affect generalization and complicate interpretation. Moreover, the traditional regression parameters are not sufficient to provide information on the relative contribution of each operating parameter. Explainable artificial intelligence (XAI) methods, especially SHapley Additive exPlanations (SHAP), provide a systematic way to assess model predictions by quantifying the contribution of each feature (Bora & Sharma, 2026; Le et al., 2023). The combination of predictive modelling and SHAP analysis can thus be used to gain insights into the relationship between the fuel-cell operating conditions and electrical power output. This information can be valuable for the identification of influential variables and for the development of transparent data-driven tools for maritime hydrogen-energy systems.

In this context, the present study proposes a machine-learning-based framework for prediction of electrical power output of a hydrogen fuel-cell energy system. A set of 119 measured observations, including temperature, electrical power setpoints, auxiliary consumption, and operational status, is used. The data set is from a stationary fuel-cell facility and is analyzed as a sample of data that can be used as a foundation for the development of predictive methods that can be applied to maritime applications. Chronological training and testing partitions are used to implement and compare two regression algorithms, LR and XGBoost. The coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) are used to assess the predictive performance. Correlation and distribution analyses are performed to investigate the relationship between variables, and parity and residual plots are used to check the agreement of predictions and systematic errors. Further, SHAP-based interpretation is used to examine the impact of each feature on the power output prediction. The study aims to develop a short, interpretable regression model for fuel-cell power prediction and to give methodological insights for the integration of hydrogen fuel-cell technologies in future sustainable maritime energy systems. Figure 1 depicts the methodological flow chart showing the ML implementation.

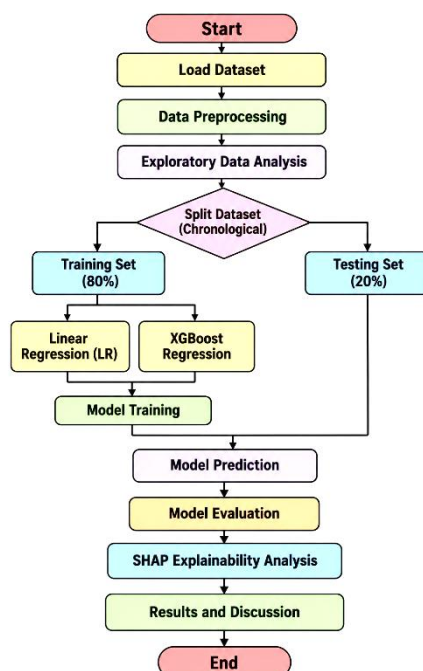


Figure 1. Flow chart for methodology used in this study

2. Materials and Methods

2.1. Data Source and Description

The present study has used experimental data that were publicly available from the Hydrogen-Aeolic Energy with Optimised eElectrolysers Upstream of Substation (HAEOLUS) project, coordinated by SINTEF (Norway) under the European Union's (EU) Horizon 2020 research programme (Grant No. 779469). The original data, titled HAEOLUS Project Fuel Cell Data, was released on Zenodo in June 2024 and can be found at <https://doi.org/10.5281/zenodo.11395583> (SINTEF (Norway), 2024). This data includes the operational measurements of a stationary hydrogen fuel-cell system that is part of a renewable hydrogen-energy demonstration

project. These measurements were used to create a data-driven power-prediction framework that could be applied to hydrogen-based maritime electrical systems. The original data consisted of 10 CSV files with temperature data, power generation setpoints, power consumption setpoints, auxiliary consumption, operating status, hydrogen consumption, and electrical power supplied to the grid. The individual records were synced with their timestamps and consolidated into a single dataset. Incomplete records and duplicate times were removed, leaving 119 observations. The regression analysis was performed without the hydrogen-consumption measurements. Seven input variables were taken into account: ambient temperature, fuel-cell gas temperature, coolant outlet temperature, power-generation setpoint, power-consumption setpoint, auxiliary power consumption, and fuel-cell operating status. Power supplied to the grid was chosen as the dependent variable and measured.

2.2. Data Preprocessing and Exploratory Analysis

Data preprocessing was done in Python with the help of Pandas and Scikit-learn libraries. The observations were sorted by time, and those with invalid times or missing target values were discarded. The continuous variables were imputed with the median and scaled using standard scaling, and the fuel-cell operating status was one-hot encoded (Alasadi & Bhaya, 2017; Malvoni et al., 2017). To avoid data leakage, pre-processing transformations were only fitted to the training data. Predictors were not allowed to include timestamp-derived variables or cumulative operating hours. Pearson correlation coefficients were used to perform exploratory data analysis for linear relationships between numerical variables. Scatterplots, histograms, and correlation coefficients were used to create a combined correlation and distribution matrix. In addition, a correlation heatmap was developed to determine relationships between operating parameters and electrical output (Hauke & Kossowski, 2011; Suherman et al., 2025).

2.3. Machine Learning Models

Two regression algorithms, linear regression (LR) and Extreme Gradient Boosting (XGBoost), were implemented to predict fuel-cell electrical power. LR was employed as a baseline model to describe the relationship between operating parameters and electrical output through a linear combination of predictors (Pak & Oh, 2010; Pandit et al., 2021):

$$[\hat{y} = \beta_0 + \sum_{j=1}^p \beta_j x_j] \quad (1)$$

where ($\hat{y}$) represents predicted electrical power, (β_0) is the intercept, (β_j) denotes the regression coefficient, and (x_j) represents an input feature.

XGBoost was employed to capture nonlinear relationships and interactions among operating variables (Darmawan et al., 2023). The model combines sequentially developed regression trees (Chen & Guestrin, 2016; Phan et al., 2021):

$$[\hat{y}_i = \sum_k 1^K f_k(\mathbf{x}_i)] \quad (2)$$

where (K) denotes the number of trees and (f_k) represents the prediction from the k th tree. The XGBoost model was configured using 120 estimators, maximum tree depth of 2, learning rate of 0.04, subsampling ratio of 0.85, column-sampling ratio of 0.90, and L2 regularization coefficient of 8. A fixed random seed of 42 was employed for reproducibility.

2.4. Model Training and Performance Evaluation

The dataset was split chronologically into 80% training (95 observations) and 20% testing (24 observations) while maintaining the temporal order of the measurements. The same input variables were used in both regression models, and the same holdout dataset was used to evaluate both. The coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) were used to evaluate the predictive performance. For observations with measured output that is not zero, mean absolute percentage error (MAPE) was also computed. Ideal prediction and $\pm 10\%$ deviation lines were plotted on parity plots of measured and predicted values. Systematic prediction errors and differences between training and testing performance were investigated by examining the residual-versus-predicted plots and residual distributions.

2.5. Explainable Machine Learning Using SHAP

SHapley Additive exPlanations (SHAP) were used to explain the effect of the operating variables on the electrical power prediction. XGBoost was implemented by TreeExplainer. SHAP values were computed to test observations and evaluate the contribution of each feature in the training set (Hoa & Lien, 2025; Mokhtari et al., 2020). Mean absolute SHAP values were used to calculate global feature importance, and beeswarm and feature-dependence plots were used to visualize the direction and distribution of feature importance. Waterfall plots were also created to help interpret individual predictions. All analyses were carried out in Python, and graphical outputs were exported as PNG files.

3. Results and discussion

3.1. Statistical Characteristics and Correlation Analysis

Table 1 summarizes the statistical features of the fuel-cell operating parameters and electrical power output, and Fig. 2 shows the correlation matrix for the relationships between the investigated parameters. The ambient temperature ranged from 3.40 to 9.00, with an average of 5.07 ± 1.03 , while the fuel-cell gas temperature had a relatively small range of 16.20–18.90 with an average of 17.33 ± 0.97 . The coolant outlet temperature, on the other hand, had a wider range of 18 to 48.70 with an average of 35.06 ± 8.29 .

These variations suggest that the temperature environment of the fuel-cell system was not the same for the whole period of investigation, which could affect the electrical performance of the system.

The setpoints for power generation and power consumption had large variations, with mean values of 60.34 ± 36.33 and 34.82 ± 42.11 , respectively. The distribution of auxiliary consumption was very scattered, with a mean of 32.74 ± 179.02 and the highest recorded as 1957.44. The much lower median of 7.46 suggests that it is strongly skewed to the right, indicating that there were some consumption events with higher values that affected the average. Likewise, the power supplied to the grid ranged from 0 to 65.68 with a mean of 12.03 ± 16.72 and a median of zero. The difference between the mean and the median suggests that the data set has periods of time in which the electrical output is zero and periods of time in which the electrical output is active.

The results of the correlation analysis in Fig. 2 showed that the power-consumption setpoint had the highest positive linear correlation with grid power (0.71). This suggests that the electrical setpoint is closely associated with the power delivered by the fuel-cell system. The power-generation setpoint had a weaker positive correlation of 0.28, and the fuel-cell gas temperature and coolant outlet temperature had negative correlations of -0.19 and -0.16, respectively. The linear correlation of ambient temperature (0.00) and auxiliary consumption (0.04) with grid power was negligible. However, weak correlation does not mean that a variable is not predictive, especially if there are nonlinearities or different operating modes. With respect to the independent variables, a high negative correlation (-0.82) was found between fuel-cell gas temperature and the power-generation setpoint. Ambient temperature also showed a negative correlation with the fuel-cell gas temperature (-0.52), and the setpoints of the generation and consumption were positively correlated (0.56). These relationships indicate that there are significant interactions between operating parameters, so that linear and nonlinear regression models should be evaluated. The distribution of the auxiliary consumption and the occurrence of zero power operating conditions suggest in particular that the response may not be adequately represented by a single linear relationship.

Table 1. Descriptive statistics of operating variables

Parameter	Mean $\pm$ SD	Median (Q1–Q3)	Min–Max
Ambient temperature	5.07 ± 1.03	4.90 (4.50–5.50)	3.40–9.00
Fuel-cell gas temperature	17.33 ± 0.97	16.80 (16.50–18.50)	16.20–18.90
Coolant outlet temperature	35.06 ± 8.29	34.50 (28.90–42.00)	18.00–48.70
Power generation setpoint	60.34 ± 36.33	40.00 (30.00–100.00)	20.00–100.00
Power consumption setpoint	34.82 ± 42.11	23.14 (0.00–100.00)	0.00–100.00
Auxiliary consumption	32.74 ± 179.02	7.46 (0.42–31.72)	0.00–1957.44
Power supplied to grid (target)	12.03 ± 16.72	0.00 (0.00–24.27)	0.00–65.68

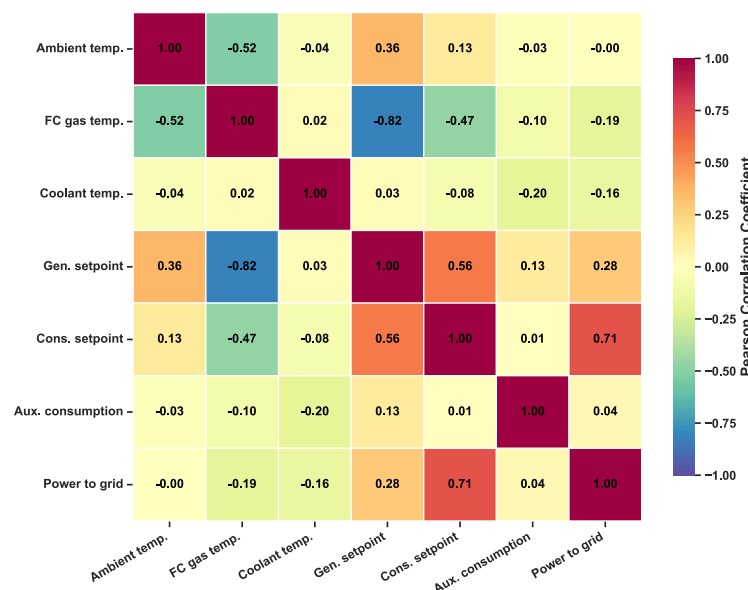


Figure 2. Correlation matrix of fuel-cell operating parameters and electrical power output

3.2. Comparative Predictive Performance

The same training and test sets were used to assess the predictive ability of both LR and XGBoost, which contained 95 and 24 observations, respectively. The statistics are summarized in Table 2, and the parity plots of the measured and predicted grid power are shown in Fig. 3(a,b). Residuals of LR and XGBoost are also shown in Figs. 4 and 5, respectively. During training, the LR model showed a very good fitting performance with R^2 of 0.9960, MSE of 1.247, RMSE of 1.116, and MAE of 0.810. As seen in Fig. 3(a), most

of the training predictions were close to the ideal 1:1 line and within the limits of the $\pm 10\%$ deviation for appreciable nonzero power values. This means that the linear model is able to capture the major relationships in the training observations. However, the testing results showed that there was significant deterioration, where the R^2 decreased to -33.8364 , and MSE increased to 4310.074. In turn, the RMSE and MAE rose to 65.651 and 19.722, respectively. The testing R^2 is very negative, meaning that LR did much worse than it would have done if it had always predicted the mean of the testing set.

Further insight into this discrepancy can be gained from the residual analysis in Fig. 4(a). The training residuals were still clustered about zero, but the testing observations contained a very large positive residual with a very negative predicted output. Such behavior strongly suggests that the linear relationship fitted to the data is not a good fit under operating conditions that are not well represented by the data. The resulting residual distribution in Fig. 4(b) also shows that the testing errors were highly asymmetric, with a large error outside the range of the residuals seen in training. Thus, the high training accuracy of LR was not found to be correlated with the predictive performance on the chronological holdout dataset. XGBoost, on the other hand, consistently predicted with high accuracy in both training and testing phases. The model achieved training R^2 , MSE, RMSE, and MAE values of 0.9740, 8.024, 2.833, and 1.212, respectively. During testing, its performance remained strong, with $R^2 = 0.9770$, MSE = 2.845, RMSE = 1.687, and MAE = 1.273. The testing RMSE was decreased by about 97.4% with XGBoost, which showed significant generalization to the test operating conditions when compared with LR.

As seen in the parity plot in Fig. 3(b), the predictions of the XGBoost were fairly close to the ideal regression line, and most testing observations had small deviations from their measured value. But some high-power training observations were underrated, with predictions coming in around 50 marks whilst the measured power went up to nearly 66. This is typical of how tree-based regression models work, as they tend to extrapolate only as far as what they've seen in their training data. However, the relatively low testing errors suggest that XGBoost captured the pertinent operating relationships better than LR. These are also confirmed by the residual plots in Fig. 5(a,b). The distribution of the XGBoost residuals was mostly centered around zero (Fig. 5(a)), with relatively small deviations for both training and test data. There were a few larger residuals at higher levels of power prediction, suggesting that prediction accuracy might be different throughout the range of operation (Fig. 5(b)). The residual distribution was significantly more compact than that achieved with LR, and it was more stable and less sensitive to the extreme prediction errors noted with the linear model. One more observation from Table 2 is that of the mean absolute percentage error (MAPE). The training and testing MAPE values of LR were 11.28% and 5.89%, respectively, while the training and testing MAPE values of XGBoost were 9.67% and 16.56%, respectively. LR yielded a lower testing MAPE, but this was computed for only those observations that had nonzero measured output. It thus omits the zero-power observations that caused the very large errors in the LR tests. Therefore, MAPE cannot be used alone to make a representative comparison of the overall prediction accuracy for this dataset. Overall, XGBoost showed significantly better testing results across all metrics— R^2 , MSE, RMSE, MAE, and residual behavior. The different behavior of the two algorithms indicates that nonlinear interactions, operating-state transitions, and electrical load variations are important in determining the fuel-cell power output. While LR is simple and efficient to compute, XGBoost can model these relationships in a more flexible way via its ensemble of regression trees. It is a more suitable predictive model for the present data set because its performance is consistent across the training and testing observations investigated.

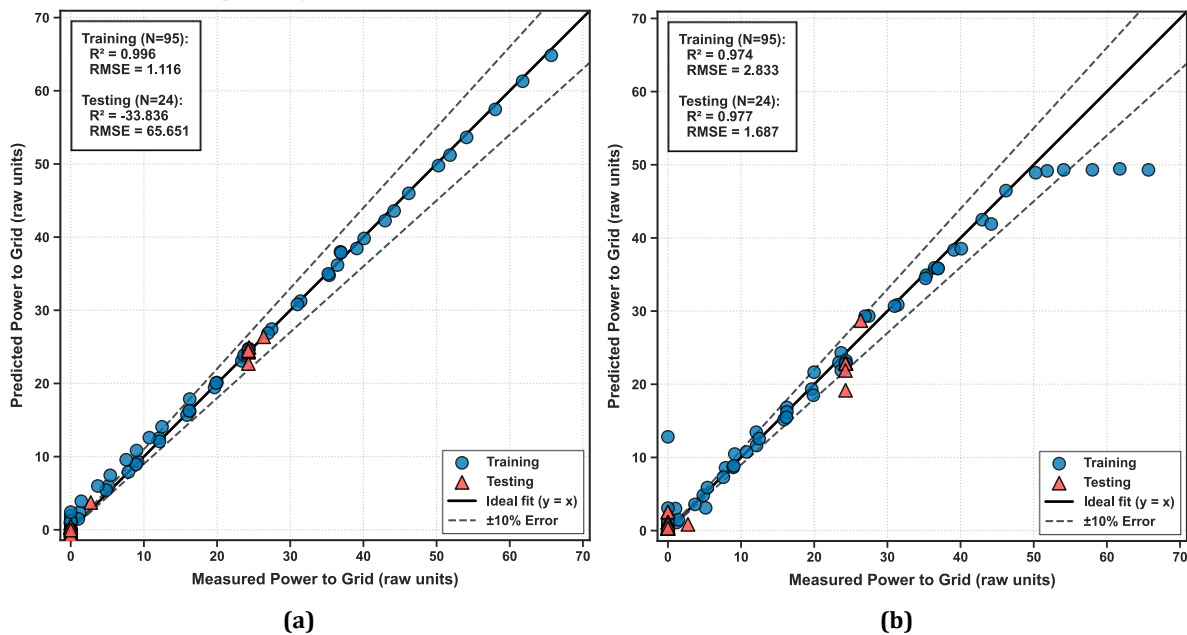


Figure 3. Experimental versus predicted grid power obtained using (a) LR and (b) XGBoost

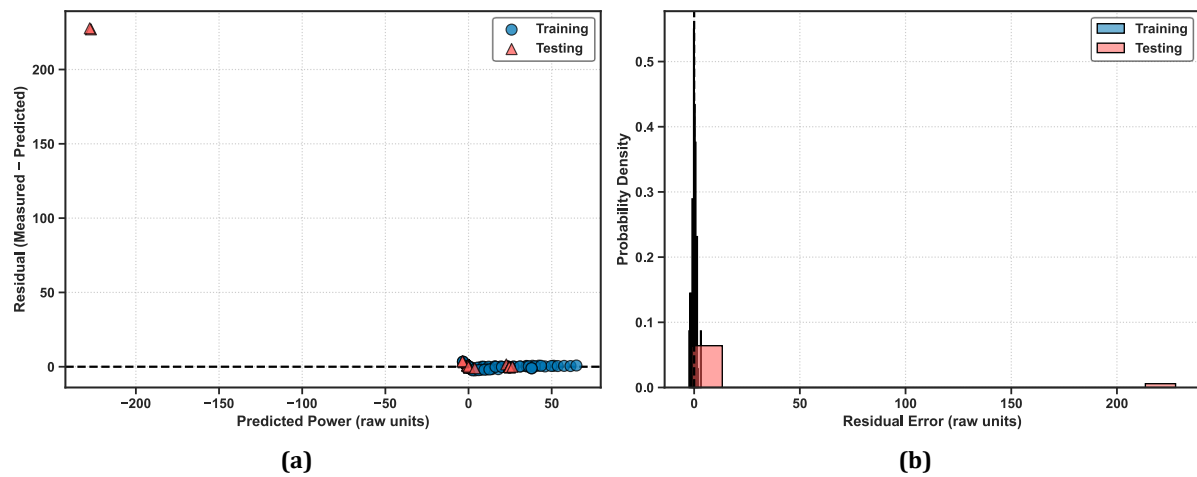


Figure 4. Residual analysis of the LR model: (a) residuals versus predicted grid power and (b) residual error distribution

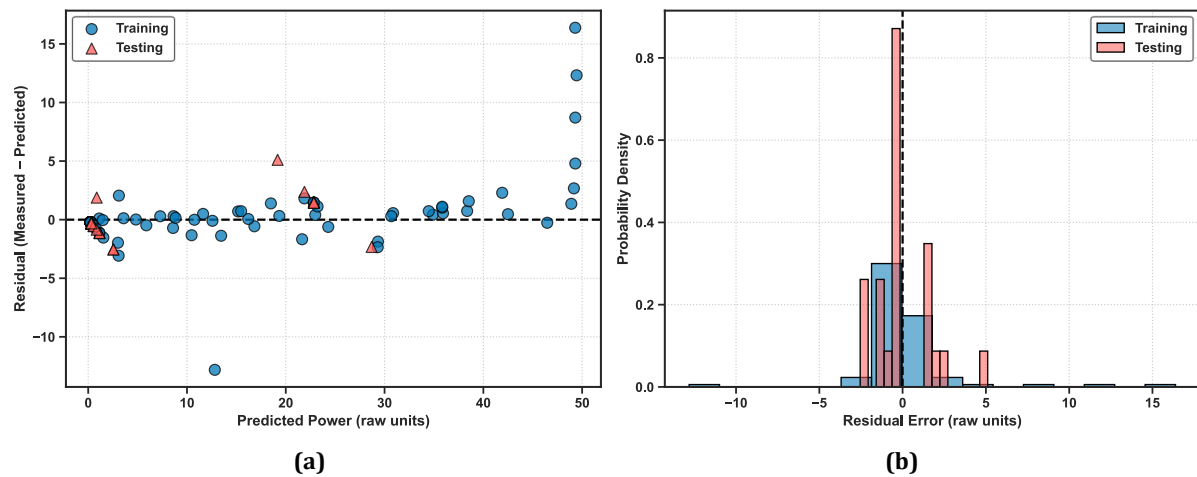


Figure 5. Residual analysis of the XGBoost model: (a) residuals versus predicted grid power and (b) residual error distribution

Table 2. Comparative performance of LR and XGBoost models for fuel-cell power prediction

Model	Dataset	n	R ²	MSE	RMSE	MAE	MAPE (%)
LR	Training	95	0.9960	1.247	1.116	0.810	11.28
LR	Testing	24	-33.8364	4310.074	65.651	19.722	5.89
XGBoost	Training	95	0.9740	8.024	2.833	1.212	9.67
XGBoost	Testing	24	0.9770	2.845	1.687	1.273	16.56

3.3. SHAP-Based Interpretation of Fuel-Cell Power Prediction

To further explore the XGBoost model, the SHapley Additive exPlanations (SHAP) model was used to find the operating parameters that affect the electrical power predicted by the model. The local feature contributions, SHAP dependence relationships, global feature-impact distribution, and mean absolute SHAP importance are shown in Figs. 6–9, respectively. These analyses help to interpret the individual predictions and the total effect of the variables investigated. A SHAP waterfall analysis is shown in Fig. 6, which shows the prediction for an individual test observation when the baseline model prediction, $E[f(X)]$, was 12.798 and the final model prediction, $f(x)$, was 22.839. The most positive contribution came from auxiliary consumption (+ about 11.41). Fuel-cell gas temperature and coolant outlet temperature also had a positive effect, around 1.30 and 0.38, respectively. On the other hand, the power-consumption setpoint had a reduction effect of about 3.12 on the predicted output. Approximately +0.07 was the contribution of the operating-status indicator with code 7, while the effects of ambient temperature and generation setpoint were negligible. These findings show the individual impacts of various factors on moving a particular prediction away from the model's average prediction.

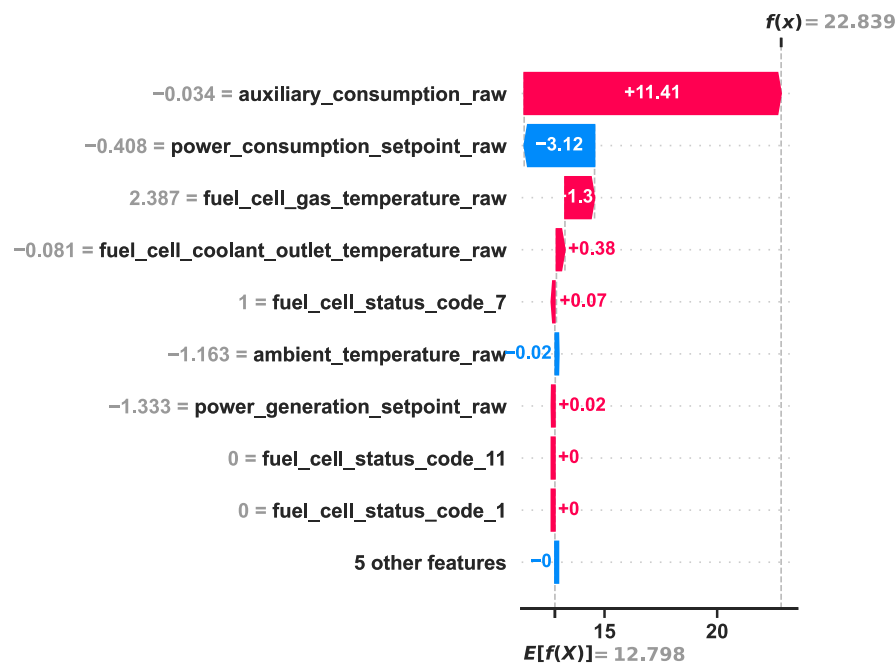


Figure 6. SHAP waterfall analysis illustrating individual feature contributions to an XGBoost power prediction

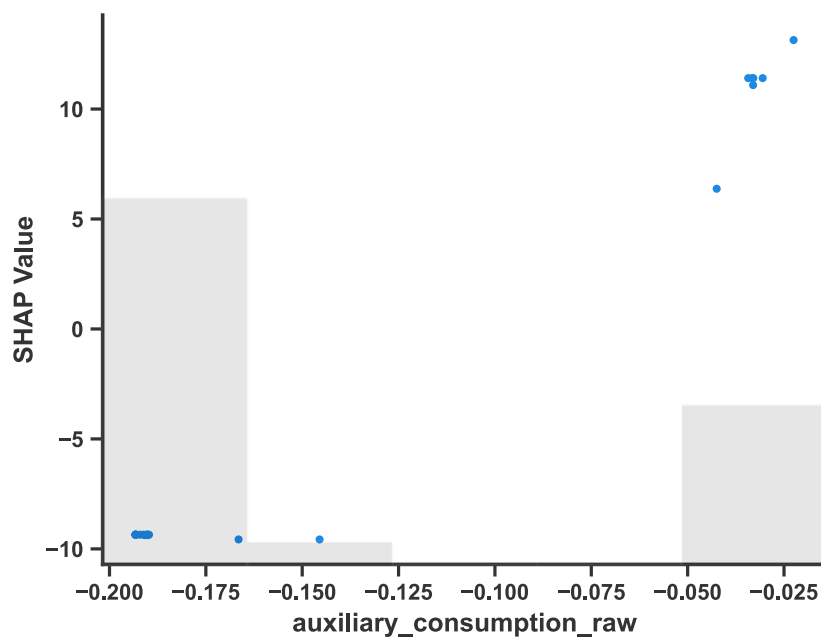


Figure 7. SHAP dependence analysis of auxiliary consumption for the XGBoost model

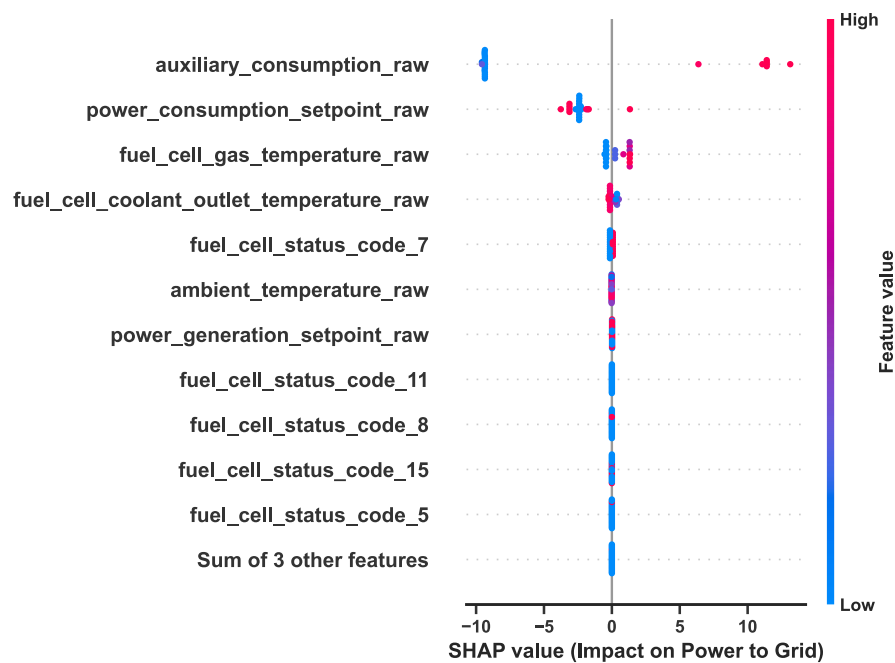


Figure 8. SHAP beeswarm plot showing the magnitude and direction of feature contributions to XGBoost predictions

The relationship between the auxiliary consumption and its SHAP contribution is further explored in Fig. 7. The plot reveals a significant range of SHAP values in the SHAP value space spanned by the auxiliary-consumption range, with some values being strongly negative (around -9) and some being strongly positive (above $+10$). The pattern of observations plotted indicates a nonlinear, possibly threshold response, and not a consistent linear response. Notably, the feature values on the horizontal axis seem to be transformed, and the values displayed should not be interpreted as physical auxiliary-power measurements without taking into account the preprocessing transformation. The observed relationship shows that the auxiliary consumption's contribution to prediction is dependent on the operating range of the auxiliary consumption and may be related to the interactions with other system variables. The distribution of global SHAP shown in Fig. 8 is another good indicator of the relative impact and direction of each feature. The distribution of SHAPs is the most spread out for auxiliary consumption, ranging around from about -9 to over $+10$. Some of the highest values of the features correspond to some of the highest positive contributions, while some of the lowest values of the features correspond to some of the lowest negative contributions. The power-consumption setpoint also shows significant variations in its contributions, with mostly negative effects for a few higher feature values. The gas temperature in the fuel cells has a relatively small distribution around zero, and the coolant outlet temperature, ambient temperature, power-generation setpoint, and operating-status indicators have relatively small effects. These distributions illustrate that the contribution of individual variables to the predictions varies from operating observation to observation. The mean absolute SHAP values in Fig. 9 quantitatively determine the global importance ranking of the input features. The top two contributors were auxiliary consumption (mean absolute SHAP value of 9.82) and the power-consumption setpoint (2.45). Fuel-cell gas temperature was the third, with a contribution of 0.72, followed by coolant outlet temperature (0.22) and the operating-status indicator corresponding to code 7 (0.11). The contributions of ambient temperature and the power-generation setpoint were much smaller at about 0.02, and the other operating-status indicators were negligible.

The high proportion of auxiliary consumption is especially significant as it had a direct linear correlation with grid power of only 0.04 (see Fig. 2). This contrast suggests that the variable has predictive information not well represented by the general linear correlation. This type of behavior can occur because of nonlinear operating relationships, from electrical setpoints, or as a result of unusually high consumption observations. Additionally, the relatively high auxiliary consumption variability from Table 1 indicates that extreme operating events could affect the calculated SHAP importance. So, the SHAP results should be seen as predictive associations of the model and not as actual evidence that higher auxiliary consumption leads to higher net electrical power. In general, the SHAP analysis shows that the measured thermal variables in the studied data set have a lower impact on the XGBoost predictions than the electrical operating variables.

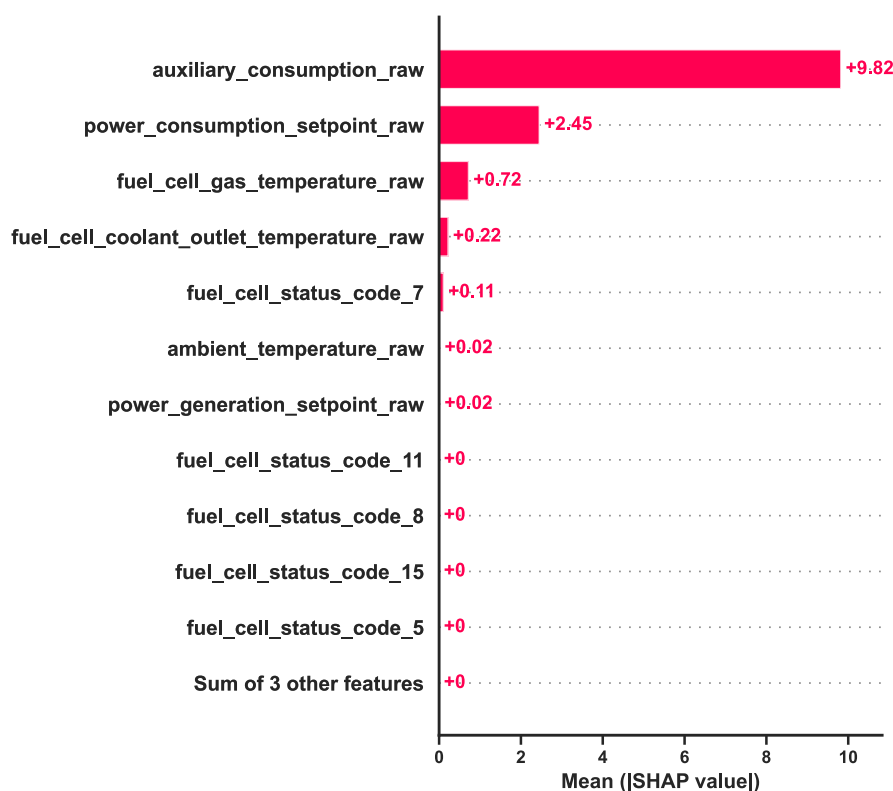


Figure 9. Global feature importance derived from mean absolute SHAP values for the XGBoost model

4. Conclusion

This study investigated the applicability of linear regression (LR) and Extreme Gradient Boosting (XGBoost) to electrical power prediction using 119 operational observations from a stationary hydrogen fuel-cell system. A chronological division allocated 95 observations for training and 24 for testing, while correlation analysis, residual diagnostics, and SHapley Additive exPlanations (SHAP) supported interpretation of the prediction results. Although LR produced an excellent training fit ($R^2 = 0.9960$; RMSE = 1.116), its testing performance deteriorated markedly ($R^2 = -33.8364$; RMSE = 65.651), reflecting sensitivity to operating conditions outside the relationships adequately represented during training. Conversely, XGBoost maintained comparable training and testing accuracy, with R^2 values of 0.9740 and 0.9770 and a testing RMSE of 1.687. Its more compact residual distribution confirmed improved robustness on the evaluated holdout segment. SHAP analysis ranked auxiliary consumption and the power-consumption setpoint as the most influential predictive variables, with mean absolute SHAP values of 9.82 and 2.45, respectively. The strong SHAP importance of auxiliary consumption, despite its weak linear correlation with output, suggests nonlinear associations that warrant further investigation. Future research should validate the framework using larger datasets from actual hydrogen-powered vessels, diverse duty cycles, and independent operating periods under realistic marine environmental and loading conditions. Additional work could incorporate fuel-cell degradation, battery interaction, hydrogen consumption, and uncertainty-aware forecasting to support practical maritime energy-management decisions. The present results establish a promising predictive approach without constituting shipboard validation.

References

- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16). <https://doi.org/10.3923/jeasci.2017.4102.4107>
- Amat Ventayol, A., Lam, J. S. L., Bai, X., & Chen, Z. S. (2025). Comparative life cycle assessment of hydrogen internal combustion engine and fuel cells in shipping. *International Journal of Hydrogen Energy*, 109. <https://doi.org/10.1016/j.ijhydene.2025.02.150>
- Bora, B. J., & Sharma, P. (2026). Towards sustainable and explainable dual-fuel diesel engine modeling: a SHAP-Based evaluation of Ammonia, Biogas, and hydrogen combustion dynamics. *Energy Conversion and Management*, 351, 121014. <https://doi.org/10.1016/j.enconman.2025.121014>
- Cavanini, L., Majecki, P., Grimble, M. J., & Van Der Molen, G. M. (2023). Energy Management Control of Hydrogen Fuel Cell Powered Ships. *Proceedings of the American Control Conference*, 2023-May. <https://doi.org/10.23919/ACC55779.2023.10156221>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-August-2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Coraddu, A., Tamburello, S., Löffler, C., Ceyhun, H. E., van Biert, L., & Oneto, L. (2025). State-of-Art Energy Management Strategies for Hybrid Fuel Cell Applications for Ships. In *Fuel Cell and Hydrogen Technologies in Maritime Transportation*. https://doi.org/10.1007/978-3-031-86110-9_6
- Darmawan, H., Yuliana, M., & Hadi, M. Z. S. (2023). GRU and XGBoost Performance with Hyperparameter Tuning Using GridSearchCV and Bayesian Optimization on an IoT-Based Weather Prediction System. *International Journal on Advanced Science, Engineering and Information Technology*, 13(3). <https://doi.org/10.18517/ijaseit.13.3.18377>
- Dzida, M., & Nguyen, L. H. (2026). A critical review of pathways, fuels, and challenges to achieving decarbonization strategies in the maritime sector. *Journal of Transportation and Sustainable Technologies*, 1(1). <https://doi.org/10.55228/jtst.2026.110249>
- Golbabaei, M. H., Saeidi Varnoosfaderani, M., Zare, A., Salari, H., Hemmati, F., Abdoli, H., & Hamawandi, B. (2022). Performance Analysis of Anode-Supported Solid Oxide Fuel Cells: A Machine Learning Approach. *Materials*, 15(21). <https://doi.org/10.3390/ma15217760>

- Hauke, J., & Kossowski, T. (2011). Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *QUAGEO*, 30(2), 87–93. <https://doi.org/10.2478/v10117-011-0021-1>
- Hoa, T. T. N., & Lien, M. H. (2025). Regional Balancing Strategy and Vietnam–Japan Maritime Security Cooperation in Shaping the Emerging Security Order in The South China Sea. *Journal of Cultural Analysis and Social Change*, 10(2). <https://doi.org/10.64753/jcasc.v10i2.1712>
- IMO. (2023). 2023 IMO Strategy on Reduction Of GHG Emissions From Ships. Resolution MEPC.377(80). International Maritime Organisation. In IMO.
- Inal, O. B. (2024). Decarbonization of shipping: Hydrogen and fuel cells legislation in the maritime industry. *Brodogradnja*, 75(2). <https://doi.org/10.21278/brod75205>
- Klass, L., Kabza, A., Sehnke, F., Strecker, K., & Hölzle, M. (2023). Lifelong performance monitoring of PEM fuel cells using machine learning models. *Journal of Power Sources*, 580. <https://doi.org/10.1016/j.jpowsour.2023.233308>
- Kumar, G., Punetha, D., & Pandey, S. K. (2025). Performance analysis of polymer electrolyte membrane fuel cells: Insights from sensor-based monitoring and machine learning. *Journal of Power Sources*, 659. <https://doi.org/10.1016/j.jpowsour.2025.238337>
- Le, A. T., Pandey, A., Sirohi, R., Sharma, P., Chen, W.-H., Pham, N. D. K., Tran, V. D., Nguyen, X. P., & Hoang, A. T. (2023). Precise Prediction of Biochar Yield and Proximate Analysis by Modern Machine Learning and SHapley Additive exPlanations. *Energy & Fuels*, 37(22), 17310–17327. <https://doi.org/10.1021/acs.energyfuels.3c02868>
- Lee, H., Jung, I., Roh, G., Na, Y., & Kang, H. (2020). Comparative analysis of on-board methane and methanol reforming systems combined with HT-PEM fuel cell and CO₂ capture/liquefaction system for hydrogen fueled ship application. *Energies*, 13(1). <https://doi.org/10.3390/en13010224>
- Malvoni, M., De Giorgi, M. G., & Congedo, P. M. (2017). Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia*, 126. <https://doi.org/10.1016/j.egypro.2017.08.293>
- Mokhtari, K. El, Higdon, B. P., & Basar, A. (2020). Interpreting financial time series with SHAP values. *CASCON 2019 Proceedings - Conference of the Centre for Advanced Studies on Collaborative Research - Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*.
- Nguyen, M. D., Nguyen, P. Q. P., Luong Ha, C. Q., Dinh, X. M., Cao, V. S., & Dat Do, H. (2025). Development of Intelligent Marine Logistics Models Using Machine Learning. *JOIV : International Journal on Informatics Visualization*, 9(1), 72. <https://doi.org/10.62527/joiv.9.1.3666>
- Pak, S. Il, & Oh, T. H. (2010). Correlation and simple linear regression. *Journal of Veterinary Clinics*, 27(4). https://doi.org/10.1007/978-3-319-89993-0_6
- Pandit, P., Dey, P., & Krishnamurthy, K. N. (2021). Comparative Assessment of Multiple Linear Regression and Fuzzy Linear Regression Models. *SN Computer Science*, 2(2). <https://doi.org/10.1007/s42979-021-00473-3>
- Panja, T., & Meharia, P. (2024). Enhancing Microbial Fuel Cell Performance Prediction and IoT Integration Using Machine Learning and Renewable Energy. *Journal of Strategic Innovation and Sustainability*, 19(1). <https://doi.org/10.33423/jsis.v19i1.6787>
- Penga, J., Vidović, T., Radica, G., & Penga, Ž. (2024). Analysis of Hybrid Ship Machinery System with Proton Exchange Membrane Fuel Cells and Battery Pack. *Applied Sciences (Switzerland)*, 14(7). <https://doi.org/10.3390/app14072878>
- Phan, Q. T., Wu, Y. K., & Phan, Q. D. (2021). A hybrid wind power forecasting model with xgboost, data preprocessing considering different nwps. *Applied Sciences (Switzerland)*, 11(3). <https://doi.org/10.3390/app11031100>
- Serbin, S., Burunsuz, K., Chen, D., & Kowalski, J. (2022). Investigation of the Characteristics of a Low-Emission Gas Turbine Combustion Chamber Operating on a Mixture of Natural Gas and Hydrogen. *Polish Maritime Research*, 29(2). <https://doi.org/10.2478/pomr-2022-0018>
- Shrote, A. B., Kiran Kumar, K., Kaur, C., Al Ansari, M. S., Singh, P., Hazela, B., & Madhu, G. C. (2024). Monitoring of operational conditions of fuel cells by using machine learning. *EAI Endorsed Transactions on Internet of Things*, 10. <https://doi.org/10.4108/eetiot.5377>
- SINTEF (Norway). (2024). HAEOLUS Project Fuel Cell Data. In *Hydrogen-Aeolic Energy with Optimised eLectrolysers Upstream of Substation (HAEOLUS) project*. Zenodo .
- Su, D., Zheng, J., Ma, J., Dong, Z., Chen, Z., & Qin, Y. (2023). Application of Machine Learning in Fuel Cell Research. In *Energies (Vol. 16, Number 11)*. <https://doi.org/10.3390/en16114390>
- Suherman, A. F., Lisnaeni, P. P., Izqiatullailiyah, S. A., Herlinawati, T., & Ahman. (2025). A Comparative Analysis of Spearman and Pearson Correlation Using SPSS. *OPTIMA: Journal of Guidance and Counseling*, 5(1).
- Szemer, M., Buchanec, S., Prokop, T., & Brus, G. (2025). Topology-informed machine learning for efficient prediction of solid oxide fuel cell electrode polarization. *Energy and AI*, 20, 100495. <https://doi.org/10.1016/j.egyai.2025.100495>
- Wang, Y., Seo, B., Wang, B., Zamel, N., Jiao, K., & Adroher, X. C. (2020). Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. In *Energy and AI (Vol. 1)*. <https://doi.org/10.1016/j.egyai.2020.100014>
- Xie, P., Asgharian, H., Guerrero, J. M., Vasquez, J. C., Araya, S. S., & Liso, V. (2024). A two-layer energy management system for a hybrid electrical passenger ship with multi-PEM fuel cell stack. *International Journal of Hydrogen Energy*, 50. <https://doi.org/10.1016/j.ijhydene.2023.09.297>
- Yang, K., Liu, J., Wang, Y., Shi, X., Wang, J., Lu, Q., Ciucci, F., & Yang, Z. (2022). Machine-learning-assisted prediction of long-term performance degradation on solid oxide fuel cell cathodes induced by chromium poisoning. *Journal of Materials Chemistry A*, 10(44), 23683–23690. <https://doi.org/10.1039/D2TA03944C>
- Ye, M., Sharp, P., Brandon, N., & Kucernak, A. (2022). System-level comparison of ammonia, compressed and liquid hydrogen as fuels for polymer electrolyte fuel cell powered shipping. *International Journal of Hydrogen Energy*, 47(13). <https://doi.org/10.1016/j.ijhydene.2021.12.164>

