

RESEARCH

Open Access



Enhanced fault detection in polymer electrolyte fuel cells via integral analysis and machine learning

Ester Melo¹, Julio Barzola-Monteses^{2,3}, Holguer H. Noriega⁴ and Mayken Espinoza-Andaluz^{5*}

*Correspondence:
masespin@espol.edu.ec

¹ Escuela Superior Politécnica del Litoral, 09-01-5863 Guayaquil, Ecuador

² Centro de Estudios en Tecnologías Aplicadas, Universidad Bolivariana del Ecuador, 092405 Durán, Ecuador

³ Artificial Intelligence and Information Technology Research Group, University of Guayaquil, 090514 Guayaquil, Ecuador

⁴ Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, 20133 Milan, Italy

⁵ Centro de Energías Renovables y Alternativas, Facultad de Ingeniería Mecánica y Ciencias de la Producción, Escuela Superior Politécnica del Litoral, 09-01-5863 Guayaquil, Ecuador

Abstract

The growing energy demand and population raising require alternative, clean, and sustainable energy systems. During the last few years, hydrogen energy has proven to be a crucial factor under the current conditions. Although the energy conversion process in polymer electrolyte fuel cells (PEFCs) is clean and noiseless since the only by-products are heat and water, the inside phenomena are not simple. As a result, correct monitoring of the health situation of the device is required to perform efficiently. This paper aims to explore and evaluate the machine learning (ML) and deep learning (DL) models for predicting classification fault detection in PEFCs. It represents a support for decision-making by the fuel cell operator or user. Seven ML and DL model classifiers are considered. A database comprising 182,156 records and 20 variables arising from the fuel cell's energy conversion process and operating conditions is considered. This dataset is unbalanced; therefore, techniques to balance are applied and analyzed in the training and testing of several models. The results showed that the logistic regression (LR), k-nearest neighbor (KNN), decision tree (DT), random forest (RF), and Naive Bayes (NB) models present similar and optimal trends in terms of performance indicators and computational cost; unlike support vector machine (SMV) and multi-layer perceptron (MLP) whose performance is affected when the data is balanced and even presents a higher computational cost. Therefore, it is a novel approach for fault detection analysis in PEFC that combines the interpretability of different ML and DL algorithms while addressing data imbalance, so common in the real world, using resampling techniques. This methodology provides clear information for the model decision-making process, improving confidence and facilitating further optimization; in contrast to traditional physics-based models, paving the way for data-driven control strategies.

Keywords: Classification problems, Deep learning models, Fault detection, Machine learning models, PEFC

Introduction

With global energy demand expected to increase between 25 and 58% by 2050, finding clean, sustainable alternatives to fossil fuels is more critical than ever (Ruijven et al. 2019). This situation worsened with the COVID-19 epidemic because mitigation

practices and measures had gradual impacts on energy demand and consumption (Jiang et al. 2021). For example, in Saudi Arabia, a virus lockdown in 2020 caused a 16% increase in electricity consumption compared to the level in 2019 (Aldubyan and Krarti 2022). However, academic literature indicates that spatial and temporal variations make each region face its challenges in the energy sector (Jiang et al. 2021).

During the last years, hydrogen energy has proven to be a crucial factor under the current conditions (Yue et al. 2021). Hydrogen is a suitable energy carrier that can be used at different stages of the energy cycle. An electrochemical device that converts chemical energy into electrical energy in a noiseless and efficient manner is the fuel cell (FC) (Saikia et al. 2018). In general, hydrogen has managed to establish itself in some specific niches. A notable example is forklifts, where hydrogen fuel cells have gained popularity due to their efficiency and performance. While hydrogen vehicles are already commercially available in several countries, indicating significant progress in the adoption of this technology. In the field of domestic use, fuel cells have also experienced notable growth, with 225,000 heating systems sold. This shows that FC technology is starting to gain traction in the market, which is a crucial step towards widespread adoption (Staffell et al. 2019).

Even though it seems that the fuel cell and their different classification is a recent development, this technology originated in the nineteenth century with the first studies of the Welsh physicist and jurist Sir William Grove on gaseous batteries, whose results were published in 1843 (Grove 1839). In addition, another study was presented by Christian Friedrich Schönbein in Switzerland (Klell et al. 2023). Within the group of fuel cells, there is a special place for polymer electrolyte fuel cells (PEFCs), which have demonstrated prospective energy conversion due to their high energy efficiency, minimal operating temperature, soft emissions, and quick beginning (Calili-Cankir et al. 2022). So, this technology can power zero-emission vehicles to provide clean, reliable electricity in remote areas; this makes its global impact revolutionary (Ogawa et al. 2018).

Although the energy conversion process in a PEFC is clean, since the only by-products are heat and water, the inside transport phenomena are complex. As a result, correct monitoring of the health situation of the device is required to perform efficiently (Carrette et al. 2000). Inside a PEFC, the energy conversion is given by the electrochemical reactions in the anode and cathode reaction zones. Inlet temperature, relative humidity, and inlet pressure are some variables that must be considered to evaluate the device's behavior. Also, thermal and water management directly affects PEFC health and behavior (Melo et al. 2020; Nguyen and White 1993).

The main constitutive element affected by the operating hours is the electrolyte, i.e., a Nafion membrane, which allows only the passing of positive ions. At the same time, the electrical energy is used in the external circuit. The membrane can be subjected to excess water (flooding) or low humidity (dry), and then the PEFC would fault. Moreover, PEFCs are exposed to different defects that can reduce performance and even result in total failure (Rama et al. 2008). PEFC technology is still immature and should be more reliable and sturdier than presently (Lee et al. 2018, 2017; Mekhilef et al. 2012). Device lifetime is the primary challenge to a successful deployment of PEFCs. Thus, diagnosis and fault detection systems (FDD) are fundamental to guarantee the reliable operation of PEFCs (Li et al. 2019).

This research aims to explore and evaluate the capabilities of the machine and deep learning models for predicting classification fault detection in PEFCs, which will support decision-making by the fuel cell operator or user. For this purpose, seven machine and deep learning model classifiers are considered. Experiments are performed using a database of 182,156 records and 20 variables arising from the fuel cell's energy conversion process and operating conditions.

The rest of the paper is organized as follows. "[Literature review](#)" section summarizes works related to this study. "[Methodology](#)" section presents dataset features, resampling techniques, and a brief explanation of the ML and DL techniques to be applied. "[Experimental results](#)" section shows the experimental evaluations and discussion of the results. Finally, "[Discussion](#)" section includes some final comments and directions for future work.

Literature review

In the Fuel Cell technology and within the specific PEFC's area, advances in Machine learning (ML) and deep learning (DL) have facilitated the development of fuel cell design, material choice, correlations, system control with FC, power management, and monitoring of operation health of FC (Wang et al. [2020](#)).

ML is the most broadly utilized subclass of AI, while DL uses multi-layered neural networks to learn from enormous amounts of information. According to a US Department of Energy study, ML, DL, and artificial intelligence (AI) have increased interest from energy development and material designers. The number of copyrights of these areas in the energy field in 2000–2017 has increased considerably (US Department of Energy [2023](#)).

In that sense, many ML and DL methods can be used to perform the analysis of PEFC. Through this study, the most widely used will be analyzed in detail.

One of the first methods of ML to be analyzed in this study is the so-called Logistic Regression (LR). In several studies, LR is used in part with other ML methods in the analysis of PEFC. For instance, in Eslamibidgoli et al. ([2021](#)), researchers presented a convolutional neural network method (ConvNets) for the high throughput testing of spread electron microscopy images at the ink phase. The first steps are data sampling and data augmentation at multiple levels and using a network model pre-trained on a large generic natural images dataset (ImageNet) for transfer learning. A final part of the methodology is training an LR model on attributes obtained from transfer learning. According to the results, preparing the training set with Selective Search (SS) followed by a pre-trained model for feature extraction and training the LR model for classification was a fast and accurate process for the presented problem. Finally, it was concluded that convolutional neural networks are skilled at quick distinction between non-good samples and the optimal ones. Moreover, the algorithms can be trained for fabrication processes, e.g., recognition of the slurry during the fabrication of Li-ion electrodes.

Another case in which LR is employed with other ML methods is presented in Xing et al. ([2022](#)). In the mentioned study, a Fault Detection and Isolation (FDI) algorithm was developed on a matrix that describes the connection between faults and residuals called the fault signature matrix. The efficiency of the proposed method is proven by comparing the diagnostic results in this work and those obtained by support

vector machine (SVM) and LR. In addition, the authors presented a novel data-driven approach based on sensor pre-selection and artificial neural networks (ANN). Many steps were accomplished in this study: first, a sensitivity analysis was applied to get the features of sensor data in the time domain and frequency domain. Later, a filter procedure was used to exclude the sensors with poor response to the changes in system states. Ultimately, the investigational data supervised by the outstanding sensors are employed to verify the model using the ANN model. Levenberg–Marquardt (LM) algorithm, resilient propagation (RP) algorithm, and scaled conjugate gradient (SCG) algorithm are utilized in the neural network training, respectively.

In recent years, other cases of LR usage have been reported in the literature. For instance, in Chen et al. (2024), the researchers used logistic regression models to evaluate the effects of continuing environmental air contamination and use of solid fuel on postmenopausal females with 10-year high cardiovascular disease (CVD) risk. Another recent study is explained in Modanloo et al. (2024) where many approaches are used for the manufacturing of microchannels in bipolar plates (BPPs) of a proton exchange membrane fuel cell (PEMFC) using the stamping process. During the processes, a regression model (RM) is used to predict the filling rate of the microchannel that comes from the response surface method (RSM).

Specifically, SVM is another ML method of analysis. An example of SVM's use in the PEFC analysis is presented in Li et al. 2014a, where an online fault diagnosis of twenty PEFC stacks was implemented. The proposed approach used SVM and a fluidic model to diagnose the faults in water management. The health states of the stack can be found using the fluidic model. This method is then dedicated to identifying the training data used to train the dimension reduction model, Fisher discriminant analysis (FDA), and the SVM classifier.

Another application case of SVM is presented in Han and Chung (2017). In this work, researchers gave a hybrid model blending an SVM model with an empirical equation about the polarization curves of a PEMFC under different working conditions. Various processes, such as validating, testing, and training, were applied to the hybrid model. In this way, working data was obtained. The training part determined the model coefficients and hyper-parameters of the model. The output predicted by the trained hybrid model, i.e., the polarization curves, were similar to the measurements for both the validation and the testing data set. Finally, they demonstrated that the hybrid model compensated the large prediction errors in the high voltage range of the polarization curves obtained by the singular SVM model. A final recommendation by the study exposed that this mixed model can predict the cell voltages of PEM fuel cells with high accuracy even when the main running variables (the oxygen relative humidity, the hydrogen and oxygen bay temperatures, the stack temperature, and the current density) vary.

Recently scholars in Quan et al. (2024) used SVM as a comparison method to test the diagnostic accuracy of an enhanced fault diagnosis method for fuel cell systems using a kernel extreme learning machine optimized with an improved sparrow search algorithm. The results based on the observations of the paper were 10.4% higher than support vector machine (SVM), and the operation time is only slightly higher than that of the SVM model.

In (Ma et al. 2022) scientists designed a driving pattern recognition with a model predictive control (DPR-MPC) to be used in a fuel cell hybrid electric vehicle (FCHEV) to tackle a technological challenge in FCHEV which is the power allocation between proton exchange membrane fuel cell (PEMFC) and lithium-ion battery remains. Among the methods, three SVMs were utilized to construct the recognizer, and the fuel cell efficiency was introduced into the cost function. Finally, the results of this study showed that the proposal avoided PEMFC and reduced fuel consumption by 6.67%.

The diagnostic accuracy of the proposed method is 10.4%, 5.7%, 4.8%, 4.2%, 3.0%, 1.8% higher than support vector machine (SVM), back propagation neural network (BPNN), Kernel Extreme Learning Machine (KELM), genetic algorithm-based KELM (GA-KELM), particle swarm optimization-based KELM (PSO-KELM) and SSA-KELM, respectively, and the operation time is only slightly higher than that of the SVM model and KELM model (Quan et al. 2024).

A third case of the ML approach is K-Nearest Neighbor (KNN). The use of KNN in the analysis of PEFC is presented in Onanena et al. (2012). Here, the authors assessed an algorithm based on KNN procedures and a multiclass linear discriminant analysis (LDA) classifier to identify drying and flooding on a PEMFC. This paper reports a form appreciation based on a judgment approach for fault-diagnosing 20 fuel cell stacks using Electrochemical Impedance Spectroscopy (EIS). To implement this methodology, researchers followed different steps: first, measurable features were extracted. Later, a fault categorization, a correlation-based feature selection, was executed, getting only significant features. Lastly, an assignment of each observation to one of the predefined defect classes was completed. According to the research, the results found with the two proposed feature extraction methods (FE1, FE2) and the two classifiers (LDA, KNN) correspond to the true class (false positive rate, true positive rate). Finally, this study concludes the KNN classifier with the FE1 features performs the maximum classification rate (99.6%).

A different case of using a KNN classifier in the context of other methods is explained in Detti et al. (2017). In this study, authors approach a PEMFC, drying out, and flooding diagnosis based on pattern recognition. The KNN method is used with KMEAN methods to find faults in PEMFC. The Total Harmonic Distortion (THD), computed using the Fast Fourier Transform of the voltage signal, is used as a signifier for an overall Pattern Recognition approach, i.e., fault detection and classification through a classification approach. In overview, the discovery of faults through the variation in the total harmonic distortion level and the frequency spectrum is initially applied. Then, based on supervised and non-supervised classification, fault documentation is performed, with an upright sorting rate of 84% and 98.5%.

In the last years, other studies about using the K-nearest neighbor (KNN) approach have been reported in Fuel Cell literature. For example, in Saxena et al. (2024) scholars propose a hybrid KNN-SVM machine learning approach for solar power forecasting to increase the exactitude of forecasting solar power from solar farms, and based on the results of this study, the precision of the predictions by 98%. One of the positive consequences of this method mix is the rise in the reliability of power system operators. On the other hand, in Awasthi et al. (2024) the researchers found that the use of KNN enhanced the grouping of faults in distribution networks with several generators.

The KNN approach was based on the derivation of a database of fault events through a steady-state and short-circuit analysis. Apart from the efficacy shown by the KNN-based method, the approach considered the source impedance variation during faults, easing the fault type classification.

An alternative ML method is the Decision Tree (DT). A case study is examined in Santamaria et al. (2020) where difficult associations between pressure drops in PEMFC reactant channels and liquid phase behavior were disclosed. This study uses a supervised DT algorithm to develop new diagnostic tools to confirm flow-field design. The trained network predicts pressure drop down a reactant channel when a projected image of liquid phase distribution in the channel is used as an input. The results show significant improvement over prior efforts utilizing wetted area values to characterize fuel cell performance. This can result in high prediction error since the same water area can have numerous pressure signatures depending on its distribution.

Another example of DT usage is explained in Santamaria et al. (2021), where this method takes part of the overall methodology. This work presents a novel technique for two-phase data collection processing and its use in an ML algorithm. DT regression correlates the liquid allocations in reactant canals with the two-phase flow pressure drop along a 2.4 mm × 3.00 mm transparent channel by inserting liquid through a gas-diffusion layer (GDL), during which air was streamed through the channels. The DT models completed the pressure drop prediction with 90% accuracy, working with the liquid distributions as inputs and the related pressure drop data as outputs. The techniques were used in applications for fluid saturation estimation based on pressure and flow-field design.

Recently, other scholars found interesting results about the use of DT. One case to mention is shown in Lu et al. (2024). As part of a study on counter-flow mass transfer characteristics and performance optimization of commercial large-scale proton exchange membrane fuel cells (PEMFC), scholars shaped the PEMFC flow field model with a decision tree. In the scholar's opinion, the reason for using DT was due to the ability to threaten complex datasets like in unknow distributions generating flexibly, and automatic adaptation of the dataset to specific characteristics. The model was trained with a dynamic dataset of 20% testing and 80% training features.

Another case of the use of DT is presented in Hai et al. (2024). In this study, research modeled a solid oxide fuel cell power plant combined with an absorption-ejection refrigeration cycle. The DT is used in the optimization section, where it was used with other machine learning methods including support vectors, and neural networks to reduce cost and computational time. With this optimization part, the optimization of this cycle shows that the ratio between the output hydrogen product and feedstock is enhanced up to 68%, with a cost rate reduced up to 9.7 dollars per hour.

A complementary ML method is the Random Forest (RF). A case about using RF and other ML methods is analyzed in the context of PEMFC technology. For instance, in Vaz et al. (2023), scholars look for new and alternative classes of membranes within the cells. Sulfonated polyimides (SPIs) based hydrocarbon membranes have been exploited in the work mentioned. Both supervised and unsupervised ML approaches are developed to predict the proton conductivity of SPIs; for instance, a Random Forest regression (RFR) model detected an extra set of features that can foresee proton conductivity with

acceptable error. With this knowledge, correspondence about the features of the proton conductivity class labels has been explored. Then, the design of novel SPI polymer electrolyte membranes is possible while relating proton transport at the ionomer stage with factors such as the inter-chain interactions and morphology of the microstructure.

Another case of RF application is presented in Huo et al. (2021), where the RF algorithm, together with convolutional neural networks (CNN), are used as the basis of a performance forecast approach to reduce the number of needless experiments for the membrane electrode assembly (MEA) in the PEMFCs. In this research, the RF algorithm is used to choose the important components as the input of the model, as proved previously in other studies, enhancing the quality of the training dataset. On the other hand, CNN is used to create the performance forecast model, in which the I-V polarization curve is the output. All these steps are conducted because achieving the I-V polarization curve is a complex process, including thermodynamic, electrochemical, and hydrodynamic fields.

Another case of ML process is proposed in Zheng et al. (2017), where the research uses Bayesian methods to detect flooding and drying up in the PEMFC with records from the EIS in an offline manner. In this study, a naive Bayesian classifier was the methodology elected. Six operating modes (normal mode, moderate flooding, minor flooding, light flooding, moderate drying, and minor drying) with twelve input variables. A real PEMFC experiment dataset is adopted to confirm the approach's results. According to this experimentation, results show that the suggested model can effectively recognize the flooding fault of the fuel cell with over 99% accuracy under load-varying conditions.

On the other hand, an interesting method of DL is Multi-Layer Perceptron (MLP). This is commonly used in the analysis of PEFC technology. In (Vaz et al. 2023), the study used MLP with another surrogate model called Response Surface Analysis (RSA) together with optimization algorithms like NSGAI and Particle Swarm Optimization (PSO) to find out the optimal cathode catalyst layer (CL) parameters, i.e., platinum loading, the weight ratio of ionomer to carbon, the weight ratio of Pt to carbon, and porosity of the cathode CL. First, MLP, RSA, and PSO were combined into a single-objective optimization where the cell performance was maximized. Consequently, a multi-objective optimization was run with MLP integrated with the NSGAI algorithm, where the objective function was maximized while the overall PEMFC stack price was minimized.

Scholars in Quan et al. (2024) present another case of MLP application. In this study, authors find new and alternative ways to model PEMFC systems, which are nonlinear systems with multiple inputs (drawn current, the gas pressures at the anode and cathode side, and the humidity of these gases) and a single output, i.e., the cell voltage. The experimental data for identification is limited. Thus, this study investigates ANN as an MLP network with different numbers of unseen neurons. Following the other part of the study, a black-box model based on semi-empirical models available in the literature is informed. Six experimental campaigns were carried out for parameter identification and model validation to optimize the PEMFC behavior.

Not long ago, other scientists had other studies using MLP techniques in fuel cells. One case of this is shown (Ghorbanzade Zaferani et al. 2024) where the power density and voltage behavior of 13 glucose fuel cells were optimized through multi-layer perceptron (MLP) together with other techniques such as response surface methodology

(RSM) and machine learning (specifically, Artificial Neural Network- ANN). Indeed, many tests were developed by using RSM followed by a mixed MLP-ANN. Finally, the researchers presented multi-objective optimization for the power density and the voltage of the fuel cells, obtaining the best optimum mode of each type of FC. Based on the results of MLP performance, the authors state that the proposed models forecast the fuel cells' power density and voltage behavior.

As part of the Neural Network (NN) method of different hybrid electric cars, in Liu et al. (2024) scholars used Multi-Layer Perceptron (MLP) controllers with numerous hidden layers. In this study, an adjustable seek-out variable that is the battery SoC as an input to the NN-based approach of the ending state of the Equivalent Consumption Minimization Strategy (ECMS), was introduced. Based on the results, scholars have stated the high adaptability of different charging/discharging modes of the battery in conventional networks considering state limitations. Moreover, the proposed NN method has reduced computation costs and time by about 95% with a very low loss of optimality in comparison with dynamic programming approaches. Additionally, the NN method demonstrated that it could save fuel by more than 3% on average.

In summary, different methods excel in various aspects associated with PEFC analysis, including fault detection, material selection, design optimization, and performance prediction. For example, LR and DT are well-suited for tasks such as fault diagnosis and pressure drop prediction, while KNN and RF excel in feature extraction and classification. MLP works well for modeling and optimization. However, combining multiple methods often leads to improved performance and accuracy. Therefore, exploring and evaluating automatic and deep learning models to determine and compare their performance against the tasks involved in PEFC analysis represents a significant contribution to facilitating the understanding of the behavior of this type of cell, as well as the development of new analysis methods.

Methodology

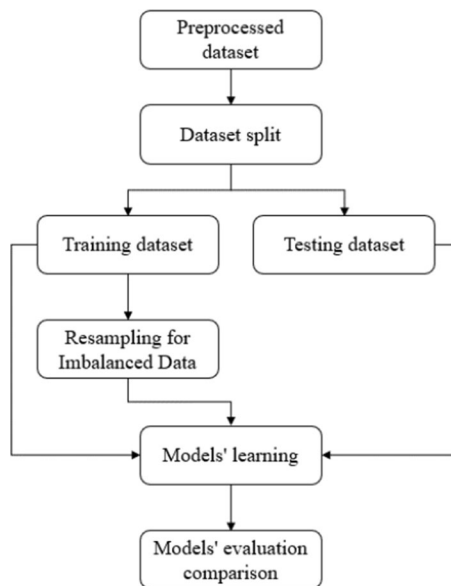
Description of the features and data set split

The original database comprises 182,156 records and 20 variables arising from the fuel cell's energy conversion process and operating conditions (Mao et al. 2017). However, for this study and considering the literature and results presented in the preprocessing stage (Melo et al. 2022; Mao et al. 2021), only 12 variables or features were used. In detail, the preprocessing included the following stages: cleaning, Integration or exploratory data analysis (EDA), transformation, and reduction. These steps allowed us to eliminate inconsistent, incomplete, and erroneous data; as well as reduce the size of the data without losing important information using the filter method that was evaluated with the ANOVA test. Consequently, the data were prepared in such a format that it can be suitable for analysis of ML or DL models (Melo et al. 2022). Additionally, a categorical variable named "state" was created, which contains the state of health of the PEFC: normal or faulty. The description and unit of measurement of each of the variables are described in Table 1.

According to the types of health states of the PEFC, 94% of the data corresponds to normal operating conditions, and the remaining 6% corresponds to the failure state. The database maintained an imbalance that favored the normal state. Tyagi & Mittal

Table 1 Recorded measurements from the PEFC preprocessed database

Assigned name	Description	Unit
Current	Current	A
Anode in p1	Anode inlet pressure from stack 1	mbar
Anode in p2	Anode inlet pressure from stack 2	mbar
Cathode air in	Cathode air inlet flow	SLPM
Cathode in t2	Cathode inlet temperature from stack 2	°C
Cathode in t1	Cathode inlet temperature from stack 1	°C
Water in p2	Primary water inlet pressure from stack 2	mbar
Water in p1	Primary water inlet pressure from stack 1	mbar
Water in f2	Primary water inlet flow from stack 2	SLPM
Water in f1	Primary water inlet flow from stack 1	SLPM
Water in tp	Water inlet temperature	°C
State	State of health of the fuel cell	N/A

**Fig. 1** Flow chart of fault classification procedures

(Tyagi and Mittal 2019) explain that although real-world datasets are unbalanced, they negatively affect the accuracy of class predictions in classification problems. This bias can be managed by either oversampling the minority classes or undersampling the majority class.

Therefore, the unbalanced database was divided into training and test data before applying the ML and DL classification algorithms. Then, techniques were used to balance the database, and the training and testing procedure was repeated. This is to compare the results of the classification algorithms in both the unbalanced and balanced databases. For a better understanding, Fig. 1 summarizes the applied methodology.

Resampling techniques

Synthetic minority over-sampling technique (SMOTE)

SMOTE refers to a sampling method in which the minority classes of a feature set are oversampled from synthetic examples to rebalance the original training data set. Following the KNN logic, new classes are created by interpolating between several minority class instances that lie within feature space (Tyagi and Mittal 2019; Fernandez et al. 2018). Due to its simplicity and robustness when applied to different types of problems (Azad et al. 2022; Demidova and Klyueva 2017; Rupapara et al. 2021), SMOTE has become a standard in learning from unbalanced data (Fernandez et al. 2018). Specifically, in recent years, the application of this method has been extended to fuel cell fault detection problems, improving the accuracy of learning model evaluation (Wang et al. 2021; Zhang et al. 2020; Xiao et al. 2022).

The application of this method in the present study consisted of using the imbalance-learn library of the scikit-learn package in Python. This library contains the technique which was applied to the training data set. Afterward, the different classification models were used considering the balanced training set.

Near-miss under-sampling technique

Near-Miss is a subsampling method that removes records corresponding to the majority class from the training data set to reduce the distribution bias. Like SMOTE, this method follows the KNN logic for subsampling (Peng et al. 2019). Likewise, it has proven to be a robust technique in different applications on classification problems with large-scale unbalanced datasets (Bao et al. 2016; Mqadi et al. 2021). Despite its contribution to improving the performance of classification learning models, its scope is still limited in the field of renewable energies (Kulkarni et al. 2021). There is no evidence of its application to fuel cell fault detection problems.

As in the previous method, the imbalance-learn Python library was used for the Near-Miss application. In this library, the Near-Miss technique has three versions: (a) the first version selects the positive samples with the smallest mean distance to the nearest samples of the negative class; (b) the second version selects the positive samples with the smallest mean distance to the farthest samples of the negative class; (c) and the third version is based on keeping the nearest neighbors of each negative sample and then selecting the positive samples based on the mean distance between them and their nearest neighbors (Lemaître et al. 2016). The third version is applied to the training data set in this case.

Machine learning approach

Based on the review of the state of the art shown in the previous session, the following algorithms have been selected for this study:

Logistic regression (LR)

The logic of LR is based on the two possible outcomes, usually represented as "success" or "failure", or in math language "1" or "0". Moreover, the sigmoid function which is an S-shape curve is used to determine the connection between the independent

variables and the probability of the dependent variable. Then, this relationship becomes a linear connection, applying a natural logarithm of odds ratio, that is the proportion of the probability of success and the probability of failure (Madushani et al. 2023).

The LR is a technique used for statistical modeling in which the probability, P_1 , of the dichotomous outcome event is related to a set of explanatory variables in the form (Eslamibidgoli et al. 2021; Schumacher et al. 1996; Vach et al. 1996):

$$\text{logit}(P_1) = \ln\left(\frac{P_1}{1-P_1}\right) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \dots + \beta_nx_n = \sum_{i=1}^n \beta_ix_i \quad (1)$$

In Eq. (1), β_0 is the intercept and $\beta_1, \beta_2, \dots, \beta_n$ are the coefficients associated with the independent variables x_1, x_2, \dots, x_n . For the coefficients $\beta_0, \beta_1, \dots, \beta_n$, the maximum likelihood estimation (MLE) is the most common method used.

The LR computes the variations in the logarithm of odds of the response variable instead of the shifts in the response variable itself. There are many kinds of independent (explanatory) variables. They could be dichotomous, discrete, continuous, or combinatorial. Usually, LR does not suppose linearity of the relationship between the explanatory variables and the variable response, and Gaussian-distributed independent variables are not required. The regressed correlation between the explanatory variables and their response is not linear because the logarithm of odds is linearly related to the explanatory variables. The chance of an event as a function of the explanatory variables is nonlinear, as derived from Eq. (1):

$$P_1(x) = \frac{1}{1 + e^{-\text{logit}(P_1(x))}} = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_ix_i)}} \quad (2)$$

For the specific case of Eq. (2), $\beta_0 = -\frac{\mu}{s}$ known as the intercept, here μ is the location parameter, i.e., the midpoint of the curve, where $\rho(\mu) = 0.5$, and s is a scale parameter. Moreover, $\beta_1 = \frac{1}{s}$ is a rate parameter.

The LR will impose the probability values of $P_1(x)$ to lie between 0 and 1 ($P_1 \rightarrow 0$ as the right-hand side of Eq. (2) approaches $-\infty$, and $P_1 \rightarrow 1$ as it approaches $+\infty$).

For example, if you want to predict whether a PEFC is working correctly or not, LR analyzes the PEFC data and calculates the probability that it is functioning correctly. That is, it is like a calculator that answers 0 (probable failure) and 1 (correct operation).

Support vector machine (SVM)

SVM is a classification method proposed by Platt (Platt 1998), which has been accepted in diagnosing fuel cell failures from a binary perspective. In other words, as Li et al. (Li et al. 2014b) indicate, the SVM is based on a hyperplane that looks for an optimal point to separate the data into two classes and, in turn, maximizes the margin between the hyperplane and the training records closest to each category (support vectors).

The binary SVM consists of training and performing processes based on collected $(N_1 + N_2)$ labeled samples. $z_1, z_2, \dots, z_{N_1+N_2}$ from classes z_1 and $z_{2*g_n} \in \{-1, 1\}$ is the class label of the sample z_n (-1 for class 1, 1 for class 2). Then, it solves the following quadratic problem:

$$\min J(x) = \frac{1}{2} \sum_{n=1}^{N_1+N_2} \sum_{m=1}^{N_1+N_2} x_n x_m g_n g_m k(z_n, z_m) - \sum_{n=1}^{N_1+N_2} x_n \tag{3}$$

$$s.t. \sum_{n=1}^N x_n g_n = 0, 0 \leq x_n \leq D \text{ for } n = 1, 2, \dots, N.$$

The constraints expression below of $\min J(x)$ is a set of linear restrictions, typically contending with optimization problems where x_n and g_n are variables or parameters. Moreover, N is the number of variables to be optimized, and D is the upper limit for each x_n . On the other hand, $\sum_{n=1}^N x_n g_n$ is the linear combination of the two variables. Altogether, are required to guarantee specific requirements are encountered.

where $x = [x_1, x_2, \dots, x_{N_1+N_2}]^T$, $k(z_n, z_m)$ is a kernel function. Save support vectors: $z_1^s, z_2^s, \dots, z_S^s$ and corresponding g_n and x_n , which are denoted by $\{g_n^s\}$ and $\{x_n^s\}$. Support vectors are those samples whose corresponding $x_n > 0$.

Sometimes, SVM can represent the input data like spaces with higher-dimensional characteristics to deal with an easier and separate hyperplane. This approach is called the kernel trick. linear, polynomial, radial basis function (RBF), and sigmoid kernels are cases of usual kernel functions.

For a new sample z , its class label is determined by:

$$g = \text{sign} \left\{ \sum_{n=1}^S x_n^s g_n^s k(z_n^s, z) + b \right\} \tag{4}$$

where $b = \frac{1}{2} \sum_{j=1}^S (g_j^s - \sum_{n=1}^S x_n^s g_n^s k(z_n^s, z_j^s))$

From Eq. (4), it can be stated that g_j^s is the output label related to the j -th support vector. On the other hand, x_n^s and g_n^s are characterized by the n -th data point and the n -th support vector, respectively. The term $k(z_n^s, z_j^s)$ is the kernel function that computes the relationships between the data points z and the support vectors.

Once SVM is trained, it can be used for categorization by determining based on the location of the test data among the sides of the hyperplane.

In simple terms, if you want to classify different types of faults in PEFC, SVM draws an imaginary line between the different types of faults to separate them. It is like a judge who decides what type of failure a PEFC has based on its data.

In the case of fuel cells, the classes correspond to the state of health (SoH) of the cell and can be healthy or faulty. Specifically, in this study, the SVM was configured in a supervised way from the variable created for the classification. The performance and efficiency analysis of the algorithm focuses on the database for testing.

K-nearest neighbor (KNN)

According to Demidova and Klyueva (2017), KNN should be the prime approach chosen between different classification methods based on the lack of information about a set of data distributions. This study proposed a non-parametric method for pattern classification, which developed into the KNN rule. Complementing the last starting point, in Cover and Hart (1967), scholars prearranged some formal features of the KNN rule. Moreover, researchers in Fukunaga and Hostetler (1975) introduced the

Bayer error rate, and in Dudani (1976), distance-weighted approaches were applied instead.

This method generally tries to find the k closest points in the data to a set of query points. The KNN classifier usually uses the Euclidean distance between specified training samples and a test sample (Eslamibidgoli et al. 2021).

$$d(x_i, x_l) = \sqrt{(x_{i1} - x_{l1})^2 + (x_{i2} - x_{l2})^2 + \dots + (x_{ip} - x_{lp})^2} \quad (5)$$

In Eq. (5), x_i is an input sample with p features, which are $x_{i1}, x_{i2}, \dots, x_{ip}$. Moreover, n is the total number of input samples ($i = 1, 2, \dots, n$). Figure 2 depicts an example of KNN.

For example, if you want to know what type of failure an FC has, KNN looks for the FCs with faults most similar to the one you are observing and, based on its data, predicts what type of fault it has. It's like asking your neighbors what type of fault their FCs have to find out what type of fault yours has.

Decision tree (DT)

The DT is an approach when multifaced decision situations are broadly used in many industrial applications. When a complex function may be represented by DT algorithms classing new overlooked data and, thus, gets an endowed generalization feature. Therefore, this is one of the benefits of DTs, which is their ability to shatter a complicated decision into a set of simpler decisions based on the desired objectives (Han and Chung 2017). Another of the advantages of DT approaches is that their results can classify problems robustly with nature, non-parametric, and high computational effectiveness (Detti et al. 2017).

From a diagram point of view, a DT is also a flow diagram arrangement in which each internal node represents an "assessment" of a property, each branch represents the result of

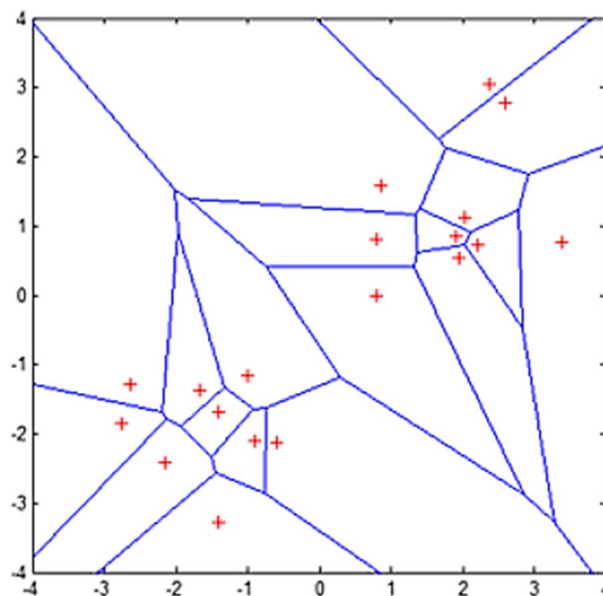


Fig. 2 K-nearest neighbor example: Voronoi tessellation showing Voronoi cells of 19 samples marked with a "+" Figure taken from Azad et al. (2022)

the trial, and each leaf node represents a decision taken after calculating all features called class labels—the paths from leaf to classification rules.

A DT consists of three types of nodes: (Santamaria et al. 2021) end nodes, classically characterized by triangles; decision nodes, classically represented by squares; and chance nodes, typically denoted by circles. Figure 3 shows a schematic of decision trees.

From a mathematical point of view, based on training vectors $x_i \in R^l, i = 1, \dots, l$ and a label vector $y \in R^l$, DT partitions the feature space recursively so that samples with the same label or target value are grouped. Then, Q_m with n_m samples could represent the data at node m . For each split $\theta = (j, t_m)$ consisting of a feature j and threshold t_m , the data partition will be into $Q_m^{left}(\theta)$ and $Q_m^{right}(\theta)$ subsets, where:

$$Q_m^{left}(\theta) = \{(x, y) | x_j \leq t_m\}$$

$$Q_m^{right}(\theta) = Q_m \setminus Q_m^{left}(\theta) \tag{6}$$

The split of node m is determined by an impurity function $H()$. It depends on the problem being solved (classification or regression). In this case, the objective is a classification outcome taking on values $0, 1, \dots, K-1$, for node m let be the proportion of class k observations. The prediction probability for the region is P_{mk} , if m is a terminal node:

$$P_{mk} = \frac{1}{n_m} \sum_{y \in Q_m} I(y = k) \tag{7}$$

Common impurity functions are Gini Eq. (8) and Log loss or Entropy Eq. (9):

$$H(Q_m) = \sum_k P_{mk}(1 - P_{mk}) \tag{8}$$

$$H(Q_m) = - \sum_k P_{mk} \log(P_{mk}) \tag{9}$$

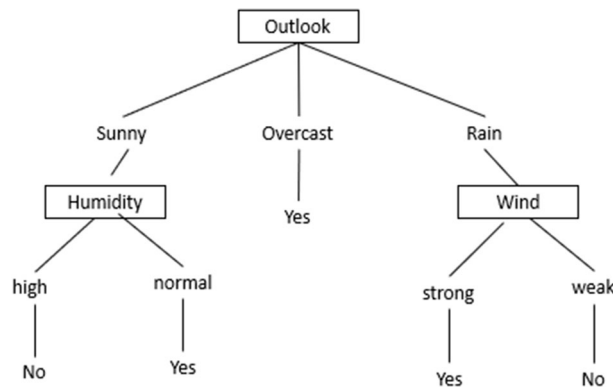


Fig. 3 Decision Tree Based Approaches in Data Mining. Image taken from Zhang et al. (2020)

A simple example is that if you want to decide whether an FC needs maintenance or not, a DT helps you decide by asking questions about the FC data, such as voltage, current, and temperature. It's like a flow chart that guides you toward the best decision.

In practice, Decision Trees are often used as fostering additions for more complex methods such as Random Forests or Gradient Boosting Machines, which blend multiple Decision Trees to enhance their performance.

Random forest (RF)

The basis of an RF is the base learner, a binary tree built using recursive partitioning. The classification and regression tree methodology (CART) is the approach to construct the base learner. This procedure has many dual divisions, recursively separating the structure into near-homogeneous or homogeneous terminus joins (Vaz et al. 2023).

The RF is usually a set of thousands of trees, where each single tree raises using a fuller sample boot of the initial data. In Breiman's RF model, every tree is lodged on the foundation of a training sample set and a random variable, the random variable corresponding to the K_{th} tree is denoted as Θ_k , between any two of the random variables, resulting in a classifier $h(X, \Theta_k)$ where X is the input vector. After a specific number of running (indeed k times), a classification sequence is obtained: $h_1(x), h_2(x), \dots, h_k(x)$ to compute more than one arrangement model system. The final decision function is (Meiler et al. 2012):

$$H(x) = \operatorname{argmax} \sum_{i=1}^k I(h_i(x) = Y) \tag{10}$$

where $H(x)$ is a combination of the classification models, h_i is a single DT model, Y is the input variable, and $I()$ is the indicator function. Each tree can choose the best classification path depending on the input, as shown in Fig. 4.

In RF, the margin function is used to quantify the extent to which the average number of polls at X, Y for the right class exceeds that for the wrong category. The margin function is defined as (Meiler et al. 2012):

$$mg(X, Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \tag{11}$$

Then from Eq. (11), it can be stated that the larger the margin value, the higher the accuracy of the classification prediction and the more confidence in classification.

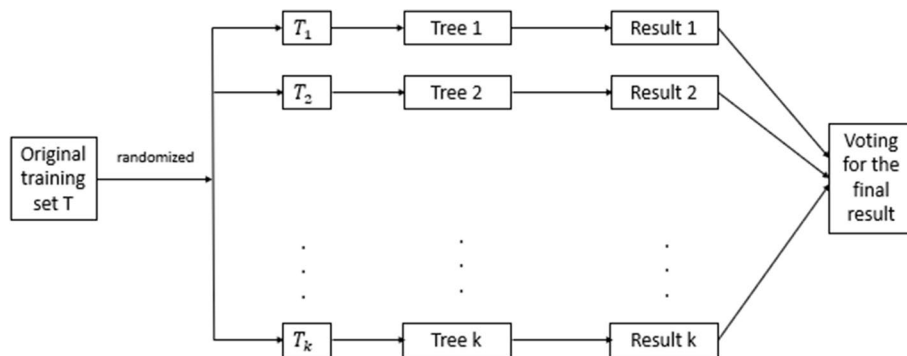


Fig. 4 Random Forest. Figure taken from Peng et al. (2019)

To illustrate, if you want to predict when an FC will fail, RF combines many decision trees to get a more accurate prediction. It's like having a group of experts giving you different opinions on when the FC will fail.

Gaussian Naive Bayes (NB)

Gaussian NB is an alternative Naive Bayes method based on Gaussian normal distribution and continuous data. To understand this method, it is important to remember that Naive Bayes is a cluster of supervised ML classification algorithms based on the Bayes theorem. It is an easy taxonomy technique but has extraordinary functionality. The use is mainly when the dimensionality of the input data is high. Complex classification problems can also be implemented by using the Naive Bayes Classifier.

Then, the logic of Gaussian Naive Bayes (NB) starts with Bayes' theorem that supplies an approach to the probability computation of a hypothesis founded on the likelihood of observed evidence. Mathematically, Bayes' theorem is expressed as:

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (12)$$

where $P(A|B)$ is the probability of event A given event B; $P(B|A)$ is the probability of event B given A; $P(A)$ and $P(B)$ are the probabilities of events A and B, respectively. Later, when operating with continuous data, a hypothesis frequently presumed is that the continuous values linked with every class are issued according to a normal (or Gaussian) distribution. The probability of the structure is assumed to be (Wang et al. 2020):

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}} \quad (13)$$

In Eq. (13), x is the variable, y is the class, μ is the mean, and σ is the variance. This variance could be assumed to be independent of Y (i.e., σ_i) or independent of X_i (i.e., σ_k), or both cases.

For example, if you want to know if an FC has a specific problem, NB analyzes the FC's data and calculates the probability that it has the problem. It's like a detective putting together clues to solve a case.

Deep learning approach

Multi-layer perceptron (MLP)

In recent years, MLP has also been used in the diagnosis of the health status of fuel cells. This is due to their ability to approximate nonlinear input/output relationships. This behavior is characteristic of fuel cells, so it is considered an appropriate method for diagnosing their condition and estimating the main characteristics influencing their performance (Napoli et al. 2013; Priya et al. 2018).

In this case, it is known that the behavior of the data is nonlinear (Melo et al. 2022). Therefore, the MLP was applied as a binary classification algorithm where the input corresponded to the sixteen previously mentioned variables, and the output was the variable constructed to know the state of fuel cell health. As in previous algorithms, the analysis focuses on the test data.

The logic of MLP starts with a feature vector X as input, followed by the computation of the network outputs by the feedforward propagation. Then, each layer i , can be expressed with the activation vector h as:

$$\begin{aligned} h_i^{(1)} &= \varphi^{(1)} \left(\sum_j \omega_{ij}^{(1)} x_j + b_i^{(1)} \right) \\ h_i^{(2)} &= \varphi^{(2)} \left(\sum_j \omega_{ij}^{(2)} h_j^{(1)} + b_i^{(2)} \right) \\ y_i &= \varphi^{(3)} \left(\sum_j \omega_{ij}^{(3)} h_j^{(2)} + b_i^{(3)} \right) \end{aligned} \quad (14)$$

From Eq. (14), it can be clarified that $h_i^{(1)}$ and $h_i^{(2)}$ represent the activation layer of each layer. Moreover, the terms $\varphi^{(1)}$ and $\varphi^{(2)}$ represent the activation function applied to the weighted sum of the inputs. In addition, $\omega_{ij}^{(1)}$ and $\omega_{ij}^{(2)}$ embody the weight of the connection between the j -th network in the current layer and the i -th network in the first hidden layer. On the other hand, x_j symbolizes the output of each input layer. Finally, $b_i^{(1,2,3)}$ are the bias associated with the i -th network in the first layer.

MLP distinguishes $\varphi^{(1)}$ and $\varphi^{(2)}$ because different layers may have different activation functions. Also, each layer can contain multiple units, so all its units' activations can be represented with an activation vector $h()$. The same represents each layer's weights with a weight matrix $W()$. Each layer also has a bias vector $b()$:

$$\begin{aligned} h^{(1)} &= \varphi^{(1)} \left(W^{(1)} x + b^{(1)} \right) \\ h^{(2)} &= \varphi^{(2)} \left(W^{(2)} h^{(1)} + b^{(2)} \right) \\ y &= \varphi^{(3)} \left(W^{(3)} h^{(2)} + b^{(3)} \right) \end{aligned} \quad (15)$$

Finally, all the training examples are combined into a single matrix X and could store each layer's hidden units for all the training examples as a matrix $H()$. Each row contains the hidden units, for one example. This is written considering the transposes as follows:

$$\begin{aligned} H^{(1)} &= \varphi^{(1)} \left(X W^{(1)T} x + 1 b^{(1)T} \right) \\ H^{(2)} &= \varphi^{(2)} \left(H^{(1)} W^{(2)T} + 1 b^{(2)T} \right) \\ Y &= \varphi^{(3)} \left(H^{(2)} W^{(3)} + 1 b^{(3)T} \right) \end{aligned} \quad (16)$$

From Eq. (16) the equations have the same shape as equations from (14) but in matrix form.

In other words, if you want to create an automatic system to detect FC faults, MLP is an artificial network that learns to detect faults by analyzing many examples of FC data with different types of faults. It's like a "teacher" who teaches you how to detect different types of FC failures.

Model testing (MT)

Kuhn & Johnson (Bao et al. 2016) recommend using sensitivity and specificity to evaluate the binary classification models since these measures allow us to know the precision of the models. Likewise, they use the receiver operating characteristic (ROC) curve to represent the relationship between these measures. On the one hand, the sensitivity or recall, or "true positives rate", describes the proportion at which the samples with the event of interest are correctly predicted. The specificity or "true negative rate" is the proportion at which the models without events are correctly predicted. Likewise, it is possible to calculate the precision from the true positives. This measure consists of dividing the proportion of true positives by the total of predicted positives.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (17)$$

$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (18)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (19)$$

where: TP and TN represent true positive and true negative, respectively. FP and FN represent false positive and false negative, respectively.

Another useful measure for evaluating this type of algorithm is the F score. According to Sokolova et al. (Sokolova et al. 2006), this measure favors algorithms with higher sensitivity and challenges those with higher specificity. It is calculated from the precision and recovery of the test, applying additional weights. The F-score result ranges from 0 to 1, with 1 being perfect accuracy and recall and 0 being the lowest and least favorable value.

$$F - \text{Score} = \frac{(\beta^2 + 1) * \text{precision} * \text{recall}}{\beta^2 * \text{precision} + \text{recall}} \quad (20)$$

From Eq. (20), it also can be said that *Recall* is the ratio between true positive predictions and the total number of existent positive instances in the dataset. Sometimes, it is called sensitivity. Moreover, the values of β^2 permits the adjustments between precision or sensitivity (recall) if $\beta^2=1$ means the F-Scores is the mean of the harmonics among the precision and recall, i.e., equal weight for both. Other values mean more emphasis on one of the two characteristics.

Finally, the computational time is another important factor to consider when comparing and evaluating the performance of classification models (Wang et al. 2020; Li et al.

2014b; Priya et al. 2018; Mao and Jackson 2016). For this reason, in this study, the training phase’s computational time was considered another performance measure.

Experimental results

This section details the performance evaluation of the applied classification algorithms based on some measures common to the methods. Based on these, the most appropriate method for detecting faults in the state of health of the fuel cell is analyzed.

Performance evaluation measures

According to the performance of the metrics, most of the algorithms achieve high precision (greater than 98%) and F1-Score (greater than 99%) in both cases (Healthy and Faulty). In detail, LR, SVM, KNN, DT, RF, and MLP have similar performance. However, NB has slightly lower performance compared to the other algorithms.

On the other hand, it is observed that the precision and F1-Score decrease in all algorithms, especially for the "Faulty" class. Specifically, the reduction in performance is more significant for NB, KNN, and MLP. In contrast, DT and RF still achieve reasonable accuracy and F1-Score (greater than 90%) for the “Faulty” class.

Overall, the performance evaluation measurements indicate that all the algorithms can correctly evaluate the cell’s health under normal operating and fault conditions. However, it is necessary to highlight that KNN, DT, RF, and MLP present better performance when determining the state of the cell concerning the other models, which suggests analyzing other metrics to guarantee their reliability. Table 2 shows the results of the measurements by model.

Receiver operating characteristic (ROC) curve

In this sense, comparing the precision of the binary classification models on the originally unbalanced data and the balanced data with different sampling techniques was considered. According to the results shown in Table 3, the precision of LR, KNN, and NB follows a similar trend regardless of the data, which makes them models with stable performance and possible generalization. On the other hand, DT and RF stand out for their perfect precision on the unbalanced data. They keep the same on the balanced data with the ROS and SMOTE-OS techniques and do not significantly

Table 2 Performance evaluation measures of binary classification algorithms

	Healthy			Faulty		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
LR	1	1	1	1	0.98	0.99
SVM	1	1	1	1	0.95	0.98
KNN	1	1	1	1	1	1
DT	1	1	1	1	1	1
RF	1	1	1	1	1	1
NB	1	1	1	1	0.98	0.99
MLP	1	1	1	1	1	1

The table shows the results of different binary classification algorithms applied to data sets in the context of the state of health of PEFC (healthy and faulty). Algorithms evaluated include LR, SVM, KNN, DT, RF, NB, and MLP. The evaluation metrics used are Precision, Revocation, and F1-Score

Table 3 Models' accuracy comparison with sampling for imbalanced data

	Accuracy	ROS Accuracy	SMOTE-OS Accuracy	RUS Accuracy	NearMiss-US Accuracy
LR	0.998	0.998	0.999	0.998	0.988
SVM	0.997	0.998	0.998	0.997	0.059
KNN	0.999	0.999	0.999	0.999	0.998
DT	1	0.999	1	0.999	0.997
RF	1	1	1	0.999	0.999
NB	0.998	0.998	0.998	0.998	0.999
MLP	0.999	0.998	0.997	0.997	0.059

The table presents the results of comparing the accuracy of different models for balanced and unbalanced data in the context of PEFCs. The models evaluated include LR, SVM, KNN, DT, RF, NB, and MLP. The resampling techniques are ROS, SMOTE-OS, RUS, and NearMiss-US

vary on the balanced data with the other sampling techniques. Finally, the precision of the SVM seems to be affected when the model is applied to balanced data with the NearMiss-U technique, whose value is considerably reduced from 0.997 to 0.059, unlike the other techniques where the precision differs with only 0.001. In the same way, the accuracy (0.999) of MLP is reduced to 0.059 when the NearMiss-U technique balances the data.

Overall, the high accuracy of most models on balanced data indicates their potential for accurate fault detection in PEFC. Likewise, comparing performance on balanced and unbalanced data sets highlights the importance of data balance for analysis accuracy. However, it is worth noting that in the real world, obtaining high-quality, balanced data sets for PEFC can be challenging.

In addition to the exposed indicators, the ROC curve that relates the sensitivity and specificity was constructed by applying the models to the unbalanced data. Once the ROC curve of the different ML models was obtained, the AUC was obtained, the results of which are illustrated in Fig. 5. The ROC curve for all algorithms is above the diagonal line, which indicates that all models have a performance better than chance. The highest AUC is obtained by the LR, KNN, DT, and RF models, with a value close to 1. The SVM and NB models have a slightly lower AUC, but still good performance. Specifically, again the SVM model presents the lowest value compared to the other models that exceed 0.976 in the AUC.

In practice, a high AUC indicates that the model is capable of detecting PEFC faults with high accuracy. Therefore, comparing the AUC of different algorithms helps to identify the best-performing model for fault detection. Likewise, the AUC can be used to select an optimal classification threshold that balances the sensitivity and specificity of the model.

It proceeded in the same way as the DL model, MLP. As presented in Fig. 6, the relationship between the true positive and false positive ratios results in a perfect AUC of 0.998. Consequently, MLP and KNN, DT, and RF (ML models) present the best AUC results. Specifically, the MLP model seems to be suitable for fault detection in PEFC based on the AUC. However, it is important to consider the limitations of the MLP model, such as its low interpretability and the possibility that its performance may not generalize to other data sets.

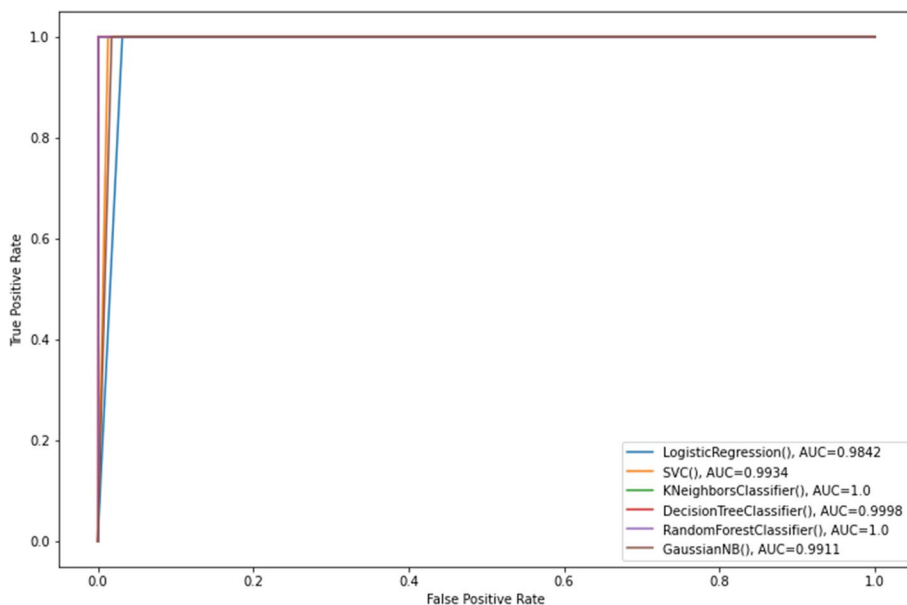


Fig. 5 ROC curves comparison with the imbalanced database for ML models. The figure shows the ROC Curve and AUC for different binary classification algorithms applied to a PEFC data set. The algorithms evaluated include LR, SVM, KNN, DT, RF, and NB. The AUC is a measure of a model's ability to distinguish between the "Healthy" and "Faulty" classes. An AUC of 1 indicates perfect performance, while an AUC of 0.5 indicates random performance

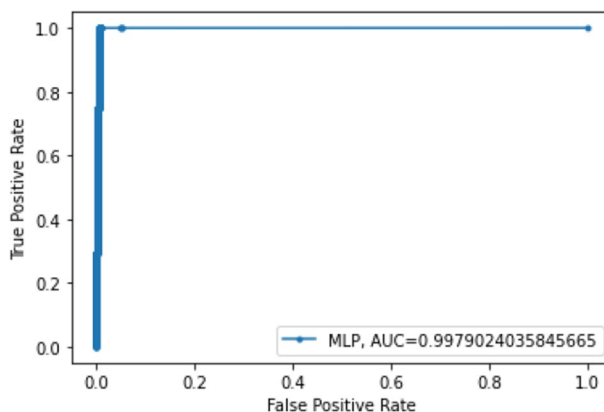


Fig. 6 ROC curve with the imbalanced database for the DL model. The figure shows the ROC Curve and AUC for the MLP algorithm applied to a PEFC data set

Time score

According to the results shown in Table 4, the algorithm that generates the lowest computational cost and can predict in a more timely manner the failures in the health status of the cell is the NB, which is consistent with that stated by other authors (Wang et al. 2020; Li et al. 2014b; Hatti et al. 2006; Kamal and Yu 2014; Lin et al. 2019). In contrast, SVM and MLP demand higher computational costs, so they would not be suitable for online fault detection.

In general, training time is an important factor to consider when choosing a machine learning algorithm for fault detection in PEFC. However, it is necessary to consider that

Table 4 Training time of binary classification algorithms

Algorithm	Training Time (s)
LR	1.47
SVM	115.25
KNN	0.48
DT	0.37
RF	11.46
NB	0.04
MLP	134.59

The table shows the results of comparing the training time of different binary classification algorithms for data in the context of PEMFCs. The algorithms evaluated include LR, SVM, KNN, DT, RF, NB, and MLP. The evaluation metric used is training time in seconds

this time may vary with the size and complexity of the data set used. The latter leads to exploring techniques to reduce training time, such as hyperparameter optimization or the use of specialized hardware.

Discussion

The results show that the LR, KNN, DT, RF, and NB models present similar and optimal trends in their ability to determine the healthy and failed state of the PEMFC correctly. The precision, recall, F1-score, and accuracy values exceeded 0.98 in their application to the balanced and unbalanced data. In practical terms, this stability becomes an important factor when considering its possible expansion to other fuel cell data sources of different functionality, for example, the solid oxide fuel cell (SOFC). In addition to the optimal performance, these models presented a lower computational cost, with training times less than or equal to 11.46 s, contributing to their effectiveness and efficiency.

On the other hand, although the SVM and MLP algorithm presents optimal results in terms of its ability to classify healthy conditions of the fuel cell, it does not remain stable when determining cell failures. Likewise, when applied to unbalanced data, it presents better precision than when used to data balanced by techniques such as NearMiss-US, which reflects its generalization weakness. Lastly, its application requires a much higher computational cost than the other models, with training times exceeding 100 s, which makes it unfeasible for PEMFC fault diagnosis.

Table 5 summarizes the results of the models according to the performance of the indicators and the computational cost.

These results show that, while some ML methods like LR and DT offer better interpretability, many DL models, such as MLPs, can be "black boxes," making it difficult to understand how they arrive at their predictions. This can hinder trust and acceptance in practical applications. Also, in the future, integrating ML and DL models into existing PEMFC systems and infrastructure requires careful consideration of factors like computational requirements, real-time operation, and potential safety implications.

Conclusions

This article presents an approach to PEMFC stack fault diagnosis that relies on different ML and DL techniques to provide a comprehensive view of the performance and computational cost that each technique demands. The procedure is carried out from

Table 5 Comparison of performance evaluation measures of classification algorithms

Model	Sensitivity	Precision	F-Score	Computational cost
LR	High	High	High	Low
SVM	High	High	High	High
KNN	High	High	High	Low
DT	High	High	High	Low
RF	High	High	High	Low
NB	High	High	High	Low
MLP	High	High	High	High

The table shows the results of comparing the performance metrics of different binary classification algorithms for data in the context of PEFCs. The algorithms evaluated include LR, SVM, KNN, DT, RF, NB, and MLP. The metrics evaluated are precision, sensitivity, F1-Score, and computational cost

a previously processed data set whose scope includes the extraction of characteristics and the category of the state of health of the cell. Then, considering that the data provided more information about a healthy state than failure conditions, it was decided to apply the algorithms to the unbalanced and balanced data using different sampling techniques. The results showed that the LR, KNN, DT, RF, and NB models present similar and optimal trends in terms of performance indicators and computational cost, unlike SMV and MLP, whose performance is affected when the data is balanced and even presents a higher computational cost.

From the theoretical aspect, the advantage provided by the methodology of the present study is that it can be scalable and adapted to other classification problems with fuel cells. Also, these results have practical implications in the energy sector, on the one hand, the application of ML and DL methods for fault detection and prediction in PEFCs has the potential to significantly improve their diagnostics and prognostics. This can lead to more effective preventive maintenance, increased cell availability, and reduced operational costs. On the other hand, ML and DL-based approaches can be valuable tools for optimizing PEFC design and selecting materials with improved properties. Consequently, more efficient, durable, and cost-effective cells can be achieved.

However, it is necessary to mention that training and validating ML and DL models require large amounts of high-quality data. Collecting and organizing such data can be challenging, especially for complex tasks like PEFC analysis. Future research should address the data availability challenges by developing efficient techniques for data acquisition and processing, and the configuration of parameters, optimization algorithms, and specific settings; since in this study models with default configurations were considered. Also, in future studies, it is expected to validate experimental results from case studies and develop robust and scalable methodologies for integrating ML and DL models into industrial environments.

Therefore, it is a very promising diagnostic proposition to diagnose the faults associated with PEMFC. In the future, extending the proposed approach to other types of health conditions is recommended, increasing the number of classes and different kinds of fuel cells.

Abbreviations

AI	Artificial intelligence
ANN	Artificial neural networks
BPNN	Back propagation neural network
CNN	Convolutional Neuronal Network
DL	Deep learning
DT	Decision tree
ECMS	Equivalent Consumption Minimization Strategy
EIS	Electrochemical Impedance Spectroscopy
FC	Fuel Cell
FCHEV	Fuel cell hybrid electric vehicle
FDA	Fisher discriminant analysis
FDI	Fault Detection and Isolation
GA-KELM	Genetic algorithm-based KELM
GDL	Gas-diffusion layer
KELM	Kernel extreme learning machine
KNN	K-Nearest Neighbors
LDA	Linear discriminant analysis
LM	Levenberg–Marquardt
LR	Logistic regression
ML	Machine learning
MLP	Multi-layer perceptron
NB	Naive Bayes
NN	Neural Network
PEFCs	Polymer electrolyte fuel cells
PSO-KELM	Particle swarm optimization-based KELM
RF	Random forest
RP	Resilient propagation
RSM	Response surface methodology
SMOTE	Synthetic minority over-sampling technique
SoH	State of health
SVM	Support vector machine

Author contributions

Conceptualisation: EM., JB-M. and ME-A.; methodology: EM., JB-M. and ME-A.; software: EM. and JB-M.; validation: HN. and ME-A.; formal analysis: JB-M. and HN.; investigation: EM.; resources: JB-M. and ME-A.; data curation: EM. and JB-M.; writing—original draft preparation: EM. and HN.; writing—review and editing: EM, JB-M. and HN.; visualization: EM. and HN.; supervision: JB-M. and ME-A. All authors have read and approved the final manuscript.

Funding

Financial support is obtained internally from ESPOL.

Data availability

Data are available from the corresponding author upon request.

Declarations**Competing interests**

The authors declare no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Received: 8 January 2024 Accepted: 20 February 2024

Published online: 26 February 2024

7. References

- Aldubyan M, Krarti M (2022) Impact of stay home living on energy demand of residential buildings: Saudi Arabian case study. *Energy* 238:121637. <https://doi.org/10.1016/j.energy.2021.121637>
- Awasthi S, Singh G, Ahamad N (2024) Classifying electrical faults in a distribution system using k-nearest neighbor (knn) model in presence of multiple distributed generators. *J Inst Eng India Ser B*. <https://doi.org/10.1007/s40031-024-00994-4>
- Azad C, Bhushan B, Sharma R, Shankar A, Singh KK, Khamparia A (2022) Prediction model using SMOTE, genetic algorithm and decision tree (PMSGD) for classification of diabetes mellitus. *Multimedia Syst* 28:1289–1307. <https://doi.org/10.1007/s00530-021-00817-2>

- Bao L, Juan C, Li J, Zhang Y (2016) Boosted Near-miss Under-sampling on SVM ensembles for concept detection in large-scale imbalanced datasets. *Neurocomputing* 172:198–206. <https://doi.org/10.1016/j.neucom.2014.05.096>
- Calli-Cankir F, Ismail MS, Ingham DB, Hughes KJ, Ma L, Pourkashanian M (2022) Air-breathing versus conventional polymer electrolyte fuel cells: a parametric numerical study. *Energy* 250:123827. <https://doi.org/10.1016/j.energy.2022.123827>
- Carrette L, Friedrich KA, Stimming U (2000) Fuel cells: principles, types, fuels, and applications. *ChemPhysChem* 1:162–193. [https://doi.org/10.1002/1439-7641\(20001215\)1:4%3c162::AID-CPHC162%3e3.0.CO;2-Z](https://doi.org/10.1002/1439-7641(20001215)1:4%3c162::AID-CPHC162%3e3.0.CO;2-Z)
- Chen N, Ma L-L, Zhang Y, Yan Y-X (2024) Association of household solid fuel use and long-term exposure to ambient air pollution with estimated 10-year high cardiovascular disease risk among postmenopausal women. *Environ Pollut* 342:123091. <https://doi.org/10.1016/j.envpol.2023.123091>
- Cover T, Hart P (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13:21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Demidova L, Klyueva I (2017) SVM classification: Optimization with the SMOTE algorithm for the class imbalance problem. In: 2017 6th Mediterranean Conference on Embedded Computing (MECO). pp 1–4. <https://doi.org/10.1109/MECO.2017.7977136>
- Deti AH, Jemei S, Morando S, Steiner NY (2017) Classification Based Method Using Fast Fourier Transform (FFT) and Total Harmonic Distortion (THD) Dedicated to Proton Exchange Membrane Fuel Cell (PEMFC) Diagnosis. In: 2017 IEEE Vehicle Power and Propulsion Conference (VPPC). pp 1–6. <https://doi.org/10.1109/VPPC.2017.8331040>
- Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics. SMC-6*, 325–327 (1976). <https://doi.org/10.1109/TSMC.1976.5408784>.
- Eslamibidgoli MJ, Tipp FP, Jitsev J, Jankovic J, Eikerling MH, Malek K (2021) Convolutional neural networks for high throughput screening of catalyst layer inks for polymer electrolyte fuel cells. *RSC Adv* 11:32126–32134. <https://doi.org/10.1039/D1RA05324H>
- Fernandez A, Garcia S, Herrera F, Chawla NV (2018) SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *J Artif Intell Res* 61:863–905. <https://doi.org/10.1613/jair.1.11192>
- Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans Inf Theory* 21:32–40. <https://doi.org/10.1109/TIT.1975.1055330>
- Ghorbanzade Zaferani SP, Amiri MK, Sarmasti Emami MR, Zahmatkesh S, Hajjaghahi-Keshтели M, Panchal H (2024) Prediction and optimization of sustainable fuel cells behavior using artificial intelligence algorithms. *Int J Hydrogen Energy* 52:746–766. <https://doi.org/10.1016/j.ijhydene.2023.03.335>
- Grove WR (1839) XXIV. On voltaic series and the combination of gases by platinum. *London Edinburgh Dublin Philo Magaz J Sci* 14:127–130. <https://doi.org/10.1080/14786443908649684>
- Hai T, Alenzi FA, Mohammed AH, Goyal V, Marjan RK, Quzwain K, Mohammed Metwally AS (2024) Solid oxide fuel cell energy system with absorption-ejection refrigeration optimized using a neural network with multiple objectives. *Int J Hydrogen Energy* 52:954–972. <https://doi.org/10.1016/j.ijhydene.2023.07.115>
- Han I-S, Chung C-B (2017) A hybrid model combining a support vector machine with an empirical equation for predicting polarization curves of PEM fuel cells. *Int J Hydrogen Energy* 42:7023–7028. <https://doi.org/10.1016/j.ijhydene.2017.01.131>
- Hatti M, Tioussi M, Nouibat W (2006) Static modelling by neural networks of a PEM fuel cell. In: IECON 2006 - 32nd Annual Conference on IEEE Industrial Electronics. pp 2121–2126. <https://doi.org/10.1109/IECON.2006.347589>
- Huo W, Li W, Zhang Z, Sun C, Zhou F, Gong G (2021) Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection. *Energy Convers Manag* 243:114367. <https://doi.org/10.1016/j.enconman.2021.114367>
- Jiang P, Fan YV, Klemeš JJ (2021) Impacts of COVID-19 on energy demand and consumption: challenges, lessons and emerging opportunities. *Appl Energy* 285:116441. <https://doi.org/10.1016/j.apenergy.2021.116441>
- Kamal M, Yu D (2014) Fault diagnosis for fuel cell stack using independent MLP neural network. Presented at the 2nd International Conference on Innovations in Engineering and Technology (ICCET'2014), Penang <https://doi.org/10.15242/ie.e0914059>.
- Klell M, Eichseder H, Trattner A (2023) Fuel Cells. In: Klell, M., Eichseder, H., and Trattner, A. (eds.) *Hydrogen in automotive engineering: production, storage, application*. pp 137–192. Springer Fachmedien, Wiesbaden. https://doi.org/10.1007/978-3-658-35061-1_6.
- Kulkarni Y, Ramamritham K, Somu N (2021) EnsembleNTLdetect: An Intelligent Framework for Electricity Theft Detection in Smart Grid. In: 2021 International Conference on Data Mining Workshops (ICDMW). pp 527–536 (2021). <https://doi.org/10.1109/ICDMW53433.2021.00070>.
- Lee W-Y, Park G-G, Sohn Y-J, Kim S-G, Kim M (2017) Fault detection and diagnosis methods for polymer electrolyte fuel cell system. *Trans Korean Hydrogen New Energy Soc* 28:252–272. <https://doi.org/10.7316/KHNES.2017.28.3.252>
- Lee W-Y, Kim M, Oh H, Sohn Y-J, Kim S-G (2018) A review on prognostics of polymer electrolyte fuel cells. *Trans Korean Hydrogen New Energy Soc* 29:339–356. <https://doi.org/10.7316/KHNES.2018.29.4.339>
- Lemaître, G., Nogueira, F., Aridas, C.: Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res*. 18, (2016).
- Li Z, Giurgea S, Outbib R, Hissel D (2014a) Online diagnosis of PEMFC by combining support vector machine and fluidic model. *Fuel Cells* 14:448–456. <https://doi.org/10.1002/fuce.201300197>
- Li Z, Outbib R, Hissel D, Giurgea S (2014b) Data-driven diagnosis of PEM fuel cell: a comparative study. *Control Eng Pract* 28:1–12. <https://doi.org/10.1016/j.conengprac.2014.02.019>
- Li Z, Outbib R, Giurgea S, Hissel D, Giraud A, Couderc P (2019) Fault diagnosis for fuel cell systems: a data-driven approach using high-precision voltage sensors. *Renew Energy* 135:1435–1444. <https://doi.org/10.1016/j.renene.2018.09.077>
- Lin R-H, Xi X-N, Wang P-N, Wu B-D, Tian S-M (2019) Review on hydrogen fuel cell condition monitoring and prediction methods. *Int J Hydrogen Energy* 44:5488–5498. <https://doi.org/10.1016/j.ijhydene.2018.09.085>
- Liu B, Wei X, Sun C, Wang B, Huo W (2024) A controllable neural network-based method for optimal energy management of fuel cell hybrid electric vehicles. *Int J Hydrogen Energy* 55:1371–1382. <https://doi.org/10.1016/j.ijhydene.2023.10.215>

- Lu G, Liu M, Su X, Zheng T, Luan Y, Fan W, Cui H, Liu Z (2024) Study on counter-flow mass transfer characteristics and performance optimization of commercial large-scale proton exchange membrane fuel cells. *Appl Energy* 359:122743. <https://doi.org/10.1016/j.apenergy.2024.122743>
- Ma Y, Li C, Wang S (2022) Multi-objective energy management strategy for fuel cell hybrid electric vehicle based on stochastic model predictive control. *ISA Trans* 131:178–196. <https://doi.org/10.1016/j.isatra.2022.04.045>
- Madushani S, Sandamal K, Meddage DPP, Pasindu HR, Gomes P (2023) Evaluating expressway traffic crash severity by using logistic regression and explainable & supervised machine learning classifiers. *Transport Eng* 13:100190. <https://doi.org/10.1016/j.treng.2023.100190>
- Mao L, Jackson L (2016) Comparative study on prediction of fuel cell performance using machine learning approaches. Loughborough University.
- Mao L, Jackson L, Dunnett S (2017) Fault diagnosis of practical polymer electrolyte membrane (PEM) fuel cell system with data-driven approaches. *Fuel Cells* 17:247–258. <https://doi.org/10.1002/fuce.201600139>
- Mao L, He K, Jackson L, Wu Q (2021) Chapter 7 - Application of artificial neural networks in polymer electrolyte membrane fuel cell system prognostics. In: Mellal, M.A. and Pecht, M.G. (eds.) *Nature-Inspired Computing Paradigms in Systems*. pp 93–109. Academic Press. <https://doi.org/10.1016/B978-0-12-823749-6.00005-2>
- Meiler M, Hofer EP, Nuhic A, Schmid O (2012) An Empirical Stationary Fuel Cell Model Using Limited Experimental Data for Identification. *J Fuel Cell Sci Technol*. <https://doi.org/10.1115/1.4007195>
- Mekhilef S, Saidur R, Safari A (2012) Comparative study of different fuel cell technologies. *Renew Sustain Energy Rev* 16:981–989. <https://doi.org/10.1016/j.rser.2011.09.020>
- Melo E, Encalada Á, Espinoza M (2020) Behavior of a Polymer Electrolyte Fuel Cell from a Statistical Point of View Based on Data Analysis. Presented at the November 1. https://doi.org/10.1007/978-3-030-62833-8_10
- Melo E, Peñafiel J, Barzola-Monteses J, Espinoza M (2022) An initial approach about data preprocessing techniques applied to polymer electrolyte fuel cells: a case study. Presented at the January 1. https://doi.org/10.1007/978-981-16-4126-8_6
- Modanloo V, Mashayekhi A, Akhond B (2024) A comparative analysis of predictive models for estimating the formability of stamped titanium bipolar plates for proton exchange membrane fuel cells. *Int J Hydrogen Energy* 56:894–902. <https://doi.org/10.1016/j.ijhydene.2023.12.242>
- Mqadi NM, Naicker N, Adeliyi T (2021) Solving misclassification of the credit card imbalance problem using near miss. *Math Probl Eng* 2021:e7194728. <https://doi.org/10.1155/2021/7194728>
- Napoli G, Ferraro M, Sergi F, Brunaccini G, Antonucci V (2013) Data driven models for a PEM fuel cell stack performance prediction. *Int J Hydrogen Energy* 38:11628–11638. <https://doi.org/10.1016/j.ijhydene.2013.04.135>
- Nguyen TV, White RE (1993) A water and heat management model for proton-exchange-membrane fuel cells. *J Electrochem Soc* 140:2178. <https://doi.org/10.1149/1.2220792>
- Ogawa T, Takeuchi M, Kajikawa Y (2018) Comprehensive analysis of trends and emerging technologies in all types of fuel cells based on a computational method. *Sustainability* 10:458. <https://doi.org/10.3390/su10020458>
- Onanena R, Oukhellou L, Côme E, Candusso D, Hissel D, Aknin P (2012) Fault-diagnosis of PEM fuel cells using electrochemical spectroscopy impedance. *IFAC Proc Vol* 45:651–656. <https://doi.org/10.3182/20120902-4-FR-2032.00114>
- Peng M, Zhang Q, Xing X, Gui T, Huang X, Jiang Y-G, Ding K, Chen Z (2019) Trainable undersampling for class-imbalance learning. *Proc AAAI Conf Artif Intell* 33:4707–4714. <https://doi.org/10.1609/aaai.v33i01.33014707>
- Platt JC (1998) Sequential minimal optimization: a fast algorithm for training support vector machines
- Priya K, Sathishkumar K, Rajasekar N (2018) A comprehensive review on parameter estimation techniques for Proton Exchange Membrane fuel cell modelling. *Renew Sustain Energy Rev* 93:121–144. <https://doi.org/10.1016/j.rser.2018.05.017>
- Quan R, Liang W, Wang J, Li X, Chang Y (2024) An enhanced fault diagnosis method for fuel cell system using a kernel extreme learning machine optimized with improved sparrow search algorithm. *Int J Hydrogen Energy* 50:1184–1196. <https://doi.org/10.1016/j.ijhydene.2023.10.019>
- Rama P, Chen R, Andrews J (2008) A review of performance degradation and failure modes for hydrogen-fuelled polymer electrolyte fuel cells. *Proc Inst Mech Eng A J Power Energy* 222:421–441. <https://doi.org/10.1243/09576509JPE603>
- Rupapara V, Rustam F, Shahzad HF, Mehmood A, Ashraf I, Choi GS (2021) Impact of SMOTE on imbalanced text features for toxic comments classification using RWVC model. *IEEE Access* 9:78621–78634. <https://doi.org/10.1109/ACCESS.2021.3083638>
- Saikia K, Kakati B, Boro B, Verma A (2018) Current Advances and Applications of Fuel Cell Technologies. In: *Recent Advancements in Biofuels and Bioenergy Utilization*. pp 303–337. https://doi.org/10.1007/978-981-13-1307-3_13
- Santamaria AD, Mortazavi M, Chauhan V, Benner J, Philbrick O, Clemente R, Jia H, Ling C (2020) Applications of artificial intelligence for analysis of two-phase flow in PEM fuel cell flow fields. *ECS Trans* 98:279. <https://doi.org/10.1149/09809.0279ecst>
- Santamaria AD, Mortazavi M, Chauhan V, Benner J, Philbrick O, Clemente R, Jia H, Ling C (2021) Machine learning applications of two-phase flow data in polymer electrolyte fuel cell reactant channels. *J Electrochem Soc* 168:054505. <https://doi.org/10.1149/1945-7111/abfa5c>
- Saxena N, Kumar R, Rao YKSS, Mondloe DS, Dhapekar NK, Sharma A, Yadav AS (2024) Hybrid KNN-SVM machine learning approach for solar power forecasting. *Environ Challenges* 14:100838. <https://doi.org/10.1016/j.envc.2024.100838>
- Schumacher M, Roßner R, Vach W (1996) Neural networks and logistic regression: part I. *Comput Stat Data Anal* 21:661–682. [https://doi.org/10.1016/0167-9473\(95\)00032-1](https://doi.org/10.1016/0167-9473(95)00032-1)
- Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Sattar, A. and Kang, B. (eds.) *AI 2006: Advances in Artificial Intelligence*. pp 1015–1021. Springer, Berlin, Heidelberg (2006). https://doi.org/10.1007/11941439_114
- Staffell I, Scamman D, Abad AV, Balcombe PE, Dodds P, Ekins P, Shah NR, Ward K (2019) The role of hydrogen and fuel cells in the global energy system. *Energy Environ Sci* 12:463–491. <https://doi.org/10.1039/C8EE01157E>
- Tyagi S, Mittal S (2020) Sampling Approaches for Imbalanced Data Classification Problem in Machine Learning. In: Singh, P.K., Kar, A.K., Singh, Y., Kolekar, M.H., and Tanwar, S. (eds.) *Proceedings of ICRIC 2019*. pp 209–221. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-29407-6_17

- US Department of Energy: Spotlight: Artificial Intelligence, <https://www.energy.gov/technologytransitions/articles/spotlight-artificial-intelligence>. Accessed 14 Jun 2023
- Vach W, Roßner R, Schumacher M (1996) Neural networks and logistic regression: part II. *Comput Stat Data Anal* 21:683–701. [https://doi.org/10.1016/0167-9473\(95\)00033-X](https://doi.org/10.1016/0167-9473(95)00033-X)
- van Ruijven BJ, De Cian E, Sue Wing I (2019) Amplification of future energy demand growth due to climate change. *Nat Commun* 10:2762. <https://doi.org/10.1038/s41467-019-10399-3>
- Vaz N, Choi J, Cha Y, Kong J, Park Y, Ju H (2023) Multi-objective optimization of the cathode catalyst layer micro-composition of polymer electrolyte membrane fuel cells using a multi-scale, two-phase fuel cell model and data-driven surrogates. *J Energy Chem* 81:28–41. <https://doi.org/10.1016/j.jechem.2023.02.027>
- Wang Y, Seo B, Wang B, Zamel N, Jiao K, Adroher XC (2020) Fundamentals, materials, and machine learning of polymer electrolyte membrane fuel cell technology. *Energy AI* 1:100014. <https://doi.org/10.1016/j.egyai.2020.100014>
- Wang J, Yang B, Zeng C, Chen Y, Guo Z, Li D, Ye H, Shao R, Shu H, Yu T (2021) Recent advances and summarization of fault diagnosis techniques for proton exchange membrane fuel cell systems: a critical overview. *J Power Sources* 500:229932. <https://doi.org/10.1016/j.jpowsour.2021.229932>
- Xiao F, Chen T, Peng Y, Zhang R (2022) Fault diagnosis method for proton exchange membrane fuel cells based on EIS measurement optimization. *Fuel Cells* 22:140–152. <https://doi.org/10.1002/fuce.202200083>
- Xing Y, Wang B, Gong Z, Hou Z, Xi F, Mou G, Du Q, Gao F, Jiao K (2022) Data-driven fault diagnosis for pem fuel cell system using sensor pre-selection method and artificial neural network model. *IEEE Trans Energy Convers* 37:1589–1599. <https://doi.org/10.1109/TEC.2022.3143163>
- Yue M, Lambert H, Pahon E, Roche R, Jemei S, Hissel D (2021) Hydrogen energy systems: a critical review of technologies, applications, trends and challenges. *Renew Sustain Energy Rev* 146:111180. <https://doi.org/10.1016/j.rser.2021.111180>
- Zhang X, Zhou J, Chen W (2020) Data-driven fault diagnosis for PEMFC systems of hybrid tram based on deep learning. *Int J Hydrogen Energy* 45:13483–13495. <https://doi.org/10.1016/j.ijhydene.2020.03.035>
- Zheng Z, Morando S, Pera M-C, Hissel D, Llarger L, Martinenghi R, Baylon Fuentes A (2017) Brain-inspired computational paradigm dedicated to fault diagnosis of PEM fuel cell stack. *Int J Hydrogen Energy* 42:5410–5425. <https://doi.org/10.1016/j.ijhydene.2016.11.043>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.