



Machine learning-driven optimization of gas diffusion layer microstructure for PEM fuel cells

Rashen Lou Omongos^{a,b}, Diego E. Galvez-Aranda^{a,b}, Franco M. Zanotto^{a,b}, András Vernes^{c,d},
Alejandro A. Franco^{a,b,e,f,*}

^a Laboratoire de Réactivité et Chimie des Solides (LRCS), Université de Picardie Jules Verne, Hub de l'Energie, UMR CNRS 7314, 15 rue Baudelocque, 80039, Amiens, France

^b Réseau sur le Stockage Electrochimique de l'Energie (RS2E), FR CNRS 3459, Hub de l'Energie, 15 rue Baudelocque, Amiens, Cedex, 80039, France

^c AC2T Research GmbH, Viktor-Kaplan-Straße 2/C, 2700, Wiener Neustadt, Austria

^d Institute of Applied Physics, Technische Universität Wien, Wiedner Hauptstraße 8-10/134, 1040, Wien, Austria

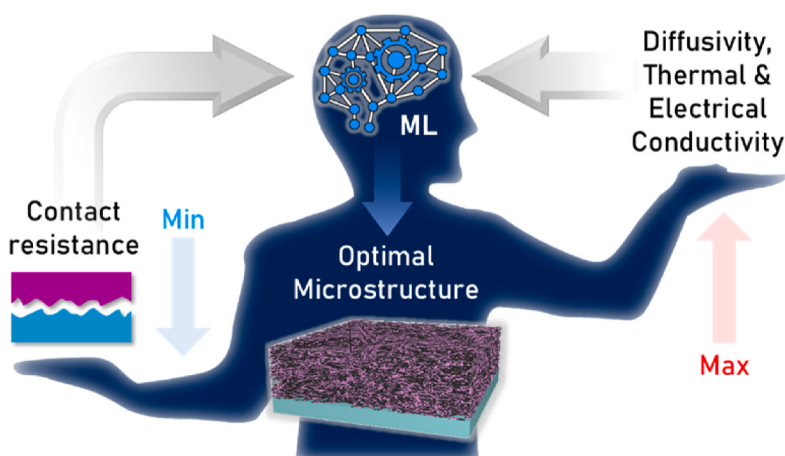
^e ALISTORE-European Research Institute, FR CNRS 3104, Hub de l'Energie, 15 rue Baudelocque, Amiens, Cedex, 80039, France

^f Institut Universitaire de France, 103 Boulevard Saint Michel, Paris, 75005, France

HIGHLIGHTS

- We developed an optimization workflow for gas diffusion layer microstructure.
- 250 gas diffusion layer microstructures were used to train a machine learning model.
- The deterministic learning predicts gas diffusion layer functional properties.
- We predicted optimal manufacturing parameters in 4 scenarios.

GRAPHICAL ABSTRACT



ABSTRACT

The Gas Diffusion Layer (GDL) is a vital component within Proton Exchange Membrane Fuel Cells (PEMFCs), playing a crucial role in mass and heat transport. Enhancing GDL microstructures directly improves transport properties, thereby leading to more efficient and durable PEMFCs. In this study, we developed a novel machine learning methodology to optimize the GDL microstructure and properties. The developed optimization framework demonstrated high efficacy, with an R^2 score $\sim 95\%$ in 6 out of 7 properties and a R^2 score $\sim 90\%$ for the GDL-Micro-Porous Layer (MPL) contact resistance. We validated our machine learning approach by comparing the predicted GDL properties to those calculated through digital characterization using physics-based methods from the stochastically generated GDL, using the optimal manufacturing parameters identified by the optimizer. Our machine learning model predicted accurately 7 GDL properties decreasing the computational cost from ~ 3 to 4 h wall time (physical model) to ~ 3 s wall time. Results show that low fiber concentration accompanied by low compression ratio

* Corresponding author. Laboratoire de Réactivité et Chimie des Solides (LRCS), Université de Picardie Jules Verne, Hub de l'Energie, UMR CNRS 7314, 15 rue Baudelocque, 80039, Amiens, France.

E-mail address: alejandro.franco@u-picardie.fr (A.A. Franco).

<https://doi.org/10.1016/j.jpowsour.2024.235583>

Received 7 August 2024; Received in revised form 22 September 2024; Accepted 3 October 2024

Available online 29 October 2024

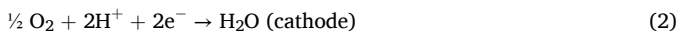
0378-7753/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

achieve maximum diffusivity and minimum GDL-MPL contact resistance. Furthermore, prioritizing maximum electrical and/or thermal conductivities while minimizing GDL-MPL contact resistance require high fiber concentration with high compression ratio. This optimization strategy shows significant potential for improving gas transport, water management, efficient current collection, and thermal regulation within PEMFCs.

1. Introduction

Effects of climate change such as scorching heat waves and rising sea levels are no longer just a looming threat but a reality we face every day [1]. Greenhouse gas emissions from traditional energy sources, such as fossil fuels, are pushing our planet to the brink. In 2023, the European Union embraced a series of commission recommendations aimed at aligning the EU's climate, energy, transport, and taxation strategies to effectively curtail net greenhouse gas emissions by a minimum of 55 % by 2030 [2]. This initiative is poised to position the EU as the pioneering climate-neutral continent by the year 2050. The primary objective of the net zero [3] is to achieve carbon neutrality by significantly reducing carbon emissions to a level that can be absorbed and stored sustainably by natural mechanisms and additional carbon dioxide removal methods, thereby ensuring a balanced atmospheric carbon footprint of zero. Presently, the energy sector is accountable for approximately three-quarters of greenhouse gas emissions, thus playing a pivotal role in mitigating the adverse impacts of climate change. Substituting environmentally damaging coal, gas, and oil-based power generation with renewable energy alternatives holds significant potential for significantly mitigating carbon emissions [3]. Hence, shifting to sustainable energy sources offer a powerful solution in creating a cleaner and healthier future for next generations.

Proton Exchange Membrane Fuel Cells (PEMFCs) has been studied for years, as a promising clean energy technology due to its high efficiency, low operating temperature, and most importantly, zero carbon emissions [4]. This technology is increasingly recognized as a valuable complement to battery electric vehicles (BEVs). While BEVs currently dominate the electric vehicle market, PEMFCs offer a viable alternative for applications where BEVs encounter limitations, such as in heavy-duty trucks [5]. PEMFCs convert hydrogen as a chemical fuel directly into electricity by employing two electrochemical reactions: oxidation of hydrogen in the anode (Eq. (1)) and reduction of oxygen in the cathode (Eq. (2)):



Electrochemical Eqs. (1) and (2) mainly take place in the heart of the PEMFC, called Membrane Electrode Assembly (MEA). MEA is composed of a proton conducting polymer, called the membrane, sandwiched between the catalyst layers (CLs), a gas diffusion layers (GDLs) and microporous layers (MPLs) on each side. The GDL is a porous material that has a good transport capability, allowing mass and heat transport within its microstructure, ultimately affecting the fuel cell performance [4]. MPL is a thin, carbon-based coating applied on top of the GDL and serves as the interface between the GDL and CL. The porous layer, made up of GDL and MPL, allow gas transport, maintaining water balance and supporting efficient electrochemical reactions within the fuel cell [6]. The mentioned porous layer is typically a bi-layer carbon-based porous material made up of a MPL with a usual thickness of 10–100 μm and a macroporous substrate, called the GDL, with a typical thickness of 150–400 μm [7], depending on the manufacturer, type and compression level. The GDL should allow the flow of reactant gases (H_2 and oxygen) and product gases to pass through it. The water formed in a cell should not choke the pores, so they are precoated with polytetrafluoroethylene (PTFE), which makes the GDL surface hydrophobic [8]. The MPL, commonly manufactured by intermixing a hydrophobic polymer (PTFE) with carbon black, enhances water removal rate [9], enhances electrical connectivity [10], relieves water accumulation by pushing the liquid

water from cathode to anode [11], and improves evaporation in the electrodes [12,13]. In the upcoming paragraph, we describe some of the key GDL factors in determining the overall performance and efficiency of PEMFC.

Porosity of the GDL directly impacts the flow of reactant gases and ensures uniform distribution to the CL and the removal of water from the fuel cell. In addition to porosity, the geometric tortuosity and pore size distribution within the GDL are also critical factors that influence the overall performance of the PEMFC [14]. Geometric tortuosity of the GDL in a PEMFC plays a crucial role in its performance. Geometric tortuosity refers to the ratio of the actual diffusive path length to the geometric thickness of the porous medium. Hence, geometric tortuosity affects the transport of gases, liquids, and heat within the fuel cell system. It influences the permeability, which in turn impact the power density and efficiency of the fuel cell. Higher geometric tortuosity can lead to longer diffusive path lengths, affecting the transport of reactants and products within the GDL [15]. Furthermore, water management within the microstructure of the GDL is fundamental for facilitating the passage of water in both vapor and liquid states through its fibers. Therefore, determining the relative diffusivity of water within the GDL microstructure is critical to mitigate flooding by modulating the fiber and binder concentrations, thereby influencing the porosity of the system. Integration of hydrophobic PTFE aids in water repellence, thereby averting flooding; however, excessive drying, which compromises GDL performance and subsequently affects the performance of the PEMFC, must be avoided. The diffusivity characteristics of the GDL are paramount in influencing the distribution of temperature and water saturation within the proton exchange membrane, consequently impacting fuel cell performance and longevity. A comprehensive understanding of diffusivity is pivotal for optimizing GDL design and function, as it directly governs the interplay among gas, liquid, and thermal dynamics within the fuel cell [16,17]. In addition, the electrical conductivity of the GDL directly impacts the power density and efficiency of the fuel cell system. Higher in-plane resistivity of the GDL can have a negative impact on the current density in certain areas, while lower resistivity can lead to improved current density. Additionally, the local current density in the CL is affected by the electrical conductivity of the GDL [18]. The thermal conductivity of the GDL in PEMFCs is intricately linked with water management and thermal regulation. As the electrochemical reactions within the MEA generate heat, effective management of thermal gradients across the single cell and fuel cell stack becomes imperative. This necessitates the provision of a pathway for heat dissipation from the MEA to the bipolar plates, ensuring optimal operating temperatures to prevent membrane degradation. Anisotropic thermal conductivity of the GDL significantly influences temperature distribution and water saturation within the proton exchange membrane, thereby impacting cell performance and durability. Notably, proper thermal management is crucial for sustaining fuel cell performance, as the temperature of the MEA profoundly influences overall efficiency [19]. Furthermore, it is observed that in GDLs with notable anisotropic thermal conductivity, the maximum and minimum temperatures in the cathode CL are primarily contingent upon average current density rather than local current density [20–24]. And ultimately, increased contact resistance between the GDL and the MPL causes great ohmic losses by acting as an electron bottleneck, translating to lower voltage output and overall efficiency. In addition, it also impedes efficient transfer of electrons between the CL and the current collector (through the GDL). This ultimately reduces the PEMFC's ability to generate electricity efficiently [19]. Hence, a low contact resistance between the GDL and the MPL is crucial for optimal PEMFC function [25,26].

There exists substantial prior research, encompassing experimental, computational, and combined experimental-computational studies that has investigated transport properties of GDLs. Extensive experimental investigations have significantly advanced our comprehension of mass transport, heat transfer and structural characteristics of GDLs, contributing valuable insights to our understanding of GDL performance. For example, Yoshimune et al. [27] developed an experimental method to evaluate the through-plane effective diffusion coefficient in GDLs using an infrared-absorption carbon dioxide sensor. They measured the Through-plane (TP) effective gas diffusivity of a commercial GDL and found to be 0.36 ± 0.02 at 25°C which they claimed to be consistent with the data presented in literature [28–36]. James et al. [37] used x-ray computed tomography reconstruction to investigate a commercially available GDL under non-homogeneous compression, claiming that non-uniform compression significantly impacts transport properties within the material, leading to a roughly twofold difference between the land and channel regions. In addition, Xia et al. [38] investigated how GDL thickness and porosity in high-temperature PEMFCs affect electron conduction and gas distribution. By optimizing these parameters (anode: $80\text{--}120\ \mu\text{m}$, cathode: $140\text{--}170\ \mu\text{m}$; porosity: $35\text{--}45\%$), the researchers achieved a remarkable 7.7 % improvement in cell performance, underscoring the critical role of precise GDL design in maximizing fuel cell efficiency. Moreover, Pfrang et al. [39] used X-ray computed tomography to investigate the thermal properties of several commercial GDLs without the MPL, concluding that the TP thermal conductivities are lower than the in-plane (IP) thermal conductivities. Furthermore, Chen et al. [40] subjected a GDL to a series of freeze-thaw cycles to assess the impact on its IP electrical resistivity under low compression, effectively assessing the impact of thermal aging on electrical performance. Results indicated weak anisotropy and greatly impacted by the transformation of carbon fiber connections within the GDL's porous structure. Notably, the initial 100 freeze-thaw cycles appeared to have the most pronounced effect on the IP electrical degradation performance, highlighting the sensitivity of the GDL to thermal stress.

Prior research also exists focusing on calculating the TP thermal conductivity and thermal contact resistance. Nitta et al. [41] investigated the influence of compression pressure on the mechanical behaviour and thermal properties of GDLs. The stress-strain response of the GDL exhibited a unique characteristic: a non-linear region followed by two distinct linear sections within the $0\text{--}5.5\ \text{MPa}$ compression range. Interestingly, the thermal conductivity of the compressed GDL remained independent of the compression, exhibiting a value of $1.18 \pm 0.11\ \text{W m}^{-1}\text{K}^{-1}$ at room temperature. Moreover, Khandelwal et al. [42] addressed the challenge of inconsistent thermal conductivity data for fuel cell materials by employing a precise steady-state technique. The researchers determined the TP thermal conductivity and contact resistance of various fuel cell components, including the dry Nafion, catalyst layer, multiple diffusion media, and the thermal contact resistance between diffusion media and a metal plate as a function of temperature and pressure. This data will aid in resolving discrepancies in literature and improve the accuracy of temperature distribution modelling in operational PEMFCs. In summary, the aforementioned studies demonstrate a concerted effort to improve the PEMFC technology by addressing current limitations through various approaches.

Regarding computational modelling, it has played a significant role in advancing our understanding of the complex phenomena taking place in PEMFCs. One approach to digitally generate GDLs microstructures involves the utilization of stochastic reconstruction methods [43]. Stochastic reconstruction methods encompass computational procedures employed to generate random or probabilistic depictions of intricate structures or systems, typically grounded in statistical principles or random processes. These techniques find widespread application across diverse domains such as materials science, image processing, and computer graphics, facilitating the simulation and analysis of complex phenomena. Stochastic reconstruction methods enable the exploration of a wide range of GDL designs, facilitating the identification of optimal

microstructural configurations that enhance the performance of GDLs, and ultimately, PEMFCs. Regarding the study of effective transport properties, the Lattice Boltzmann Method (LBM) has been proven as an outstanding approach. For example, Sarkezi-Selsky et al. [44] calculated the effect of PTFE on water transport through the GDL, concluding that increasing the amount of PTFE led to lower water saturation levels for a given capillary pressure, suggesting that increasing the hydrophobicity of the GDL reduces the risk of flooding. In addition, Wu et al. [45] employed LBM to calculate the effective transport physical properties as well as the effective thermal and electrical conductivities, diffusivity of reconstructed GDL/CL and permeability of reconstructed GDL. Tayarani-Yoosefabad et al. [7] developed a multiscale modelling tool for MPLs and MPL-coated GDLs considering various materials. Validated by experiments, the tool simulated how MPL structure and composition affect transport properties. They found out that optimizing MPL structure can improve transport within the fuel cell, and the validated modelling approach can be a valuable design tool for both MPLs and GDLs, reducing design cycle time and cost. Moreover, Hannach et al. [46] employed a numerical model to compute the effective transport properties of stochastically generated MPL and presented thermal and electrical conductivities of $0.25\ \text{W m}^{-1}\text{K}^{-1}$ and $1400\ \text{S m}^{-1}$, respectively at 58 % porosity.

Finally, several studies used machine learning (ML) to optimize the microstructure of GDL with different goals: for example, Hao et al. [47] investigated the relationship between the thickness of GDL and output power. Several studies [48–51] have employed ML followed by multi-objective optimization techniques to investigate the GDL microstructure. The ML models used were supported on convolutional neural network, extreme learning machine, kernel extreme learning machine, and response surface methodology and artificial neural network, respectively. These studies have focused on varying objectives and parameters, leading to a diverse landscape of findings. However, the previously mentioned works were restricted to optimizing only one GDL output (e.g. permeability [48], current density while minimizing temperature [49], power density [50], and again, current density [51]).

And lastly, in the battery field, our research group in Amiens, pioneered the multi-objective optimization of battery electrodes. For instance, in previous publications [52,53] our group in Amiens reported an innovative approach combining physics-based simulations of lithium ion battery electrode manufacturing, ML surrogate modelling and multi-objective Bayesian Optimization. 3D-resolved electrode microstructures generated by Coarse Grained Molecular Dynamics and Discrete Element Method (DEM) were characterized in terms of their properties and used to derive a ML surrogate model predicting how the manufacturing parameters (e.g. formulation, calendaring degree) impact the electrode properties. By embedding the ML surrogate model into a Bayesian Optimizer, it was possible to predict which manufacturing parameter values need to be adopted in order to maximize and/or minimize properties (e.g. electrode conductivity, geometric tortuosity factor, energy density). In another publication from our group in Amiens a similar strategy was followed to optimize carbon felt electrode microstructures for redox flow batteries, by combining stochastic generation of the microstructures, LBM simulation and Bayesian Optimization [54].

In this paper, we take inspiration from our previous battery works, to address the complexities inherent to the GDL microstructure, such as the electrical and thermal conductivities and diffusivity, by extending further the approach to account for numerous input parameters in contrast to previous PEMFC/GDL literature. The novelty of our work herein lies in its proposal of a multi-objective optimization of manufacturing parameters for improving the properties of the GDL microstructure (Fig. 1). The optimization process seeks to strike a balance between minimizing contact resistance between GDL and MPL and maximizing electrical conductivity, thermal conductivity and diffusivity. Through these optimizations, we aim to achieve enhanced gas transport, water management, efficient current collection, and thermal

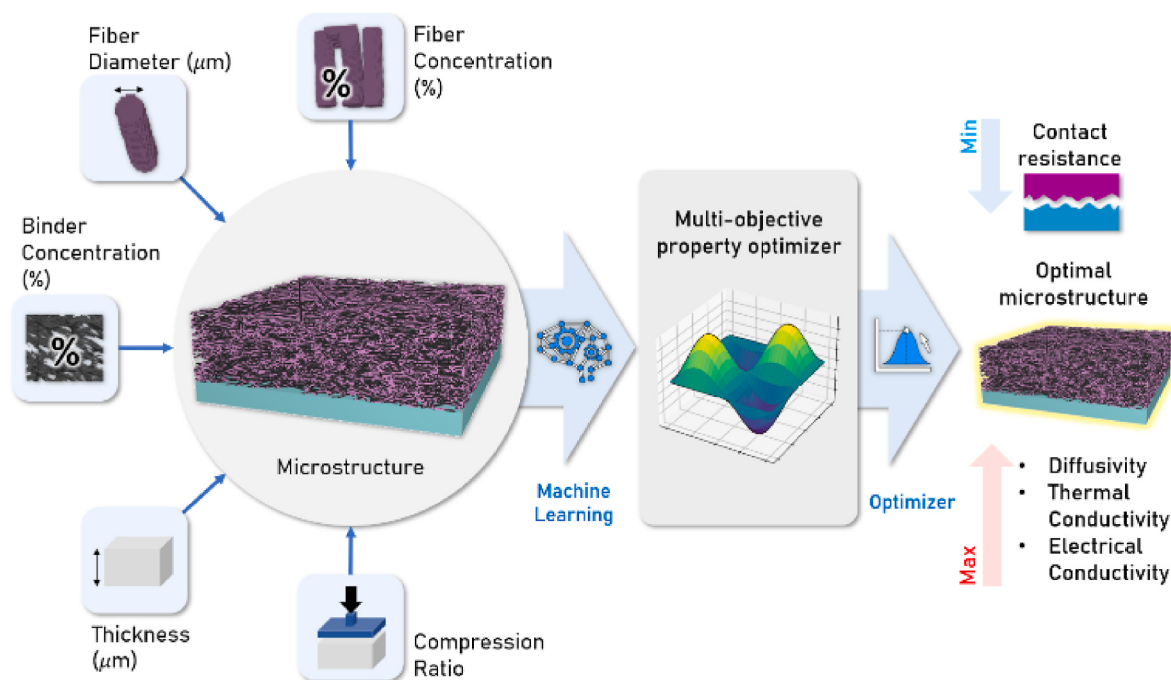


Fig. 1. Schematic representation of an electrochemical performance optimization based on fiber concentration, fiber diameter, binder concentration, thickness and compression ratio to obtain the minimum contact resistance under different scenarios.

regulation. This will facilitate the optimal distribution of reactant gases to the CL, ultimately promoting the overall efficiency and durability of the GDL and consequently, the performance and longevity of the PEMFC.

2. Methodology

This work seeks to optimize the microstructural manufacturing parameters of the GDL with the objective of achieving a reliable performance in terms of the targeted transport properties. Our workflow is composed of 4 stages: data acquisition, data preprocessing, model training and prediction & optimization.

2.1. Data acquisition

To generate the data, we employed Design of Virtual Experiments, a systematic approach to study the relationship between multiple input variables and response output variables. This enabled us to efficiently select the input parameter values to observe their effect in output properties. The selected input parameters are: fiber diameter, fiber concentration, binder concentration, compression ratio, and GDL thickness.

To achieve a comprehensive exploration of the hyperparameter space, the Latin Hypercube Sampling (LHS) [55] method was employed, generating a uniform distribution of 250 unique input parameter combinations. Through its systematic selection across each dimension, LHS mitigates clustering and ensures a more representative sample, leading to a more robust evaluation of the model. Table 1 depicts the five input parameters to be optimized and their associated range. The selection of the latter is informed by established literature values and aims to replicate the technical properties observed in commercially available GDLs.

To assess the uniformity of the distribution of the 250 data points within the hyperparameter space explored using LHS, we present 2D scatter plots in the supporting information (Fig. S1). The mentioned scatter plots depict each sample as a point, allowing for a visual inspection of how evenly these points are distributed across two

Table 1
Input parameter range of the optimization workflow.

	Minimum Value	Maximum Value	Source reference number
Fiber Diameter (μm)	6	12	[20,56]
Fiber Concentration (%)	10	25	[57]
Binder Concentration (%)	5	10	[57]
Thickness (μm)	280	320	[7]
Compression ratio	0	0.5	[37,48]

dimensions representing a combination of the input parameters. Following the generation of a comprehensive set of input parameters through LHS, the GeoDict software [58] was employed to stochastically generate microstructures for the GDL. This process involved iterating through each unique combination of fiber diameter, fiber concentration, binder concentration (PTFE), GDL thickness, and compression ratio. All remaining input parameters were held constant across all generated data points. Consequently, this approach yielded a collection of stochastically generated GDL microstructures. GDL microstructures domain sizes were set to: $1000 \mu\text{m} \times 1000 \mu\text{m} \times \text{thickness}$, and $2000 \mu\text{m} \times 200 \mu\text{m} \times \text{thickness}$ after a careful size convergence analysis performed to obtain a good compromise between computational cost and model accuracy. Further details about the convergence analysis are shown in the results section.

The snapshots presented in Fig. 2 depict the progressive microstructural development of the GDL at each processing step. The initial stage (Fig. 2a) illustrates the stochastic generation of fibers based on the designated fiber diameter and concentration. Subsequently (Fig. 2b), PTFE as the binder material was incorporated according to the specified binder concentration and contact angle. It is important to clarify that the terms fiber concentration and binder concentration used in this study refer to solid volume percentages. These values allow for the determination of the initial porosity of the stochastically generated GDL. The following step (Fig. 2c) involves compression of the structure to the predefined compression factor. Following the stochastic generation of

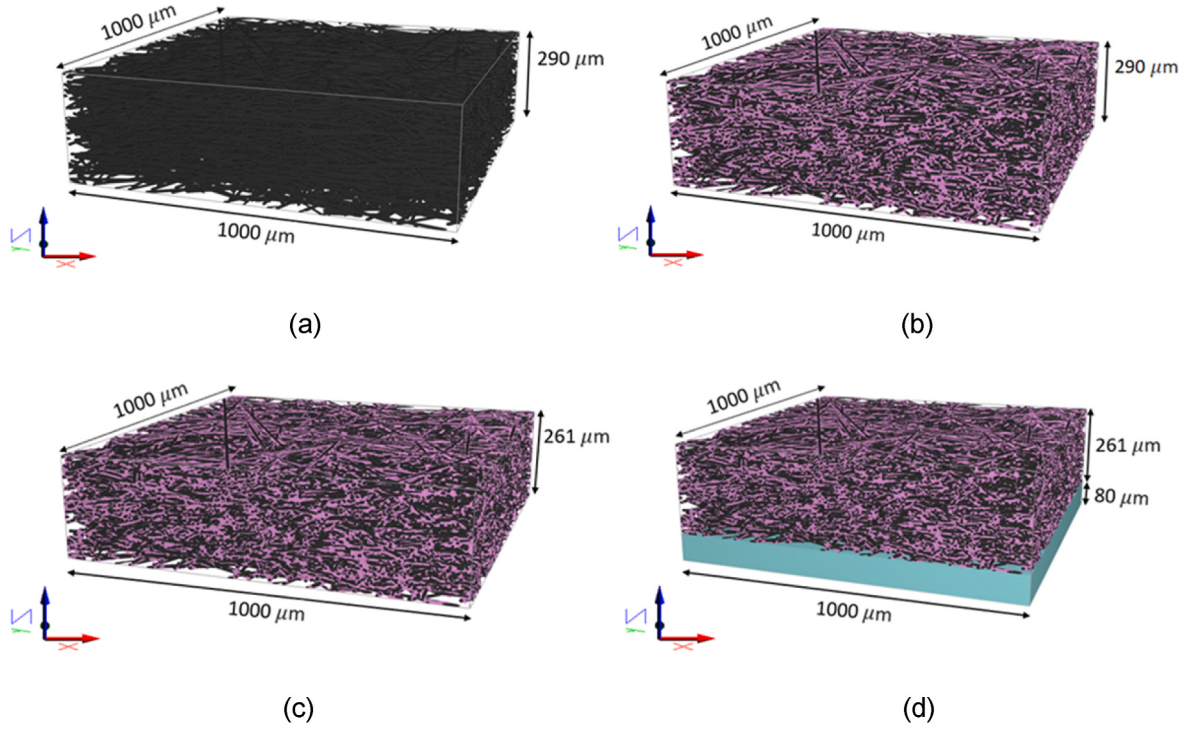


Fig. 2. (a) GDL microstructure after addition of fibers. The subsequent microstructures showcase the addition of (b) binder material (PTFE), (c) the compression of the structure, and finally, (d) the attachment of the 80 μm MPL.

the microstructure, the transport properties of these GDLs were then digitally characterized through physics-based calculations. This characterization included through-plane (TP) and in-plane (IP) electrical conductivity, thermal conductivity, and diffusivity. Details on these physics-based digital characterizations are presented in [supporting information S1-S4](#).

Finally (Fig. 2d), a MPL with a thickness of 80 μm was attached along the negative z-axis. It is noteworthy that the MPL was assumed to be a porous, homogeneous, and isotropic layer characterized by thermal and electrical conductivities of $0.25 \text{ W m}^{-1}\text{K}^{-1}$ and 1400 S m^{-1} , respectively. These values are adopted from the work of Hannach et al. [46] Subsequently, the electrical conductivity of the combined GDL-MPL structures were then reevaluated. This recalculation served as the foundation for determining the contact resistance between two such layers. The contact resistance, CR , was quantified using equation (3):

$$CR = R_{\text{Total}} - R_{\text{GDL}} - R_{\text{MPL}} \quad (3)$$

2.2. Data Pre-processing

The full dataset consists of 250 samples containing the 5 input parameters: fiber diameter (Fd), fiber concentration (Fc), binder concentration (Bc), thickness (Th) and compression ratio (Cr), and its corresponding 7 output parameters: thermal IP (TIP), thermal TP (TTP), electrical IP (EIP), electrical TP (ETP), diffusivity IP (DIP), diffusivity TP (DTP) and contact resistance (CR). Therefore, the dataset is a matrix M of dimensions 250×12 , where the first 5 columns correspond to the input parameters and the last 7 columns correspond to the output parameters.

$$M = \begin{bmatrix} Fd_1 & Fc_1 & Bc_1 & Th_1 & Cr_1 & TIP_1 & TTP_1 & EIP_1 & ETP_1 & DIP_1 & DTP_1 & CR_1 \\ Fd_2 & Fc_2 & Bc_2 & Th_2 & Cr_2 & TIP_2 & TTP_2 & EIP_2 & ETP_2 & DIP_2 & DTP_2 & CR_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Fd_{250} & Fc_{250} & Bc_{250} & Th_{250} & Cr_{250} & TIP_{250} & TTP_{250} & EIP_{250} & ETP_{250} & DIP_{250} & DTP_{250} & CR_{250} \end{bmatrix} \quad (4)$$

M was divided in two matrices, X containing the 5 input and Y con-

taining the 7 output parameters. Out of the total 250 samples, 200 are randomly taken for training, and the remaining 50 for testing. Therefore, the full dataset is divided into 4 matrices, X_{train} , Y_{train} , X_{test} and Y_{test} .

$$X_{\text{train}} = \begin{bmatrix} Fd_1 & Fc_1 & Bc_1 & Th_1 & Cr_1 \\ Fd_2 & Fc_2 & Bc_2 & Th_2 & Cr_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Fd_{200} & Fc_{200} & Bc_{200} & Th_{200} & Cr_{200} \end{bmatrix} \quad (5a)$$

$$Y_{\text{train}} = \begin{bmatrix} TIP_1 & TTP_1 & EIP_1 & ETP_1 & DIP_1 & DTP_1 & CR_1 \\ TIP_2 & TTP_2 & EIP_2 & ETP_2 & DIP_2 & DTP_2 & CR_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ TIP_{200} & TTP_{200} & EIP_{200} & ETP_{200} & DIP_{200} & DTP_{200} & CR_{200} \end{bmatrix} \quad (5b)$$

$$X_{\text{test}} = \begin{bmatrix} Fd_{201} & Fc_{201} & Bc_{201} & Th_{201} & Cr_{201} \\ Fd_{202} & Fc_{202} & Bc_{202} & Th_{202} & Cr_{202} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Fd_{250} & Fc_{250} & Bc_{250} & Th_{250} & Cr_{250} \end{bmatrix} \quad (6a)$$

$$Y_{\text{test}} = \begin{bmatrix} TIP_{201} & TTP_{201} & EIP_{201} & ETP_{201} & DIP_{201} & DTP_{201} & CR_{201} \\ TIP_{202} & TTP_{202} & EIP_{202} & ETP_{202} & DIP_{202} & DTP_{202} & CR_{202} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ TIP_{250} & TTP_{250} & EIP_{250} & ETP_{250} & DIP_{250} & DTP_{250} & CR_{250} \end{bmatrix} \quad (6b)$$

2.3. Model training

In the present work, a random forest regression method was used to learn the complex relationships between the five input and the seven output parameters. A random forest fits a number of decision tree regressors (forest) on several dataset sub-samples averaging them with the

main goal of reducing the variance, thus, making it less likely to overfit in comparison to a single decision tree solution [59]. Given the use of out-of-the-bag evaluation technique in the random forest regression method, we do not require a validation dataset [60]. The model was optimized by minimizing the mean square error (MSE).

The model was evaluated on the test dataset by calculating the MSE, mean absolute error (MAE) and R^2 score for all the 7 output parameters.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

where n is the number of samples (50 for testing), y_i is the target value and $\hat{y}_i$ is the predicted value. An Adapting Boosting, AdaBoost [61], ensemble method was used in combination with the random forest regressor to predict the 7 output properties. The AdaBoost algorithm combined with random forest regressor as weak learning has been previously used in different regression problems showing minor errors and strong predictability capacity [62–65]. Details of the ML architecture and parameters are shown in Fig. S2.

2.4. Optimization

After training the regressor algorithm, a multi-objective optimization was performed. We defined an objective function, transforming the multi-objective problem into a single function and denoted by C_f , to be minimized for the following 4 cases: 1. minimizing contact resistance (CR) and maximizing diffusivities (DIP & DTP) (Eq. (10)), 2. Minimizing contact resistance (CR) and maximizing thermal conductivities (TIP & TTP) (Eq. (11)), 3. Minimizing contact resistance (CR) and maximizing electrical conductivities (EIP & ETP) (Eq. (12)), and 4. Minimizing contact resistance (CR) and maximizing simultaneously diffusivities (DIP & DTP), thermal conductivities (TIP & TTP) and electrical conductivities (EIP & ETP) (Eq. (13)).

$$C_f^1 = \min (CR - DIP - DTP) \quad (10)$$

$$C_f^2 = \min (CR - TIP - TTP) \quad (11)$$

$$C_f^3 = \min (CR - EIP - ETP) \quad (12)$$

$$C_f^4 = \min (CR - DIP - DTP - TIP - TTP - EIP - ETP) \quad (13)$$

Such functions are dependent on the input parameters and the deterministic learning performed in the training of the model predicting the property as a function of a set of input parameters. Given the different orders of magnitude of the different properties, the deterministic learning was trained on data previously normalized allowing to fit each property value into [0, 1], avoiding any bias induced by the property values. For the optimization we used a Nelder-Mead (NM) method [66] together with a grid search (GS) approach. The NM method minimizes the objective function, C_f , by following an algorithm based on simplex, checking its evaluation at each vertex of the simplex [67]. In order to improve the optimized result and trying to decrease the chances of finding local minima, we also used a GS approach. GS is an iterative algorithm that will split the input data in various combinations exploring the output in all of them. Since our input data is a 5D dataset given the 5 input parameters, we split the range of each input parameter in 4, forming a grid of $4^5 = 1024$ points, on each of these point the NM

method is performed, keeping only the minimum value of the objective function. The ML regression model and optimization model were developed on Python using the Scikit-learn package [68].

Summarizing, in this study we employed a ML approach to optimize the GDL properties for fuel cell applications. The model, as schematized in Fig. 3, considers five input parameters and predicts 7 GDL properties as outputs, including electrical and thermal conductivities (in-plane and through-plane), diffusivity conductivities (in-plane and through-plane), and GDL-MPL contact resistance. The complexity of the problem arises from its multi-dimensional nature, with several interdependent parameters. A ML-based optimization strategy is implemented to efficiently identify optimal parameter sets. This two-step process involves: (1) establishing a robust correlation function between input and output parameters, achieved using a combination of Random Forest Regressor and Ada Boost Regressor, aiming for an R-squared score exceeding 0.9 for all predictions. (2) Utilizing the established correlation function to guide an optimizer towards optimal parameter sets based on predefined conditions. The NM method is chosen for this purpose, alongside a grid search approach to mitigate the risk of getting trapped in local minima. We then compare the GDL properties using the results from our optimization workflow and the physics-based digital characterization of stochastically generated GDL microstructure. Both used the optimal input parameters from the five cases mentioned above as input parameters.

3. Results and discussion

3.1. Domain size convergence analysis

A convergence test was conducted to verify the representativeness of the chosen domain sizes for the system under investigation. Eighteen unique domain sizes with increasing surface area were generated using ten different random seeds. These domains were evaluated based on the material properties (IP and TP electrical conductivity, thermal conductivity, and diffusivity) at 10 % compression with an 8 μm fiber diameter (Fig. 4) and a 12 μm fiber diameter (Fig. 5). Each calculation showed in Figs. 4 and 5, as well as the data showed in Table S1 and Table S2 were obtained using 10 different random seeds for each case. Considering the 12 calculated properties, 6 for the 8 μm fiber diameter and 6 for the 12 μm fiber diameter, the 2000 $\mu\text{m} \times 200 \mu\text{m}$ domain present a less than 15 % error in 9 out of 12 properties when compared to the biggest sample (4000 $\mu\text{m} \times 2000 \mu\text{m}$). Regarding the 1000 $\mu\text{m} \times 1000 \mu\text{m}$ domain size, it presents a less than 15 % error in 8 out of 12 properties when compared to the biggest sample (4000 $\mu\text{m} \times 4000 \mu\text{m}$).

Additionally, we also analysed the influence of the 10 different random seeds on the material properties. We calculated the Relative Standard Deviation (RSD), which is expressed in percentage and is obtained by multiplying the standard deviation by 100 and dividing this product by the mean. RSD measures the dispersion of a set of numbers around the mean. A higher RSD indicates a greater deviation, meaning the numbers are spread further away from the mean. On the contrary, a lower RSD indicates a smaller deviation, suggesting the numbers are closer to the mean. We used the RSD to determine the influence of different random seeds on the material properties. As showed in Table 2, all RSD values are less than 7 % except for the electrical TP in both cases with values of 11.4 % and 13.1 % for both domain sizes. In chemical analysis, a reliable value for RSD is below 10 %. Keeping that in mind, considering the purposes of this work we consider that our RSD values indicate that independently of the random seed, the samples produce statistically identical results. The RSD average along the 6 properties are 3.8 % and 4.91 %, for the 2000 $\mu\text{m} \times 200 \mu\text{m}$ and 1000 $\mu\text{m} \times 1000 \mu\text{m}$ domain sizes, respectively. Consequently, the influence of different random seeds on the material properties appears negligible.

Given the error percentage calculation when compared to the biggest sample (Table S3) and the non-overlap when different random seeds and different fiber diameter used, we decided to prepare two datasets: 2000

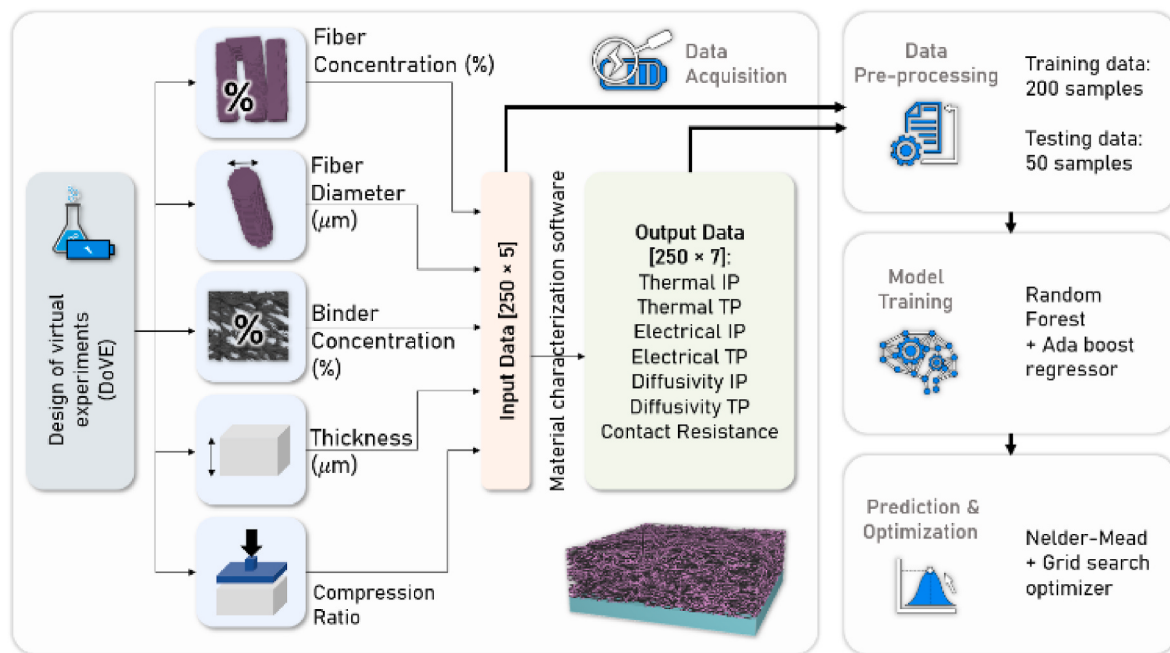


Fig. 3. Diagram illustrating our research workflow.

$\mu\text{m} \times 200 \mu\text{m}$ and $1000 \mu\text{m} \times 1000 \mu\text{m}$, since both domain sizes show the least error percentage ($<15\%$) when they are compared with the biggest domain size (Table S3) in at least 8 of the 12 functional properties. Additionally, both domain sizes ($2000 \mu\text{m} \times 200 \mu\text{m}$ and $1000 \mu\text{m} \times 1000 \mu\text{m}$) are computationally less expensive to calculate in comparison to the biggest domain size sample ($4000 \mu\text{m} \times 4000 \mu\text{m}$). As an example, characterization of a $2000 \mu\text{m} \times 200 \mu\text{m}$, $1000 \mu\text{m} \times 1000 \mu\text{m}$ and a $4000 \mu\text{m} \times 4000 \mu\text{m}$ takes 2 h, 2.5 h, and 13 h wall-time, respectively. Another reason to use both domain sizes to prepare two different models, is that in the properties' calculations (Figs. 4–5) it is evident that there are two clear different levels when checking the properties values as a function of domain size. Even though we consider as ground truth the values obtained from the biggest sample ($4000 \mu\text{m} \times 4000 \mu\text{m}$), and it is the $2000 \mu\text{m} \times 200 \mu\text{m}$ domain size that showed values closer to the biggest one, we did not discard the $1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size, and preferred to train two models using a different domain size data set for each on them and then analysed and compared how different or similar the outcomes are. Therefore, two deterministic models were trained for each domain size dataset, and then two optimizers were prepared to get optimal functional properties for both domain sizes.

3.2. Correlation vs partial correlation: input vs output

A correlation matrix is a useful tool to visualize pairwise correlations between all variables in a dataset. Its elements represent the correlation coefficient between two variables, ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no significant correlation. The partial correlation matrix, on the other hand, measures the unique association between two variables while holding constant the effects of other variables. By removing the shared variance due to other variables, partial correlations reveal the direct relationship between the variables of interest [69].

Fig. 6 depicts the correlation and partial correlation matrices using the Pearson method [70], showing how varying the input parameters (fiber diameter, fiber concentration, binder concentration, GDL thickness and compression ratio) affects the GDL performance. For additional insight, three properties were also added with respect to the database used in the training set: Porosity, Geometric Tortuosity, and Resistivity. The correlation matrix (Fig. 6a) reveals an association between porosity

and both fiber concentration and compression ratio. However, the partial correlation matrix (Fig. 6b) suggests that compression primarily drives this relationship. This implies that increasing compression leads to a decrease in porosity. Furthermore, fiber concentration shows a positive correlation with thermal conductivity, which aligns with expectations due to the conductive properties of the fibers. Conversely, fiber concentration demonstrates a negative correlation with diffusivity. This occurs because increasing the amount of solid materials (fiber and binder) within the domain reduces the available space for diffusion, thereby decreasing diffusivity. Similar to fiber concentration, binder concentration shows a negative correlation with porosity and diffusivity. This is due to the increased amount of solid materials, which leads to a reduction in both porosity and diffusivity. Additionally, binder concentration has minimal impact on thermal and electrical conductivities because of the magnitude of the PTFE loading. Furthermore, as the compression ratio increases, porosity and diffusivity decrease, while geometric tortuosity and both thermal and electrical conductivities increase. This is because compressing the domain creates more contact points between the fibers, thereby enhancing conductivity. Focusing more on the partial correlation matrix, geometric tortuosity exhibits a positive/negative correlation with fiber diameter, binder concentration, GDL thickness, and compression ratio. Conversely, in-plane thermal and electrical conductivities are positively/negatively correlated mainly with fiber diameter, indicating enhanced conduction through the fibers with increasing diameter. Through-plane conductivities, however, demonstrate a comparable dependence on fiber diameter, fiber concentration, and thickness. Binder concentration appears to primarily impact TP diffusivity, potentially due to its distribution within the GDL microstructure. Contact resistance displays the weakest correlation with GDL thickness, as expected since the thickness does not necessarily affect the contact points between the carbon fibers and the MPL layer.

In summary, increasing fiber and binder concentration generally reduces porosity while increasing geometric tortuosity. Electrical and thermal conductivities (IP and TP) tend to rise with increasing fiber diameter and fiber concentration due to its conductive nature. Resistivity and diffusivity decrease with these changes. The complete correlation (Fig. S3) and partial correlation (Fig. S4) matrices are provided in the Supplementary Information for further reference. Similar trends are also observed in the correlation (Fig. S5) and partial

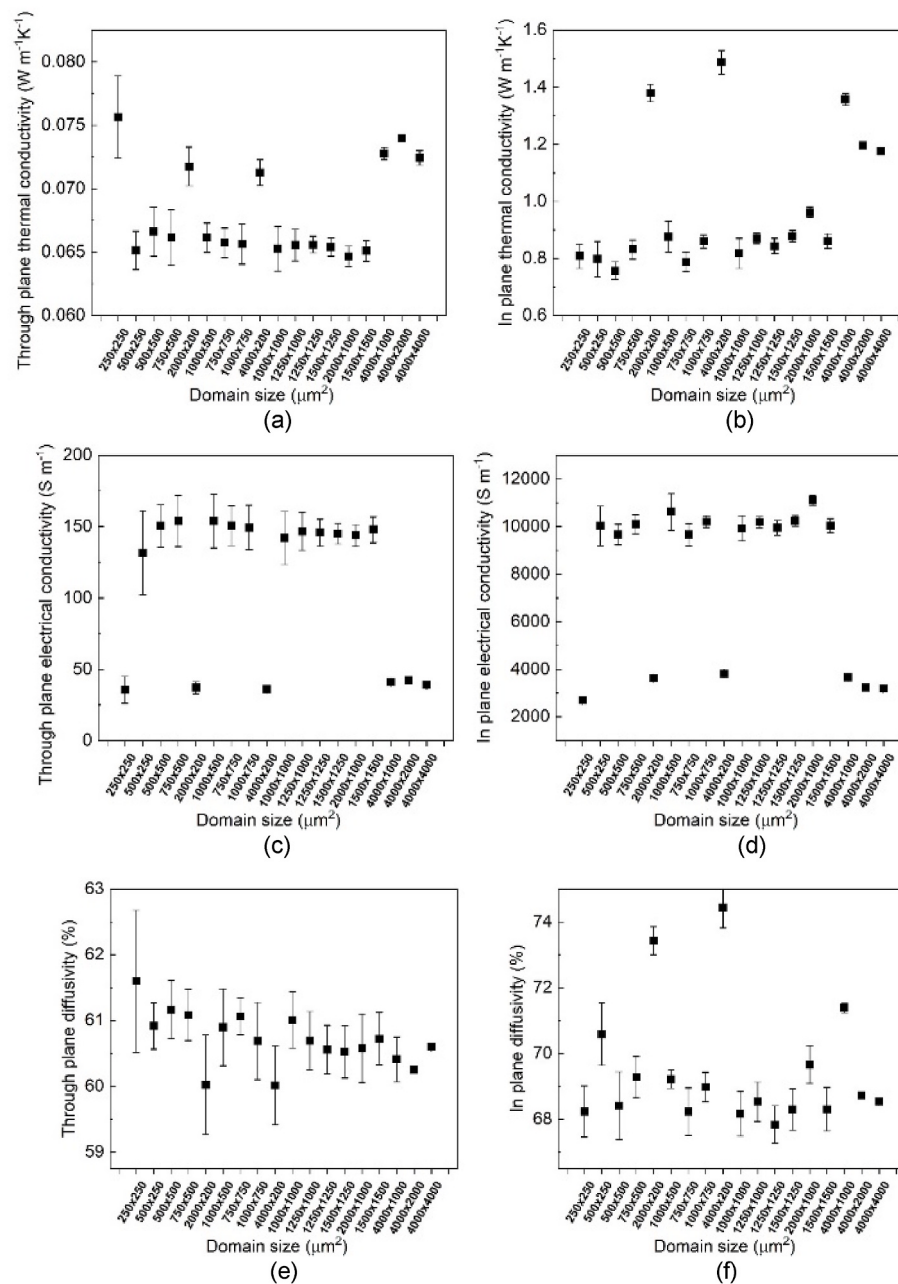


Fig. 4. Convergence analysis conducted on 3 different GDL properties (IP and TP electrical conductivity, thermal conductivity, and diffusivity) generated using ten different random seeds using 8 μm fiber diameter.

correlation (Fig. S6) matrices for the dataset with a domain size of $2000 \mu\text{m} \times 200 \mu\text{m}$. The consistency of these trends across different domain sizes suggests that the correlation and partial correlation analyses are robust and independent of the specific domain size dataset. This indicates that the underlying relationships between the variables are preserved regardless of the size of the domain, reinforcing the generalizability and reliability of the correlation analysis in this context.

3.3. Machine learning performance

Validation metrics were calculated over the testing dataset are MSE, MAE and R^2 score and are reported in [Table 3](#) for the $1000\text{ }\mu\text{m} \times 1000\text{ }\mu\text{m}$ domain size and in [Table S4](#) for the $2000\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ domain size. In addition to the current model, which consist of a random forest with an Ada boost regressor, we also calculated the validation metrics of two additional models, a pure random forest and a pure Ada boost one.

Validation metrics comparison between these 3 models for the $1000\text{ }\mu\text{m} \times 1000\text{ }\mu\text{m}$ domain size (Table S5) and $2000\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ domain size (Table S6) indicates that the model with best metrics, regardless the selected domain size, is the random forest with Ada boost, which is the one we use for the upcoming analysis. Validation metrics results are supported by the regression plots displayed in Fig. 7 for the $1000\text{ }\mu\text{m} \times 1000\text{ }\mu\text{m}$ domain size and in Fig. S7 for the $2000\text{ }\mu\text{m} \times 200\text{ }\mu\text{m}$ domain size. In general, considering all the 7 output properties we obtain a high R^2 score, with values over 0.9. These results indicate a high predictive capability of the deterministic learning to predict the 7 GDL properties from the 5 input parameters. Additionally, we randomly shuffled the training and testing dataset 30 different times to calculate the distribution of the R^2 score and determined the 95 % confidence interval (CI95) [71]. CI95 interval adds a statistical examination on how the deterministic learning prediction behave regarding the training/testing dataset. In that sense, at a certain threshold, CI95 indicates the metrics'

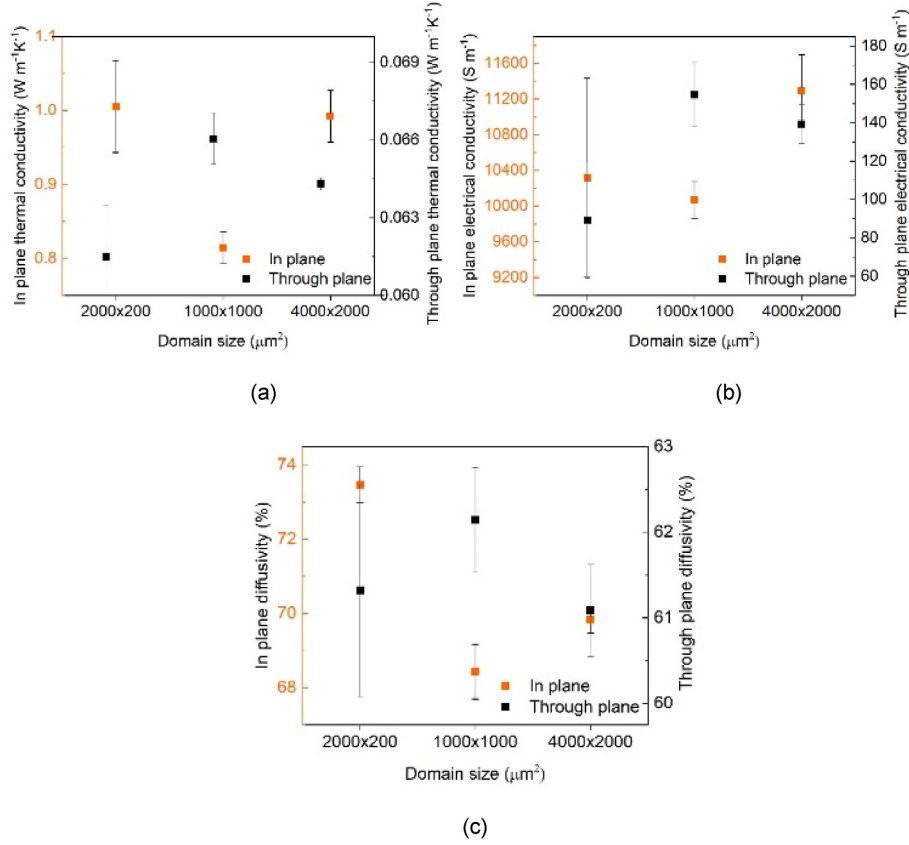


Fig. 5. Mean values of IP and TP (a) electrical conductivity, (b) thermal conductivity, and (c) diffusivity using 12 μm fiber diameter along with their corresponding error bars, which represent the standard deviation.

Table 2

RSD calculated for both domain sizes, 2000 $\mu\text{m} \times 200 \mu\text{m}$ and 1000 $\mu\text{m} \times 1000 \mu\text{m}$.

	2000 $\mu\text{m} \times 200 \mu\text{m}$ RSD (%)	1000 $\mu\text{m} \times 1000 \mu\text{m}$ RSD (%)
Thermal IP (TIP)	2.2	6.4
Thermal TP (TTP)	2.7	3.1
Electrical IP (EIP)	2.2	5.3
Electrical TP (ETP)	11.4	13.1
Diffusivity IP (DIP)	0.5	0.9
Diffusivity TP (DTP)	1.3	0.7

variability and likelihood of precision. Regarding our results, the CI95 includes values over 0.9 to 1.0, with a narrow width (< 0.03), suggesting that the deterministic learning is able to predict the different patterns of the dataset, regardless of the training/testing division, providing accurate predictions.

In order to assess the ability of the model to predict the 7 outputs, we use a cross-validation approach consisting of taking one subset as test data, and then train the model over the remaining subsets. The data set is randomly split into 5 folds. Table 4 shows the data metrics comparing training versus testing datasets. Averaged R2, MAE and MSE are calculated considering the 7 outputs of the model.

3.4. ML optimizer

The optimization results are presented in Table 5 shows the optimal 5 input manufacturing parameters for each of the 4 scenarios: C_f^1 : minimum contact resistance, maximum TP, IP diffusivity; C_f^2 : minimum contact resistance, maximum TP, IP electrical conductivity; C_f^3 : minimum contact resistance, maximum TP, IP thermal conductivity; and C_f^4 :

minimum contact resistance, maximum TP, IP electrical and thermal conductivities and diffusivity.

Building upon the established relationship between GDL manufacturing parameters and its properties, this work investigates the impact of these parameters on key GDL performance. As hypothesized, increasing fiber concentration enhances electrical conductivity but compromises porosity, thereby hindering diffusivity. Therefore, achieving optimal GDL properties necessitates a delicate balance between these parameters. Table 5 presents the optimal manufacturing parameters predicted by our developed optimizer framework for 2 domain sizes on 4 different cases. C_f^1 , C_f^2 , C_f^3 , and C_f^4 proposed a fiber diameter of 7 μm within the specified range of 6–12 μm . Notably, C_f^1 predicted a lower fiber concentration (14.22 %) compared to other cases, aligning with the expectation of maximizing diffusivity through a reduced solid concentration percentage. This finding aligns with the correlation matrices presented in Fig. 7. Additionally, it also predicted a thickness of 290 μm and a lower compression ratio of 0.1. These characteristics facilitate easier particle transport along the x and z-axes, leading to lower geometric tortuosity and ultimately, higher diffusivity. In contrast, Cases C_f^2 and C_f^3 exhibited higher fiber concentration values (17 %). This translates to increased electrical and thermal conductivities due to the presence of more conductive materials within the GDL. Furthermore, a higher compression ratio, as observed in these cases, enhances contact points between conductive materials, further enhancing conductivity. Finally, case C_f^4 displayed manufacturing parameters comparable to cases C_f^2 and C_f^3 , but with a lower fiber concentration. This approach represents a compromise, attempting to balance diffusivity and conductivity requirements. This analysis underscores the importance of carefully considering the interplay between GDL manufacturing parameters and their impact on GDL properties for

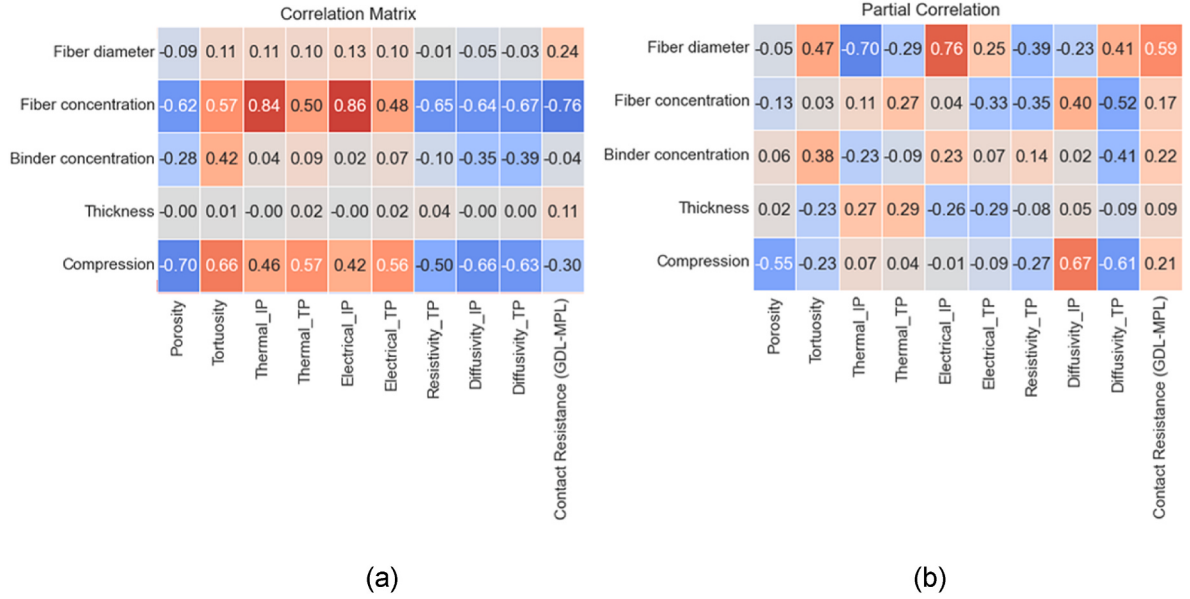


Fig. 6. Comparison between the correlation matrix (a) and the partial correlation matrix (b) of domain size $1000 \mu\text{m} \times 1000 \mu\text{m}$.

Table 3

MSE, MAE, and R^2 score calculated over the testing dataset and associated with the fitting of the GDL properties ($1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size). The 95 % confidence interval (CI95) were estimated with a total of 30 random seeds of training/testing datasets for the uncertainty of the R^2 score.

	MSE	MAE	R^2 score	CI95
Thermal IP, $\text{W m}^{-1}\text{K}^{-1}$	4.072	0.084	0.975	[0.980; 0.983]
Thermal TP, $\text{W m}^{-1}\text{K}^{-1}$	5.787	0.053	0.943	[0.968; 0.973]
Electrical IP, S m^{-1}	4.174	0.022	0.973	[0.978; 0.981]
Electrical TP, S m^{-1}	6.527	0.163	0.934	[0.962; 0.970]
Diffusivity IP, %	3.628	0.134	0.971	[0.953; 0.963]
Diffusivity TP, %	3.690	0.124	0.969	[0.947; 0.956]
Contact resistance, Ω	5.271	0.116	0.921	[0.902; 0.923]

optimal PEMFC performance. Furthermore, $2000 \mu\text{m} \times 200 \mu\text{m} \times$ thickness presented a slightly higher binder concentration for the C_f^2 and a slightly lower thickness, along with C_f^3 . This is to increase the contact points between the solid materials (carbon fiber and PTFE) in the GDL, consequently increasing electrical and thermal conductivities.

Fig. 8 presents radar charts depicting the associated GDL properties for each of the four investigated scenarios. The optimal manufacturing parameters from Table 5 were used as inputs for a stochastically generated model, and the results were compared with those obtained from a deterministic model, shown in Table 6.

Fig. 8a focuses on identifying optimal properties that maximize diffusivity while minimizing contact resistance. This balancing act presents a significant challenge. Reducing fiber concentration subsequently increasing porosity and facilitate particle transport through the GDL inherently decreases the probability of fiber in contact with MPL. This disrupts electron transmission between the two layers, consequently increasing contact resistance. As a result, C_f^1 exhibits a relatively higher contact resistance (0.14Ω) compared to the other cases (shown in Table 6). Fig. 8b and c exhibit very similar trends, which is expected given the comparable nature of the manufacturing parameters involved. A slight variation in diffusivity values between these figures is evident, with Fig. 8b showing slightly higher values than in Fig. 8c. This

difference arises because Fig. 8b prioritizes the optimization of properties that enhance electrical conductivity while minimizing contact resistance, leading to marginally higher diffusivity values. In contrast, Fig. 8c focuses on optimizing thermal conductivity while minimizing contact resistance. The numerical results corresponding to these figures are presented in Table 6. Notably, achieving high electrical and thermal conductivities necessitates a higher fiber concentration alongside a significant compression ratio. This increased compression enhances contact points between carbon fibers, ultimately facilitating mass transport. Finally, Fig. 8d showcases a slightly deviating trend compared to Fig. 8b and c. To accommodate the requirement for diffusivity, C_f^4 implements a slightly lower electrical and thermal conductivities by increasing the thickness of the layer.

It is noteworthy that the optimal GDL configuration obtained is contingent upon the prioritized GDL properties. This work explored two scenarios: (1) prioritizing specific properties (C_f^1 , C_f^2 , and C_f^3) and (2) achieving a balance between all desired properties (C_f^4). The weighting scheme within the optimization framework allows for adjustments based on the targeted GDL functionality. Consequently, this approach can generate diverse optimal solutions tailored to specific industrial applications. Indeed, real-world GDL performance depends on the strategic manipulation of properties to achieve the most desirable outcomes. In this regard, the presented optimization workflow demonstrates its adaptability to various optimization problems across different application domains.

Table 6 represents the comparison between the GDL functional properties, and the contact resistance taken from the deterministic learning and the stochastically generated model at different optimal scenarios. The distinction between the deterministic learning and the stochastically generated model lies in their methodologies. Deterministic learning is from our developed ML model, which predicts the output properties of the GDL based on input parameters used to generate its microstructure. On the other hand, the stochastically generated model is from digital characterization using physics-based methods applied to a stochastically generated GDL. Scenarios C_f^2 , C_f^3 , and C_f^4 exhibited a percent error of less than 10 %, except for the contact resistance which showed errors of less than 12 % and Electrical TP of less than 13 %. On the other hand, C_f^1 demonstrated a percent error of 37 % for the electrical TP, predicting a value of 155 S m^{-1} . As shown in Fig. 7d, which illustrates the predictive capability of deterministic learning for the

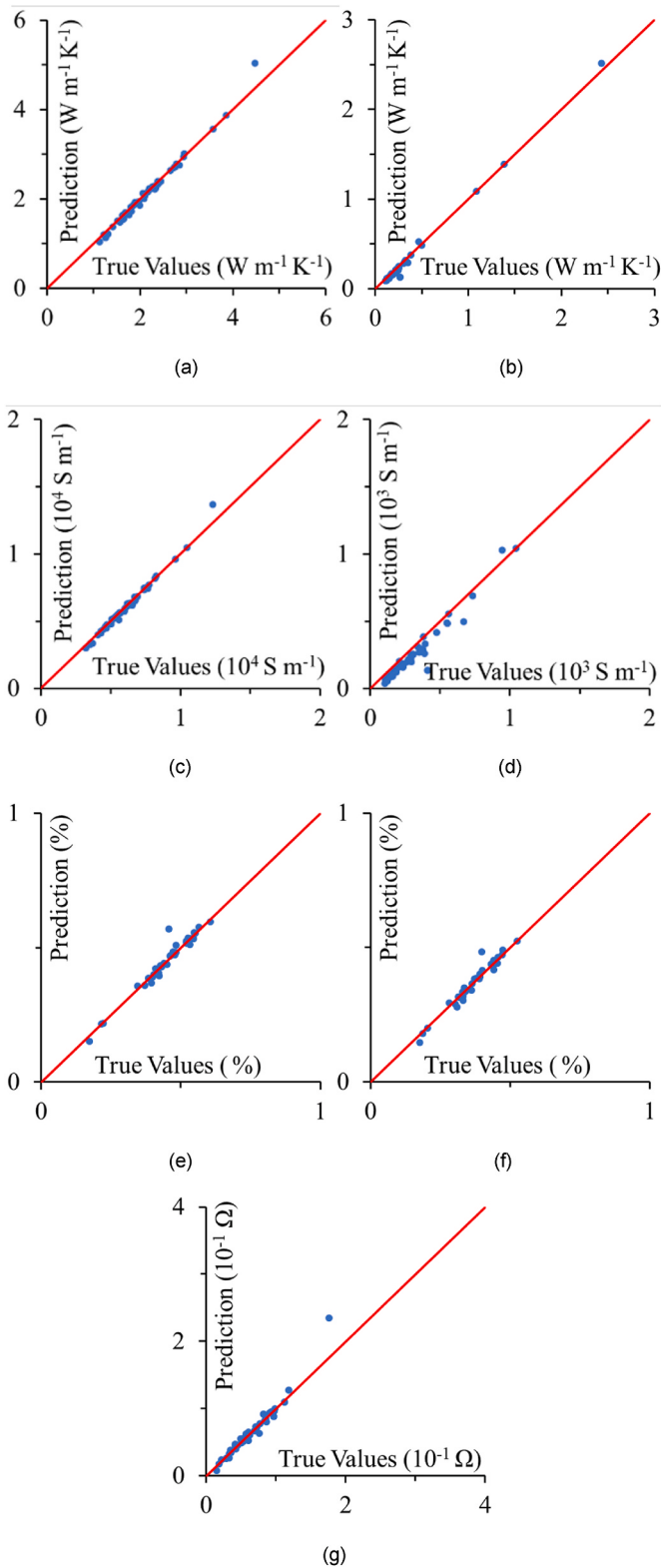


Fig. 7. True values (x-axis) vs. predictions (y-axis) values of random forest models trained for the $1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size, showing the accuracy achieved when predicting (a) Thermal IP, (b) Thermal TP, (c) Electrical IP, (d) Electrical TP, (e) Diffusivity IP, (f) Diffusivity TP, and (g) Contact Resistance.

functional property Electrical TP, there is a lack of data points in the region beyond 70 S m^{-1} . Consequently, the accuracy of predictions in this high-value range is limited. However, improvements in predictive

Table 4

Average R2, MAE and MSE scores of the cross-validation approach using 5 folds.

	Train	Test
R ²	0.9984	0.9713
MSE	0.0012	0.0199
MAE	0.0261	0.0753

accuracy are evident in other scenarios. Table 7, in general, presented a less than 10 % error in scenarios C_f^2 , C_f^3 , and C_f^4 . Furthermore, C_f^1 showed slightly higher percent error of 40 % compared to the $1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size. In general, there is a consistency between the values calculated from the physics-based model and the values predicted by the deterministic learning, showing that the deterministic learning integrated within our optimization framework is able to predict GDL functional properties, and the contact resistance between the GDL and the MPL, correctly.

4. Conclusions

This work successfully developed a multi-objective optimization approach for enhanced microstructure of the GDL. A design of virtual experiments approach was initially utilized to achieve a uniform distribution of sample points within the defined parameter space. This space encompassed key GDL manufacturing parameters such as fiber diameter, fiber concentration, binder concentration, GDL thickness, and compression ratio. The maximum and minimum values of the input range properties were defined to be in good agreement with previous literature reports. Subsequently, these sample points were used as input parameters to stochastically generate GDL microstructures. From these generated microstructures, essential output properties were digitally characterized through physics-based calculations, such as thermal and electrical conductivities, diffusivity, and contact resistance between the GDL and the MPL.

Then, a random forest + Adaboost regression model was implemented to establish a ML regression model between the output/input parameters. The main advantage of the ML model is the tremendous decrease on the computational cost but keeping a good accuracy in comparison with the stochastically generated model. Regarding the computational cost, the ML model requires $\sim 3 \text{ s}$ wall-time to predict the 7 output properties, when the physics-based calculations require $\sim 3\text{--}4 \text{ h}$ wall-time. In terms of accuracy, the ML model is able to predict the 7 output properties with a R² score $\sim 95 \%$ in 6 of the 7 properties, and a R² score $\sim 90 \%$ for the contact resistance. Given the low computational cost of the ML model ($\sim 3 \text{ s}$ wall-time per iteration), it was then integrated into an optimization framework to calculate the optimal manufacturing parameters to achieve the desired GDL properties, given certain constraints applied on the framework. The selected optimization framework consists of a Nelder-Mead (NM) method together with a grid search approach to identify the optimal set of input parameters that yielded the desired GDL functional properties. A key aspect in the optimization framework is the low computational cost of each iteration coming from the ML model, since the grid search will perform 1024 iterations plus the calculation of the NM method on each iteration. The optimization framework takes ~ 10 days to get the optimal values. The dual-method, NM method plus the grid search, is performed to overcome the risk of getting a local minimum.

Furthermore, we showed how to navigate the critical trade-off between GDL properties. Prioritizing maximum GDL diffusivity and minimum contact resistance, which exhibit an inverse correlation, necessitates the utilization of low fiber concentration and low compression ratio. On the other hand, achieving maximum electrical and/or thermal conductivities while minimizing contact resistance require high fiber concentration accompanied by high compression ratio.

Table 5Optimal manufacturing parameters predicted by the optimizer framework depending on the 4 optimal cases: C_f^1 , C_f^2 , C_f^3 , and C_f^4 .

Domain size, μm^3		Fiber Diameter, (μm)	Fiber Concentration, (%)	Binder Concentration, (%)	Thickness, (μm)	Compression ratio
$1000 \mu\text{m} \times 1000 \mu\text{m} \times \text{thickness}$	C_f^1	7.00	14.22	6.00	290	0.10
	C_f^2	7.00	17.00	6.12	290	0.35
	C_f^3	7.09	17.00	6.64	290	0.35
	C_f^4	7.00	16.70	6.00	290	0.35
$2000 \mu\text{m} \times 200 \mu\text{m} \times \text{thickness}$	C_f^1	7.00	12.37	6.00	297	0.10
	C_f^2	7.00	17.00	8.88	293	0.35
	C_f^3	7.64	17.00	6.62	290	0.35
	C_f^4	7.00	16.99	6.00	297	0.33

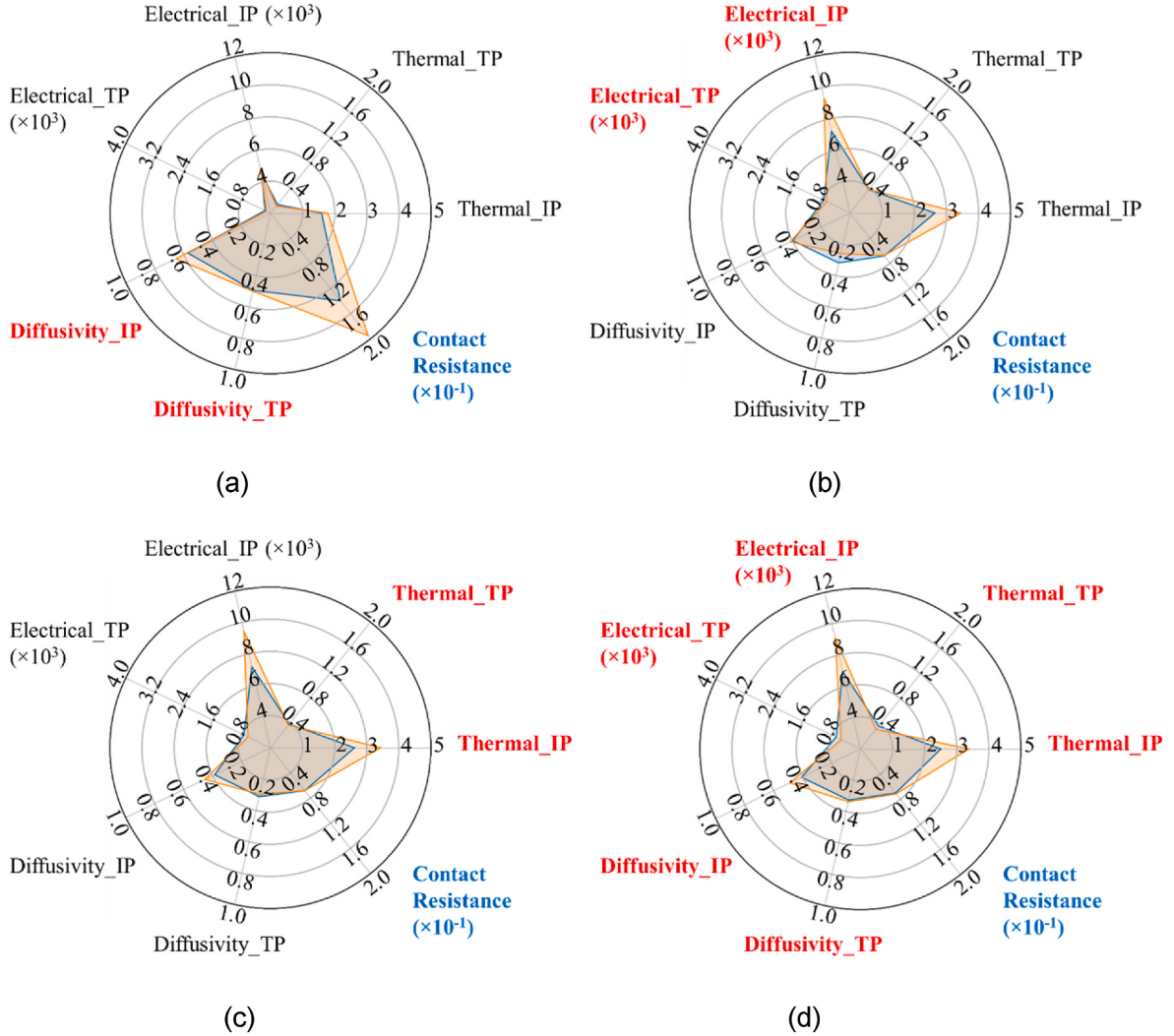


Fig. 8. Graphical representation of the 4 scenarios: (a) C_f^1 , (b) C_f^2 , (c) C_f^3 and (d) C_f^4 of the optimized GDL properties obtained through the deterministic-assisted optimization loop. A radar chart displays the optimized values. Label highlighted in blue is the minimized property, and labels highlighted in red are the maximized properties. Orange plots represent the results obtained from the $2000 \mu\text{m} \times 200 \mu\text{m}$ domain size. Blue plots represent the results obtained from the $1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

The developed optimization workflow can be used by the industries and experimentalists as it presented adaptability to various optimization constraints across different application domains to meet real-world demands. An illustrative example involves an experimentalist utilizing the developed ML model to initially validate the optimal combination of manufacturing parameters needed to achieve the desired properties of a GDL. This validation occurs prior to the actual fabrication of the GDL.

Key limitations of this study lie in the purely geometrical nature of the compression algorithm employed. It effectively reduces the pore spaces between the solid materials (fiber and binder) along the z-axis until a defined compression is achieved. Consequently, when further compression is attempted and no spaces left between the solid materials, the algorithm overlaps the solid materials, leading to a violation of mass conservation. In addition, this study does not cover the convection heat

Table 6

Comparison of GDL optimal scenarios for the $1000 \mu\text{m} \times 1000 \mu\text{m}$ domain size using the deterministic learning (A) and the stochastically generated model (B). Both used the optimal manufacturing conditions as input parameters.

		Thermal IP, W $\text{m}^{-1}\text{K}^{-1}$	Thermal TP, W $\text{m}^{-1}\text{K}^{-1}$	Electrical IP, S m^{-1}	Electrical TP, S m^{-1}	Diffusivity IP, 10^2 %	Diffusivity TP, 10^2 %	Contact Resistance, Ω
C_f^1	A	1.59	0.14	4534	155	0.57	0.49	0.14
	B	1.54	0.11	4383	97	0.59	0.50	0.12
C_f^2	A	2.64	0.37	7227	730	0.40	0.32	0.07
	B	2.51	0.35	6857	637	0.40	0.32	0.06
C_f^3	A	2.62	0.38	7171	733	0.39	0.31	0.07
	B	2.51	0.35	6852	632	0.39	0.32	0.06
C_f^4	A	2.51	0.36	6956	674	0.41	0.33	0.07
	B	2.46	0.33	6730	593	0.41	0.33	0.06

Table 7

Comparison of GDL optimal scenarios for the $2000 \mu\text{m} \times 200 \mu\text{m}$ domain size using the deterministic learning (A) and the stochastically generated model (B). Both used the optimal manufacturing conditions as input parameters.

		Thermal IP, W $\text{m}^{-1}\text{K}^{-1}$	Thermal TP, W $\text{m}^{-1}\text{K}^{-1}$	Electrical IP, S m^{-1}	Electrical TP, S m^{-1}	Diffusivity IP, 10^2 %	Diffusivity TP, 10^2 %	Contact Resistance, Ω
C_f^1	A	1.79	0.11	4857	113	0.65	0.50	0.20
	B	1.71	0.09	4627	68	0.67	0.53	0.21
C_f^2	A	3.45	0.38	9315	649	0.41	0.26	0.07
	B	3.26	0.38	8760	665	0.40	0.25	0.07
C_f^3	A	3.44	0.36	9442	617	0.45	0.30	0.07
	B	3.36	0.32	9154	551	0.47	0.31	0.06
C_f^4	A	3.38	0.31	9107	531	0.49	0.34	0.07
	B	3.26	0.29	8896	489	0.49	0.34	0.07

transfer through the domain. Future research in this area could incorporate the mechanical properties as well as the fracture of the materials within the domain under compression. This can be done using another methodology, such as DEM. For that, we can apply the approach previously developed by us [72] to predict particle fracture during electrode calendaring. Moreover, this study did not consider the anisotropy parameter which controls the carbon fiber orientation in GDL. The reasons for not including it as a manufacturing parameter for the purpose of this study are twofold. First, the computational cost is prohibitive, as we are constrained by available computational time to generate the database. Second, our focus is on manufacturing parameters commonly utilized by experimentalists and industry professionals, such as fiber diameter, fiber and binder concentrations, GDL thickness, and compression ratio. Additionally, employing alternative optimization algorithms could enable a more extensive exploration of manufacturing parameters to identify the optimal properties for generating GDL microstructures. Despite these limitations, we believe that our approach paves the way towards efficient digital tools for autonomous optimization of GDL manufacturing processes.

Data availability

The data supporting this article have been included as part of the Supplementary Information.

CRediT authorship contribution statement

Rashen Lou Omongos: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Diego E. Galvez-Aranda:** Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Franco M. Zanotto:** Writing – review & editing, Methodology, Conceptualization. **Andr s Vernes:** Writing – review & editing, Supervision. **Alejandro A. Franco:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

There are no conflicts to declare.

Acknowledgements

R.L.O, A.V. and A.A.F. acknowledge the support of the BLESSED project funded by the European Union under Marie Sklodowska-Curie Actions (Grant Agreement No. 101072578). A.V. acknowledges the InTribology project (FFG no. 906860) funded in the COMET programme by BMK and BMAW as well as Nieder sterreich and Vorarlberg, and managed by FFG. A.A.F. acknowledges the European Union's Horizon 2020 research and innovation program for the funding support through the European Research Council (grant agreement 772873, "ARTISTIC" project). A.A.F. and F.M.Z. acknowledge the European Union's Horizon Europe research and innovation program under grant agreement No. 101069686 (PULSELiON). A.A.F. acknowledges the European Research Council for the funding support through the ERC Proof-of-Concept grant No. 101069244 (SMARTISTIC project). A.A.F. and D. E. G. acknowledge the funding support of the French National Research Agency under the France 2030 program (Grant ANR-22-PEBA-0002, PEPR project "BATMAN") A.A.F. acknowledges Institut Universitaire de France for the support. R.L.O would like to thank Imelda Cardenas Sierra, Xingyu Zhang, and Sourab Barath Vijayaraghavan, from Prof. A.A. Franco's group, for their unwavering support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpowsour.2024.235583>.

References

- [1] Effects - NASA Science. <https://science.nasa.gov/climate-change/effects/> (accessed 2024-June-5).
- [2] Climate targets - European Commission, 2030, https://climate.ec.europa.eu/eu-action/climate-strategies-targets/2030-climate-targets_en. (Accessed 12 April 2024).
- [3] Nations, U. Net Zero Coalition. United Nations. <https://www.un.org/en/climatechange/net-zero-coalition> (accessed 2024-April-12).

- [4] M.M. Mench, Fuel Cell Engines, first ed., Wiley, 2008 <https://doi.org/10.1002/9780470209769>.
- [5] C. Cunan, M.-K. Tran, Y. Lee, S. Kwok, V. Leung, M. Fowler, A review of heavy-duty vehicle powertrain technologies: diesel engine vehicles, battery electric vehicles, and hydrogen fuel cell electric vehicles, *Cleanroom Technol.* 3 (2) (2021) 474–489, <https://doi.org/10.3390/cleantechnol3020028>.
- [6] V.K. Mathur, J. Crawford, Fundamentals of gas diffusion layers in PEM fuel cells, in: S. Basu (Ed.), *Recent Trends in Fuel Cell Science and Technology*, Springer New York, New York, NY, 2007, pp. 116–128, https://doi.org/10.1007/978-0-387-68815-2_4.
- [7] Z. Tavarani-Yoosefabad, J. Bellerive, E. Kjeang, Multiscale stochastic modeling of microporous layers and Bi-layer gas diffusion media for polymer electrolyte fuel cells, *J. Power Sources* 581 (2023) 233476, <https://doi.org/10.1016/j.jpowsour.2023.233476>.
- [8] S. Basu (Ed.), *Recent Trends in Fuel Cell Science and Technology*, Springer, New York, NY, 2007 [u.a.].
- [9] J.T. Gostick, M.A. Ioannidis, M.W. Fowler, M.D. Pritzker, On the role of the microporous layer in PEMFC operation, *Electrochim. Commun.* 11 (3) (2009) 576–579, <https://doi.org/10.1016/j.elecom.2008.12.053>.
- [10] D. Malevich, E. Halliopi, B.A. Peppley, J.G. Pharoah, K. Karan, Investigation of charge-transfer and mass-transport resistances in PEMFCs with microporous layer using electrochemical impedance spectroscopy, *J. Electrochem. Soc.* 156 (2) (2009) B216, <https://doi.org/10.1149/1.3033408>.
- [11] A.Z. Weber, J. Newman, Effects of microporous layers in polymer electrolyte fuel cells, *J. Electrochem. Soc.* 152 (4) (2005) A677, <https://doi.org/10.1149/1.1861194>.
- [12] J.P. Owejan, J.E. Owejan, W. Gu, T.A. Trabold, T.W. Tighe, M.F. Mathias, Water transport mechanisms in PEMFC gas diffusion layers, *J. Electrochem. Soc.* 157 (10) (2010) B1456, <https://doi.org/10.1149/1.3468615>.
- [13] A. Thomas, G. Maranzana, S. Didierjean, J. Dillet, O. Lottin, Thermal and water transfer in PEMFCs: investigating the role of the microporous layer, *Int. J. Hydrogen Energy* 39 (6) (2014) 2649–2658, <https://doi.org/10.1016/j.ijhydene.2013.11.105>.
- [14] X. Han, P. Liu, S. Fan, Y. Liu, Z. Jin, Numerical study of heterogeneous porosity in gas diffusion layers of high-temperature proton-exchange membrane fuel cells, *J. Appl. Electrochem.* 52 (12) (2022) 1733–1746, <https://doi.org/10.1007/s10800-022-01746-2>.
- [15] J. Soler, E. Hontañón, L. Daza, Electrode permeability and flow-field configuration: influence on the performance of a PEMFC, *J. Power Sources* 118 (1–2) (2003) 172–178, [https://doi.org/10.1016/S0378-7753\(03\)00081-8](https://doi.org/10.1016/S0378-7753(03)00081-8).
- [16] K. Jiao, X. Li, Water transport in polymer electrolyte membrane fuel cells, *Prog. Energy Combust. Sci.* 37 (3) (2011) 221–291, <https://doi.org/10.1016/j.peecs.2010.06.002>.
- [17] T. Tsukamoto, T. Aoki, H. Kanesaka, T. Taniguchi, T. Takayama, H. Motegi, R. Takayama, S. Tanaka, K. Komiyama, M. Yoneda, Three-dimensional numerical simulation of full-scale proton exchange membrane fuel cells at high current densities, *J. Power Sources* 488 (2021) 229412, <https://doi.org/10.1016/j.jpowsour.2020.229412>.
- [18] C.J. Bapat, S.T. Thynell, Effect of anisotropic electrical resistivity of gas diffusion layers (GDLs) on current density and temperature distribution in a polymer electrolyte membrane (PEM) fuel cell, *J. Power Sources* 185 (1) (2008) 428–432, <https://doi.org/10.1016/j.jpowsour.2008.06.016>.
- [19] H. Guo, L. Chen, S.A. Ismail, L. Jiang, S. Guo, J. Gu, X. Zhang, Y. Li, Y. Zhu, Z. Zhang, D. Han, Gas diffusion layer for proton exchange membrane fuel cells: a review, *Materials* 15 (24) (2022) 8800, <https://doi.org/10.3390/ma15248800>.
- [20] J. Yablecki, A. Nabovati, A. Bazylak, Modeling the effective thermal conductivity of an anisotropic gas diffusion layer in a polymer electrolyte membrane fuel cell, *J. Electrochem. Soc.* 159 (6) (2012) B647–B653, <https://doi.org/10.1149/2.013206jes>.
- [21] A. Hakenjos, H. Muentert, U. Wittstadt, C. Hebling, A PEM fuel cell for combined measurement of current and temperature distribution, and flow field flooding, *J. Power Sources* 131 (1–2) (2004) 213–216, <https://doi.org/10.1016/j.jpowsour.2003.11.081>.
- [22] S. He, M.M. Mench, S. Tadigadapa, Thin film temperature sensor for real-time measurement of electrolyte temperature in a polymer electrolyte fuel cell, *Sensor Actuator Phys.* 125 (2) (2006) 170–177, <https://doi.org/10.1016/j.sna.2005.05.021>.
- [23] P.J.S. Vie, S. Kjelstrup, Thermal conductivities from temperature profiles in the polymer electrolyte fuel cell, *Electrochim. Acta* 49 (7) (2004) 1069–1077, <https://doi.org/10.1016/j.electacta.2003.10.018>.
- [24] M. Wang, H. Guo, C. Ma, Temperature distribution on the MEA surface of a PEMFC with serpentine channel flow bed, *J. Power Sources* 157 (1) (2006) 181–187, <https://doi.org/10.1016/j.jpowsour.2005.08.012>.
- [25] S. Park, J.-W. Lee, B.N. Popov, Effect of carbon loading in microporous layer on PEM fuel cell performance, *J. Power Sources* 163 (1) (2006) 357–363, <https://doi.org/10.1016/j.jpowsour.2006.09.020>.
- [26] Z. Qi, A. Kaufman, Improvement of water management by a microporous sublayer for PEM fuel cells, *J. Power Sources* 109 (1) (2002) 38–46, [https://doi.org/10.1016/S0378-7753\(02\)00058-7](https://doi.org/10.1016/S0378-7753(02)00058-7).
- [27] W. Yoshimune, S. Kato, M. Inagaki, S.A. Yamaguchi, Simple method to measure through-plane effective gas diffusivity of a gas diffusion layer for polymer electrolyte fuel cells, *Int. J. Heat Mass Tran.* 191 (2022) 122887, <https://doi.org/10.1016/j.ijheatmasstransfer.2022.122887>.
- [28] W. Yoshimune, S. Kato, S. Yamaguchi, Multi-scale pore morphologies of a compressed gas diffusion layer for polymer electrolyte fuel cells, *Int. J. Heat Mass Tran.* 152 (2020) 119537, <https://doi.org/10.1016/j.ijheatmasstransfer.2020.119537>.
- [29] G.S. Hwang, A.Z. Weber, Effective-diffusivity measurement of partially-saturated fuel-cell gas-diffusion layers, *J. Electrochem. Soc.* 159 (11) (2012) F683, <https://doi.org/10.1149/2.024211jes>.
- [30] J.M. LaManna, S.G. Kandlikar, Determination of effective water vapor diffusion coefficient in pemfc gas diffusion layers, *Int. J. Hydrogen Energy* 36 (8) (2011) 5021–5029, <https://doi.org/10.1016/j.ijhydene.2011.01.036>.
- [31] P. Mangal, L.M. Pant, N. Carrigy, M. Dumontier, V. Zingan, S. Mitra, M. Secanell, Experimental study of mass transport in PEMFCs: through plane permeability and molecular diffusivity in GDLs, *Electrochim. Acta* 167 (2015) 160–171, <https://doi.org/10.1016/j.electacta.2015.03.100>.
- [32] N. Zamel, N.G.C. Astrath, X. Li, J. Shen, J. Zhou, F.B.G. Astrath, H. Wang, Z.-S. Liu, Experimental measurements of effective diffusion coefficient of oxygen–nitrogen mixture in PEM fuel cell diffusion media, *Chem. Eng. Sci.* 65 (2) (2010) 931–937, <https://doi.org/10.1016/j.ces.2009.09.044>.
- [33] C. Chan, N. Zamel, X. Li, J. Shen, Experimental measurement of effective diffusion coefficient of gas diffusion layer/microporous layer in PEM fuel cells, *Electrochim. Acta* 65 (2012) 13–21, <https://doi.org/10.1016/j.electacta.2011.12.110>.
- [34] R. Flückiger, S.A. Freunberger, D. Kramer, A. Wokaun, G.G. Scherer, F.N. Büchi, Anisotropic, effective diffusivity of porous gas diffusion layer materials for PEFC, *Electrochim. Acta* 54 (2) (2008) 551–559, <https://doi.org/10.1016/j.electacta.2008.07.034>.
- [35] R. Koresawa, Y. Utaka, Precise measurement of effective oxygen diffusivity for microporous media containing moisture by review of galvanic cell oxygen absorber configuration, *Int. J. Heat Mass Tran.* 76 (2014) 549–558, <https://doi.org/10.1016/j.ijheatmasstransfer.2014.05.005>.
- [36] R. Rashapov, F. Imami, J.T. Gostick, A method for measuring in-plane effective diffusivity in thin porous media, *Int. J. Heat Mass Tran.* 85 (2015) 367–374, <https://doi.org/10.1016/j.ijheatmasstransfer.2015.01.101>.
- [37] J.P. James, H.-W. Choi, J.G. Pharoah, X-ray computed tomography reconstruction and analysis of polymer electrolyte membrane fuel cell porous transport layers, *Int. J. Hydrogen Energy* 37 (23) (2012) 18216–18230, <https://doi.org/10.1016/j.ijhydene.2012.08.077>.
- [38] L. Xia, M. Ni, Q. He, Q. Xu, C. Cheng, Optimization of gas diffusion layer in high temperature PEMFC with the focuses on thickness and porosity, *Appl. Energy* 300 (2021) 117357, <https://doi.org/10.1016/j.apenergy.2021.117357>.
- [39] A. Pfingst, D. Veyret, G. Tsotridis, G. Janssen, X-Ray Computed Tomography of PEM Fuel Cells, 2010.
- [40] Y. Chen, C. Jiang, C. Cho, Characterization of effective in-plane electrical resistivity of a gas diffusion layer in polymer electrolyte membrane fuel cells through freeze–thaw thermal cycles, *Energies* 13 (1) (2019) 145, <https://doi.org/10.3390/en13010145>.
- [41] I. Nitta, O. Himanen, M. Mikkola, Thermal conductivity and contact resistance of compressed gas diffusion layer of PEM fuel cell, *Fuel Cell.* 8 (2) (2008) 111–119, <https://doi.org/10.1002/fuce.200700054>.
- [42] M. Khandelwal, M.M. Mench, Direct measurement of through-plane thermal conductivity and contact resistance in fuel cell materials, *J. Power Sources* 161 (2) (2006) 1106–1115, <https://doi.org/10.1016/j.jpowsour.2006.06.092>.
- [43] J.L. Doob, *Stochastic Processes*, Wiley, 1962.
- [44] P. Sarkezi-Selsky, M. Rastedt, A. Latz, T. Jahnke, *Lattice-Boltzmann Simulation of Two-phase Flows in the GDL and MPL of Polymer Electrolyte Fuel Cells*, 2019.
- [45] W. Wu, F. Jiang, Microstructure reconstruction and characterization of PEMFC electrodes, *Int. J. Hydrogen Energy* 39 (28) (2014) 15894–15906, <https://doi.org/10.1016/j.ijhydene.2014.03.074>.
- [46] M. El Hannach, R. Singh, N. Djilali, E. Kjeang, Micro-porous layer stochastic reconstruction and transport parameter determination, *J. Power Sources* 282 (2015) 58–64, <https://doi.org/10.1016/j.jpowsour.2015.02.034>.
- [47] J. Hao, T. Ma, J. Zhou, H. Wei, Y. Kong, X. Du, Structural gradient optimization of diffusion layer based on finite data mapping method for PEMFC performance improvement, *Int. J. Heat Mass Tran.* 220 (2024) 124948, <https://doi.org/10.1016/j.ijheatmasstransfer.2023.124948>.
- [48] D. Froning, J. Wirtz, E. Hoppe, W. Lehnert, Flow characteristics of fibrous gas diffusion layers using machine learning methods, *Appl. Sci.* 12 (23) (2022) 12193, <https://doi.org/10.3390/app122312193>.
- [49] L.F. Hou, H. Chen, H. Guo, F. Ye, Multi-objective optimization of gas diffusion layer structure parameters for proton exchange membrane fuel cell, *Int. J. Green Energy* 21 (4) (2024) 787–797, <https://doi.org/10.1080/15435075.2023.2218496>.
- [50] H.-W. Li, B.-X. Qiao, J.-N. Liu, Y. Yang, W. Fan, G.-L. Lu, A data-driven framework for performance prediction and parameter optimization of a proton exchange membrane fuel cell, *Energy Convers. Manag.* 271 (2022) 116338, <https://doi.org/10.1016/j.enconman.2022.116338>.
- [51] J. Wang, H. Jiang, G. Chen, H. Wang, L. Lu, J. Liu, L. Xing, Integration of multi-physics and machine learning-based surrogate modelling approaches for multi-objective optimization of deformed GDL of PEM fuel cells, *Energy and AI* 14 (2023) 100261, <https://doi.org/10.1016/j.egyai.2023.100261>.
- [52] M. Duquesnoy, C. Liu, D.Z. Dominguez, V. Kumar, E. Ayerbe, A.A. Franco, Machine learning-assisted multi-objective optimization of battery manufacturing from synthetic data generated by physics-based simulations, *Energy Storage Mater.* 56 (2023) 50–61, <https://doi.org/10.1016/j.ensm.2022.12.040>.
- [53] M. Duquesnoy, C. Liu, V. Kumar, E. Ayerbe, A.A. Franco, Toward high-performance energy and power battery cells with machine learning-based optimization of electrode manufacturing, *J. Power Sources* 590 (2024) 233674, <https://doi.org/10.1016/j.jpowsour.2023.233674>.
- [54] J. Yu, M. Duquesnoy, C. Liu, A.A. Franco, Optimization of the microstructure of carbon felt electrodes by applying the Lattice Boltzmann method and bayesian

- optimizer, *J. Power Sources* 575 (2023) 233182, <https://doi.org/10.1016/j.jpowsour.2023.233182>.
- [55] M.D. McKay, R.J. Beckman, W.J. Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics* 21 (2) (1979) 239–245, <https://doi.org/10.2307/1268522>.
- [56] L. Zhu, S. Wang, P.-C. Sui, X. Gao, Multiscale modeling of an angled gas diffusion layer for polymer electrolyte membrane fuel cells: performance enhancing for aviation applications, *Int. J. Hydrogen Energy* 46 (39) (2021) 20702–20714, <https://doi.org/10.1016/j.ijhydene.2021.03.166>.
- [57] M. Secanell, A. Jarauta, A. Kosakian, M. Sabharwal, J. Zhou, PEM fuel cells, modeling, in: R.A. Meyers (Ed.), *Encyclopedia of Sustainability Science and Technology*, Springer New York, New York, NY, 2017, pp. 1–61, https://doi.org/10.1007/978-1-4939-2493-6_1019-1.
- [58] Math2Market GmbH, J. Becker, F. Biebl, L. Cheng, E. Glatt, A. Grieser, M. Groß, S. Linden, D. Mosbach, C. Wagner, A. Weber, R. Westerteiger, A. Wiegmann, A. Neundorff, *GeoDict Releases 2024*, Math2Market GmbH: DE, 2024.
- [59] Y. Liu, Y. Wang, J. Zhang, New machine learning algorithm: random forest, in: B. Liu, M. Ma, J. Chang (Eds.), *Information Computing and Applications*, Springer, Berlin, Heidelberg, 2012, pp. 246–252, https://doi.org/10.1007/978-3-642-34062-8_32.
- [60] D.P. Mohandoss, Y. Shi, K. Suo, Outlier prediction using random forest classifier, in: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, NV, USA, 2021, pp. 27–33, <https://doi.org/10.1109/CCWC51732.2021.9376077>.
- [61] P.-B. Zhang, Z.-X. Yang, A novel AdaBoost framework with robust threshold and structural optimization, *IEEE Trans. Cybern.* 48 (1) (2018) 64–76, <https://doi.org/10.1109/TCYB.2016.2623900>.
- [62] G. Shanmugasundar, M. Vanitha, R. Çep, V. Kumar, K. Kalita, M. Ramachandran, A comparative study of linear, random forest and AdaBoost regressions for modeling non-traditional machining, *Processes* 9 (11) (2021) 2015, <https://doi.org/10.3390/pr9112015>.
- [63] Leshem, G. Traffic Flow Prediction Using Adaboost Algorithm with Random Forests as a Weak Learner.
- [64] J. Thongkam, G. Xu, Y. Zhang, AdaBoost algorithm with random forests for predicting breast cancer survivability, in: 2008 IEEE International Joint Conference on Neural Networks, IEEE World Congress on Computational Intelligence), 2008, pp. 3062–3069, <https://doi.org/10.1109/IJCNN.2008.4634231>.
- [65] H.E. Hamdaoui, S. Boujraf, N.E.H. Chaoui, B. Alami, M. Maaroufi, Improving heart disease prediction using random forest and AdaBoost algorithms, *International Journal of Online and Biomedical Engineering (IJOE)* 17 (11) (2021) 60–75, <https://doi.org/10.3991/ijoe.v17i11.24781>.
- [66] J.A. Nelder, R. Mead, A simplex method for function minimization, *Comput. J.* 7 (4) (1965) 308–313, <https://doi.org/10.1093/comjnl/7.4.308>.
- [67] Y. Ozaki, M. Yano, M. Onishi, Effective hyperparameter optimization using Nelder-Mead method in deep learning, *IPSI T Comput Vis Appl* 9 (1) (2017) 20, <https://doi.org/10.1186/s41074-017-0030-7>.
- [68] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, *Scikit-learn: machine learning in Python*, *J. Mach. Learn. Res.* 12 (85) (2011) 2825–2830.
- [69] J.P. Verma, Correlation matrix and partial correlation: explaining relationships, in: J.P. Verma (Ed.), *Data Analysis in Management with SPSS Software*, Springer, India, 2013, pp. 103–132, https://doi.org/10.1007/978-81-322-0786-3_4.
- [70] 2.6 - (Pearson) Correlation Coefficient r | STAT 462. <https://online.stat.psu.edu/stat462/node/96/> (accessed 2024-September-2).
- [71] Jiang, B.; Zhang, X.; Cai, T. Estimating the Confidence Interval for Prediction Errors of Support Vector Machine Classifiers.
- [72] J. Xu, B. Paredes-Goyes, Z. Su, M. Scheel, T. Weitkamp, A. Demortière, A.A. Franco, Computational model for predicting particle fracture during electrode calendaring, *Batteries & Supercaps* 6 (12) (2023) e202300371, <https://doi.org/10.1002/batt.202300371>.