Journal of Power Sources Advances 20 (2023) 100114

Contents lists available at ScienceDirect

Journal of Power Sources Advances

journal homepage: www.sciencedirect.com/journal/journal-of-power-sources-advances

Machine learning for optimal electrode wettability in lithium ion batteries

Amina El Malki [a,b], Mark Asch [c,**], Oier Arcelus [a,b], Abbos Shodiev [a,b], Jia Yu [a,b], Alejandro A. Franco [a,b,d,e,*]

[a] Laboratoire de Réactivité et Chimie des Solides (LRCS), UMR CNRS 7314, Université de Picardie Jules Verne, Hub de l'Energie, 15 rue Baudelocque, Amiens Cedex, 80039, 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] LAMFA, CNRS UMR 7352, Université de Picardie Jules Verne, 33 rue Saint Leu, Amiens, 80039, France
[d] ALISTORE-European Research Institute, FR CNRS 3104, Hub de l'Energie, 15 rue Baudelocque, Amiens Cedex, 80039, France
[e] Institut Universitaire de France, 103 Boulevard Saint Michel, Paris, 75005, France

ARTICLE INFO

Keywords:
Lithium ion battery
Electrolyte wettability
Machine learning
Lattice Boltzmann method

ABSTRACT

Electrode wetting is a critical step in the Lithium-Ion Battery manufacturing process. The injection of electrolyte in the electrodes' porosity requires the application of pressure-vacuum pumping strategies without warranty that the full porosity will be fully occupied with electrolyte at the end of this process step. The electrode wettability strongly depends on the contact angle between the electrolyte and the electrode, the electrode microstructure characterized by its porosity, pore network and tortuosity factor, the electrolyte viscosity and density. Computational fluid dynamics approaches such as the Lattice Boltzmann Method can provide relevant information of the filling process, yet these approaches come with significant computational cost. The use of machine learning techniques can provide surrogate models for the optimization of this multi-parameter process that depends on both chemical and physical properties. Within this context, we propose a general workflow for realizing this objective and provide detailed simulation-based experiments. These physics-informed surrogate models open the path to tractable, rapid solutions of parameter identification and design optimization problems. They also provide a general workflow for applications on other optimal battery material design problems.

1. Introduction

Energy storage is one of the most prominent problems that humanity needs to solve in the 21st century. Climate change, the availability and the price instability of fossil fuels are some of the key reasons for this. Electrochemical energy storage devices and in particular lithium-ion batteries (LIBs) are among the most suited technologies to overcome this challenge, due to their capability to be discharged and recharged multiple times, their compactness, their high energy densities and their (relative) operational simplicity [1]. Indeed, LIBs are enablers of the undergoing energy transition, particularly manifested through the electrification of the automotive sector. This triggers in turn a race towards LIB cell optimization, to increase even more their energy densities and lifetime, to decrease as much as possible their cost and to improve their recyclability [2,3]. These characteristics are to a significant extent impacted by the cell manufacturing process. Such a process involves numerous steps and parameters that need to be fine-tuned in order to achieve desired performances. The starting point being the positive and negative electrodes fabrication, through the mixing of the active material, carbon additive and binder powders in a solvent, the casting of the resulting slurry on a current collector and its drying, followed by the calendering aiming to improve the electronic percolation between the particles forming the electrode. Still, the latter should be carried out carefully to avoid the collapse of the pores that will be detrimental to the electrode wettability by the liquid electrolyte [4]. Once the cell is mounted with the two electrodes and the polymer separator between them, the electrolyte infiltration is a crucial step in the manufacturing process as it needs to ensure that most of the active material exposed to the pores is wetted with the liquid electrolyte. This is important to ensure proper lithium-ion exchange between the positive and negative electrodes in the cell. The performance of this electrolyte infiltration process strongly depends on the electrolyte properties including

* Corresponding author. Laboratoire de Réactivité et Chimie des Solides (LRCS), UMR CNRS 7314, Université de Picardie Jules Verne, Hub de l'Energie, 15 rue Baudelocque, Amiens Cedex, 80039, France.
** Corresponding author.
E-mail addresses: mark.asch@u-picardie.fr (M. Asch), alejandro.franco@u-picardie.fr (A.A. Franco).

https://doi.org/10.1016/j.powera.2023.100114
Received 26 October 2022; Received in revised form 21 February 2023; Accepted 4 March 2023
Available online 24 March 2023

viscosity, density and contact angle with the electrodes and the separator [5,6]. It also depends on the electrode microstructural properties, characterized by the overall porosity, pore size distribution and pore network characteristics [7]. This makes the optimization of this process step very difficult and therefore time consuming and potentially inefficient. Few experimental studies have been focused on understanding this process, which can be understood in view of the difficulty to characterize the infiltration in an *in operando* way [8–11]. For this reason, computational modeling has emerged as a powerful tool to understand this process with some degree of detail. The Lattice Boltzmann Method (LBM) has proven itself to be a suitable tool to simulate the infiltration process in view of its mesoscopic character, where fluids are described in a discrete manner (through particles) [12,13]. In 2021 we reported for the first time a LBM model able to simulate in three dimensions how electrolyte infiltrates in electrode microstructures arising from manufacturing mechanistic simulations and from tomography characterizations [14]. In this work we investigated in particular the influence of the electrode microstructure on the wettability performance (rate of pores saturation by the electrolyte and spatial location of unwetted volumes) and the resulting electrochemical performance. Later works [15] have extended our work by attempting to consider electrolyte flow through the binder, but with contact angle parameters for which it is difficult to determine their experimental validity. Recently, we have used our LBM model to explore the influence of different electrode architectures on the facility to wet entire cell sandwiches, showing that electrode heterogeneities along the electrode thickness (*e.g.* porosity gradients) can be beneficial for better electrolyte infiltration [16]. Furthermore, we have reported a surrogate model derived from machine learning (ML) applied to the data generated by our LBM model, able to predict the dynamic path of the electrolyte and the saturation curves in electrode microstructures in a question of seconds instead of days for the case of LBM [17]. In this work we mentioned that this approach was paving the way towards the computational screening of electrolyte/electrode pairs to find the conditions that will maximize the electrode wettability by the electrolyte. Indeed, ML's powerful capabilities have the potential to revolutionize the design of experiments and the material discovery process in batteries [18,31,32].

In the present article we present an innovative ML-based computational workflow able to optimize the electrolyte infiltration conditions, such as electrolyte properties and filling conditions (pressure). Such a workflow contributes to the general computational framework we are developing in the ARTISTIC project comprising a collection of sequentially coupled mechanistic, data-driven and hybrid-basedmodels to simulate each step of the manufacturing process [19,20,35]. In the following we start by formulating the wettability problem and recalling the LBM modeling framework for its solution. After identifying the pertinent design parameters, we present the surrogate modeling framework, based on simple ML techniques. The results section contains all the details of the study, beginning with exploratory data analysis and leading to the specific regression and classification methods used for the ML. Finally, we conclude and indicate our perspectives for continuing this work.

2. Methods

In the following, we describe our computational workflow designed to optimize the wettability of NMC111-based LIB electrodes. Such a workflow is constituted of several components, starting with a LBM model capable of predicting the impact of electrolyte properties (viscosity, density, contact angle), the electrolyte filling conditions (pressure) and the electrode properties (uncalendered vs. calendered) on the infiltration dynamics of the electrolyte in the porous electrode. The LBM can also model and predict the final saturation, defined as the ratio of volume occupied by the electrolyte divided by the volume of pores. Then ML models are set up to use the LBM-generated wettability data to perform supervised and unsupervised learning of the relations between carefully selected input parameters and pertinent output parameters [26]. Details of the LBM formulation can be found in our previous publication [14]. The ML approach is based on methods of [21,22] and constitutes a pillar of the more general concept of digital twins [23,36].

2.1. Electrode microstructures for LBM simulations

The first step for performing LBM simulations is to construct the porous electrodes. 3D micro X-ray tomography of an uncalendered and of a calendered electrode were used. This is because the calendering of the electrodes has a significant impact on the electrochemical performance of the LIB cells. For more details of this topic please see our previous publication [19]. The geometries are imported into the FIJI software for slicing and segmentation, then into the Geodict software [33] to extract pores using the watershed algorithm. The complete setup used is reported in detail in our previous publication [14]. We opted for a volume of $50 \times 50 \times 50 \mu m^3$ for the different electrode microstructures used in this study, to have a good compromise between representativeness of the electrode geometry and reasonable LBM simulation time. Average LBM simulation time is 1 h per condition, on an Intel® Core™ i7-12700 with a 32 GB of RAM. The porosity and the tortuosity factor of the uncalendered and the calendered electrodes are reported in Table 1. The porosity was calculated by using the Geodict software [33] and the tortuosity factor by using the TauFactor software [34].

The major difficulty of the electrolyte infiltration process is how to choose a set of design parameters that will produce an *optimal* wetting, as measured by the saturation, or some function of the saturation. The parameter design space has too many dimensions for exhaustive exploration, given the computational cost of a LBM simulation and the convergence difficulties for certain parameter combinations. This is the motivation for seeking a surrogate model, as described later.

2.2. Design parameters

The objective of our ML approach is to evaluate the influence of four LBM critical design parameters on the filling of a LIB electrode. It is important to point out here that the LBM model only accounts for dimensionless values. Therefore, the descriptions below contain not only the physical value, but the dimensionless value as well. The conversion between the physical value and the dimensionless value is done through the conversion factor. The four LBM parameters are:

1. Kinematic viscosity of the electrolyte, ν, of which the dimensionless value is defined as,

$$\nu = c_s^2 \, (1/\omega - 1/2), \tag{1}$$

where $\omega = 1/\tau$ with τ the relaxation time that c_s is the speed of sound in the lattice.

2. Contact angle, θ, which is defined in the LBM model by assigning the values of $G_{\mathrm{ads},\sigma}$, and $G_{\mathrm{ads},\sigma^*}$,

$$\theta = arccos\left(\frac{G_{\mathrm{ads},\sigma} - G_{\mathrm{ads},\sigma^*}}{\frac{G_{\mathrm{int}}(\rho_e - \rho_{\mathrm{dis}})}{2}}\right) \tag{2}$$

where $G_{\mathrm{ads},\sigma}$ and $G_{\mathrm{ads},\sigma^*}$ are respectively the interparticle strength be-

Table 1
Porosities and tortuosity factors of the two electrodes used as inputs in the LBM model.

Electrode microstructure	Porosity	Tortuosity factor
Uncalendered	48%	2.16
Calendered	27%	7.04

tween the nonwetting fluid and solid and the interparticle strength between the wetting fluid and solid. G_{int} represents the fluid-fluid interaction, ρ_e is the density of the electrolyte, and ρ_{dis} is the dissolved density of the nonwetting fluid in the wetting fluid.

3. Pressure, p, force applied on the inlet of the electrode sample. We assume that the electrolyte penetrates through the x-axis, of the cube (electrode sample) and is stabilized when the density distribution function meets the convergence criteria.
4. Electrolyte/air density ratio, ρ, given by

$$\rho = \rho_e / \rho_a \tag{3}$$

where ρ_e denotes the electrolyte density and ρ_a is the air density. It is important to point out here that due to the stability issue, the real density ratio between the electrolyte density and air cannot be implemented in the LBM model. For instance, the highest density ratio ρ that we can simulate is 9. In the present study, we chose a fixed value for the electrolyte density.

To summarize, the authorized ranges of values for these design variables are listed in Table 2, along with our chosen conversion factors. These ranges have been chosen to be in adequacy with experimental electrolytes [5].

2.3. Assessed features

To learn a data-driven model, training data is needed. This data can be obtained from experimental observations, or model-based simulations, or some combination of the two. Here we use the data produced by the LBM simulations as in our previous work [17].

For any supervised machine learning method, we first need to define a response variable (a function in this case) that is an unknown function of the input parameters. In our case, once an LBM simulation has been carried out, we can calculate wettability descriptors. To construct the response, calculations are based on two types of descriptors: the pore saturation

$$s_i(t) = \frac{V_{l_i}(t)}{V_i} \forall i \in [1, n] \tag{4}$$

where n is the number of pores, i the pore index, t time step, V_i volume of pore i and $V_{l_i}(t)$ volume of the electrolyte in pore i, and the overall saturation

$$s(t_j) = \frac{V_l(t_j)}{V_{\text{Tot}}} \quad t_j \in [0, T_f], \tag{5}$$

where T_f is the LBM simulation final time (corresponding to the time where electrolyte dynamics stops), $V_{\text{Tot}} = \sum_{i=1}^{n} V_i$ is the total pore volume and $V_l(t_j) = \sum_{i=1}^{n} V_{l_i}(t_j)$ is the total volume of electrolyte (see Fig. 1). All the LBM-simulated saturation curves are shown in Fig. 2.

When defining the output parameters for the ML, extreme care needs to be taken to extract reliable information describing adequately the electrolyte wetting rate, the speed of the electrode filling thus providing a solid basis for all the ML models employed. We choose three response variables, S_f, T_f and κ defined as it follows:

1. T_f: the time to reach the final saturation.
2. S_f: the final saturation of the electrode, defined as:

$$S_f = s(T_f) = \frac{V_l(T_f)}{V_{\text{Tot}}}. \tag{6}$$

3. κ: the proportion of filled pores in the electrode,

$$\kappa = \frac{n_f}{n} \tag{7}$$

where n_f is the number of filled pores in the electrode microstructure. we suppose that a pore i has a good filling when it is final saturation reaches 0.8 or higher (ie: $\kappa \geq 0.8$).

The first two variables are depicted in Fig. S1 in the supplementary materials.

Predicting these variables allow us to describe and predict the speed and the completion of the electrode infiltration based on different values of the four design variables in Table 2.

2.4. Data generation

The design of parameter samples used in this work was carefully generated using a maximum projection Latin Hypercube Sampling method (MaxProj LHS) with the Lagun online design of experiments generator [24].

This technique allows us to have a space-filling parameter range with good projection properties. The LHS method is a generalization to higher dimensions of the Latin square which is an $n \times n$ array filled with n different symbols, each occurring exactly once in each row and exactly once in each column.

Assuming a three-dimensional parameter space and n_s the number of samples of each parameter, then each sample is the only one in each axis-aligned hyperplane containing it. On the other hand, LHS method uses maximin criterion that is a commonly used approach seeking to maximize the minimum distance between any two points. This results in a well-dispersed design across all dimensions and projections. Hence, our training set is well-balanced, containing a wide range of saturation behaviors, as seen in Fig. 2 B), which will ensure the best possible training for the ML models.

3. Results and discussion

3.1. Exploratory data analysis

In order to identify useful preliminary information about the data and investigate the relationships between the features and the response variables, an exploratory data analysis should always be performed, before attempting any surrogate modeling. Furthermore, this study can

Table 2
LBM input parameters ranges used in the generation of the ML training and testing data sets. These ranges are used as input for the Lagun [24] Latin Hypercube Sampling (LHS) generator.

Conversion factors	Length	1×10^{-6} m/lu
	Density	ρ_a (kg/m^3)/lu
	Time	1×10^{-6} s/lu

Parameter	Lattice value range	Physical value range
Electrolyte/Air density ratio	$5 \leq \rho \leq 9$, $\rho_{a_{LBM}} = 1$, $5 \leq \rho_{e_{LBM}} \leq 9$ $\rho_{\text{dis}} = 0.06$)	$5 \leq \rho \leq 9$, $144 \leq \rho_a \leq 260(\text{kg/m}^3)$, $\rho_e = 1300(kg/m^3)$
Kinematic viscosity	$0.0098 \leq \nu_{LBM} \leq 0.39$ (with $0.6 \leq \omega \leq 1.89$)	$0.0098 \times 10^{-6} \leq \nu \leq 0.39 \times 10^{-6}(\text{m}^2/\text{s})$
Contact angle	$G_{\text{ads},\sigma} = -G_{\text{ads},\sigma^*}$ where : $-0.4 \leq G_{\text{ads},\sigma^*} \leq 0.4$	$70^* \leq \theta \leq 108^*$
Pressure	$0 \leq p_{LBM} \leq 0.015$	$0 \leq p \leq 3.9 \times 10^6(\text{kg m}/\text{s}^2)$

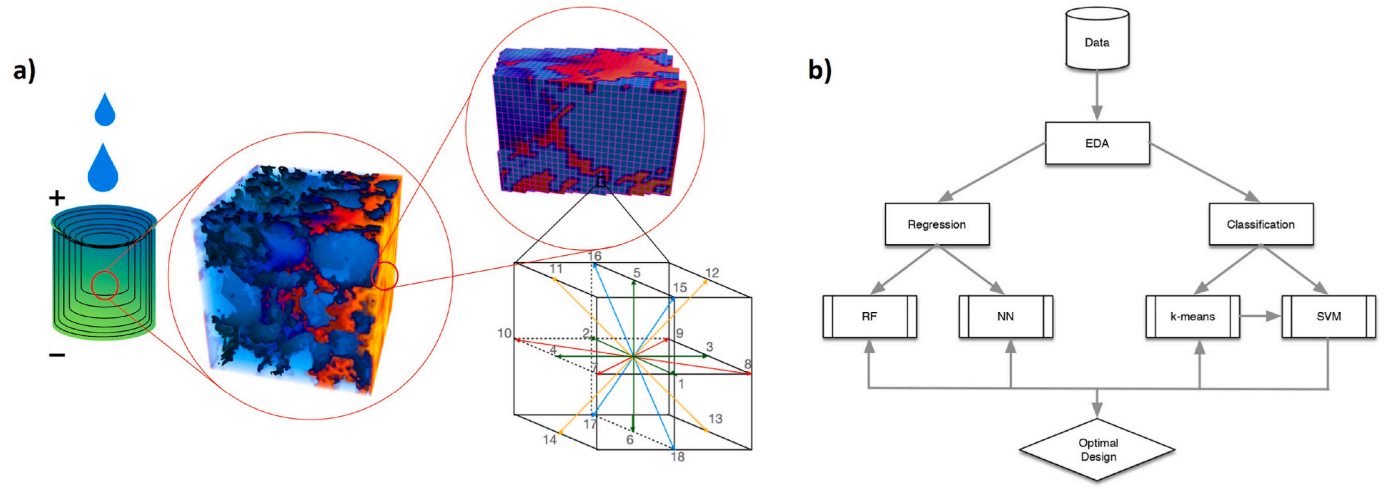

Fig. 1. a) A battery cell with a zoom into an electrode microstructure describing the electrolyte (blue) filling the porous medium initially filled with air (orange) and the D3Q19 discrete velocity lattice used in the Lattice Boltzmann Method; b) Our proposed workflow for optimal material design, demonstrated in this article in the context of electrolyte impregnation of Lithium Ion Battery Electrodes (EDA: Exploratory Data Analysis, RF: Random Forest, NN: Neural Networks, SVM: Support Vector Machine). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

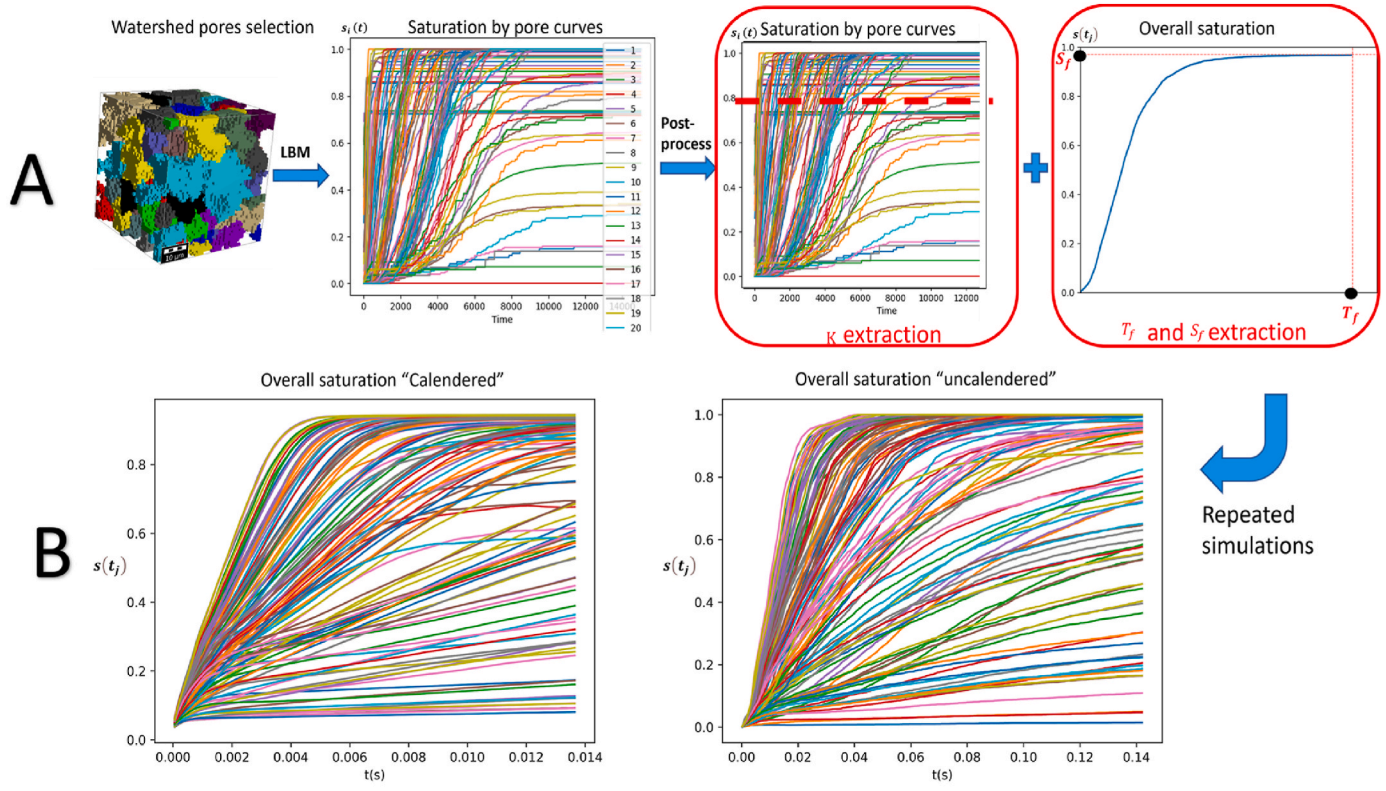

Fig. 2. A) A summary of the post processing used to extract the three output parameters. B) Overall saturation curves resulting from the complete data set used for training and testing our ML models for calendered(left), uncalendared(right) structures. The x-axis represents normalized time, the y-axis is the saturation, S(t), and each curve results from a unique LBM parameter combination.

help to detect the interactions among different variables in each of the electrode microstructures, helping us to understand the importance of the effect of calendering on the electrode wettability.

3.1.1. Scatter plots

The first step of the analysis is to display scatter plots. Fig. 3 shows a strong non-linear negative association between ν and S_f and between ν and κ, highlighting the significant impact of the electrolyte kinematic viscosity on the wetting degree (saturation) and the wetting rate (the proportion of filled pores). From one electrode microstructure to another, we observe a different order of significance, where the coefficients of correlation are respectively −0.7961 and −0.8622 for the uncalendered microstructure, versus −0.5371 and −0.587 for the calendered microstructure. Additionally, kinematic viscosity has a positive relationship with wetting time that decreases with low ν. A moderately strong correlation is furthermore noticed for both

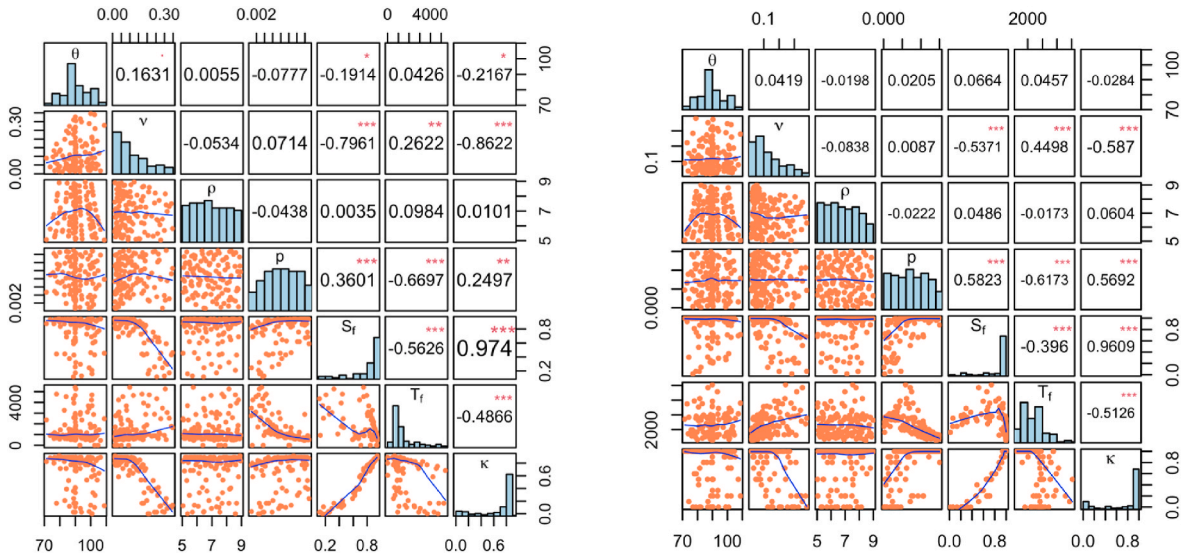

Fig. 3. Scatterplots, histograms and correlation matrix of LBM input and output data from data sets used in the training and testing of the ML models for the uncalendered (left) the calendered (right) structure. The red stars refer to the significance of the correlation coefficients. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

microstructures between p and S_f, T_f, κ revealing an interesting influence of high applied pressure on reducing the electrolyte infiltration time and completing the saturation. However, the applied pressure is of more importance in a reduced pore size volume, which is the case for the compressed microstructure, and explains the improvement of the correlation coefficients in comparison with the uncalendered electrode.

An additional correlation must not be neglected. Namely, the contact angles that have a non-substantial impact on the saturation time but have an influence on the wetting rate and degree for the uncompressed volume.

On the other hand, the analysis reveals an almost zero correlation between ρ and the responses for the calendered case, as well as for the uncalendered one.

Therefore, it is straightforward to conclude that the most influential parameter on the completion of the saturation is the electrolyte kinematic viscosity. Thus, the lower the viscosity is, the more we expect high values of saturation and the more the pores are filled. This is in addition to the crucial impact of the applied pressure on speeding up the wetting.

It is always recommended to inspect the partial correlations as well, so as to ensure the nonexistence of nuisance information in the database due to a confounding variable, which is a variable that influences both the dependent and independent variable causing then a spurious association as is discussed in the next paragraph.

3.1.2. Partial correlations

The aim of a partial correlation study is to measure the correlation between two variables, while removing the effect of the remaining controlling variables. This can provide better evidence of causality [26]. For three variables, the first order coefficient of X and Y controlling for Z, i.e. excluding the influence of Z, are calculated using the formula

$$r_{yx|z} = \left(r_{yx} - \left(r_{yz}\right)\left(r_{xz}\right)\right) \Big/ \sqrt{1 - r_{yz}^2}\sqrt{1 - r_{xz}^2} \tag{8}$$

where r_{yx}, r_{yz} , r_{xz} are the zero-order coefficients of correlation between all the possible pairs (Y,X), (Y,Z), (X,Z). Assuming n variables, equation (8) is used to calculate the n^{th}-order coefficient which is based on calculating the $(n-1)^{th}$-order coefficients. Afterwards, comparing the zero-order correlation to the n^{th}-order coefficient allows us to investigate if the association between variables is direct, spurious or intervening. Fig. S10 presents the partial correlation matrices for the two structures. The most striking observation to emerge from the comparison of correlation and partial correlation coefficients is the remarkable increase of this second, more precisely in the calendered electrode microstructure where it was first seen a nearly null association between the contact angle and all three responses, evincing then the existence of one or more suppressor variables.

Hence, θ has a significant impact on the electrode saturation behavior, a result that was confirmed in the uncalendered microstructure, Fig. S10, that reveals a spurious relation between θ and S_f and between θ and κ, on top of the improvement of the θ, T_f coefficient. Consequently, it seems that contact angle is of importance in accelerating and completing the electrolyte filling into both microstructures and with stronger interest for the calendered one. It is worth highlighting that the inversion of signs of the partial correlations can contrast physical interpretation in the contact angle observations (namely θ, S_f and T_f). It is well known that partial correlations, in spite of their capacity for detecting causal relations, can sometimes be misleading, and non-physical. The same warning is valid for viscosity and pressure relations to S_f, T_f and κ.

The correlation study has led us to deduce the most influential parameters on saturation descriptors and to see the apparent difference of correlation between the dependent and independent variables when changing the porosity percentage and pore structure. This is summarized in Fig. S11, where the partial correlation coefficients have, in addition, been hierarchically clustered to form a dendrogram, or classification tree. The exploratory analysis provides some preliminary information on the most influential parameters on wettability. The aim of this study is to feed the machine learning model with the physical intuition, in order to define the best wettability descriptors.

3.2. Regression models

Having collected the data and performed the Exploratory Data Analysis (EDA), we can then choose the ML methods that learn the complex relationships between the input and output data. Recall that there are two major categories of learning: supervised and unsupervised. In the former we have input-output data pairs, in the latter we do not have output variables, but we seek patterns or clusters in the input data. We use the two in tandem, since each can produce insights of its own. Within supervised learning, we differentiate between regression for numerical values and classification for qualitative groupings. Note that

most regression methods can also deal with discrete, categorical data.

Here, regression models are first fitted, followed by clustering and classification methods. Note that each approach can feed the other, since they produce complementary information. Both, used together, are needed for optimal material design.

3.3. Random forest method

Random forests are supervised learning models for classification and regression made out of decision trees and consist of creating a bootstrap data set with the same size as the original. A bootstrapping method includes all the tests and metrics that use random sampling with replacement and the remaining observations, the out-of-bootstrap data set, are called out-of-bag observations. The second stage involves the creation of decision trees from the bootstrapped data set using only a random subset of the explanatory variables at each step. From our bootstrapping data set, we select two or more random variables, and we choose the best variable that separates the samples. This sampling is repeated in each node with the remaining variables.

For a regression problem as in our case, the variable selection on each node is based on minimizing the mean squared error,

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

where n is the number of observations, y_i is the observed value and $\hat{y}_i$ is the predicted value.

These two methods—bootstrapping and creating a tree—are repeated until we obtain a variety of trees.

To assess the ability of the trained model to predict the three outputs, we use a cross-validation approach that consists of taking one subset for test data, and then training the model over the remaining subsets. The data set is split into 6 folds randomly and repeated 10 times for the S_f , 8 folds and 10 repeats for κ models and 15 folds and 20 repeats for the T_f model.

Fig. 4 shows the actual values versus the predicted values, respectively for S_f, T_f and κ. These fittings have relatively good R2 and RMSE values, as seen in Table 3 for the three variables for the calendered case and for both S_f and κ predictions in the uncalendared one, contrary to T_f whose accuracy is slightly lower and for which there is certainly room for improvement.

Table 3
Residuals table values representing the R-squared and the Root Mean Squared errors of the random forest models for Sf, Tf, κ. For the uncalendered and the calendered structure.

Model	Uncalendered	Calendered	Uncalendered	Calendered
	R^2	R^2	RMSE	RMSE
S_f	0.89	0.88	0.06	0.08
T_f	0.74	0.87	0.13	0.08
κ	0.89	0.88	0.08	0.12

Interestingly, random forests can provide additional information about the importance of the features in predicting the response variable based on Breiman's permutations [25]. The main idea of this method consists of permuting randomly the observations of the tested feature in the out-of-bag observations for each tree. An important increase of the out-of-bag MSE implies the importance of the corresponding variable.

It is apparent from Supplementary Fig. S3 and Fig. S4 that the electrolyte kinematic viscosity, ν, and pressure, p, are the most influential parameters on the predictions of the three responses with different orders of importance depending on the structure and the response variable. Indeed, taking for instance the results of Fig. S4, the importance index VI for a given value of ν is VI $\simeq$ 3.6 for S_f, VI $\simeq$ 0.69 for T_f and VI $\simeq$ 6.5 for κ. These values, and the rest of the VI values, are consistent with the correlation and partial correlation analysis results.

3.4. Feed-forward neural networks

The second approach we used for regression is neural networks [28]. In this work, the adopted NN architecture is a simplified, neater network with a single hidden layer. Additional tests were performed by us with

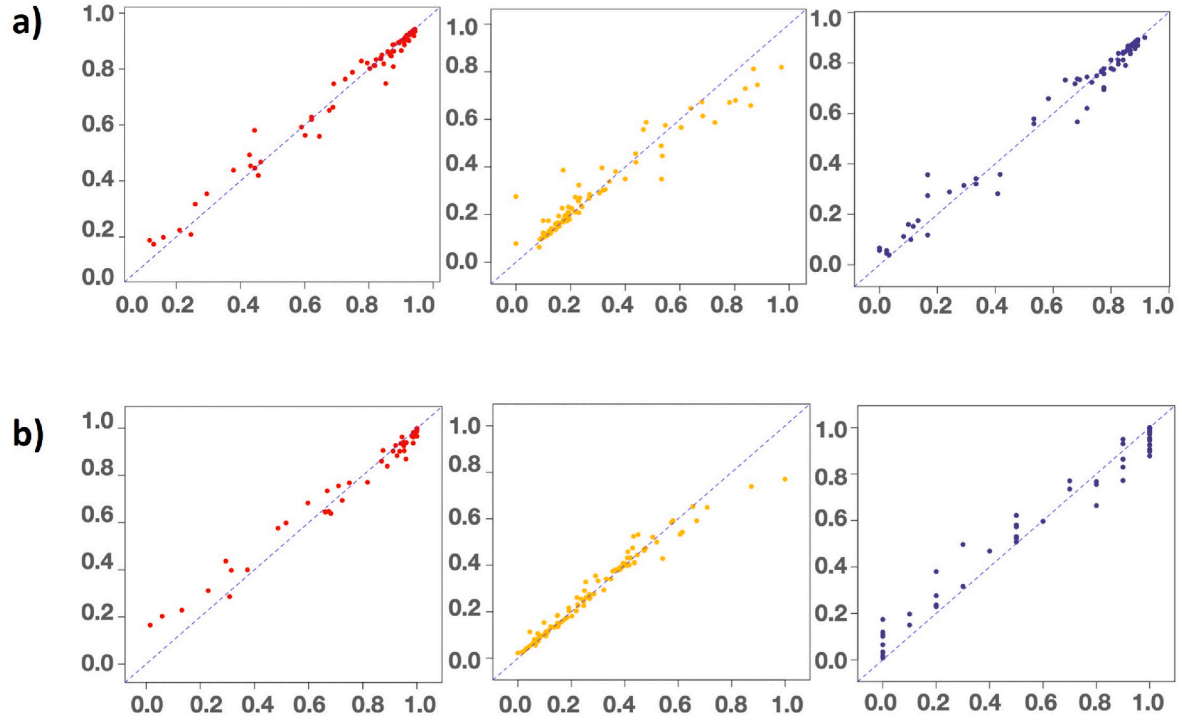

Fig. 4. Actual (x-axis) vs. predicted (y-axis) values of random forest models showing the accuracy achieved by the RF models when predicting Sf, Tf, κ (from left to right), for the (a) uncalendered (b) calendered structure.

Table 4
Residuals table values for neural networks models for Sf, Tf, κ, for the uncalendered and the calendered structure.

	Uncalendered	Calendered	Uncalendered	Calendered
	R^2	R^2	RMSE	RMSE
S_f	0.90	0.83	0.12	0.15
T_f	0.70	0.93	0.14	0.05
κ	0.81	0.76	0.11	0.16

two and three hidden layers, but no increase of the accuracy was observed. A one-layer neural network is sufficient and computationally efficient. Together with cross-validation, the model clearly provided very good prediction accuracies, as shown in Table 4 and has improved the random forest outcomes in terms of R^2 for T_f in the calendered structure. However, the random forest produces much better predictions for the final saturation, S_f, in terms of the MSE. This is clearly observed when comparing the scatter plots, Fig. 4 a) (S_f) with Supplementary Fig. S5 (S_f), and also Fig. 4 b) (S_f) with Supplementary Fig. S6 (S_f). Note that the R^2 value measures the explanatory power of the model—how much of the variance is captured—whereas the MSE is a measure of the confidence in the model, since it can be used to compute confidence intervals. This example shows that R^2 is not always a reliable estimate of the goodness-of-fit for a model and should be used with precaution.

The evidence from this part of the study suggests two different, rapid and reliable approaches to predict well-chosen electrode wettability descriptors and to assess the filling speed, homogeneity and completion by changing together the four electrolyte variables. Comparing the results of Table 3 for RF, with those of Table 4 for the NN model, and examining Fig. 4, Supplementary Fig. S5 and Fig. S6, we can make the following observations.

1. Overall, both models provide good accuracy, as measured by R^2. However, the random forest has smaller RSMEs and hence better predictive power for unseen test data. This is clear, in particular, when comparing the actual vs. predicted plots for S_f, where in spite of a good R^2, the neural network does not have good predictive power especially for lower range values.
2. The random forest model provides overall lower mean-squared errors.
3. The overall, best performance is obtained by the neural network model for T_f in the uncalendered case.
4. For both models, the performance is equally good for the calendered and uncalendered structures.

In the next section, a clustering and a classification model are presented in order to allow, just as for the regression models presented above, better decision-making on the choice of electrolyte parameters and their optimization by identifying clusters.

3.5. Classification and clustering models

In this section, two different models were evaluated in order to categorize the data considering the values of their saturation, time of saturation and the completion of the saturation. First, an unsupervised clustering model was fitted, providing an accurate partition of the data set without any labels. The *k*-means algorithm used here, found patterns, and assigned successfully the observations to the corresponding category by taking into account both the completion and the speed of the saturation. Afterwards, we used a support vector machine (SVM) technique, a supervised classification machine leaning model that takes target variables inspired by the *k*-means outcomes and uncovers the relationships between the electrolyte variables and the targetsthat are the wettability descriptors classes.

3.5.1. Unsupervised learning (k-means)

The unsupervised, *k*-means analysis aims to identify *k* groups in the data set by optimizing the distances between each point and the data with similarities and maximizing cluster separations using the metrics defined below. The Within-cluster Sum of Squares is defined as,

$$\mathrm{WSS} = \sum_{i=1}^{N} \sum_{x \in C_i} d(x, \overline{x_{C_i}})^2, \tag{10}$$

where C_i refers to the i^{th} cluster, N the size of the cluster and $\overline{x}_{C_i}$ is the cluster centroid. The compactness of the clustering is measured using the formula

$$I = BSS/TSS. \tag{11}$$

Here, BSS is the Between-cluster Sum of Squares, defined as

$$\mathrm{BSS} = \sum_{i=1}^{N} |C_i| d(\overline{x_{C_i}}, \overline{x})^2, \tag{12}$$

where $\overline{x}$ is the cluster mean and TSS stands for Total Sum of Squares. A good partition is supposed to minimize the distances between the points belonging to the same cluster and maximize the separation between the clusters. Consequently, a high I index proves a high compactness of the clustering.

An initial model was adjusted assuming an arbitrary number of clusters (three clusters) and the compactness obtained was I = 89% and I = 85.6% for the uncalendared and calendered microstructure respectively. Details can be found in Table 5.

A scree plot of the WSS as a function of the number of groups shows an additional decrease of the WSS when the number of groups increases from 3 to 4, before flattening out, as can be seen in Fig. S7. Using this observation, we improved the clustering to I = 93.9% and I = 91.4% respectively for the uncalendered and the calendered structures, considering k = 4 clusters—see Table 6 for details.

Furthermore, in an attempt to verify not only the compactness of the classification within the same group and the goodness of the separation of all the clusters as reported by the I index, but also to investigate in detail how each observation is similar to the others in the same group and to compare the different clusters, we performed a silhouette analysis which is based on a metric called the silhouette width defined as follows.

For the j^{th} observation,

$$s(j) = \frac{b(j) - a(j)}{max(a(j), b(j))}, \tag{13}$$

where a(j) is the within cluster mean distance of the point *j* to all other points in the cluster,

$$a(j) = \frac{1}{|C_i| - 1} \sum_{k \in C_i, j \neq k} d(j,k) \tag{14}$$

and b(j) is the closest neighbor distance calculating the minimal mean distance of the point *j* to all the points in the other remaining clusters.

This distance is defined for a cluster $C_{i^*} \neq C_i$ and for a point $j \in C_i$ as

$$b(j) = min_{i^* \neq i} \frac{1}{|C_{i^*}|} \sum_{k \in C_{i^*}} d(j,k). \tag{15}$$

The silhouette width provides a score between −1 and 1, with values closer to 1 indicating a well-matched sample to a cluster, 0 for an observation on the border of two groups and −1 for a sample that better fits in a neighboring cluster.

This test revealed a good classification with an average silhouette width of 0.56 for the uncalendered structure and 0.61 for the calendered structure. Moreover, the silhouette width values are positive for the whole data set for both calendered and uncalendered microstructures, as seen in Fig. S8, thus supporting the previous *k*-means compactness

Table 5
Clusters means, sizes and WSS resulting from k-means clustering with k = 3 clusters for the uncalendared and the calendered structure.

Uncalendered									
Cluster	Means							Size	WSS
	θ	ν	ρ	p	S_f	T_f	κ		
1	89.14	0.099	6.82	0.010	0.95	1752	0.94	65	4754944
2	88.61	0.21	6.50	0.005	0.75	3953	0.54	14	3504887
3	90.91	0.14	6.90	0.004	0.74	2745	0.61	50	3010435
Calendered									
Cluster	Means							Size	WSS
	θ	ν	ρ	p	S_f	T_f	κ		
1	90.25	0.097	6.87	0.009	0.87	844	0.78	86	9279600
2	92.46	0.16	7.41	0.002	0.47	4385	0.33	12	4197186
3	92.81	0.14	6.83	0.005	0.67	2418	0.55	21	5549244

Table 6
Clusters means, sizes and WSS resulting from k-means clustering with k = 4 clusters for the uncalendared and the calendered structure.

Uncalendered									
Cluster	Means							Size	WSS
	θ	ν	ρ	p	S_f	T_f	κ		
1	90.13	0.15	7.64	0.003	0.50	4643	0.36	9	1752227
2	90.11	0.10	6.86	0.01	0.87	721	0.78	68	4142408
3	93.16	0.14	6.72	0.004	0.62	844	0.51	16	2743939
4	91.96	0.11	6.93	0.006	0.81	1470	0.70	26	2032903
Calendered									
Cluster	Means							Size	WSS
	θ	ν	ρ	p	S_f	T_f	κ		
1	89.41	0.11	6.64	0.006	0.87	2195	0.84	31	899748
2	89.23	0.09	6.95	0.011	0.97	1601	0.97	45	1224068
3	91.09	0.14	6.92	0.003	0.71	2847	0.57	39	1095903
4	88.61	0.21	6.50	0.005	0.75	3953	0.54	14	3504887

results.

To better understand the structure of the clusters, and their practical performance, a parallel coordinate plot is used. This plot provides a global view of how the variables' values change from one group to another, thus confirming that data separates well the points based on their response variable values. Closer inspection of the values of the output variables provides a global idea of the most suitable data for experimental studies. The parallel coordinate plot Fig. 5 represents each observation with a line that connects the standardized values taken in each variable with colors distinguishing the different clusters. In fact, in the uncalendered case of Fig. 5, group 3 (cyan) and group 2 contain electrolyte variables that generate high values of saturation, lower filling time and remarkable percentage of filled pores. Groups 1 and 4 also present some interesting behaviors of saturation, yet with lower filling speed. The examination of Fig. 5 for the calendered microstructure, also provides some information of the most interesting groups of parameters for experimental studies, namely group 2 (green) that contains high saturation values, the lowest filling time and a high proportion of filled pores.

The above results of the unsupervised learning provide suggestions for the choice of parameter intervals giving good wettability behaviors and pave the way for the SVM method presented in the next section. In order to maximize the classification accuracy, we trained several models with different numbers of clusters and compared the results obtained.

3.5.2. Supervised classification (Support vector machine)

In this part, we use the Support Vector Machine (SVM) method for a supervised classification task.

Great care was taken both for the choice of input parameters and for data scaling. We created one set of models using all four LBM input parameters, and a second set using only the two important parameters observed in the exploratory data analysis and random forest data importance plots, namely ν and p. The three response variables were divided into k equal-length classes, or sub-intervals, with $k = 2, 3, 4$ and the class lengths are defined by

$$L_k = \frac{Y_{max} - Y_{min}}{K}, \tag{16}$$

where $Y \in \{S_f, T_f, \kappa\}$. That is, for K = 2, we classify the range of the response into 2 equal parts for each value of Y, etc.

We used SVM with a radial kernel and tuning on its parameters, together with cross-validation. In Table 8, we present the accuracy obtained for a reduced-order model, depending on only the viscosity ν and the pressure p. Additional detailed results can be found in Table 7, where we summarize the accuracies obtained for the uncalendered microstructure with a full model, and in Table 9 and Table 10 for the calendered microstructure. In the uncalendered case we compare unscaled and scaled variables, whereas in the calendered case, results are only presented for the scaled case. We use a standardization scaling to preserve the central value and deviations around it. Taken as a whole, our SVM models classified the data extremely well, with a considerable improvement when scaled, with few exceptions. This can be interpreted by removal of information by the scaling that is not a part of the noise.

The results obtained with SVM are highly accurate and have further strengthened the previous classification and broadened the comprehension of the importance of the choice of electrolyte variables' in enhancing the classification and facilitating optimal experimental decision-making.

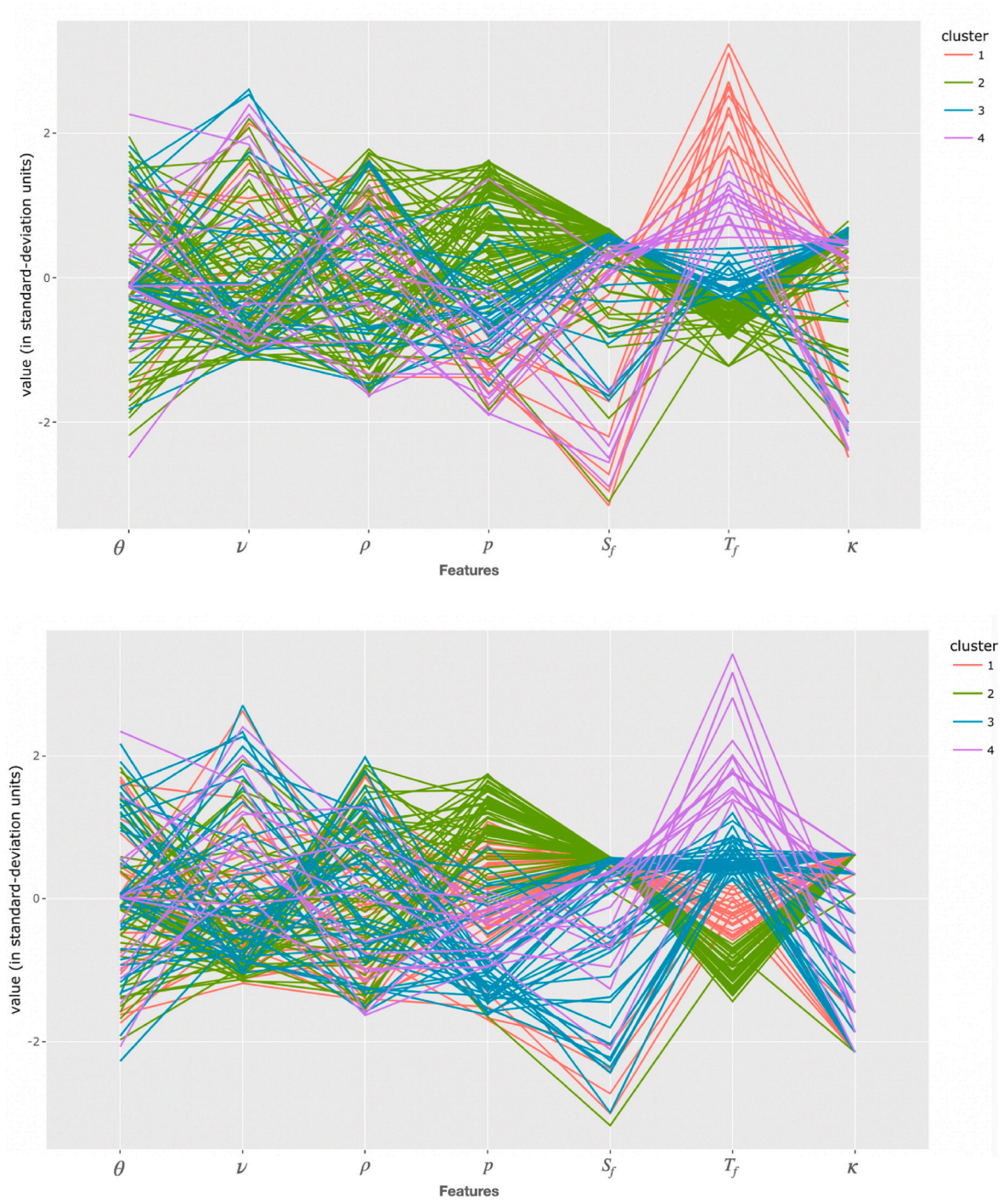

Fig. 5. Cluster plots from the k-means clustering for the uncalendared (above) and calendered(below) structure. The plots include all the clusters detected in the data set.

Table 7
Classification accuracy representing the proportion of observations in the test set that was classified correctly for SVM models taking all the four features in the classification, Y ∼ ν+ρ+p+θ, where Y ∈ {Sf, Tf, κ}, for the uncalendered structure.

Response variables	2 Classes		3 Classes		4 Classes	
	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
S_f	1	1	0.91	1	0.86	0.91
T_f	0.78	0.82	0.91	0.78	0.69	0.73
K	0.86	0.91	0.95	0.95	0.82	0.78

Table 8
Classification accuracy representing the proportion of observations in the test set that was classified correctly for SVM models taking only ν and p in the classification, Y ∼ ν + p, where Y ∈ {Sf, Tf, κ}, for the uncalendered structure.

Response variables	2 Classes		3 Classes		4 Classes	
	Unscaled	Scaled	Unscaled	Scaled	Unscaled	Scaled
S_f	1	1	1	1	0.95	0.91
T_f	0.86	0.91	0.73	0.86	0.60	0.78
κ	0.95	1	0.95	0.91	0.82	0.86

Table 9
Classification accuracy representing the proportion of observations in the test set that was classified correctly for SVM models taking all four features in the classification, Y ~ ν+ρ+p+θ, where Y ∈ {Sf, Tf, κ}, for the calendered structure.

Response variables	2 Classes	3 Classes	4 Classes
	Scaled	Scaled	Scaled
S_f	0.92	0.88	0.96
T_f	0.76	0.72	0.68
κ	0.96	0.88	0.96

Table 10
Classification accuracy representing the proportion of observations in the test set that was classified correctly for SVM models taking only ν and p in the classification, Y ~ ν + p, where Y ∈ {Sf, Tf, κ}, for the calendered structure.

Response variables	2 Classes	3 Classes	4 Classes
	Scaled	Scaled	Scaled
S_f	0.96	0.92	0.80
T_f	0.84	0.80	0.88
κ	1	0.92	0.88

We notice as well for the calendered microstructure, a good classification accuracy that was improved by reducing the number of features to two.

From the results reported above, we can make the following observations.

1. The two-variable model provides much better accuracy than the complete, four-variable model.
2. In general, the fewer the number of classes, the better the classification obtained. However, we then obtain a broader clustering. For example, the top 50 % for k = 2. A good compromise would be k = 3, where we have a finer grain, but still retain good accuracy.
3. Scaling, in general, increases the accuracy, though there are some exceptions. It is thus advisable to adjust also for the unscaled case.
4. We have equally good performance for calendered and uncalendered electrode microstructures.

4. Conclusion

Electrode wetting by liquid electrolytes is a critical step in the LIB manufacturing process. This step is recognized by LIB R&D laboratories, prototyping lines and manufacturers to be one of the most time consuming and costly steps. Indeed, the injection of electrolyte in the electrodes' porosities requires the application of pressure-vacuum pumping strategies without warranty that the full porosity will be fully occupied with electrolyte at the end of this process step. The electrode wettability strongly depends on the angle of contact between the electrolyte and the electrode, the electrode microstructure characterized by its porosity, pore network and tortuosity factor, the electrolyte viscosity and density. This is therefore a complex parameter space that needs to be optimized in order to ensure proper electrode wettability. However, this is carried out by using trial-and-error approaches which are inherently inefficient. The use of physics-informed ML techniques can provide physics-informed surrogate models for the optimization of this multi-parameter process step that depends on both chemical and physical properties. In the present article, we proposed a novel general workflow for realizing this objective and provided detailed computational simulation-based experiments. These physics-informed surrogate models pave the way to tractable solutions of parameter identification and material design optimization problems, as we already demonstrated in other battery manufacturing-related problems [29,30].

The different ML techniques reported in this article provide, in addition to good prediction capability, the mapping of the interdependencies between the four input parameters and the electrode saturation completion and speed.

Exploratory data analysis (EDA) of the training samples provided important information, where we observed separately the most influential electrolyte parameters on the saturation descriptors. Thereby, kinetic viscosity, the inlet applied pressure and the contact angle play a significant role in increasing the wetting rate and speed of the saturation for both uncalendered and calendered microstructures with different degrees of significance.

In fact, the lower the kinematic viscosity is, the higher wetting degree and wetting rate are, the higher inlet applied pressure reduces the electrolyte filling time for both microstructures. In addition, some extra effects are noticed differently from one microstructure to another, such as the impact of the kinematic viscosity on the wetting time and the pressure on the wetting rate and degree. In addition, the contact angle has influence on the wetting rate, wetting degree and filling time, which are more important in the calendered case, as revealed by the correlation and partial correlation analyses.

Thanks to the EDA, we obtained a general idea on the way the electrolyte kinematic viscosity, the inlet applied pressure and the contact angle are influencing the electrode saturation behavior. In order to find the best parameter combination for experimental use, the ML models implemented in this study provide a good prediction and classification of the saturation rate and degree and the electrolyte filling time, even for data that are not used for the training.

Fig. S2 shows a typical example of the impact of the electrolyte parameters to generate either a perfect filling (set 1) or a bad one (set 2). The values are reported in Table 11. We notice that with a slight variation of θ and ρ and a significant change of ν and p, we obtain remarkable differences in the responses. These minor variations in the choice of the electrolyte parameters make a significant impact on the wettability completion, therefore, simple hand decision-making cannot be accurate in all cases.

The approach described in this paper proposes a universal workflow for data analysis and surrogate modeling of the electrolyte infiltration parameter by optimizing design choices. This workflow can also be applied to almost any material design problem, in particular for batteries. It suffices to replace our data by other data, our underlying model by a suitable model, and then follow the steps of the workflow. Last but not least, we provide in the following some general comments that can be taken as advice for the successful implementation of our workflow.

First, once the data is collected, the EDA is essential for:

1. Familiarization with the data and choice of response variables;
2. Elimination of any unusual, or erroneous data points;
3. Preliminary identification of the most influential features, or parameters, and the relations—or lack of relations—between the features themselves, and between the features and the response variables;
4. Reduction of the complexity by identification of co-linear variables, and confounding variables.

The next steps are regression and classification. Note that classification, especially unsupervised clustering, can be considered as being a part of EDA, since it can help in determining response variables. In our

Table 11
Table of parameter design corresponding to the optimal and incomplete wettability in Fig. S2.

Features					Responses		
Set	ν	ρ	p	θ	S_f	T_f	κ
1	0.0256	6.2307	0.0075	89.88	1	1733s	1
2	0.2152	6.8974	0.0021	90.33	0.229	2675s	0

analysis, we have purposefully selected simple approaches because of their broad applicability, ease of computation and facility of interpretation—no black boxes here. We chose, and recommend, for regression:

1. Random forests, because of their established robustness and their capacity to rank explanatory variables by their importance;
2. Neural networks, for their extreme versatility and their universal approximation properties.

We chose, and recommend, for clustering and classification:

1. k-means for initial unsupervised clustering and identification of groups of properties;
2. SVM for refined, supervised clustering that provides a surrogate model.

The models implemented in this work were tested on unseen data to validate the process and have provided accurate results - see section 2.7 in the supplementary material for more details on the machine learning models details. In the final step, which of course will be context-dependent, we can exploit all the surrogate models found above for optimal manufacturing process design and manufacturing process planning.

This study has provided valuable insights into wettability optimization. Future work could expand on these findings by addressing this problem through Bayesian optimization.

Author contributions

A.A.F. and M.A. conceived the approach and supervised the research. A.E.M. performed all the computations and developed the machine learning models. A.S. and O.A. provided theoretical and software support. All authors contributed to the discussion and the writing of the paper. A.A.F. obtained the funding, managed and supervised the project.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Available upon request. In a latter stage, data and codes will be made available in the ARTISTIC project online calculator: https://www.erc-artistic.eu/computational-portal

Acknowledgments

A.A.F., A.E.M. and A. S. acknowledge 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 O.A. acknowledge partial funding from the Région Hauts de France and from the European Union's Horizon 2020 research and innovation program under grant agreement 957189 (BIG-MAP). A. A.F. acknowledges Institut Universitaire de France for the support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.powera.2023.100114.

References

[1] G.E. Blomgren, The development and future of lithium ion batteries, J. Electrochem. Soc. 164 (Number 1) (2016) A5019–A5025.
[2] A. Masias, J. Marcicki, W.A. Paxton, Opportunities and challenges of lithium ion batteries in automotive applications, ACS Energy Lett. 6 (February 2021) 621–630.
[3] A.R. Shekhar, M.H. Parekh, V.G. Pol, Worldwide ubiquitous utilization of lithium-ion batteries: what we have done, are doing, and could do safely once they are dead? J. Power Sources 523 (2022), 231015.
[4] E.N. Primo, M. Chouchane, M. Touzin, P. Vazquez, A.A. Franco, Understanding the calendering processability of Li(Ni0.33Mn0.33Co0.33)O2-based cathodes, J. Power Sources 488 (2021), 229361.
[5] A. Davoodabadi, J. Li, Y. Liang, D.L. Wood, T.J. Singler, C. Jin, Analysis of electrolyte imbibition through lithium-ion battery electrodes, J. Power Sources 424 (2019) 193–203.
[6] A. Davoodabadi, C. Jin, I.I.I. David L. Wood, T.J. Singler, J. Li, On electrolyte wetting through lithium-ion battery separators, Extreme Mech. Lett. 40 (2020), 100960.
[7] A. Davoodabadi, J. Li, H. Zhou, D.L. Wood, T.J. Singler, C. Jin, Effect of calendering and temperature on electrolyte wetting in lithium-ion battery electrodes, J. Energy Storage 26 (2019), 101034.
[8] T. Knoche, V. Zinth, M. Schulz, J. Schnell, R. Gilles, G. Reinhart, In situ visualization of the electrolyte solvent filling process by neutron radiography, J. Power Sources 331 (2016) 267–276.
[9] F.J. Günter, J.B. Habedank, D. Schreiner, T. Neuwirth, R. Gilles, G. Reinhart, Introduction to electrochemical impedance spectroscopy as a measurement method for the wetting degree of lithium-ion cells, J. Electrochem. Soc. 165 (2018) A3249–A3256.
[10] W.J. Weydanz, H. Reisenweber, A. Gottschalk, M. Schulz, T. Knoche, G. Reinhart, M. Masuch, J. Franke, R. Gilles, Visualization of electrolyte filling process and influence of vacuum during filling for hard case prismatic lithium ion cells by neutron imaging to optimize the production process, J. Power Sources 380 (2018) 126–134.
[11] A. Schilling, P. Gümbel, M. Möller, F. Kalkan, F. Dietrich, K. Dröder, X-Ray based visualization of the electrolyte filling process of lithium ion batteries, J. Electrochem. Soc. 166 (December 2018) A5163. –A5167.
[12] S.G. Lee, D.H. Jeon, B.M. Kim, J.H. Kang, C.-J. Kim, Lattice Boltzmann simulation for electrolyte transport in porous electrode of lithium ion batteries, J. Electrochem. Soc. 160 (2013) H258–H265.
[13] D.H. Jeon, Wettability in electrodes and its impact on the performance of lithium-ion batteries, Energy Storage Mater. 18 (2019) 139–147.
[14] A. Shodiev, E. Primo, O. Arcelus, M. Chouchane, M. Osenberg, A. Hilger, I. Manke, J. Li, A.A. Franco, Insight on electrolyte infiltration of lithium ion battery electrodes by means of a new three-dimensional-resolved lattice Boltzmann model, Energy Storage Mater. 38 (2021) 80–92.
[15] M. Lautenschlaeger, B. Prifling, B. Kellers, J. Weinmiller, T. Danner, V. Schmidt, A. Latz, Understanding electrolyte filling of lithium-ion battery electrodes on the pore scale using the lattice Boltzmann method, Batteries. Supercaps. 5 (2022).
[16] A. Shodiev, F.M. Zanotto, J. Yu, M. Chouchane, J. Li, A.A. Franco, Designing electrode architectures to facilitate electrolyte infiltration for lithium-ion batteries, Energy Storage Mater. 49 (2022) 268–277.
[17] A. Shodiev, M. Duquesnoy, O. Arcelus, M. Chouchane, J. Li, A.A. Franco, Machine learning 3D-resolved prediction of electrolyte infiltration in battery porous electrodes, J. Power Sources 511 (2021), 230384.
[18] T. Lombardo, M. Duquesnoy, H. El-Bouysidy, F. Årén, A. Gallo-Bueno, P. B. Jørgensen, A. Bhowmik, A. Demortière, E. Ayerbe, F. Alcaide, M. Reynaud, J. Carrasco, A. Grimaud, C. Zhang, T. Vegge, P. Johansson, A.A. Franco, Artificial intelligence applied to battery research: hype or reality? Chem. Rev. 122 (2022) 10899–10969.
[19] A.C. Ngandjong, T. Lombardo, E.N. Primo, M. Chouchane, A. Shodiev, O. Arcelus, A.A. Franco, Investigating electrode calendering and its impact on electrochemical performance by means of a new discrete element method model: towards a digital twin of Li-Ion battery manufacturing, J. Power Sources 485 (2021), 229320.
[20] T. Lombardo, F. Caro, A.C. Ngandjong, J.-B. Hoock, M. Duquesnoy, J.C. Delepine, A. Ponchelet, S. Doison, A.A. Franco, The ARTISTIC online calculator: exploring the impact of lithium-ion battery electrode manufacturing parameters interactively through your browser, Batteries. Supercaps. 5 (2022), e202100324.
[21] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer-Verlag, New York, 2013.
[22] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer-Verlag, New York, 2013.
[23] M. Asch, A Toolbox for Digital Twins: from Model-Based to Data-Driven, SIAM, Philadelphia, 2022.
[24] SafranTech, Lagun Computer Code Exploration Platform, 2021.
[25] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[26] J. Pearl, in: Causality. Models, Reasoning, and Inference, second ed., Cambridge University Press, , UK, 2009.
[28] A. Pinkus, Approximation theory of the MLP model in neural networks, Acta Numer. 8 (1999) 143–195.
[29] M. Duquesnoy, T. Lombardo, M. Chouchane, E.N. Primo, A.A. Franco, Data-driven assessment of electrode calendering process by combining experimental results, in silico mesostructures generation and machine learning, J. Power Sources 480 (2020), 229103.
[30] M. Duquesnoy, T. Lombardo, A.C. Ngandjong, J. Xu, F. Haudiquez, H. Oularbi, A. A. Franco, Functional data-driven framework for fast forecasting of electrode slurry rheology simulated by molecular dynamics, npj Comput Mater 8 (2022) 161.

[31] J. Li, K. Lim, H. Yang, Z. Ren, S. Raghavan, C. Po-Yen, T. Buonassisi, X. Wang, AI applications through the whole life cycle of material discovery, Matter 3 (2020) 393–432.
[32] D. Heaven, Why deep-learning AIs are so easy to fool, Nature 574 (2019) 163–166.
[33] ([Online]), https://www.geodict.com/geodict-software/geodict-base-modules.html.
[34] S.J. Cooper, A. Bertei, P.R. Shearing, J.A. Kilner, N.P. Brandon, TauFactor: an open-source application for calculating tortuosity factors from tomographic data, Software 5 (2016) 203–210.
[35] M. Duquesnoy, C. Liu, D. Zapata 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.
[36] F.M. Zanotto, D.Z. Dominguez, E. Ayerbe, I. Boyano, C. Burmeister, M. Duquesnoy, M. Eisentraeger, J. Florez Montaño, A. Gallo-Bueno, L. Gold, F. Hall, N. Kaden, B. Muerkens, L. Otaegui, Y. Reynier, S. Stier, M. Thomitzek, A. Turetskyy, N. Vallin, J. Wessel, X. Xu, J. Abbasov, A.A. Franco, Data specifications for battery manufacturing digitalization: Current status, challenges, and opportunities, Batteries & Supercaps 5 (9) (2022), e202200224.