Highlights

Performance of upcycled tyre-derived carbon in ethaline-glycol ionic electrolytes for sustainable supercapacitors: modelling cyclic voltammetry using standard and machine-learning methods.

Maria A. Sandoval-Riofrio, Kavisha Jayathunge

-
-
-

Performance of upcycled tyre-derived carbon in ethaline-glycol ionic electrolytes for sustainable supercapacitors: modelling cyclic voltammetry using standard and machine-learning methods.

Maria A. Sandoval-Riofrio[a,*,1], Kavisha Jayathunge[a,2]

[a] Bournemouth University, Talbot Campus, Fern Barrow, Poole BH12 5BB, Dorset, UK

ARTICLE INFO

Keywords:

ABSTRACT

We evaluate multiple machine learning model configurations for predicting specific capacitance, showing that incorporating an area-based loss consistently improves accuracy and stability across runs. The best-performing model achieves an average error of 3.75%, compared with a baseline error of 15.49%. Analysis of the model's learned weights indicates that electrolyte temperature is the dominant factor influencing specific capacitance, followed closely by active mass and electrolysis voltage.

1. Deep learning analysis

Machine learning models can be thought of as universal function approximators. That is, provided enough data to learn from, they are able to model any continuous, closed domain function to an arbitrary degree of accuracy (Geuchen et al., 2025). Such models consist of interconnected "neurons" arranged in layers. These connections are dense, meaning each neuron is connected to all other neurons in the previous and next layer, giving rise to the common "fully connected" layer terminology. We will refer to a single layer as $F_i^{n,m}$ where i is the position of the layer, n is the input size (the number of input neurons expected), and m is the output size. The layer computes:

$$\hat{y} = F_i^{n,m}(x) = W_i x + b_i \tag{1}$$

where x is an n-dimensional vector, $\hat{y}$ is an m-dimensional vector, W_i is an $n \times m$ weight matrix and b_i is a constant term. Layers may be chained sequentially, as long as the inner dimensions of adjacent layers are equal, i.e. W_i and W_{i+1} can be multiplied. Such a sequence of connected layers is commonly referred to as a Multilayer Perceptron (MLP). An important element of the layers of the MLP is the activation function, which is applied to the output of each neuron, and introduces non-linearity into the model. This is essential for an MLP to fit the definition of a universal function approximator as defined earlier, because any sequence of fully connected layers must necessarily collapse into a single linear transform, which is not powerful enough to represent more complicated functions. The Rectified Linear Unit function (ReLU) is a popular activation function (Agarap, 2019) designed for this purpose and is defined as follows:

$$\mathrm{ReLU}(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \tag{2}$$

Thus, the MLP must have a ReLU activation function between each hidden layer, giving the final model structure as laid out in Figure 1. Obtaining a $\hat{y}$ prediction from an MLP completes the "forward pass" part of training. The prediction is then evaluated against the target value y (also called the ground truth value) using a cost function L – see Equations 8 - 11 for how it is defined for this application. The partial derivative of L is calculated with respect to each parameter $w_i^{n,m}$ in W_i:

$$\frac{\partial L}{\partial w_i^{n,m}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w_i^{n,m}} \tag{3}$$

Computing all such partials completes the "backwards" pass of the training step – a process known as backpropagation – and gives us $\frac{\partial L}{\partial W_i}$, which tells the model how much, and the direction in which W should change in order to minimise L. Repeating these forward and backwards passes over the MLP over and over again brings the weights closer to their "ideal" values – ones for which L is minimised. In other words, the optimsing algorithm uses the difference between the output of the model and its target to update W in the correct direction. This is what is meant by training a machine learning model.

1.1. Previous work

The application of machine learning systems to the prediction of specific capacitance values is varied – some previous investigations involve predicting specific capacitance directly from experimental conditions (Kumar Yogesh et al., 2025), while others use these experimental conditions to predict hysteresis curves (Ravichandran et al., 2024); the

*Corresponding author

msandoval2@bournemouth.ac.uk (M.A. Sandoval-Riofrio); kjayathunge@bournemouth.ac.uk (K. Jayathunge)

https://www.linkedin.com/in/maria-a-sandoval/ (M.A. Sandoval-Riofrio); https://staffprofiles.bournemouth.ac.uk/display/kjayathunge (K. Jayathunge)

ORCID(s): 0000-0003-0680-4539 (M.A. Sandoval-Riofrio); 0009-0006-5162-5731 (K. Jayathunge)

[1] Department of Engineering

[2] National Centre for Computer Animation

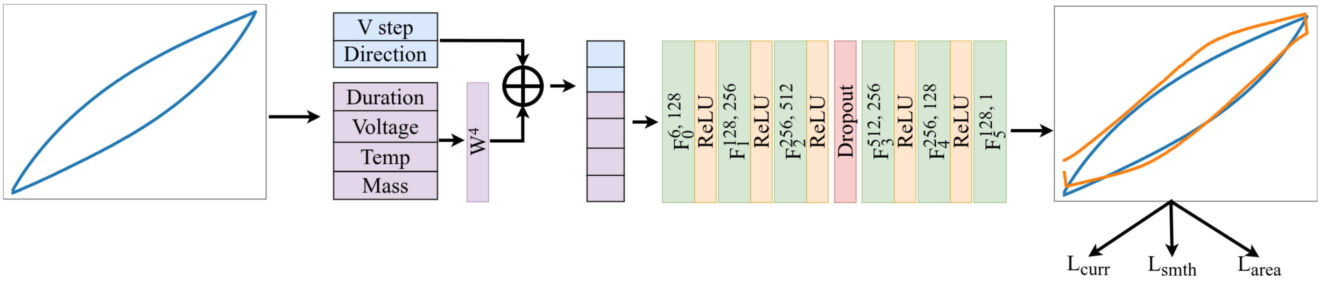

Figure 1: An overview of the training pipeline and model architecture, incorporating dropout after the largest layer to discourage overfitting. The input is a 6-dimensional vector consisting of a potential step, scan direction, and several experimental conditions under which the electrode material was sysnthesised; the output is a prediction of the current value at that potential. Four trainable weights W^4 provide a scaling factor for each of the conditions. Ground truth hysteresis curves shown in blue and model prediction shown in orange.

latter approach allows the indirect calculation of specific capacitance C_{sp}, which is proportional to the area between charge and discharge curves. It is this latter approach that is used by the present study, because predicting the full charge-discharge dynamics of cyclic voltammetry allows the model to generalise better to a wider range of conditions (Deebansok et al., 2024) – especially important given the limited data available for this study. Using this approach, C_{sp} is given by:

$$C_{sp} = \frac{\int I_{pred}(E)\,dE}{2 \cdot v \cdot m \cdot \Delta V} \tag{4}$$

where I_{pred} is the function that is approximated by the MLP, E is the potential, v is the scanrate, m is the total mass of the active material in the electrodes, and ΔV is the change in voltage over the whole charge/discharge cycle.

1.2. Dataset

The complete dataset is made up of current samples from 5 cyclic voltammetry cycles – see Table 1. These were collected via potentiostat with a scanrate of $100\,\mathrm{mV} \cdot \mathrm{s}^{-1}$. Each cycle consists of 264 rows, with columns for current (I / A), voltage step (E / V), active electrode mass (M / mg), temperature (T / °C), electrolysis duration (D / h), electrolysis voltage (V_e / V), and scan direction (S). The last feature tells the model if the particular point is on the charge or discharge curve and is represented by either +1 (charging) or -1 (discharging).

Notice that the various inputs span many orders of magnitude – this is a problem for machine learning models. Using raw features such as those presented in Table 1 mean that ones with large absolute values dominate the weights of the model, even though they may not be as important, and vice versa, which can lead to lower performance (Kim et al., 2025). For this reason, all features are scaled across all cycles such that they have zero mean and unit standard deviation. To transform an individual instance of a feature d_i (e.g. electrode mass, temperature) to its scaled value $d_{norm,i}$:

$$d_{norm,i} = \frac{d_i - \mu(D)}{\sigma(D)} \tag{5}$$

Table 1
Characteristics of each cycle that constitutes the dataset used for training our model. The scanrate was $100\,\mathrm{mV} \cdot \mathrm{s}^{-1}$ in all cases.

Cycle	M / mg	T / °C	D / h	V_e / V	Electrolyte
BCMS2	0.11	800	2	2	Ethaline
BCMS3	0.11	750	2	2	Ethaline
BCMS5	0.18	650	2	2	Ethaline
BCMS7	0.25	700	2	2.8	Ethaline
BCMS9	0.25	750	2	2.8	Ethaline

where D is the collection of all the values for this feature, μ is the mean of values in D, and σ is their standard deviation. These scaled values are used in all subsequent machine learning operations, and final outputs are de-normalised using the inverse of Equation 5 for display and comparison with experimental values. Current values were also transformed in this way. However, the potential step E was rescaled slightly differently: rather than ensuring zero mean and unit variance, we rescale them such that the minimum potential across the whole dataset is -1 and the maximum is +1:

$$\begin{aligned} E_{mid} &= \frac{E_{max} + E_{min}}{2} \\ E_{half-range} &= \frac{E_{max} - E_{min}}{2} \\ E_{norm} &= \frac{E - E_{mid}}{E_{half-range}} \end{aligned} \tag{6}$$

where E_{min} and E_{max} are the minimum and maximum potential values *across the whole dataset*. This scaling ensures that all cycles share a common and bounded potential domain, making it easier for the model to lear consistent relationships across different experimental conditions.

The input to the model is therefore a 6 dimensional, normalised feature vector that consists of a potential step and experimental conditions under which the electrode material was synthesised, and the output is the current value at the potential step.

1.3. Experimental design and model architecture

We employ leave-one-out cross validation when training: for any one run, the model is trained only on points from 4 of the cycles; points from the 5^{th} are held out for validation. Here, validation means only evaluating the forward pass of the model and not updating the weights of the model. It is a test to determine if the model is able to generalise to "unseen" data, and each cycle is subjected to this treatment in turn. We repeat training runs five times independently to build confidence that the model's output is reliable.

Due to the small size of the dataset, we introduce a simple heuristic to improve the training stability and to encourage the model to learn meaningful relationships between experimental conditions and the voltammetric response. This is realised by introducing four trainable parameters that act as scaling factors for the experimental conditions that are used when synthesising the black carbon – active mass (M), electrolysis temperature (T), electrolysis duration (D) and electrolysis voltage (V_e). These parameters allow the model to explicitly learn the relative importance of each condition rather than solely relying on implicit weighting within the first layer of the MLP.

As illustrated in Figure 1, the learned scaling parameters are applied to the condition vector prior to concatenation with the potential step (E) and scan direction (S), resulting in the 6-dimensional input that is expected by the MLP. Scaling is applied in the following way: the learned weights are first divided by a temperature parameter τ, which controls the sharpness of the relative weighting, and subsequently normalised by the softmax function to give the scaling factor s_i for each learned weight:

$$s_i = \frac{\exp(\alpha_i/\tau)}{\sum_{j=1}^{4} \exp(\alpha_j/\tau)} \tag{7}$$

where α_i is any one of the four learned weights. The resulting normalised weights are multiplied element-wise by the condition vector, yielding the weighted condition vector.

The rest of the model is a multilayered perceptron with 4 hidden layers in between a 6-neuron input layer and a 1-neuron output layer. Please refer to Figure 1 for the full model architecture. Note the dropout layer, which randomly (with some probability p) deactivates a subset of neurons during training. This is done so that the model does not rely too heavily on any single neuron or set of neurons, thereby reducing overfitting and improving generalisation to new data (Mou et al., 2018). Wider layers with more neurons benefit more from dropout as they are more at risk from overfitting (Srivastava et al., 2014), hence the placement between F_2 and F_3.

As mentioned in the Section 1, the objective function scores the output of the MLP so it can be used in gradient calculations, which in turn are used to inform updates to the model's weights. Because the model predicts the current at a particular voltage step, part of the loss function is the mean squared error (MSE) between the predicted and actual value. This is given by:

$$L_{curr} = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \tag{8}$$

where $\hat{y}_i$ is the predicted current value, y_i is the true value, and N is the number of predicted points. Further, in order to maintain a smooth curve that is physically consistent with the behaviour of real hysteresis curves, we impose a penalty on the second derivative of the predicted current with respect to potential, discouraging sharp local curvature and jitter:

$$L_{smth} = \left\| \frac{\partial^2 I_{pred}}{\partial E^2} \right\|_2^2 \tag{9}$$

where I_{pred} is the predicted current and E is the potential. Finally, we also use the area of the predicted curve in another MSE loss calculation to encourage the model to produce the correct net area exhibited by the experimental curve:

$$L_{area} = \frac{1}{M} \sum_{j=1}^{M} (\hat{a}_j - a_j)^2 \tag{10}$$

where M is the number of cycles, a_j is the actual area between the curves for a given cycle, and $\hat{a}_j$ is the area calculated using predicted current values. The losses are combined for final objective for the MLP:

$$L = L_{curr} + L_{smth} + L_{area} \tag{11}$$

In order to assess the importance of these loss functions (as well as other model settings), an ablation study was conducted to measure the impact these settings on the final prediction. A comparison of C_{sp} predicted by models under these experimental settings is given in Figure 2, and a comparison of their relative errors is given in Table 2. The base MLP, henceforth called "Base" is only tasked with the L_{curr} objective. The following experimental modifications were added to the Base model:

1. "+ D" – inclusion of a **D**ropout layer, as discussed in Section 1.3
2. "+ S" – addition of the **S**moothing L_{smth} objective to L
3. "+ A" – addition of the **A**rea L_{area} objective to L
4. "+ W" – inclusion of trainable **W**eights for each of the synthesis conditions

1.4. Results and discussion

Results were collected for each experimental setting described in Section 1.3. Figure 2 reports the averaged, normalised C_{sp} values computed over five independent

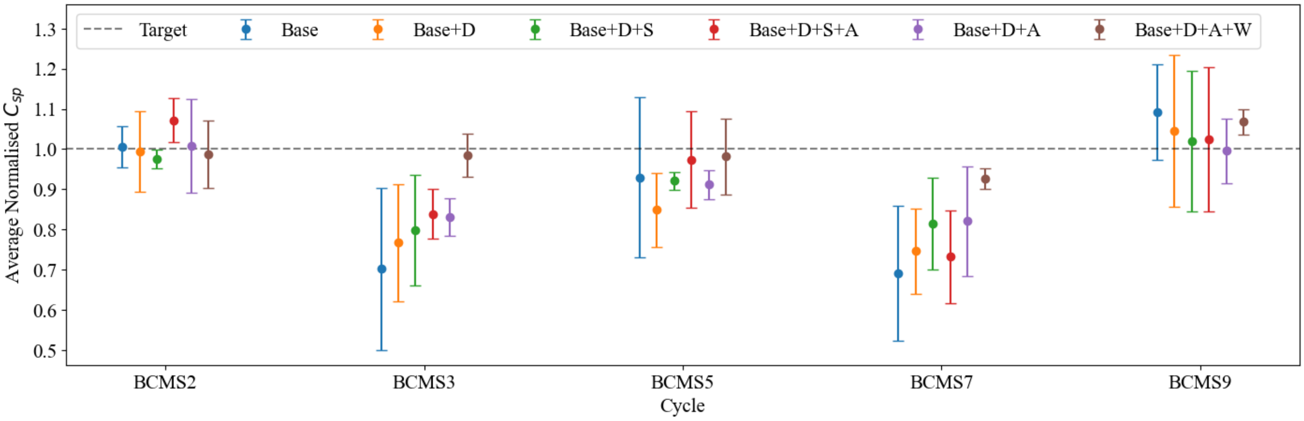

Figure 2: A plot reporting the average normalised C_{sp} over 5 runs per cycle per experimental setting. Error bars indicate 95% confidence intervals over runs.

Table 2
Average error (relative %) for each cycle in leave-one-out cross validation over 5 independent runs. Base+D+A exhibits the least overall error across all the settings.

Cycle	Base	Base +D	Base +D +S	Base +D +S +A	Base +D +A	Base +D +A +W
BCMS2	0.62	0.58	2.54	7.17	0.77	1.22
BCMS3	29.82	23.28	20.15	16.11	16.93	1.53
BCMS5	7.04	15.05	7.88	2.61	8.83	1.77
BCMS7	30.87	25.37	18.50	26.78	17.92	7.42
BCMS9	9.12	4.64	1.91	2.42	0.39	6.83
Average	15.49	13.79	10.20	11.02	8.97	**3.75**

Table 3
Relative weights assigned by the proposed model to the various molten salt parameters. The exclusion of one curve from the training gives a slightly different set of weights for each experiment, but overall, the model was most sensitive to temperature (T), electrolysis voltage (V_e) and active mass (M). Electrolysis duration (D) was found to be the least important input in predicting C_{sp}.

Experiment	M	T	D	V_e
BCMS2	0.193	0.294	0.173	0.339
BCMS3	0.266	0.264	0.237	0.233
BCMS5	0.254	0.285	0.175	0.286
BCMS7	0.223	0.301	0.202	0.275
BCMS9	0.214	0.413	0.170	0.203
Average	0.230	0.311	0.191	0.267

runs. Predicted C_{sp} were normalised by their corresponding ground-truth values to facilitate comparison across settings, as C_{sp} spans several orders of magnitude. Error bars indicate the 95% confidence interval over the five runs.

Overall, the Base+D+A+W configuration (corresponding to the model incorporating dropout, condition scaling, and the area loss of Equation 10) exhibits the samllest deviation from the C_{sp} baseline, on average. This setting also yields comparatively narrower confidence intervals, suggesting improved stability across runs. Incorporating the area-based loss term resulted in improved performance across all curves. As shown in Figure 2, models that include this loss term tend to produce predictions that are closer to the target baseline. It should be noted that the smoothness loss term (Equation 9) seems to be somewhat at odds with the area loss. This effect is reflected in Table 2, which presents the relative error percentage between the predicted and true C_{sp} values across all the experimental settings. The addition of area loss to Base+D+S (10.2% error) results in the error increasing slightly to 11.02%. Removing the smoothness loss (corresponding to the Base+D+A configuration) yields a lower overall error (8.97%). This suggests that the two loss terms may impose competing constraints on the learned representations, which could partially limit their combined effectiveness. However, the most effective strategy incorporates condition scaling to the latter configuration (giving Base+D+A+W), and yielding an average error of 3.75%.

The inclusion of trainable scaling values for the four experimental conditions (M, T, V and D) offers insights into the relative importance of each of these in determining C_{sp}. Table 3 demonstrates that on average, the model assigns most importance (31.1% relative weight) to the temperature of the molten salt (T). This is followed by electrolysis voltage (V_e) at 26.7%, and active material mass (M) at 23.0%. Finally, duration of electrolysis (D) recieves the lowest relative weight (19.1%). This is unsurprising, as this setting did not vary at all across the dataset (see Table 1), so could not explain any of the variance in the predicted output.

A key limitation of the experimental setup lies in the limited diversity of the input features. Although each of the five cycles contains 264 samples, all samples within a given cycle share the same experimental conditions; the only varying factor is the voltage step. Moreover, the voltage progression itself is identical across all cycles. Thus, even though a unique feature vector exists for each point, the dataset comprises only 5 unique condition vectors across 1320 samples. This extremely low condition variance substantially constrains the expressive capacity available to the model and limits the generalisability of the learned representations. In this context, the fact that the model exhibits meaningful performance at all is somewhat surprising, and underscores the need for more diverse experimental conditions (i.e., a wider range of these conditions) in future studies.

2. Conclusion

The conclusions section should come in this section at the end of the article, before the Author contributions statement and/or Conflicts of interest statement.

References

Agarap, A.F., 2019. Deep Learning using Rectified Linear Units (ReLU). doi:10.48550/arXiv.1803.08375, arXiv:1803.08375.

Deebansok, S., Deng, J., Le Calvez, E., Zhu, Y., Crosnier, O., Brousse, T., Fontaine, O., 2024. Capacitive tendency concept alongside supervised machine-learning toward classifying electrochemical behavior of battery and pseudocapacitor materials. Nature Communications 15, 1133. doi:10.1038/s41467-024-45394-w.

Geuchen, P., Jahn, T., Matt, H., 2025. Universal approximation with complex-valued deep narrow neural networks. Constructive Approximation 62, 361–402. doi:10.1007/s00365-025-09713-8.

Kim, Y.S., Kim, M.K., Fu, N., Liu, J., Wang, J., Srebric, J., 2025. Investigating the impact of data normalization methods on predicting electricity consumption in a building using different artificial neural network models. Sustainable Cities and Society 118, 105570. doi:10.1016/j.scs.2024.105570.

Kumar Yogesh, G., Nandi, D., Yeetsorn, R., Wanchan, W., Devi, C., Pratap Singh, R., Vasistha, A., Kumar, M., Koinkar, P., Yadav, K., 2025. A machine learning approach for estimating supercapacitor performance of graphene oxide nano-ring based electrode materials. Energy Advances 4, 119–139. doi:10.1039/D4YA00577E.

Mou, W., Zhou, Y., Gao, J., Wang, L., 2018. Dropout Training, Data-dependent Regularization, and Generalization Bounds, in: Proceedings of the 35th International Conference on Machine Learning, PMLR. pp. 3645–3653.

Ravichandran, A., Raman, V., Selvaraj, Y., Mohanraj, P., Kuzhandaivel, H., 2024. Machine Learning-Based Prediction of Cyclic Voltammetry Behavior of Substitution of Zinc and Cobalt in BiFeO3/Bi25FeO40 for Supercapacitor Applications. ACS Omega 9, 33459–33470. doi:10.1021/acsomega.3c10485.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 15, 1929–1958.