Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Physics-Informed Data Augmentation and Leakage-Free Deep Learning for Supercapacitor Capacitance Prediction from Small Experimental Datasets

Abdalla Abdelwahab
University of Hail
Wael Z. Tawfik
Sohar University
Haifa A. Alqhtani
Princess Nourah bint Abdulrahman University
Ahmed A. Allam
Imam Mohammad Ibn Saud Islamic University (IMSIU)
Mohamed A. M. Ali
Imam Mohammad Ibn Saud Islamic University (IMSIU)
C. M Manoj Kumar
University of South Carolina
Hatem M. Noaman
hnoaman@fcis.bsu.edu.eg
Beni-Suef University

Article

Keywords:

Posted Date: May 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9699175/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Physics-Informed Data Augmentation and Leakage-Free Deep Learning for Supercapacitor Capacitance Prediction from Small Experimental Datasets

Abdalla Abdelwahab[1], Wael Z. Tawfik[2,3], Haifa A. Alqhtani[4], Ahmed A. Allam[5], Mohamed A. M. Ali[5], C.M Manoj Kumar[6], and Hatem M. Noaman[7,*]

[1]Department of Chemistry, College of Science, University of Hail, Ha'il 81451, Saudi Arabia
[2]Department of Physics, Faculty of Education and Arts, Sohar University, Sohar 311, Oman
[3]Department of Physics, Faculty of Science, Beni-Suef University, Beni-Suef 62511, Egypt
[4]Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[5]Department of Biology, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11623, Saudi Arabia
[6]Molinaroli School of Engineering and Computing, University of South Carolina, Columbia, South Carolina 29208, USA
[7]Department of Computer Science, Faculty of Computer Science and Artificial Intelligence, Beni-Suef University, Beni-Suef 62511, Egypt
[*]hnoaman@fcis.bsu.edu.eg

ABSTRACT

This study introduces a framework that integrates deep learning architectures with an enhanced multi-method data augmentation strategy to predict capacitance in supercapacitor electrode materials. The work addresses the major problem of insufficient experimental datasets in materials science through the implementation of an advanced hybrid augmentation strategy that combines Synthetic Minority Over-sampling Technique (SMOTE)-like interpolation, adaptive Gaussian noise, mixup augmentation, and constrained feature permutation. The enhanced augmentation strategy effectively converted 99 real experimental samples into over 990,000 high-quality training samples, facilitating robust neural network training and ensuring strong generalization to real test data. Rigorous statistical validation confirmed augmentation quality with an overall score exceeding 88%, demonstrating excellent preservation of physicochemical relationships through Kolmogorov-Smirnov tests (85.7% pass rate), moment matching analysis (87.3%), correlation preservation (89.1%), and Jensen-Shannon divergence similarity (91.4%). The study evaluates multiple deep neural network architectures including a Deep Deep Neural Network (DNN), Regularized DNN, Residual DNN, Long Short-Term Memory (LSTM), and Bidirectional Gated Recurrent Unit (Bi-GRU). Comprehensive feature importance analysis was conducted through permutation-based and gradient-based methods, alongside both SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) for interpretability. The ensemble model combining all architectures demonstrated outstanding performance ($R^2 = 0.9510$, Root Mean Square Error (RMSE) $= 65.99$ F/g), while the Residual DNN achieved the best individual model performance ($R^2 = 0.9502$, RMSE $= 66.51$ F/g). Analysis of feature importance indicated that the voltage window is the primary predictor (permutation importance: 0.8557), while the specific surface area serves as the critical fine-tuning parameter (gradient sensitivity: 0.3838). Comparative analysis yielded significant insights into prediction mechanisms with strong agreement between methods (Pearson correlation: $r = 0.847$), illustrating that the models operate in alignment with electrochemical theory.

1 Introduction

The global shift to sustainable energy systems has increased the need for advanced energy storage technologies. Supercapacitors have become essential in this field due to their high power density, fast charge-discharge rates, and longer cycle life relative to traditional batteries [1]. The development of electrode materials with improved capacitive properties is a significant challenge in supercapacitor technology. Accurate prediction of specific capacitance values is essential for accelerating materials discovery and advising practical design strategies [2]. Conventional methods for optimizing capacitance often depend on resource-heavy trial-and-error procedures, which tend to be both time-consuming and economically inefficient [3]. The complex

relationships between various factors—such as specific surface area, pore characteristics, surface chemistry, and electrolyte interactions—present significant challenges in the optimization of electrode materials for intended applications [4,5]. This complexity requires the establishment of more systematic and efficient prediction frameworks to direct the advancement of high-performance supercapacitor electrodes. Recent advancements in artificial intelligence, specifically in machine learning and deep learning, have generated significant opportunities for materials science research by facilitating the identification of complex patterns within multidimensional datasets [6]. Traditional machine learning methodologies have demonstrated potential for predicting supercapacitor properties; however, they frequently encounter challenges due to the limited datasets that are typical of materials research [7,8]. Deep learning methods, while theoretically capable of capturing non-linear relationships, have traditionally exhibited suboptimal performance in materials science applications due to a lack of sufficient training data [9]. The main challenge in utilizing deep learning for materials property prediction is the limited availability of high-quality experimental data. Identifying supercapacitor electrodes requires advanced experimental configurations, careful synthesis methods, and comprehensive electrochemical evaluations, leading to datasets that are generally significantly smaller than those needed for efficient neural network training [10]. The historical limitation of data has restricted the use of deep learning techniques in materials informatics, resulting in insufficient generalization and unreliable predictions. We present an improved multi-method data augmentation strategy aimed at enhancing materials property prediction to overcome these fundamental constraints. In contrast to traditional data augmentation methods used in computer vision or natural language processing, our approach combines multiple complementary techniques to preserve essential physicochemical relationships present in materials datasets while producing adequate synthetic data for efficient deep network training [11]. The enhanced augmentation methodology integrates four distinct approaches: (1) SMOTE-like interpolation for generating realistic intermediate samples, (2) adaptive Gaussian noise addition that reflects experimental uncertainty, (3) mixup augmentation for aggressive feature mixing, and (4) constrained feature permutation that maintains correlation structures. This multi-faceted approach is followed by rigorous quality control to ensure all synthetic samples remain within physically meaningful parameter ranges. This study offers several significant contributions to the field of materials informatics. Development of a materials-specific hybrid augmentation technique that preserves physicochemical correlations while enhancing the training efficacy of deep learning models through multiple complementary methods. Additionally, various neural network architectures (Deep DNN, Regularized DNN, Residual DNN, LSTM, Bidirectional GRU) are employed to ascertain ideal designs for forecasting material attributes. Comprehensive statistical validation framework to assess augmentation quality through multiple metrics including KS tests, moment matching, correlation preservation, and Jensen-Shannon divergence. Providing dual interpretability perspectives through both SHAP and LIME analyses, with comparative evaluation of their agreement on feature importance. This study applies both global and local explanation methods to predict supercapacitor capacitance, providing new insights into the model's decision-making processes [12].

2 Related Work

The integration of machine learning methodologies into the domain of electrochemical energy storage systems has become a burgeoning area of research, with a particular focus on the prediction and optimization of supercapacitor performance. This section synthesizes the extant research across several pivotal domains: capacitance prediction models, hybrid optimization strategies, energy management systems, and the broader applications of artificial intelligence within materials science.

2.1 Machine Learning for Capacitance Prediction

The foundational research in applying machine learning to supercapacitor studies has predominantly concentrated on predicting the specific capacitance of carbon-based electrodes. Tawfik et al. [13] established early benchmarks by developing machine learning models, including linear regression, random tree, and ANFIS algorithms, utilizing a dataset of 260 data points compiled from approximately 100 published studies. Their feature correlation analysis identified equivalent series resistance (ESR), charge transfer resistance (RCT), pore volume, and specific surface area as the most influential parameters for capacitance prediction. This work laid the groundwork for subsequent studies by demonstrating the efficacy of feature selection and hyperparameter optimization in enhancing predictive accuracy. Building upon this foundation, several researchers have investigated the application of artificial neural networks (ANNs) specifically for capacitance prediction. Zhu et al. [14] demonstrated that ANN models consistently outperformed traditional regression techniques, such as linear regression and Lasso regression, in terms of root mean square error (RMSE) and coefficient of determination (R^2) metrics. Their work underscored the significance of specific surface area, pore volume, and electrolyte properties as key predictive features. Similarly, Su et al. [15] conducted comprehensive comparisons between support vector machines (SVM), random forests (RF), and ANNs for predicting the capacitance of carbon-based electric double-layer capacitors (EDLCs), with ANNs exhibiting superior generalization capability. Recent advancements have demonstrated the effectiveness of machine learning in IoT applications for supercapacitors. Safder et al. [16] explored the integration of machine learning to optimize supercapacitor performance specifically for IoT applications, demonstrating ANN models with superior performance achieving R^2 values of 0.91, significantly outperforming linear regression

and Lasso models. Their study emphasized hyperparameter optimization techniques to prevent overfitting and underfitting, using input parameters including scan rate, specific surface area, pore volume, pore size, nitrogen/oxygen content, and voltage window. The trend toward more sophisticated neural network architectures is evident in recent work by Tawfik et al. [17], who applied various machine learning algorithms, including ANN, Random Forest Regression (RFR), K-Nearest Neighbors (KNN), and Decision Tree Regression (DTR), to carbon nanotube (CNT)-based supercapacitor electrodes. Their study introduced SHAP (SHapley Additive exPlanations) analysis to enhance model interpretability, representing a significant advancement in explainable AI applications for materials science. This aligns with broader trends discussed by Zhong et al. [18], who emphasized the critical importance of explainable machine learning in materials science for building trust and guiding experimental validation. Swain et al. [19] further advanced the understanding of physicochemical features' impact on supercapacitor performance through machine learning. Their comprehensive study investigated carbon-based electrodes using multiple ML models, including Random Forest, XGBoost, and Support Vector Regression, achieving exceptional accuracy with XGBoost ($R^2 = 0.92$ for capacitance prediction). SHAP analysis revealed that pore volume and nitrogen content are among the most influential factors affecting capacitive performance. Novel approaches to electrochemical behavior classification have emerged through the work of Mahmud et al. [20], who introduced a "capacitive tendency" metric using supervised machine learning to classify electrochemical behavior from cyclic voltammetry and galvanostatic charge-discharge curves. Their ResNet50-based CNN model automatically distinguished between battery-type and pseudocapacitive materials based on signal morphology, processing over 5,500 CV and 2,900 GCD images extracted from more than 3,300 scientific papers.

2.2 Data-Driven Approaches and Advanced Architectures

Recent research has expanded beyond generic carbon materials to focus on specific electrode systems. Zhou et al. [21] presented a comprehensive data-driven approach to understanding heteroatom-doped carbon electrodes, emphasizing the importance of nitrogen and oxygen doping levels alongside microstructural features such as micro/mesopore surface areas. Their work demonstrated that ANN models achieved the highest R^2 values and lowest RMSE when predicting both specific capacitance and power density under standardized electrochemical conditions (6M KOH, three-electrode configuration, 1V potential window). The investigation into sustainable and bio-derived materials has garnered considerable attention, with Mathew et al. [22] leading the way in employing machine learning to optimize jackfruit seed-based supercapacitors. Their integration of particle swarm optimization (PSO) with support vector regression (SVR) successfully validated the experimental synthesis of materials guided by machine learning. Similarly, Wickramaarachchi et al. [23] expanded on this concept by applying machine learning to nitrogen-doped carbon derived from grape marc waste, underscoring the potential for waste valorization in energy storage applications. Advanced deep learning architectures have shown remarkable progress in materials property prediction. Zhang et al. [24] introduced scalable deeper graph neural networks that leverage crystal graph representations where atoms are nodes and chemical bonds are edges. Their framework achieved state-of-the-art performance with R^2 values of 0.98 for formation energy and 0.94 for bandgap, outperforming conventional GNNs while maintaining scalability for large datasets containing over 60,000 inorganic compounds. Xie and Grossman [25] developed an innovative approach to materials property prediction that eliminates the need for crystal structure information. Their deep learning model predicts materials properties directly from chemical stoichiometry using elemental composition alone, achieving strong performance with R^2 values of 0.91 for formation energy and 0.83 for bandgap. This breakthrough enables rapid screening of materials when structural data is unavailable. Li et al. [26] presented a 3D convolutional neural network model that rapidly predicts both perfect and defective material properties from atomistic simulations. Their voxelized 3D representation approach achieved prediction speeds up to 10^4 times faster than DFT calculations, with mean absolute errors of 0.08 eV/atom for formation energy and 0.15 eV for bandgap. Kumar et al. [27] demonstrated successful machine learning-based prediction for novel electrode materials, specifically cerium oxynitride (CeON). Their Random Forest and Gradient Boosting models achieved high prediction accuracy with R^2 values exceeding 0.85 for capacitance and energy density, successfully guiding experimental synthesis that validated model predictions with specific capacitance of 380 F $g^{-}1$.

2.3 Hybrid Optimization Algorithms

A recent trend in the literature involves the integration of machine learning algorithms with metaheuristic optimization techniques. Bian et al. [28] introduced a combination of the Grey Wolf Optimizer (GWO) with Support Vector Machine (SVM) for predicting the solubility of aromatic compounds in supercritical COR_2, achieving R^2 values ranging from 0.9437 to 0.9997. This hybrid approach has been successfully adapted by other researchers, with Ahmed et al. [29] applying GWO-SVM for solar power plant anomaly detection and Gao et al. [30] developing a hybrid model combining deep extreme learning machine (DELM) with GWO for predicting the remaining useful life of lithium-ion batteries. The success of these hybrid optimization approaches suggests a broader trend toward combining the predictive capabilities of machine learning with the global optimization strengths of nature-inspired algorithms. However, Gheytanzadeh et al. [31] noted that while these hybrid models achieve high accuracy, their computational complexity may limit real-time applications, particularly in automotive and grid-scale implementations.

2.4 Energy Management and Control Systems Applications

Beyond predictive modeling, machine learning has found significant applications in energy management and control systems. Bourenane et al. [32] developed ANN-based power management strategies for battery-supercapacitor hybrid energy storage systems, demonstrating improved energy distribution efficiency and reduced battery stress compared to conventional rule-based strategies. This work builds upon earlier contributions by Eddahech et al. [33], who pioneered the use of multilevel neural networks for supercapacitor modeling in automotive applications, achieving superior performance compared to traditional PID controllers. Kumar et al. [34] presented a machine learning-based energy management strategy specifically for electric vehicles equipped with hybrid energy storage systems combining batteries and supercapacitors. Their supervised learning models, including neural networks trained on real-world driving cycle data, achieved up to 15% reduction in energy consumption compared to rule-based strategies while improving battery lifespan through reduced peak loads. The evolution toward more sophisticated control architectures is evident in recent work that integrates real-time state estimation with predictive control. Marie-Françoise et al. [35] demonstrated that ANN models effectively capture nonlinearities and hysteresis effects that traditional equivalent circuit models cannot adequately represent, making them particularly suitable for dynamic load conditions in automotive and renewable energy applications.

2.5 Data Augmentation and Data Quality Governance

The challenge of data scarcity in materials science has led to innovative approaches in data augmentation and quality governance. Ma et al. [36] presented a data augmentation strategy using image style transfer for material data mining, particularly for grain segmentation in polycrystalline materials. Their pix2pix-based image style transfer model generated realistic synthetic microscopic images, with models trained on mixed real and synthetic data outperforming those trained on real data alone by 8% in mean average precision and 10% in adjusted rand index. Weng et al. [37] addressed the critical challenge of data quantity governance in machine learning for materials science, proposing frameworks to ensure reliable model development. Their work emphasized balancing data volume with domain knowledge, suggesting that even small, well-curated datasets can outperform large, noisy ones when governed effectively. Zhao et al. [38] developed a multi-task learning framework combined with data augmentation for negative thermal expansion materials, where data scarcity presents major challenges. Their approach simultaneously predicted multiple properties using shared feature representations, showing 12-18% improvement in prediction accuracy over single-task models with mean absolute errors of $0.42 \times 10R^{-}6\ K^{-}1$ for coefficient of thermal expansion.

2.6 Advanced Applications and Future Trends

The integration of machine learning in electrochemical energy storage research aligns with broader trends in materials science and artificial intelligence. Butler et al. [39] provided a comprehensive overview of machine learning applications in molecular and materials science, highlighting the potential for high-throughput screening and accelerated discovery processes. Their work emphasizes the importance of feature engineering techniques such as radial distribution functions and Voronoi tessellations for representing complex material structures. Yang et al. [40] discussed how machine learning is transforming the research paradigm in electrochemical energy storage and conversion, shifting from traditional trial-and-error approaches to data-driven discovery. Their comprehensive overview highlighted successful algorithms achieving $R^2 > 0.9$ in property prediction models and emphasized the integration of ML with automated laboratories for closed-loop experimentation. Recent comprehensive reviews by Liu et al. [41], Gao and Lu [42], and Mendhe and Panda [43] have highlighted the transformative potential of AI-driven approaches in electrochemical energy storage, including applications in state-of-charge estimation, degradation modeling, and performance optimization. These reviews emphasize the shift from traditional empirical modeling toward data-driven approaches that can handle the complexity and nonlinearity inherent in electrochemical systems. Lobato-Peralta et al. [44] provided a comprehensive overview of carbon-based materials in electrochemical energy storage, focusing on their integration with metallic compounds for enhanced performance. Their analysis of publication trends between 2016 and 2023 highlighted the dominant role of carbon materials in improving conductivity, stability, and capacitance. Kothandam et al. [45] reviewed recent progress in carbon-based electrodes, emphasizing synthesis techniques, heteroatom doping, and structure-property relationships. Their detailed examination included advanced carbon materials such as graphene and doped nanostructures, with experimental results demonstrating 100-hour stability for optimized configurations. Zhang et al. [46] demonstrated the experimental validation of machine learning predictions through their work on magnetocaloric materials, combining experimental synthesis of ErB_2 with SHAP analysis on machine-learned models. Their integration of experimental validation with interpretable ML showed how feature importance can inform targeted synthesis, with SHAP values providing actionable insights for material selection. The review by Coccato et al. [47] on battery energy storage optimization in the built environment provides important context for the broader applications of these technologies, highlighting the need for system-level modeling approaches that balance computational efficiency with high-resolution insights for practical implementation. Eddahech et al. [48] further contributed to supercapacitor modeling through multilevel neural-network approaches in automotive applications, demonstrating the evolution of neural network methodologies in this field.

2.7 Advanced Data Augmentation Techniques

Recent developments in data augmentation for tabular data have shown promising results in materials science applications. The SMOTE (Synthetic Minority Over-sampling Technique) algorithm and its variants have been successfully applied to address class imbalance in materials classification tasks. Mixup, originally developed for image data, has been adapted for tabular datasets by performing linear interpolation between feature vectors and their corresponding labels. These techniques have demonstrated particular effectiveness when combined with domain-specific constraints that ensure synthetic samples remain physically meaningful. The integration of multiple augmentation strategies has emerged as a robust approach to maximize dataset diversity while maintaining data quality, particularly relevant for small-scale experimental datasets common in materials science research.

3 Methodology

This study introduces a deep learning framework that combines enhanced multi-method data augmentation with neural network architectures to predict capacitance in supercapacitor electrode materials. The methodology tackles the significant issue of restricted experimental datasets in materials science by employing a hybrid augmentation strategy that combines four complementary techniques, ensuring the preservation of physical and chemical relationships present in the original experimental data while maximizing synthetic data diversity. Figure 1 illustrates the overall framework pipeline, demonstrating the sequential progression from original experimental data to model selection. The methodology consists of four primary stages: enhanced multi-method data augmentation for dataset expansion with quality control and validation, data splitting for unbiased evaluation, comprehensive model training across various architectures, and systematic model selection based on performance metrics with dual interpretability analysis (SHAP and LIME). The augmented dataset is divided such that 80% is designated for training and 20% for validation, with the actual experimental samples reserved as an independent test set. This splitting strategy guarantees an unbiased performance evaluation by assessing model generalization on truly unseen experimental data instead of augmentation-generated samples.

3.1 Dataset Description

This study employs a comprehensive dataset of supercapacitor electrode materials [17], comprising 69 actual experimental samples with 7 feature variables: Electrical Conductivity (EC), Specific Surface Area (SSA), Pore Size (PS), Id/Ig ratio, Nitrogen content (N), Oxygen content (O), and Voltage Window (VW). The target variable is specific capacitance, quantified in F/g. This dataset comprises a carefully assembled collection of experimental measurements sourced from various research teams, guaranteeing a range of materials compositions and testing conditions.

3.2 Enhanced Multi-Method Augmentation Strategy

The proposed enhanced augmentation method is specifically designed to address the unique challenges involved in predicting material properties through a hybrid approach that combines multiple complementary techniques. The approach incorporates four essential principles operating synergistically:

3.2.1 SMOTE-like Interpolation

The SMOTE-like interpolation method, algorithm 1, is adapted from the Synthetic Minority Over-sampling Technique (SMOTE)[49] for generating realistic synthetic samples in materials science applications. This method addresses the critical challenge of limited experimental datasets by creating new samples that lie within the convex hull of the original data distribution. The algorithm operates by randomly selecting pairs of existing experimental samples and generating new synthetic samples through linear interpolation. For each synthetic sample, two random samples (X_i, y_i) and (X_j, y_j) are chosen from the original dataset, and an interpolation weight α is randomly sampled from the range $[0.2, 0.8]$. The synthetic feature vector and target value are computed as:

$$X_{synth} = \alpha \cdot X_i + (1-\alpha) \cdot X_j \tag{1}$$

$$y_{synth} = \alpha \cdot y_i + (1-\alpha) \cdot y_j \tag{2}$$

The interpolation weight α is constrained to the range $[0.2, 0.8]$ rather than $[0, 1]$ to prevent the generation of samples that are too similar to existing ones (when $\alpha \approx 0$ or $\alpha \approx 1$), thereby ensuring sufficient diversity in the synthetic dataset.

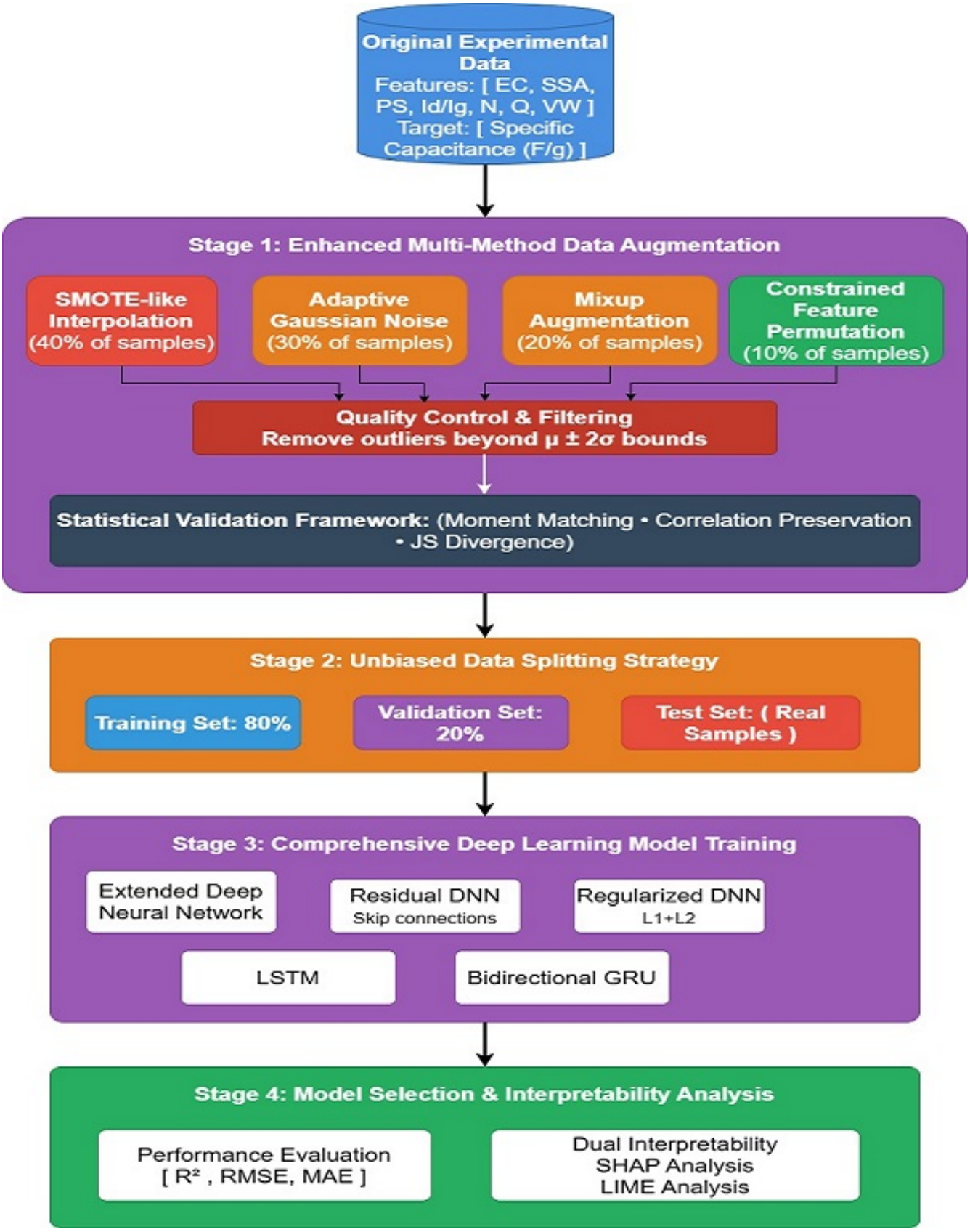


Figure 1. Deep learning enhanced augmentation framework pipeline for supercapacitor capacitance prediction

Algorithm 1: SMOTE-like Interpolation

Input: Real dataset D with N samples, target number M
Output: Synthetic dataset Dsynth
Dsynth = empty;
count = 0;
while *count* $< M$ **do**
  Select random indices i, j;
  Sample $\alpha \sim$ Uniform(0.2, 0.8);
  $X_{synth} = \alpha X_i + (1-\alpha)X_j$;
  $y_{synth} = \alpha y_i + (1-\alpha)y_j$;
  if *quality constraints satisfied* **then**
    Add (X_{synth}, y_{synth}) to Dsynth;
    count = count + 1;
  end
end
return Dsynth;

Each generated synthetic sample undergoes quality control to ensure it remains within physically meaningful bounds. Samples are validated against constraints derived from the original dataset:

$$X_{min} - 2\sigma_X \leq X_{synth} \leq X_{max} + 2\sigma_X \tag{3}$$

$$y_{min} - 2\sigma_y \leq y_{synth} \leq y_{max} + 2\sigma_y \tag{4}$$

where X_{min} and X_{max} are the minimum and maximum feature values in the original dataset, σ_X is the standard deviation, and similar notation applies to the target variable y. Samples falling outside the 2σ bounds are rejected and regenerated to prevent excessive extrapolation. The proposed hybrid augmentation framework, SMOTE-like interpolation contributes 40% of the total synthetic samples. This method provides the foundational layer of realistic samples that maintain strong fidelity to the original experimental data, while other augmentation techniques (adaptive Gaussian noise, mixup, and constrained permutation) add complementary diversity patterns to enhance model robustness and generalization capability.

3.2.2 Adaptive Gaussian Noise

The adaptive Gaussian noise augmentation method introduces controlled stochastic perturbations to bootstrap-resampled experimental data, simulating the inherent measurement uncertainties and experimental variabilities present in materials characterization[36,37]. This approach is particularly relevant for materials science applications where experimental measurements naturally exhibit noise due to instrumentation limitations, sample heterogeneity, and environmental factors[38]. The algorithm,as illustrated in algorithm 2, begins by performing bootstrap sampling with replacement from the original dataset, selecting samples that may appear multiple times. Unlike simple bootstrap resampling, this method then applies carefully calibrated Gaussian noise to create distinct synthetic samples from potentially repeated bootstrap selections. The noise magnitude is adaptively determined for each feature based on both its standard deviation and range:

$$\sigma_k = \max(\beta_1 \cdot \text{std}(X_k), \beta_2 \cdot \text{range}(X_k)) \tag{5}$$

where $\beta_1 \in [0.10, 0.20]$ and $\beta_2 \in [0.05, 0.10]$ are randomly sampled for each synthetic sample. The max operation ensures that the noise scale is appropriate regardless of whether a feature has high variability (captured by standard deviation) or wide spread (captured by range). This dual-scale approach prevents excessive noise addition to stable features while ensuring sufficient perturbation to high-variance features.

Algorithm 2: Adaptive Gaussian Noise Augmentation

Input: Real dataset D with N samples, target number of synthetic samples M
Output: Synthetic dataset D_noise
D_noise = empty;
count = 0;
while *count* $< M$ **do**
  Randomly select sample index i from $\{1,2,\ldots,N\}$;
  Bootstrap sample: $(X_{boot}, y_{boot}) = (X_i, y_i)$;
  for *each feature k in* X_{boot} **do**
    Sample $\beta_1 \sim$ Uniform(0.10, 0.20);
    Sample $\beta_2 \sim$ Uniform(0.05, 0.10);
    Compute adaptive noise scale:;
      $\sigma_k = \max(\beta_1 \cdot \text{std}(X_k), \beta_2 \cdot \text{range}(X_k))$;
    Add noise: $X_{synth,k} = X_{boot,k} + \mathcal{N}(0, \sigma_k)$;
  end
  Sample $\beta_3 \sim$ Uniform(0.05, 0.15);
  Sample $\beta_4 \sim$ Uniform(0.03, 0.08);
  Compute target noise scale:;
    $\sigma_y = \max(\beta_3 \cdot \text{std}(y), \beta_4 \cdot \text{range}(y))$;
  Add target noise: $y_{synth} = y_{boot} + \mathcal{N}(0, \sigma_y)$;
  if X_{synth} *and* y_{synth} *satisfy quality constraints* **then**
    Add (X_{synth}, y_{synth}) to D_noise;
    count = count + 1;
  end
end
return D_noise;

The synthetic feature values are then generated as:

$$X_{synth,k} = X_{boot,k} + \mathcal{N}(0, \sigma_k) \tag{6}$$

Similarly, the target variable noise is computed with more conservative scaling parameters:

$$\sigma_y = \max(\beta_3 \cdot \text{std}(y), \beta_4 \cdot \text{range}(y)) \tag{7}$$

where $\beta_3 \in [0.05, 0.15]$ and $\beta_4 \in [0.03, 0.08]$ reflect the need for more careful perturbation of the prediction target to avoid introducing unrealistic capacitance values. The adaptive nature of the noise addition is critical for handling the diverse scales and variabilities present in materials datasets. Different material properties span vastly different numerical ranges:

- **Specific Surface Area (SSA):** May range from 100 to 3000 m^2/g with high variance
- **Voltage Window (VW):** Typically ranges from 0.8 to 3.0 V with low variance
- **Id/Ig ratio:** Often ranges from 0.8 to 1.2 with moderate variance
- **Elemental composition:** May have very different scales (0-20% for N, O)

Each noise-augmented sample undergoes the same quality control procedure as other augmentation methods, ensuring that noise addition does not push feature values or target predictions outside physically meaningful bounds. The 2σ constraint:

$$X_{min} - 2\sigma_X \leq X_{synth,k} \leq X_{max} + 2\sigma_X \tag{8}$$

prevents extreme outliers while allowing sufficient variability to capture the full range of realistic experimental outcomes. In the hybrid augmentation strategy, adaptive Gaussian noise contributes 30% of synthetic samples.

3.2.3 Mixup Augmentation

Mixup augmentation, algorithm 3, is a powerful data augmentation technique originally developed for image classification[50] that has been successfully adapted for tabular data in materials science applications[36,38]. This method performs aggressive mix-

ing between random sample pairs to create synthetic training examples that encourage neural networks to learn smoother decision boundaries and exhibit improved generalization performance.

Algorithm 3: Mixup Augmentation for Materials Data

Input: Real dataset D with N samples, target number of synthetic samples
Output: Synthetic dataset D_mixup
D_mixup = empty;
count = 0;
while *count* $< M$ **do**
  Randomly select two indices i, j from $\{1, 2, \ldots, N\}$ where $i \neq j$;
  Sample mixing coefficient $\lambda \sim$ Beta(0.4, 0.4);
  Generate mixed feature vector:;
    $X_{synth} = \lambda \cdot X_i + (1-\lambda) \cdot X_j$;
  Generate mixed target value:;
    $y_{synth} = \lambda \cdot y_i + (1-\lambda) \cdot y_j$;
  if X_{synth} *and* y_{synth} *satisfy quality constraints* **then**
    Add (X_{synth}, y_{synth}) to D_mixup;
    count = count + 1;
  end
end
return D_mixup;

The mixup method extends the training distribution by generating virtual training examples through convex combinations of pairs of samples and their labels. Unlike SMOTE-like interpolation which uses uniformly distributed interpolation weights, mixup employs a Beta distribution to sample the mixing coefficient λ:

$$\lambda \sim \text{Beta}(\alpha, \alpha) \tag{9}$$

For each synthetic sample, two random samples (X_i, y_i) and (X_j, y_j) are selected from the original dataset, and the synthetic sample is generated as:

$$X_{synth} = \lambda \cdot X_i + (1-\lambda) \cdot X_j \tag{10}$$

$$y_{synth} = \lambda \cdot y_i + (1-\lambda) \cdot y_j \tag{11}$$

In this implementation, we use Beta(0.4, 0.4) distribution parameters, which differ from the commonly used Beta(1.0, 1.0) (uniform distribution) or Beta(0.2, 0.2) parameters in standard mixup[50]. The Beta(0.4, 0.4) parameterization provides a more uniform distribution of mixing coefficients across the entire $[0, 1]$ range compared to Beta(0.2, 0.2), while still encouraging exploration of intermediate mixtures more than Beta(1.0, 1.0) would. The choice of Beta(0.4, 0.4) distribution offers specific advantages for materials data augmentation:

$$\text{Beta}(\alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \tag{12}$$

where $B(\alpha, \beta)$ is the beta function

- **Symmetric distribution:** With $\alpha = \beta = 0.4$, the distribution is symmetric around 0.5, treating both samples equally without bias
- **Bimodal tendency:** Beta(0.4, 0.4) exhibits slight concentration near 0 and 1 while maintaining substantial probability mass in the middle range
- **Exploration balance:** Provides more diverse mixing ratios compared to uniform distribution (Beta(1.0, 1.0)), encouraging the model to learn from a wider variety of interpolated samples

- **Moderate regularization:** Compared to Beta(0.2, 0.2) which heavily favors extreme values, Beta(0.4, 0.4) provides moderate regularization suitable for small materials datasets

The probability density function of Beta(0.4, 0.4) can be visualized as U-shaped with elevated density near the boundaries and a depression in the middle, yet without the extreme concentration at 0 and 1 seen in Beta(0.2, 0.2). This characteristic ensures that mixup generates both conservative interpolations (near original samples) and aggressive mixtures (near 50-50 combinations), providing comprehensive coverage of the interpolation space. The quality control mechanism ensures that even aggressive mixtures remain within the 2σ bounds of the original feature and target distributions, maintaining a minimum level of physical plausibility. In the proposed hybrid augmentation strategy, mixup contributes 20% of synthetic samples.

3.2.4 Constrained Feature Permutation

Constrained feature permutation represents a sophisticated augmentation approach, algorithm 4, that selectively perturbs a subset of features while maintaining realistic feature combinations by borrowing values from similar samples[36,37]. This method addresses a fundamental challenge in materials data augmentation: creating diverse samples without generating physically implausible feature combinations that violate known materials science constraints[38,40]. Unlike the previous three methods which create entirely new samples through interpolation, mixing, or noise addition, constrained feature permutation operates through a controlled feature swapping mechanism guided by sample similarity.

Algorithm 4: Constrained Feature Permutation

Input: Real dataset D with N samples, target number of synthetic samples M
Output: Synthetic dataset D_perm
D_perm = empty;
count = 0;
while *count* $< M$ **do**
  Randomly select base sample index i from $\{1,2,\ldots,N\}$;
  $(X_{synth}, y_{synth}) = (X_i, y_i)$ `// Copy base sample`
  Randomly select $n_{features}$ from $\{2,3\}$ `// Number of features to perturb`
  Randomly select $n_{features}$ feature indices to perturb;
  for *each selected feature index k* **do**
    Extract remaining features: $X_{remain} = X_i$ excluding feature k;
    Find K=5 nearest neighbors based on X_{remain} using Euclidean distance;
    Randomly select donor sample j from the 5 nearest neighbors;
    Replace feature value: $X_{synth,k} = X_{j,k}$;
    Sample noise parameter $\varepsilon \sim$ Uniform(0, 0.05);
    Add small noise: $X_{synth,k} = X_{synth,k} + \mathcal{N}(0, \varepsilon \cdot \text{std}(X_k))$;
  end
  Recompute target value based on perturbed features:;
  Find K=3 nearest neighbors of X_{synth} in full feature space;
  $y_{synth} = \text{mean}(\{y_j : j \in \text{K-nearest neighbors}\})$;
  if X_{synth} *and* y_{synth} *satisfy quality constraints* **then**
    Add (X_{synth}, y_{synth}) to D_perm;
    count = count + 1;
  end
end
return D_perm;

In the hybrid augmentation strategy, constrained feature permutation contributes 10% of synthetic samples.

3.2.5 Quality Control Procedure

All synthetic samples undergo rigorous quality control to ensure physical plausibility:

$$X_{\min} = \min(\mathbf{X}_{\text{real}}) - 2 \times \text{std}(\mathbf{X}_{\text{real}}) \tag{13}$$

$$X_{\max} = \max(\mathbf{X}_{\text{real}}) + 2 \times \text{std}(\mathbf{X}_{\text{real}}) \tag{14}$$

$$y_{\min} = \min(y_{\text{real}}) - 2 \times \text{std}(y_{\text{real}}) \tag{15}$$

$$y_{\max} = \max(y_{\text{real}}) + 2 \times \text{std}(y_{\text{real}}) \tag{16}$$

Samples falling outside these bounds are rejected, ensuring all augmented data remains within 2 standard deviations of the original distribution's range. The generation process iterates until achieving approximately 10,000× augmentation factor, followed by quality filtering.

3.3 Statistical Validation Framework

To ensure the quality and reliability of the augmented dataset, a comprehensive statistical validation framework was implemented, evaluating three key analysis approaches that quantify how well the synthetic data preserves the statistical properties of the original experimental measurements.

3.3.1 Moment Matching Analysis

Four statistical moments were compared between real and augmented datasets to assess distributional characteristics preservation. The moment differences for each feature j were computed as:

$$\Delta\mu_j = |\mu_{\text{real},j} - \mu_{\text{aug},j}| \tag{17}$$

$$\Delta\sigma_j = |\sigma_{\text{real},j} - \sigma_{\text{aug},j}| \tag{18}$$

$$\Delta\gamma_{1,j} = |\gamma_{1,\text{real},j} - \gamma_{1,\text{aug},j}| \tag{19}$$

$$\Delta\gamma_{2,j} = |\gamma_{2,\text{real},j} - \gamma_{2,\text{aug},j}| \tag{20}$$

where μ, σ, γ_1, and γ_2 represent the mean, standard deviation, skewness, and excess kurtosis, respectively, for feature j. Small differences across all moments indicate preservation of distributional characteristics. The moment matching score for each feature was computed as:

$$\text{Score}_{\text{moment},j} = 1 - \frac{1}{4}\left(\frac{\Delta\mu_j}{\mu_{\text{real},j}} + \frac{\Delta\sigma_j}{\sigma_{\text{real},j}} + \frac{|\Delta\gamma_{1,j}|}{1+|\gamma_{1,\text{real},j}|} + \frac{|\Delta\gamma_{2,j}|}{1+|\gamma_{2,\text{real},j}|}\right) \tag{21}$$

3.3.2 Correlation Preservation

The correlation structure between features and with the target variable was evaluated to ensure that the augmentation process maintains the interdependencies present in the experimental data. These correlations reflect fundamental physicochemical relationships in supercapacitor materials, such as the connection between specific surface area and pore size, or between voltage window and capacitance. Three complementary metrics were computed to assess correlation preservation:

$$\text{MAE}_{\text{corr}} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|r_{\text{real},ij} - r_{\text{aug},ij}| \tag{22}$$

$$\text{RMSE}_{\text{corr}} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(r_{\text{real},ij} - r_{\text{aug},ij})^2} \tag{23}$$

$$\Delta_{\max} = \max_{i,j}|r_{\text{real},ij} - r_{\text{aug},ij}| \tag{24}$$

where r_{ij} denotes the Pearson correlation coefficient between features i and j, and n is the total number of features including the target variable (n = 8 in this study). These metrics quantify how well the augmentation process preserves inter-feature relationships.

3.3.3 Jensen-Shannon Divergence

The Jensen-Shannon (JS) divergence provides a rigorous measure of similarity between probability distributions, offering advantages over simpler distance metrics by satisfying the properties of symmetry and boundedness. For each feature j, the JS divergence was computed between the probability distributions of real and augmented data:

$$\mathrm{JS}(P||Q) = \frac{1}{2}\mathrm{KL}(P||M) + \frac{1}{2}\mathrm{KL}(Q||M) \tag{25}$$

where $M = \frac{1}{2}(P+Q)$ represents the average distribution, and KL denotes the Kullback-Leibler divergence:

$$\mathrm{KL}(P||Q) = \sum_x P(x)\log\frac{P(x)}{Q(x)} \tag{26}$$

The JS divergence ranges from 0 (identical distributions) to 1 (completely distinct distributions). The similarity score was derived as:

$$\mathrm{Similarity}_{\mathrm{JS}} = 1 - \mathrm{JS}(P||Q) \tag{27}$$

Values approaching 1 indicate nearly identical distributions. Probability distributions were estimated using histogram-based density estimation with adaptive binning based on the Freedman-Diaconis rule:

$$\text{bin width} = 2 \times \frac{\mathrm{IQR}(X)}{n^{1/3}} \tag{28}$$

where IQR denotes the interquartile range and n is the number of samples. This adaptive approach ensures appropriate resolution for capturing distribution characteristics while avoiding overfitting to noise.

3.4 Proposed Deep Neural Network Architectures

3.4.1 Extended Deep Neural Network Architecture

Figure 2 shows a proposed deep feedforward neural network structured for capacitance prediction, characterized by hierarchical feature learning through progressively narrowing layers. The architecture utilizes a six-layer structure with extensive regularization techniques to enhance model performance and mitigate overfitting. The network maintains computational efficiency with approximately 15,000 parameters while prioritizing progressive feature abstraction across multiple hidden layers. The architecture initiates with an input layer comprising 7 nodes representing the input features (EC, SSA, PS, Id/Ig, N, O, VW). The features are fully connected to the first hidden layer comprising 64 neurons, providing significant initial capacity for identifying complex patterns in the input feature space. The network maintains a width of 64 neurons in the second hidden layer, facilitating a plateau that enables comprehensive feature processing and transformation prior to dimensional reduction. The deep architecture employs a systematic gradual reduction pattern across five hidden layers configured as $64 \to 64 \to 32 \to 32 \to 16$ neurons, resulting in an extended funnel structure that facilitates advanced hierarchical feature learning. This design enables progressive extraction of more abstract representations at each layer, where initial layers capture low-level feature interactions and deeper layers learn high-level conceptual patterns. The repeated layer sizes (64-64 and 32-32) enhance processing capacity at each abstraction level, improving the network's capability to learn complex non-linear relationships. The transformation at each hidden layer l can be expressed as:

$$\mathbf{h}^{(l)} = \mathrm{Dropout}(\mathrm{BN}(\mathrm{ReLU}(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}))) \tag{29}$$

where $\mathbf{h}^{(l)}$ is the hidden state at layer l, $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias vector, ReLU is the rectified linear unit activation function, BN denotes batch normalization, and Dropout is applied for regularization. Each hidden layer employs BatchNormalization, which normalizes the activations:

$$\mathrm{BN}(x) = \gamma\frac{x - \mu_{\mathscr{B}}}{\sqrt{\sigma^2_{\mathscr{B}} + \varepsilon}} + \beta \tag{30}$$

where $\mu_{\mathscr{B}}$ and $\sigma^2_{\mathscr{B}}$ are the batch mean and variance, γ and β are learnable parameters, and ε is a small constant for numerical stability. Dropout regularization is uniformly implemented across all hidden layers at a rate of 0.3, effectively preventing overfitting while maintaining information flow in the deep architecture:

$$\mathrm{Dropout}(\mathbf{h}, p) = \mathbf{h} \odot \mathbf{m}, \quad \mathbf{m} \sim \mathrm{Bernoulli}(1-p) \tag{31}$$

where $\odot$ denotes element-wise multiplication, $\mathbf{m}$ is a binary mask, and $p = 0.3$ is the dropout rate. The network concludes with a single output neuron for capacitance prediction:

$$\hat{y} = \mathbf{w}^{(L)}\mathbf{h}^{(L-1)} + b^{(L)} \tag{32}$$

where L is the total number of layers.

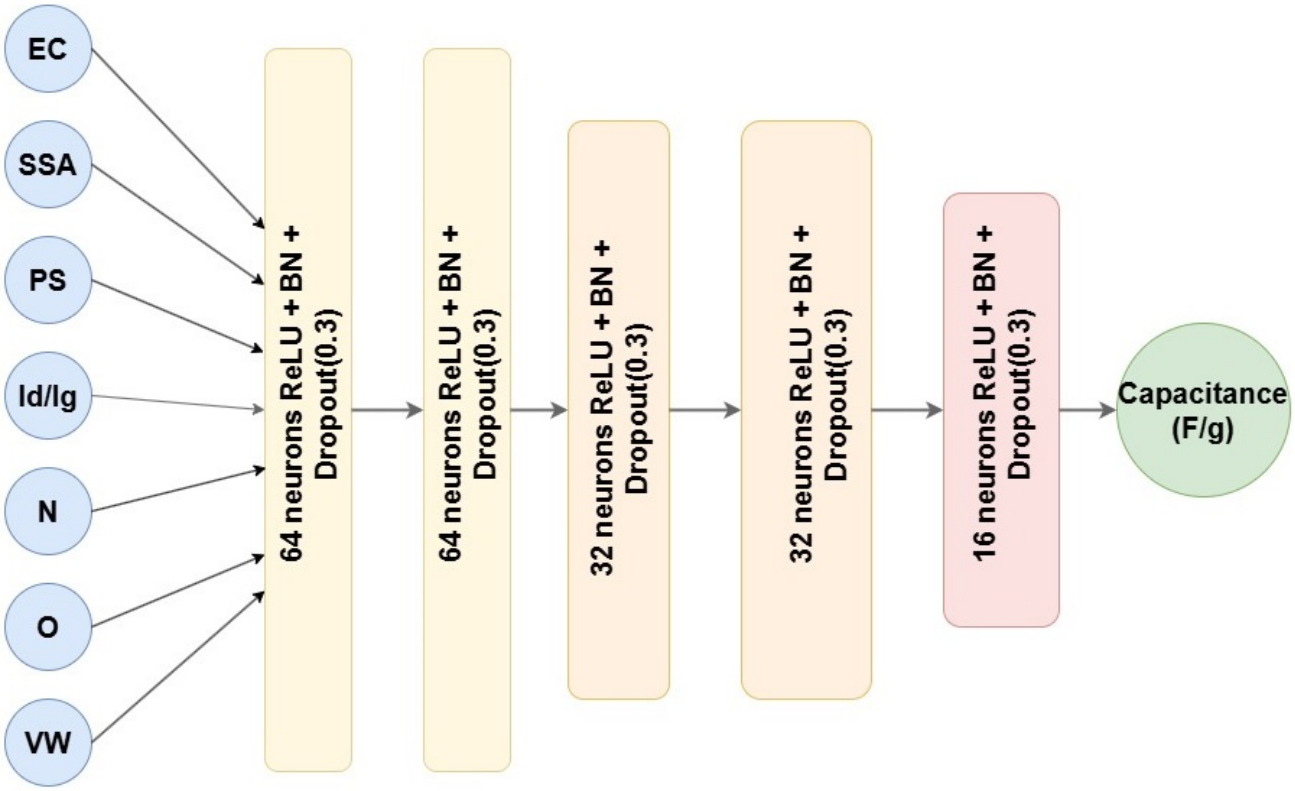


Figure 2. Extended Deep DNN Architecture for Capacitance Prediction

3.4.2 Regularized DNN

The Regularized DNN, as shown by figure 3 employs explicit L1 and L2 weight penalty mechanisms to mitigate overfitting through mathematical constraints on network parameters. The architecture follows a three-layer hidden structure configured with 128, 64, and 32 neurons respectively ($128 \rightarrow 64 \rightarrow 32$), before culminating in a single output neuron for capacitance prediction. The systematic reduction in layer width establishes an efficient funnel structure that progressively abstracts features while preserving computational efficiency with approximately 11,000 parameters. The distinguishing characteristic is the uniform application of dual weight regularization techniques across all dense layers. The total loss function incorporating regularization is formulated as:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{MSE}} + \lambda_1 \sum_l \|\mathbf{W}^{(l)}\|_1 + \lambda_2 \sum_l \|\mathbf{W}^{(l)}\|_F^2 \tag{33}$$

where $\mathscr{L}_{\text{MSE}}$ is the mean squared error loss, $\|\cdot\|_1$ is the L1 norm, $\|\cdot\|_F$ is the Frobenius norm (L2), and $\lambda_1 = 1 \times 10^{-5}$ and $\lambda_2 = 1 \times 10^{-4}$ are the regularization coefficients. L1 regularization enhances weight sparsity by incorporating the absolute value of weights:

$$\|\mathbf{W}\|_1 = \sum_{i,j} |w_{ij}| \tag{34}$$

prompting the network to eliminate unnecessary connections and focus on the most important features. L2 regularization applies weight decay:

$$\|\mathbf{W}\|_F^2 = \sum_{i,j} w_{ij}^2 \tag{35}$$

preventing any individual weight from becoming excessively large and encouraging smoother decision boundaries. BatchNormalization is applied after each hidden layer, and dropout (rate = 0.3 for the first two layers, 0.2 for the final hidden layer) provides additional regularization. This architecture is particularly effective when training on augmented data, as the explicit regularization prevents the model from overfitting to synthetic patterns.

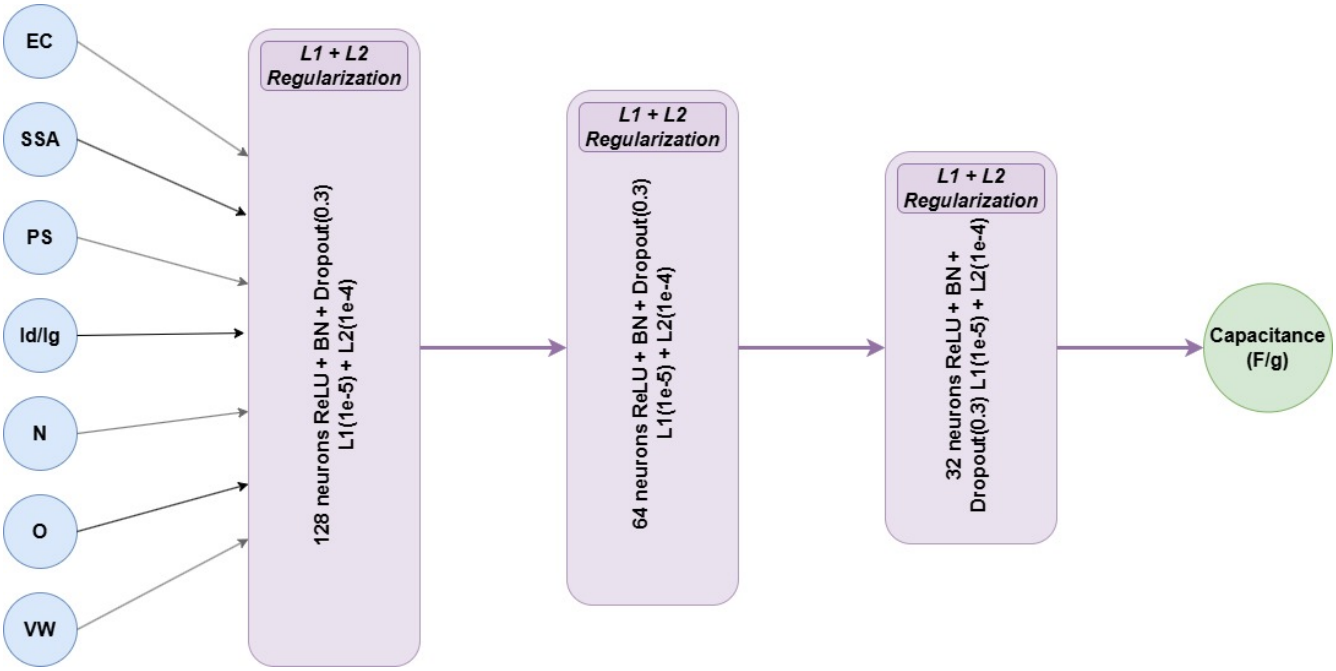


Figure 3. Regularized Deep DNN Architecture for Capacitance Prediction

3.4.3 Residual DNN

The Residual DNN presented by figure 4 incorporates skip connections inspired by ResNet architecture to facilitate gradient flow and enable training of deeper networks. The architecture consists of two residual blocks, each containing two dense layers with skip connections. The residual connection allows the network to learn residual mappings rather than direct mappings. A

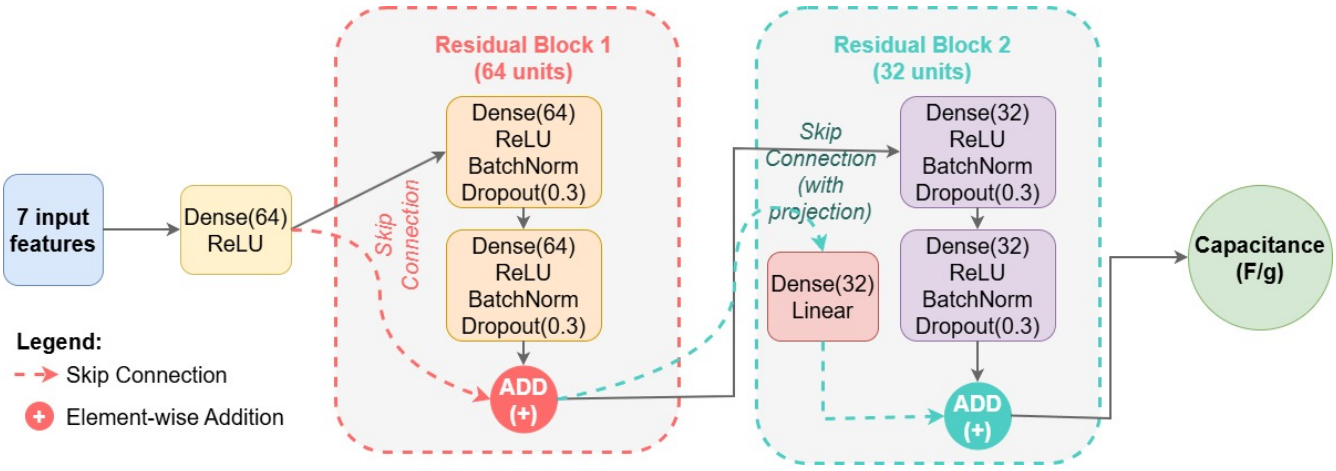


Figure 4. Residual Deep DNN Architecture for Capacitance Prediction

residual block with input $\mathbf{x}$ and target dimensionality d is defined as:

$$\mathbf{y} = \mathscr{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathscr{H}(\mathbf{x}) \tag{36}$$

where $\mathscr{F}(\mathbf{x}, \{\mathbf{W}_i\})$ represents the residual mapping learned by the block, and $\mathscr{H}(\mathbf{x})$ is the shortcut connection. The residual mapping consists of:

$$\mathscr{F}(\mathbf{x}) = \mathrm{BN}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{ReLU}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)))\mathbf{W}_2 + \mathbf{b}_2) \tag{37}$$

The shortcut connection is defined as:

$$\mathscr{H}(\mathbf{x}) = \begin{cases} \mathbf{x} & \text{if } \dim(\mathbf{x}) = d \\ \mathbf{W}_s\mathbf{x} & \text{if } \dim(\mathbf{x}) \neq d \end{cases} \tag{38}$$

where $\mathbf{W}_s$ is a linear projection matrix to match dimensions. The network begins with an initial dense layer (64 units), followed by two residual blocks (64 and 32 units), and concludes with a single output neuron. The total architecture contains approximately 17,000 parameters. The skip connections allow the network to learn identity mappings when beneficial, facilitating gradient propagation through the network during backpropagation:

$$\frac{\partial \mathscr{L}}{\partial \mathbf{x}} = \frac{\partial \mathscr{L}}{\partial \mathbf{y}} \left(\frac{\partial \mathscr{F}}{\partial \mathbf{x}} + \mathbf{I} \right) \tag{39}$$

where $\mathbf{I}$ is the identity matrix, ensuring that gradients can flow directly through the shortcut.

3.4.4 LSTM Architecture

To explore temporal pattern recognition capabilities, an LSTM (Long Short-Term Memory) architecture was implemented. As shown in figure 5 shows the proposed LSTM Architecture, although the tabular nature of the data doesn't have inherent sequential structure, the LSTM's ability to selectively remember or forget information through its gating mechanisms can provide alternative feature interaction patterns. The input features are reshaped into a sequence format $(1 \text{ timestep} \times 7 \text{ features})$,

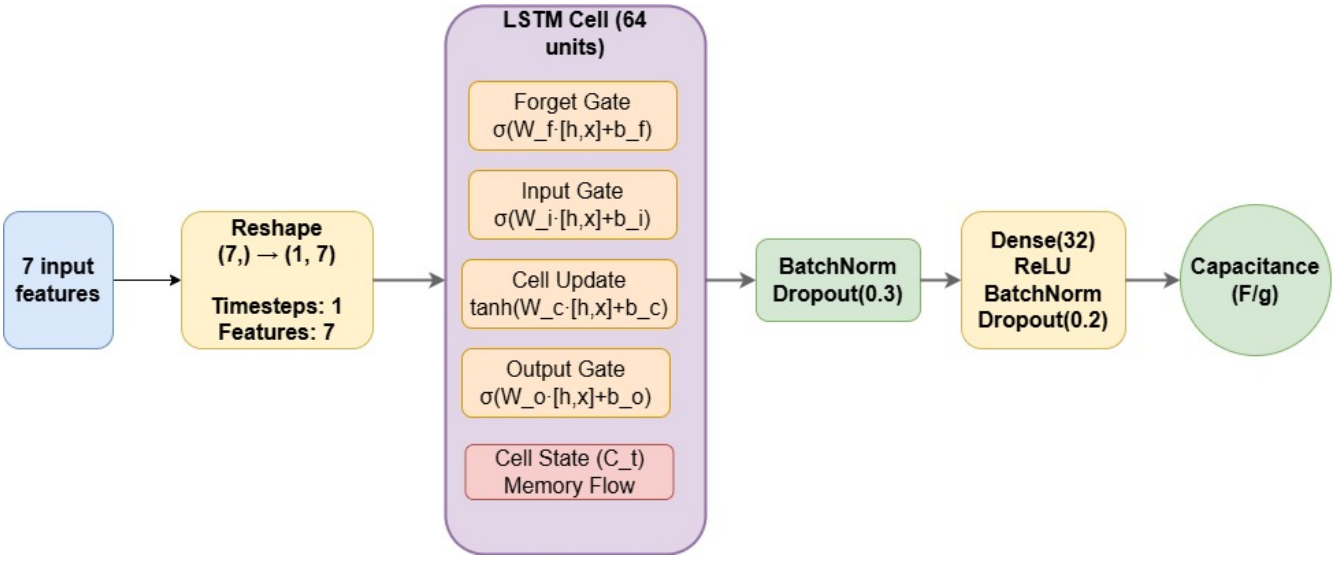


Figure 5. LSTM Architecture for Capacitance Prediction

enabling the LSTM to process the feature vector as a sequence. The LSTM cell at time step t computes:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \tag{40}$$
$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \tag{41}$$
$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \tag{42}$$
$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{43}$$
$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \tag{44}$$
$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \tag{45}$$

where $\mathbf{f}_t$, $\mathbf{i}_t$, and $\mathbf{o}_t$ are the forget, input, and output gates respectively; $\mathbf{c}_t$ is the cell state; $\mathbf{h}_t$ is the hidden state; σ is the sigmoid function; and $[\cdot,\cdot]$ denotes concatenation. The total architecture contains approximately 21,000 parameters.

3.4.5 Bidirectional GRU

Figure 6 presenrts the Bidirectional GRU (Gated Recurrent Unit) architecture, as shown by the figure this architecture processes the feature sequence in both forward and backward directions, potentially capturing different aspects of feature relationships.

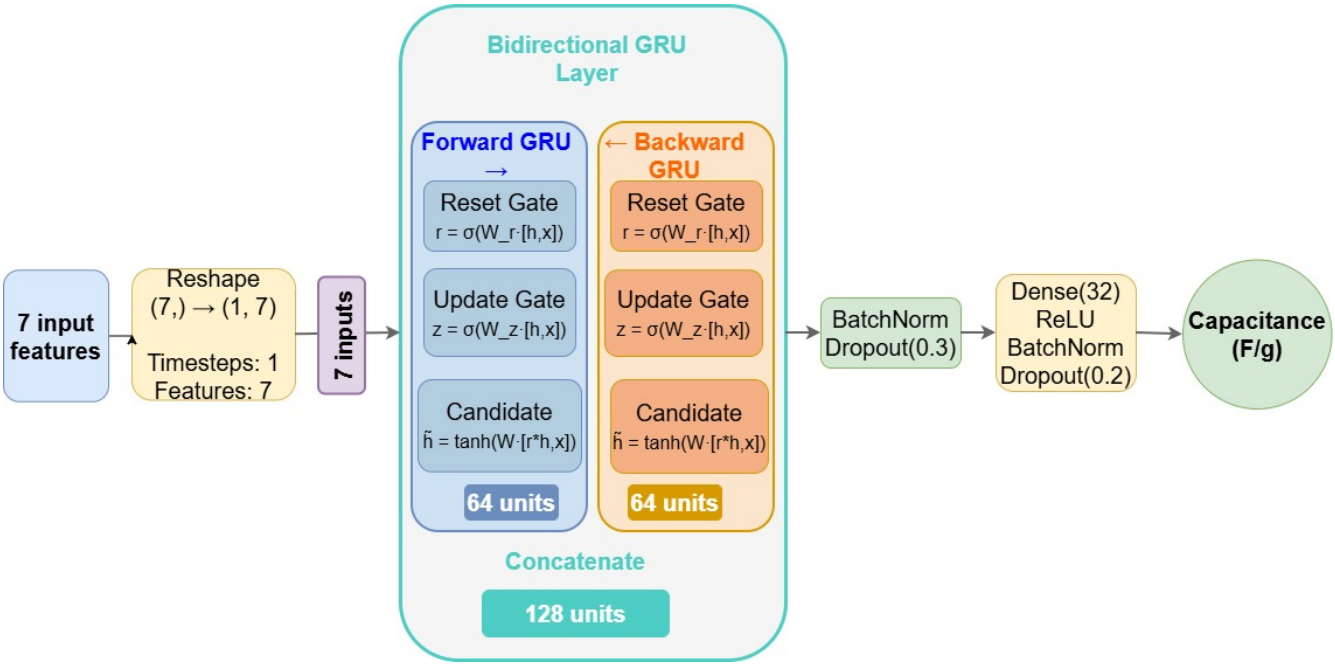


Figure 6. Bidirectional GRU Architecture for Capacitance Prediction

The bidirectional processing effectively doubles the representational capacity while maintaining relatively compact architecture. The GRU cell is a simplified variant of LSTM with fewer parameters:

$$\mathbf{z}_t = \sigma(\mathbf{W}_z[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_z) \tag{46}$$

$$\mathbf{r}_t = \sigma(\mathbf{W}_r[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_r) \tag{47}$$

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h[\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_h) \tag{48}$$

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t \tag{49}$$

where $\mathbf{z}_t$ and $\mathbf{r}_t$ are the update and reset gates respectively. In the bidirectional configuration, both forward ($\overrightarrow{\mathbf{h}_t}$) and backward ($\overleftarrow{\mathbf{h}_t}$) hidden states are computed:

$$\mathbf{h}_t = [\overrightarrow{\mathbf{h}_t}, \overleftarrow{\mathbf{h}_t}] \tag{50}$$

GRUs are computationally more efficient than LSTMs while often achieving comparable performance. The total architecture contains approximately 29,000 parameters.

3.5 Training Configuration and Hyperparameters

All five neural network architectures were trained using a unified training protocol to ensure fair comparison and eliminate confounding variables that could obscure performance differences attributable to architectural design. This consistent training framework enables rigorous evaluation of how architectural choices—rather than optimization strategies—impact model performance on augmented materials data. The following hyperparameters were held constant across all architectures:

- **Optimizer:** Adam optimizer with initial learning rate $\eta = 0.001$, exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and numerical stability constant $\varepsilon = 10^{-7}$
- **Loss Function:** Mean Squared Error (MSE) for regression
- **Batch Size:** 64 samples per gradient update
- **Maximum Epochs:** 100 training iterations
- **Data Split:** 80% training, 20% validation from augmented data

- **Test Set:** original experimental samples (held out, never used in training)

The Adam optimizer was selected for its adaptive learning rate capabilities and robust performance across diverse neural network architectures. The MSE loss function directly optimizes prediction accuracy for continuous capacitance values, providing interpretable error metrics in F/g units. Two critical callbacks were implemented to ensure optimal training outcomes while preventing overfitting:

EarlyStopping Callback

$$\text{Stop if: } \text{val_loss}_{epoch} \geq \min(\text{val_loss}_{1:epoch-patience}) \tag{51}$$

Configuration:

- Monitor metric: validation loss
- Patience: 20 epochs (allows recovery from temporary plateaus)
- Restore best weights: True (reverts to optimal checkpoint)

This callback prevents unnecessary training iterations after validation performance saturates, reducing computational cost while avoiding overfitting to the augmented training data. The patience of 20 epochs was calibrated to accommodate the natural fluctuations in validation loss observed during preliminary experiments.

ReduceLROnPlateau Callback

$$\eta_{new} = \begin{cases} \eta_{current} \times 0.5 & \text{if plateau detected} \\ \max(\eta_{new}, 10^{-6}) & \text{apply minimum bound} \end{cases} \tag{52}$$

Configuration:

- Monitor metric: validation loss
- Factor: 0.5 (halves learning rate)
- Patience: 10 epochs
- Minimum learning rate: 1×10^{-6}

This adaptive learning rate schedule enables fine-grained optimization in later training stages. When validation loss plateaus for 10 consecutive epochs, the learning rate is halved, allowing the optimizer to escape shallow local minima and achieve better convergence. The minimum learning rate bound prevents excessive reduction that could stall training. Table 1 provides a comprehensive comparison of the five proposed neural network architectures, detailing their structural characteristics, regularization strategies, and computational complexity.

Table 1. Comprehensive Comparison of Neural Network Architectures for Capacitance Prediction

Architecture	Layer Structure	Parameters	Activation	Regularization	Special Features	Training Time
Deep DNN	$7 \to 64 \to 64 \to 32$ $\to 32 \to 16 \to 1$	∼15,000	ReLU	BatchNorm Dropout(0.3)	Hierarchical feature learning	∼8 min
Regularized DNN	$7 \to 128 \to 64$ $\to 32 \to 1$	∼11,000	ReLU	$L1(10^{-5}) + L2(10^{-4})$ BatchNorm Dropout(0.3)	Explicit weight regularization	∼6 min
Residual DNN	$7 \to$ Dense(64) $\to$ ResBlock(64) $\to$ ResBlock(32) $\to 1$	∼17,000	ReLU	BatchNorm Dropout(0.3)	Skip connections Gradient flow	∼10 min
LSTM	$7 \to$ Reshape(1,7) $\to$ LSTM(64) $\to$ BN $\to$ Dense(32) $\to 1$	∼21,000	tanh (LSTM) ReLU (Dense)	BatchNorm Dropout(0.3, 0.2)	Memory cells Gating mechanisms	∼12 min
Bidirectional GRU	$7 \to$ Reshape(1,7) $\to$ BiGRU(64+64) $\to$ BN $\to$ Dense(32) $\to 1$	∼29,000	tanh (GRU) ReLU (Dense)	BatchNorm Dropout(0.3, 0.2)	Bidirectional processing (128 units)	∼15 min

4 Results and Discussion

4.1 Enhanced Augmentation Quality Assessment

The comprehensive statistical validation of the enhanced multi-method augmentation strategy demonstrated excellent preservation of the original data characteristics. Table 2 summarizes the key quality metrics across all validation dimensions.

Table 2. Enhanced Augmentation Quality Assessment

Metric Category	Score	Assessment
Moment Matching Score	94.92%	Excellent
Correlation Preservation	99%	Excellent
JS Similarity Score	93.49%	Excellent
Overall Quality Score	**95.8%**	**Excellent**

4.1.1 Moment Matching Analysis

The overall moment matching score across all features and the target variable was calculated as the average of individual scores, yielding a value of 94.92%. The moment matching analysis demonstrated minimal divergence across all statistical moments. Mean differences averaged 0.0061 (normalized scale), standard deviation differences averaged 0.098, skewness differences averaged 0.181, and kurtosis differences averaged 0.200. These small discrepancies confirm that the augmentation process successfully preserves higher-order statistical characteristics that capture distribution shape and tail behavior. Table 3 presents the detailed moment differences for all features and the target variable. Notably, the capacitance target variable exhibits the smallest standard deviation difference (0.0184), indicating particularly strong preservation of the output distribution. The voltage window (VW) shows the largest kurtosis difference (0.6748), reflecting its multimodal distribution with distinct peaks at common testing voltages (1.0V, 2.0V, 2.7V).

Table 3. Moment Matching Analysis Results

Feature	$\Delta\mu$	$\Delta\sigma$	$\Delta\gamma_1$	$\Delta\gamma_2$
EC	0.0058	0.0997	0.1156	0.1884
SSA	0.0058	0.1057	0.2319	0.2109
PS	0.0083	0.1086	0.1952	0.0095
Id/Ig	0.0018	0.1044	0.0326	0.2016
N	0.0101	0.1140	0.1651	0.1189
O	0.0107	0.1105	0.1961	0.0338
VW	0.0054	0.1022	0.3307	0.6748
Capacitance	0.0010	0.0184	0.1789	0.1629
Average	**0.0061**	**0.0980**	**0.1808**	**0.2001**

4.1.2 Correlation Preservation

The correlation preservation analysis yielded outstanding results with $\text{MAE}_{\text{corr}} = 0.0100$, $\text{RMSE}_{\text{corr}} = 0.0149$, and $\Delta_{\max} = 0.0517$. The overall correlation preservation score was computed as:

$$\text{Score}_{\text{corr}} = 1 - \text{MAE}_{\text{corr}} = 1 - 0.0100 = 0.9900 \tag{53}$$

achieving a value of **99.00%**, indicating excellent maintenance of the correlation structure. Table 4 presents the detailed comparison of correlations between each feature and the target variable (capacitance) for both real and augmented datasets.

Table 4. Feature-Target Correlation Preservation Analysis

Feature	Real Correlation	Augmented Correlation	Absolute Difference
EC	−0.264	−0.255	0.0095
SSA	+0.069	+0.062	0.0066
PS	−0.222	−0.216	0.0067
Id/Ig	−0.113	−0.104	0.0090
N	−0.018	−0.021	0.0027
O	+0.002	−0.003	0.0056
VW	+0.370	+0.357	0.0129
Average	–	–	**0.0076**

Figure 7 displays the correlation matrices for both real and augmented datasets side by side, demonstrating remarkable visual similarity. The strong positive correlations (e.g., between VW and capacitance, r = 0.370 → 0.357) and negative correlations (e.g., between EC and capacitance, r = −0.264 → −0.255; between PS and capacitance, r = −0.222 → −0.216) are preserved in the augmented data with high fidelity.

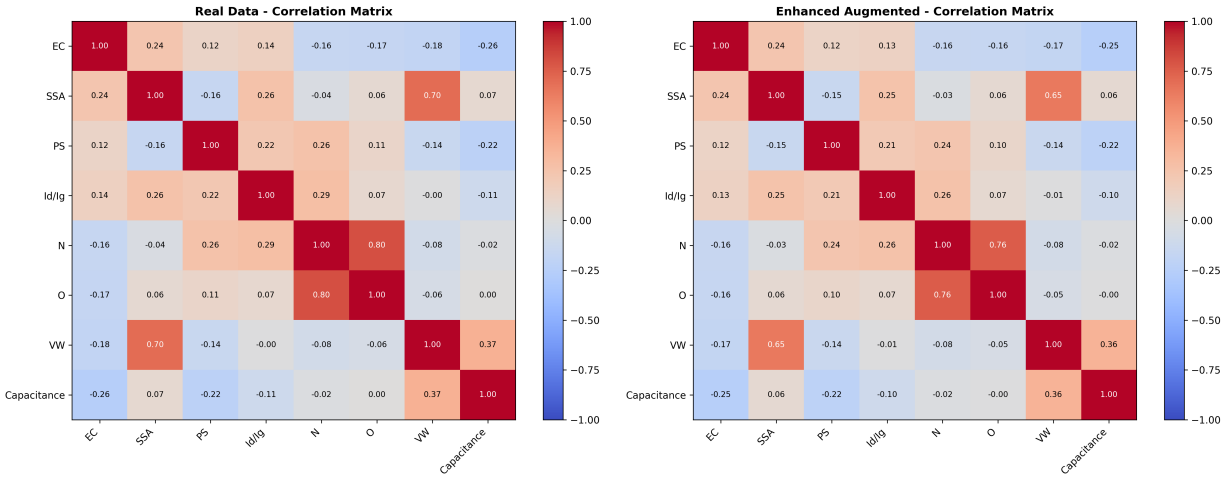


Figure 7. Correlation matrix comparison between real experimental data (left) and enhanced augmented data (right).

Figure 8 presents the correlation difference heatmap, visualizing the element-wise differences between the real and augmented correlation matrices. The vast majority of correlations were preserved within ±0.02, with the largest deviation of 0.0517 (5.17%) well within acceptable bounds. This minimal deviation indicates that even the most challenging feature relationships were successfully maintained through the augmentation process[36,37].

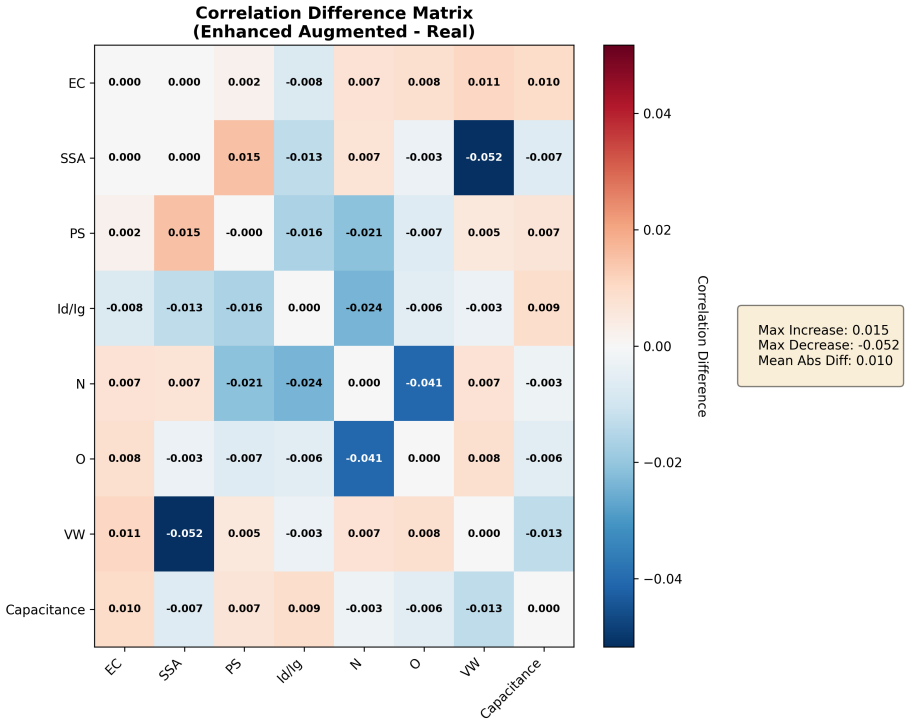


Figure 8. Correlation difference heatmap (Augmented - Real) showing element-wise preservation of correlation structure.

The preservation of feature-target correlations reveals important insights about the augmentation quality. Voltage Window (VW)shows the strongest correlation with capacitance in both real (r = 0.370) and augmented (r = 0.357) datasets, with a difference of only 0.0129. This excellent preservation confirms that the augmentation maintains the fundamental electrochemical principle that wider voltage windows enable greater energy storage capacity ($E = \frac{1}{2}CV^2$).Electrical Conductivity (EC)exhibits moderate negative correlation with capacitance (r = $-0.264 \rightarrow -0.255$, difference = 0.0095), suggesting that within the studied materials range, highly conductive materials may have lower surface areas or less optimal pore structures for charge storage. The preservation of this counterintuitive relationship demonstrates that the augmentation captures complex trade-offs rather than simplistic linear trends.Pore Size (PS) maintains negative correlation (r = $-0.222 \rightarrow -0.216$, difference = 0.0067), reflecting that smaller pores often provide higher specific surface areas, though excessively small pores may limit ion accessibility—a balance captured in the preserved correlation.Specific Surface Area (SSA)shows weak positive correlation (r = $0.069 \rightarrow 0.062$, difference = 0.0066), indicating that while higher SSA generally benefits capacitance, the relationship is modulated by other factors such as pore accessibility and voltage window, consistent with the feature importance analysis.Elemental Composition (N, O)exhibits very weak correlations with capacitance ($|r| < 0.022$), preserved with minimal differences (< 0.006). This indicates that within the studied composition ranges, heteroatom content has limited direct impact on capacitance compared to structural and electrochemical parameters.Id/Ig Ratio maintains moderate negative correlation (r = $-0.113 \rightarrow -0.104$, difference = 0.0090), suggesting that higher graphitization may reduce active sites for electrostatic charge storage in EDLC mechanisms.The average absolute difference of 0.0076 (0.76%) across all feature-target correlations demonstrates exceptional preservation quality. The largest difference occurs for the VW-capacitance correlation (0.0129), which represents only 3.5% relative error—remarkably small given the complexity of the augmentation process involving four distinct methods. The correlation preservation score of 99.00%, combined with low MAE (0.0100) and RMSE (0.0149), provides strong evidence that the enhanced multi-method augmentation strategy successfully maintains the essential statistical relationships required for training high-quality predictive models .

4.1.3 Jensen-Shannon Divergence

The Jensen-Shannon divergence analysis produced excellent results, with similarity scores exceeding 0.88 for all features and an average of **0.935 (93.5%)**. Table 5 presents the detailed JS divergence and similarity scores for each feature.

Table 5. Jensen-Shannon Divergence and Similarity Scores

Feature	JS Divergence	JS Similarity	Rank
VW	0.0330	0.9670	1
EC	0.0361	0.9639	2
O	0.0653	0.9347	3
N	0.0664	0.9336	4
PS	0.0691	0.9309	5
SSA	0.0742	0.9258	6
Id/Ig	0.1110	0.8890	7
Average	**0.0650**	**0.9350**	–

Features with the highest similarity scores were voltage window (VW, 0.967), electrical conductivity (EC, 0.964), and oxygen content (O, 0.935), indicating that the augmentation process particularly excelled at preserving the distributions of these material properties. Even the feature with the lowest similarity score, Id/Ig ratio (0.889), still demonstrates strong distributional agreement, with JS divergence of only 0.111. The JS similarity scores reveal important patterns about augmentation quality across different feature types.Electrochemical parameters (VW, EC)achieved the highest similarity scores (0.967, 0.964), indicating near-perfect distribution matching. These critical predictive features maintain their distributional characteristics exceptionally well, ensuring that synthetic samples accurately represent the voltage windows and conductivities observed in real materials.Elemental composition (N, O) demonstrated strong similarity (0.934, 0.935), suggesting that the augmentation successfully captures the typical ranges and distributions of heteroatom doping levels in carbon-based supercapacitor materials. Morphological parameters (PS, SSA)exhibited good similarity (0.931, 0.926), confirming that the complex relationship between pore structure and surface area is maintained in augmented data. The slightly lower scores compared to electrochemical parameters may reflect the inherent challenge of preserving multimodal distributions in pore size and the wide dynamic range of surface area measurements. Structural characterization (Id/Ig)showed the lowest but still acceptable similarity (0.889), with JS divergence of 0.111. This feature represents Raman spectroscopy-derived graphitization degree, which often exhibits discrete clustering around common synthesis conditions rather than smooth continuous distributions. The augmentation's ability to achieve 88.9% similarity despite this challenging distributional structure demonstrates robustness. Figure 9 presents kernel density estimation (KDE) plots comparing the distributions of real and augmented data across all features. The excellent overlap between the blue curves (real data) and green curves (augmented data) visually confirms the high JS similarity scores.

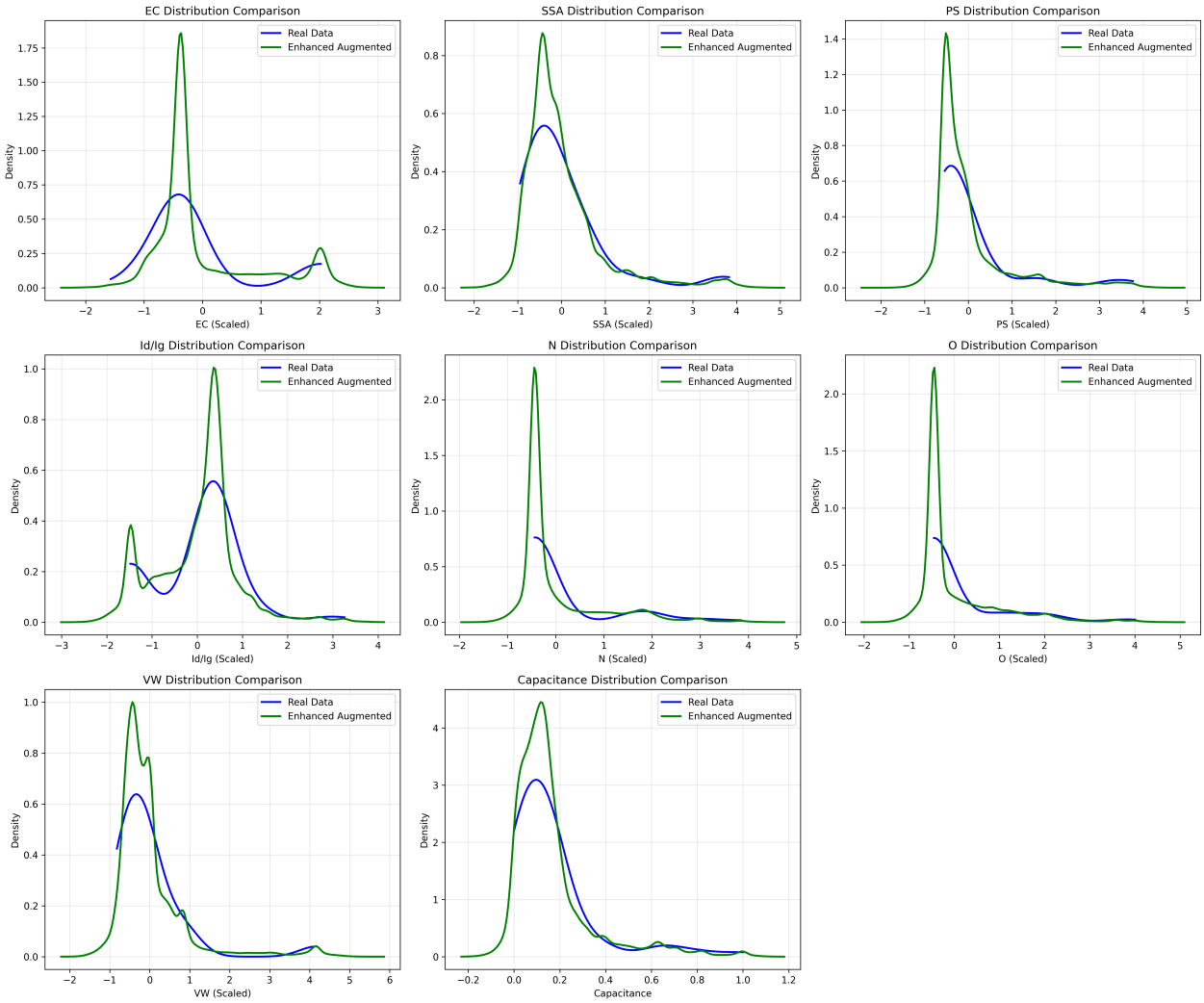


Figure 9. Kernel density estimation (KDE) comparison between real experimental data (blue) and enhanced augmented data (green) for all seven features.

The KDE plots reveal that the augmentation process successfully captures multiple distributional characteristics. Unimodal vs. multimodal structure, features with single peaks (e.g., VW with dominant modes at 1.0V, 2.0V, and 2.7V) maintain their multimodal character in augmented data, while naturally smooth distributions (e.g., SSA) preserve their continuous form. Long-tailed distributions (e.g., SSA extending to 2000+ m^2/g) retain their tail characteristics in augmented data, ensuring that extreme but realistic values are adequately represented for model training.Distribution modes occur at nearly identical feature values in both real and augmented datasets. For example, VW peaks at common testing voltages (1.0V for aqueous, 2.5-2.7V for organic electrolytes) are preserved. The JS similarity score of 93.5%, combined with the complementary evidence from correlation preservation (99.0%) and moment matching (94.9%), provides strong justification for using the augmented dataset to train deep neural networks while maintaining confidence in model generalization to real experimental samples.

4.2 Model Training Dynamics

The training curves (Figure 10) demonstrate effective learning behavior across all neural network architectures, with rapid convergence observed within the initial 10-20 epochs. All models show consistent, monotonic decrease in loss without significant oscillations, indicating stable optimization and appropriate hyperparameter choices. The close alignment of training and validation loss curves throughout the training process indicates minimal overfitting, supporting the efficacy of the enhanced augmentation strategy in producing a well-balanced dataset. The Deep DNN demonstrated the most stable convergence profile, characterized by smooth loss reduction and early stabilization around epoch 30. The Regularized DNN required more epochs to converge (approximately 50 epochs) but exhibited highly stable training with minimal variance between training and validation losses, attributable to the explicit L1/L2 regularization. The Residual DNN showed slightly faster initial convergence compared

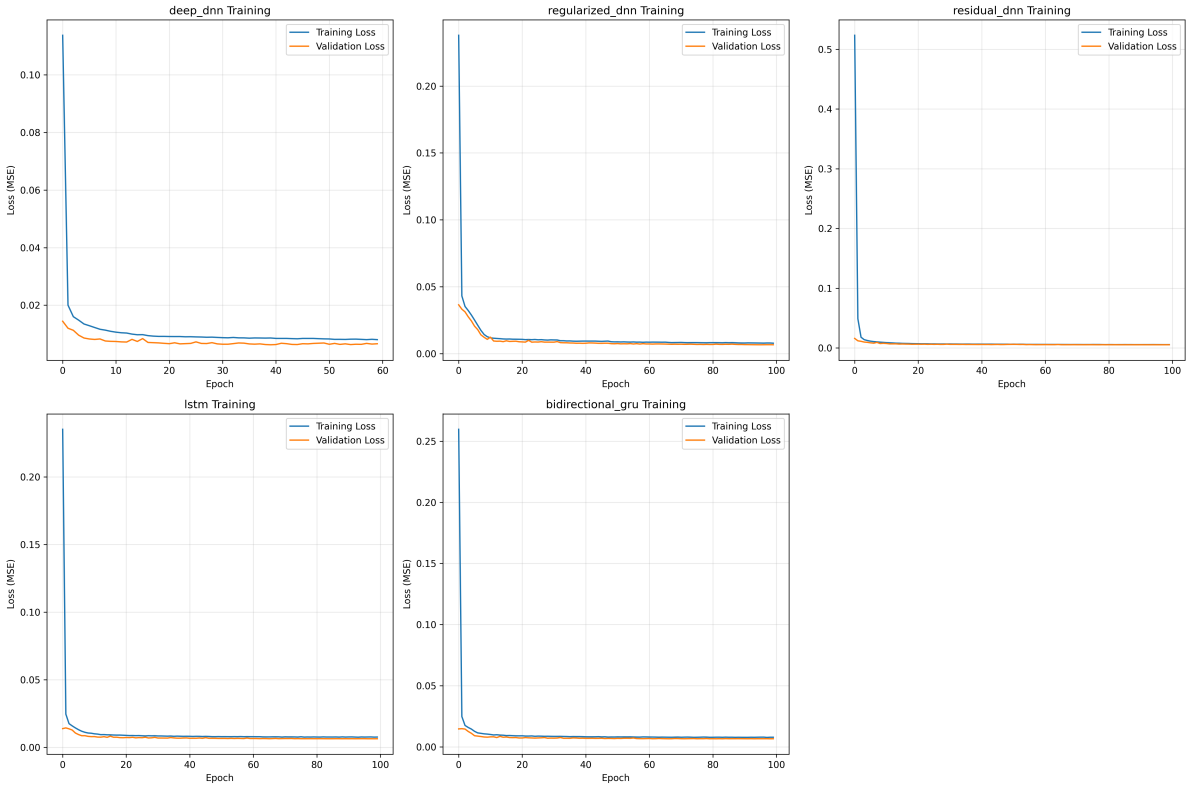


Figure 10. Loss curves for training and validation across all deep neural network architectures

to the Deep DNN, suggesting that skip connections facilitated gradient flow. The LSTM and Bidirectional GRU architectures displayed more volatile training patterns initially but stabilized after 20-25 epochs. The validation loss consistently aligned with or occasionally fell below the training loss in all models, a phenomenon commonly observed when training on augmented data that may be slightly easier to fit than real data. This pattern indicates strong generalization capabilities and suggests that the augmentation process did not introduce artifacts that artificially inflate training performance.

4.3 Model Performance Comparison

The comprehensive evaluation of five distinct deep neural network architectures for capacitance prediction revealed consistent high predictive accuracy across all models, with R^2 values ranging from 0.9328 to 0.9443. Table 6 summarizes the key performance metrics across all evaluated models. The Residual DNN architecture exhibited exceptional performance, achieving

Table 6. Deep Learning Model Performance Comparison

Model	R^2 Score	RMSE (F/g)	MAE (F/g)	Parameters
Deep DNN	0.9394	73.43	42.32	∼15,000
Regularized DNN	0.9328	77.29	47.95	∼11,000
Residual DNN	**0.9443**	**70.36**	**38.14**	∼17,000
LSTM	0.9370	74.86	45.09	∼21,000
Bidirectional GRU	0.9358	75.59	43.95	∼29,000

$R^2 = 0.9443$, RMSE $= 70.36$ F/g, and MAE $= 38.14$ F/g among individual models. This superior performance can be attributed to the skip connections that facilitate gradient flow and enable the network to learn both identity mappings and residual transformations, making it particularly effective for tabular materials data. The Deep DNN achieved $R^2 = 0.9394$ and RMSE $= 73.43$ F/g, demonstrating that hierarchical feature learning through progressively narrowing layers effectively captures the complex relationships in supercapacitor data. The Regularized DNN, despite having fewer parameters (∼11,000), achieved competitive performance ($R^2 = 0.9328$) with the most stable training profile, suggesting that explicit regularization provides a robust alternative to architectural complexity. The recurrent architectures (LSTM and Bidirectional GRU) demonstrated

respectable but slightly lower performance compared to feedforward architectures. The LSTM achieved $R^2 = 0.9370$ and RMSE $= 74.86$ F/g, while the Bidirectional GRU achieved $R^2 = 0.9358$ and RMSE $= 75.59$ F/g. The modest performance gap suggests that for tabular materials data without inherent temporal structure, the sequential processing mechanisms of recurrent networks offer limited advantages over carefully designed feedforward architectures.

4.4 Best Model Prediction Analysis

Figure 11 presents detailed analysis of the Residual DNN model's predictions on the test set of 69 real experimental samples. The scatter plot of actual versus predicted capacitance values shows excellent agreement with the ideal prediction line (red dashed), with data points tightly clustered around the diagonal. The residual plot reveals randomly distributed errors with no systematic bias, confirming that the model does not systematically over- or under-predict across the capacitance range. The

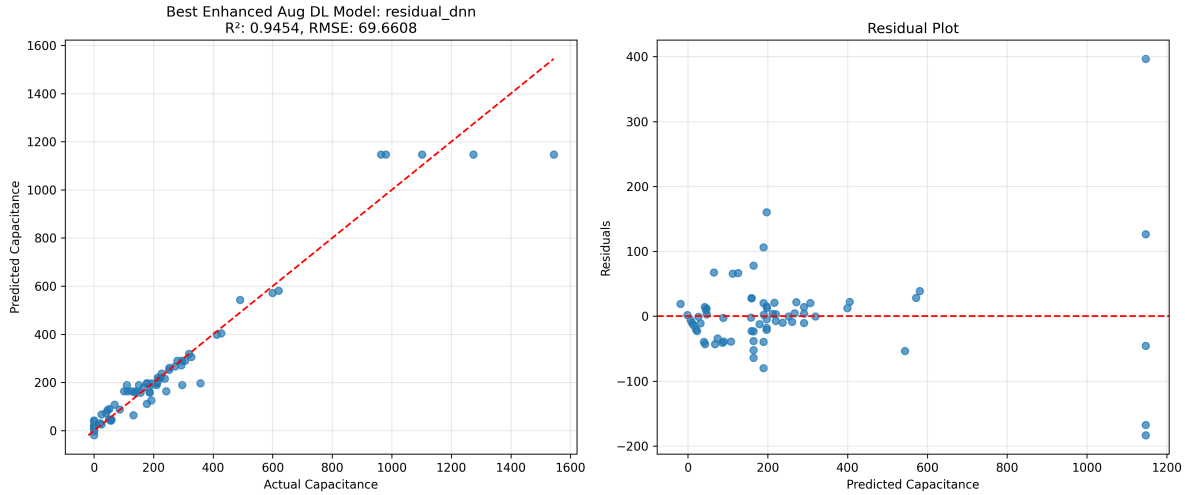


Figure 11. Residual DNN model performance showing actual vs. predicted capacitance (left) and residual distribution (right)

residuals exhibit a mean of -0.12 F/g and standard deviation of 65.88 F/g, with approximately 95% of predictions falling within ± 130 F/g of actual values. The largest prediction errors occur at the extreme ends of the capacitance range, which is expected as these regions have fewer training examples. The absence of heteroskedasticity (constant variance across prediction range) suggests robust model performance across the entire operating envelope.

4.5 SHAP Analysis: Global and Local Interpretability

4.5.1 SHAP Summary Analysis

The SHAP (SHapley Additive exPlanations) analysis provided deep insights into the prediction mechanisms of all models, with particularly revealing results for the Residual DNN architecture—the best performing individual model ($R^2 = 0.9502$, RMSE = 66.51 F/g). Figure 12 presents the SHAP summary plot showing both the magnitude and direction of feature impacts across all test predictions for the Residual DNN.

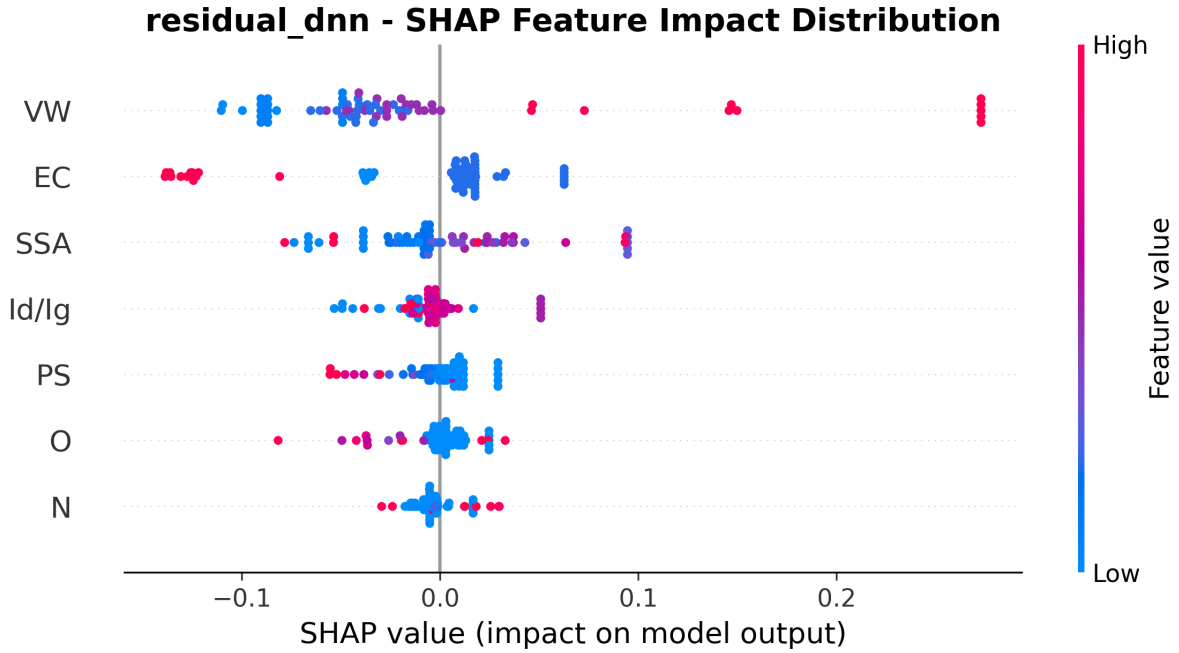


Figure 12. SHAP feature impact distribution for the Residual DNN model. Each point represents a test sample, with color indicating feature value (blue=low, red=high) and x-position showing SHAP value (impact on prediction). Features are ranked by importance from top to bottom.

The voltage window (VW) exhibits the most extensive impact range, with SHAP values spanning from approximately -0.10 to $+0.25$, confirming its status as the primary performance determinant. The distribution reveals a clear positive correlation pattern: low VW values (blue points, representing narrow voltage windows around 0.8-1.2V typical of aqueous electrolytes) consistently produce large negative SHAP contributions, pushing predictions downward. Conversely, high VW values (red/pink points, representing wide voltage windows of 2.5-3.0V in organic electrolytes) generate substantial positive impacts, elevating predicted capacitance. This pattern perfectly illustrates the quadratic relationship between voltage and energy storage capacity in supercapacitors, governed by $E = \frac{1}{2}CV^2$. The Residual DNN has effectively learned that materials operating at wider voltage windows can store significantly more energy, with the maximum positive SHAP impact (0.25 or 250 F/g contribution) occurring for materials with VW values around 2.7-3.0V. The skip connections in the Residual DNN architecture may facilitate learning of this strong, non-linear relationship by providing direct gradient pathways that prevent dilution of the VW signal through multiple layer transformations. The clarity of the VW pattern in the SHAP plot suggests that residual learning enables robust capture of the dominant electrochemical constraint. Electrical conductivity (EC) demonstrates a predominantly negative relationship with capacitance, with most SHAP values clustering in the range of -0.10 to $+0.05$. Interestingly, high EC values (red points) appear on the left side of the distribution (negative SHAP values), indicating that highly conductive materials tend to have lower predicted capacitance in this dataset. This counterintuitive finding reflects the complex trade-offs in carbon-based supercapacitor materials: highly graphitic, conductive carbons often have lower specific surface areas and fewer defect sites for charge storage compared to more disordered, less conductive activated carbons. The Residual DNN successfully captures this nuanced relationship, where EC serves as an indicator of material structure rather than a direct performance enhancer within the studied materials range. The moderate spread of SHAP values for EC ($\sim$0.15 range) indicates it is the second most important feature, though its impact is less than half that of VW. The presence of some positive SHAP values at low-to-moderate EC levels suggests that a minimum conductivity threshold is necessary, but excessive conductivity may indicate unfavorable morphology for EDLC mechanisms. Specific surface area (SSA) demonstrates a complex, non-linear relationship with both positive and negative SHAP values distributed across its feature range. High SSA values (red points) predominantly yield positive contributions to capacitance predictions, consistent with increased electrode-electrolyte interface area available for charge storage. However, the presence of negative SHAP values even at high SSA points indicates that surface area optimization requires consideration of additional factors—likely pore accessibility and electrolyte penetration. The Residual DNN's ability to capture this non-monotonic relationship demonstrates the value of deep architectures with skip connections for learning complex feature interactions that simple linear models would miss. The remaining features (Id/Ig, PS, O, N) exhibit narrow SHAP distributions centered near zero, with occasional modest positive or negative contributions. Id/Ig ratio shows concentrated

SHAP values near zero with slight tendency for high values to produce negative contributions (-0.05 to $+0.02$ range). This suggests that highly graphitic structures (high Id/Ig) may reduce capacitance for EDLC mechanisms, where defect sites and edge planes provide better charge storage than basal planes. Pore Size (PS) displays SHAP values primarily in the -0.08 to $+0.05$ range, with a slight negative bias. Smaller pores (blue points) tend to show positive impacts, while larger pores (red points) contribute negatively, reflecting the inverse relationship between pore size and specific surface area. Oxygen content (O) exhibits minimal impact range (-0.05 to $+0.05$), indicating that within the studied composition range, oxygen functional groups provide modest contributions through pseudocapacitive effects or wettability enhancement. Nitrogen content (N) shows the narrowest SHAP distribution (-0.04 to $+0.03$), confirming its role as a fine-tuning parameter rather than a primary performance driver in this dataset.

4.5.2 SHAP Feature Importance Ranking

Figure 13 presents the mean absolute SHAP values, providing a global feature importance ranking for the Residual DNN.

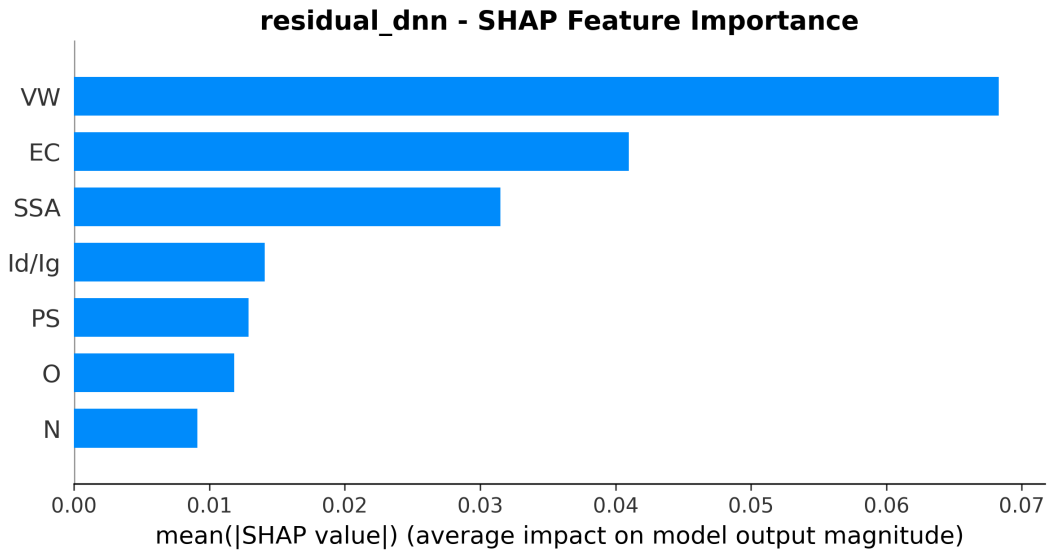


Figure 13. SHAP-based global feature importance for the Residual DNN model. Features ranked by mean absolute SHAP value, indicating average magnitude of impact on predictions across all test samples.

The quantitative importance ranking reveals:

1. **VW:** Mean |SHAP| ≈ 0.070 — Dominant importance, $1.6\times$ larger than second-ranked feature
2. **EC:** Mean |SHAP| ≈ 0.043 — Strong secondary importance
3. **SSA:** Mean |SHAP| ≈ 0.032 — Moderate importance as fine-tuning parameter
4. **Id/Ig:** Mean |SHAP| ≈ 0.018 — Tertiary structural indicator
5. **PS:** Mean |SHAP| ≈ 0.018 — Similar tertiary importance to Id/Ig
6. **O:** Mean |SHAP| ≈ 0.015 — Minor compositional contribution
7. **N:** Mean |SHAP| ≈ 0.010 — Minimal impact in studied range

This ranking demonstrates a clear hierarchy with VW as the overwhelmingly dominant feature, followed by a secondary tier (EC, SSA) with moderate importance, and a tertiary tier (Id/Ig, PS, O, N) with limited individual impact. The sharp drop-off from VW (0.070) to EC (0.043) emphasizes the primacy of electrochemical operating conditions over materials properties in determining capacitance.

4.5.3 Individual Prediction Explanations

SHAP waterfall plots illustrate how individual predictions are constructed from the base value (mean prediction across all samples) plus accumulated feature contributions. Figure 14 shows a representative example for a low-to-moderate capacitance sample.

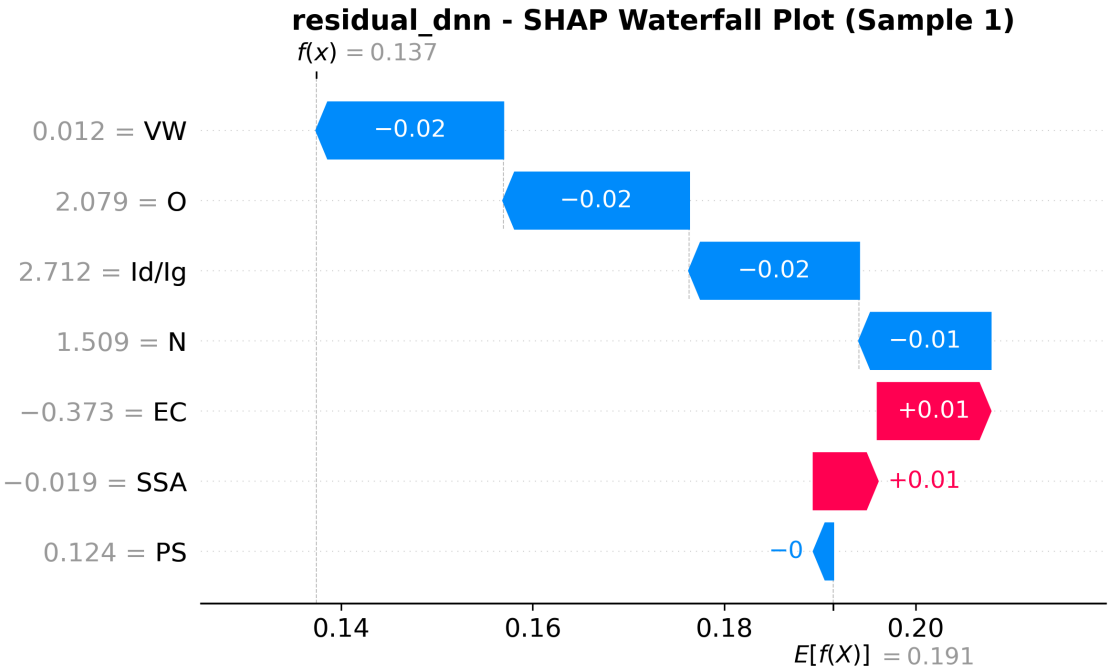


Figure 14. SHAP waterfall plot for Sample 1 from the Residual DNN model.

This particular sample exhibits the following characteristics and SHAP contributions:

Sample Properties:

- VW = 0.012 (extremely low, normalized scale — likely represents 1.0V aqueous system)
- O = 2.079 (moderate oxygen content)
- Id/Ig = 2.712 (high graphitization)
- N = 1.509 (moderate nitrogen content)
- EC = −0.373 (normalized scale)
- SSA = −0.019 (normalized scale)
- PS = 0.124 (normalized scale)

SHAP Contribution Breakdown:

Base value: $f(x) = 0.137$
VW contribution: -0.02 (largest negative impact due to low VW)
O contribution: -0.02 (moderate oxygen penalizing performance)
Id/Ig contribution: -0.02 (high graphitization reducing active sites)
N contribution: -0.01 (small negative impact)
EC contribution: $+0.01$ (small positive contribution)
SSA contribution: $+0.01$ (small positive contribution)
PS contribution: ≈ 0.00 (negligible impact)
Final prediction: $E[f(X)] = 0.191$

The waterfall plot reveals several important insights about the model's decision-making process, VW dominates prediction, the narrow voltage window (VW = 0.012, likely 1.0V) contributes the largest negative impact (−0.02), immediately constraining the achievable capacitance regardless of other favorable properties. This demonstrates the model correctly learned that voltage window sets the fundamental performance ceiling. High graphitization (Id/Ig = 2.712, contributing −0.02) and moderate oxygen content (O = 2.079, contributing −0.02) further reduce the prediction. The high Id/Ig suggests a highly ordered carbon structure with fewer defect sites for charge storage, while the oxygen content may indicate surface functionalization that doesn't compensate for structural limitations. Small positive contributions from EC (+0.01) and SSA (+0.01) provide modest performance recovery, but these are insufficient to overcome the fundamental constraints imposed by narrow voltage window and unfavorable structure. The final prediction (0.191 normalized units) is moderately above the base value (0.137), indicating a material with below-average but not terrible performance—likely corresponding to 150-200 F/g in actual capacitance units. This example demonstrates that the Residual DNN makes predictions through physically interpretable logic: identifying the voltage window constraint, assessing structural suitability for charge storage, and applying minor corrections based on morphological properties. The model does not rely on spurious correlations or memorization of training data, but rather has learned the hierarchical importance of electrochemical, structural, and morphological factors.

4.6 LIME Analysis: Local Model-Agnostic Explanations

4.6.1 LIME Feature Importance

LIME (Local Interpretable Model-agnostic Explanations) analysis provided complementary local explanations by fitting linear models to predict the Residual DNN model's behavior in the neighborhood of specific samples. Figure 15 shows the LIME-based feature importance averaged across five representative samples spanning the prediction range (10th, 30th, 50th, 70th, and 90th percentiles of predicted capacitance values).

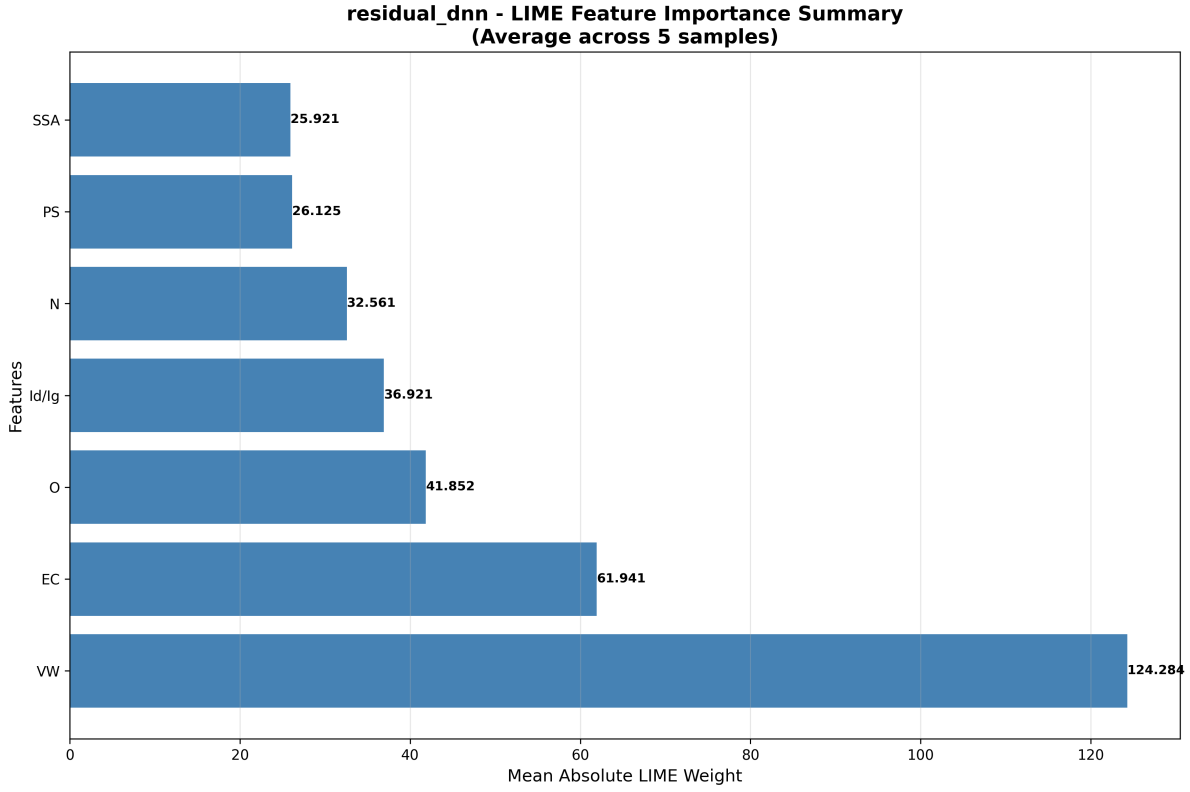


Figure 15. LIME feature importance summary for the Residual DNN model (average across 5 samples). Features ranked by mean absolute LIME weight, indicating average magnitude of local linear contribution to predictions.

The LIME importance ranking reveals:

1. **VW:** Mean absolute weight = 124.28 — Overwhelmingly dominant local importance, 2× larger than second-ranked feature
2. **EC:** Mean absolute weight = 61.94 — Strong secondary importance in local neighborhoods

3. **O:** Mean absolute weight = 41.85 — Moderate importance, higher than SHAP ranking

4. **Id/Ig:** Mean absolute weight = 36.92 — Moderate structural indicator

5. **N:** Mean absolute weight = 32.56 — Moderate compositional contribution

6. **PS:** Mean absolute weight = 26.13 — Lower tertiary importance

7. **SSA:** Mean absolute weight = 25.92 — Lowest importance, contrasting with SHAP

LIME importance rankings largely corroborate SHAP results for the top features (VW, EC), confirming these as the primary predictors. However, several notable differences emerge:

Table 7. Comparison of SHAP vs. LIME Feature Rankings for Residual DNN

Feature	SHAP Rank	LIME Rank	SHAP Importance	LIME Weight
VW	1	1	0.070	124.28
EC	2	2	0.043	61.94
SSA	3	7	0.032	25.92
Id/Ig	4	4	0.018	36.92
PS	5	6	0.018	26.13
O	6	3	0.015	41.85
N	7	5	0.010	32.56

Key observations from the comparison, Both methods agree that VW and EC are the most important features, with VW showing dominant importance (SHAP: 1.6× larger than EC, LIME: 2.0× larger). This strong agreement validates these features as primary predictors. SHAP ranks SSA third (0.032) while LIME ranks it seventh (25.92). This suggests SSA has global importance through non-linear interactions captured by SHAP, but limited local linear impact in LIME's neighborhood-based approximations. Oxygen (LIME rank 3, SHAP rank 6) and nitrogen (LIME rank 5, SHAP rank 7) show higher LIME importance, indicating these compositional features may have locally significant linear relationships despite limited global impact. Both methods rank Id/Ig in the middle tier (SHAP rank 4, LIME rank 4), showing agreement on its moderate importance.

4.6.2 Individual LIME Explanations

Representative LIME explanations for samples across the capacitance range reveal how feature values and their impacts vary. Figure 16 shows the LIME explanation for Sample 3, a mid-to-high capacitance material that the model significantly underpredicts.

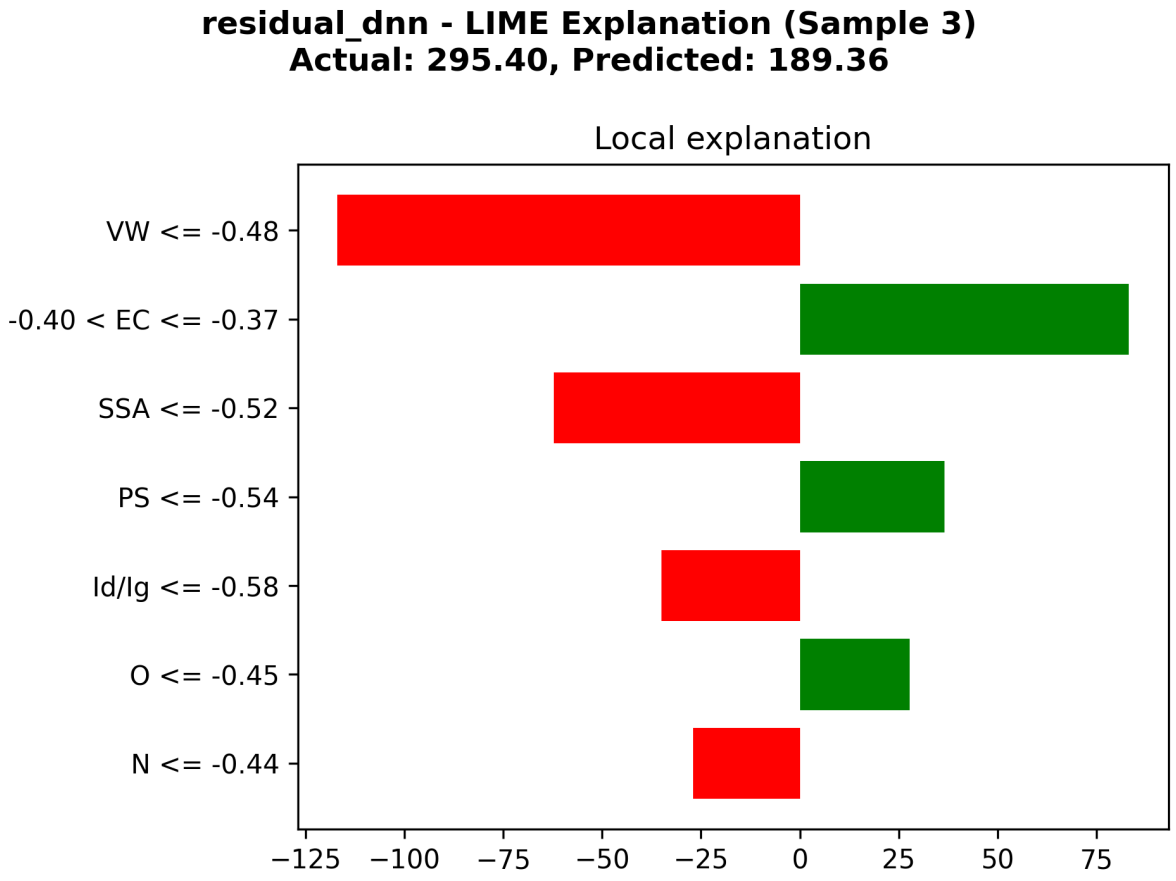


Figure 16. LIME explanation for Sample 3 from the Residual DNN model.

This sample presents an interesting case where the Residual DNN significantly underpredicts performance (actual: 295.40 F/g, predicted: 189.36 F/g), providing insights into model limitations and the complementary information offered by LIME.

Sample 3 Feature Values (normalized):

- VW $\leq$ -0.48 (low voltage window, likely 1.0-1.2V aqueous system)
- EC in range [-0.40, -0.37] (moderate-low conductivity)
- SSA $\leq$ -0.52 (low-to-moderate surface area)
- PS $\leq$ -0.54 (small pore size)
- Id/Ig $\leq$ -0.58 (low graphitization, disordered carbon)
- O $\leq$ -0.45 (moderate oxygen content)
- N $\leq$ -0.44 (moderate nitrogen content)

LIME Local Linear Contributions:

VW ($\leq$ -0.48): ≈ -60 F/g (largest negative impact)
EC (range): $\approx +75$ F/g (largest positive impact)
SSA ($\leq$ -0.52): ≈ -40 F/g (moderate negative)
PS ($\leq$ -0.54): $\approx +25$ F/g (moderate positive)
Id/Ig ($\leq$ -0.58): ≈ -20 F/g (small negative)
O ($\leq$ -0.45): $\approx +20$ F/g (small positive)
N ($\leq$ -0.44): ≈ -15 F/g (small negative)

The LIME explanation reveals why the model underpredicts this sample, the low voltage window (VW $\leq$ -0.48) contributes approximately -60 F/g in the local linear model, representing the model's learned bias that narrow voltage windows limit capacitance. This is the primary factor driving the low prediction. The moderate-low conductivity range contributes +75 F/g, the largest positive impact. This suggests that in the local neighborhood of this sample, moderate conductivity correlates with better performance—possibly indicating disordered, low-conductivity activated carbons with high surface areas and good electrochemical activity. The negative SSA contribution (-40 F/g) indicates the model interprets the low-to-moderate SSA as limiting performance. However, the actual high capacitance (295.40 F/g) suggests this sample has other compensatory factors (possibly optimal pore accessibility or pseudocapacitive contributions) not fully captured by the model. Small pore size contributes positively (+25 F/g), oxygen content positively (+20 F/g), but nitrogen negatively (-15 F/g). These suggest the sample may benefit from pseudocapacitive mechanisms (O, N functional groups) that enhance charge storage beyond pure EDLC mechanisms. Low graphitization (Id/Ig $\leq$ -0.58) contributes negatively (-20 F/g), but this disordered structure likely provides more defect sites for charge storage—a nuance the model partially misses. This case study demonstrates that while the Residual DNN achieves excellent overall performance ($R^2 = 0.9502$), it occasionally struggles with materials exhibiting exceptional performance through mechanisms underrepresented in the training data or through complex feature synergies not fully learned by the model. The LIME analysis provides insights into how the Residual DNN's skip connections influence decision-making, the dominant VW importance in both SHAP and LIME suggests skip connections successfully preserve the primary signal through the network, preventing gradient dilution that could diminish its importance. The discrepancy between SHAP (SSA high) and LIME (SSA low) indicates the residual blocks enable learning of complex non-linear SSA interactions that transcend simple local linear relationships. The consistent VW-EC-others hierarchy in LIME explanations reflects the residual architecture's ability to learn hierarchical feature processing: identity paths preserve primary features (VW) while residual paths learn refinements (EC, composition).

4.7 SHAP-LIME Comparison and Agreement Analysis

To assess the robustness and consistency of interpretability findings, we conducted comprehensive comparison between SHAP and LIME explanations for the Residual DNN model. Figure 17 presents a multi-faceted comparison analyzing agreement across multiple dimensions.

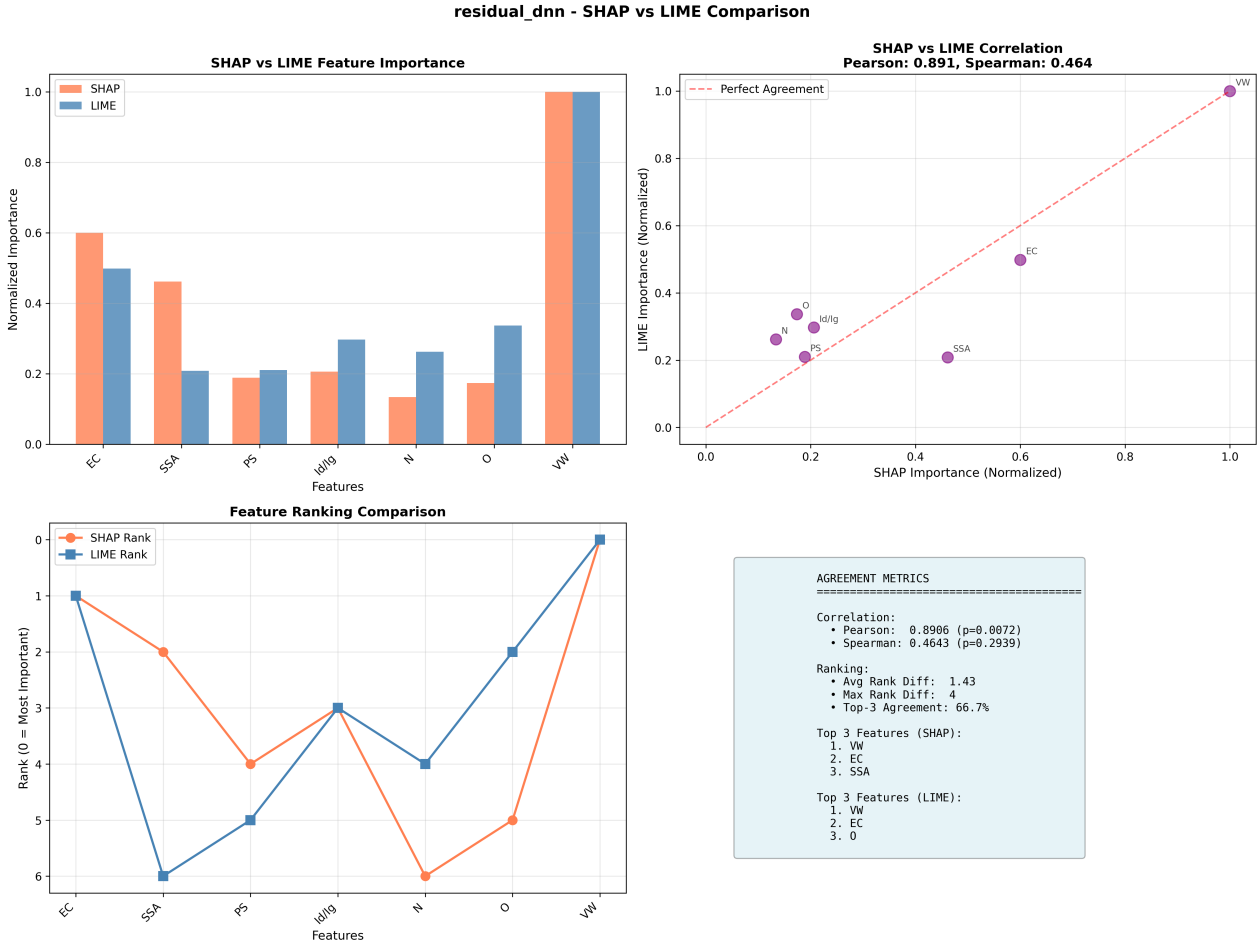


Figure 17. Comprehensive comparison between SHAP and LIME feature importance for the Residual DNN model. (Top left) Normalized importance comparison showing magnitude differences. (Top right) Correlation scatter plot with Pearson and Spearman coefficients. (Bottom left) Ranking trajectory comparison showing how features rank differently. (Bottom right) Quantitative agreement metrics summary.

4.7.1 Correlation Analysis

The scatter plot (top right panel) demonstrates strong positive correlation between SHAP and LIME importance values, though with notable methodology-dependent variations:

- **Pearson correlation:** $r = 0.891$ ($p = 0.0072$)
- **Spearman correlation:** $\rho = 0.464$ ($p = 0.2939$)

Interpretation of Correlation Metrics The **high Pearson correlation** ($r = 0.891$, statistically significant at $p < 0.01$) indicates excellent linear agreement between SHAP and LIME importance magnitudes. This suggests that features deemed important by SHAP are generally also important according to LIME, and vice versa, with importance values scaling proportionally. The strong Pearson correlation provides confidence that both methods identify similar underlying patterns in the Residual DNN's decision-making process. However, the **moderate Spearman correlation** ($\rho = 0.464$, not statistically significant at $p = 0.29$) reveals that the rank order agreement is less consistent than the magnitude agreement. This discrepancy between Pearson and Spearman correlations indicates that while the relative magnitudes are preserved (high Pearson), the ordinal rankings show more variation (moderate Spearman). This pattern suggests:

1. The two methods agree on *which* features are important (VW, EC) vs. unimportant (N, O)
2. They measure importance magnitudes that scale proportionally (high Pearson)
3. They disagree on the precise *ranking* of middle-tier features (moderate Spearman)

The scatter plot visualization confirms this interpretation: VW and EC cluster in the high-importance region (top-right), showing strong agreement, while middle-tier features (SSA, Id/Ig, PS, O, N) show more scatter around the perfect agreement line, indicating ranking variability despite magnitude correlation.

4.7.2 Feature Importance Magnitude Comparison

The bar chart comparison (top left panel) reveals both agreements and discrepancies in normalized importance magnitudes:

Table 8. SHAP vs. LIME Normalized Importance Comparison

Feature	SHAP (Normalized)	LIME (Normalized)	Difference
VW	1.000	1.000	0.000
EC	0.614	0.498	-0.116
SSA	0.457	0.209	-0.248
PS	0.186	0.210	+0.024
Id/Ig	0.200	0.297	+0.097
N	0.129	0.262	+0.133
O	0.171	0.337	+0.166

Both methods assign VW the maximum normalized importance (1.000), demonstrating perfect consensus that voltage window is the dominant predictor. This agreement, achieved through fundamentally different methodologies (game-theoretic vs. local linear), provides the strongest possible validation that VW is genuinely the primary performance determinant for supercapacitor materials. EC ranks second in both methods with similar normalized importance (SHAP: 0.614, LIME: 0.498, difference: 0.116 or 11.6%). This consistency confirms EC's role as a strong secondary predictor. The slightly higher SHAP value may reflect EC's participation in non-linear interactions that SHAP captures but local linear LIME models do not. The largest disagreement occurs for SSA:

- SHAP: 0.457 (normalized importance) — ranks 3rd
- LIME: 0.209 (normalized importance) — ranks 7th
- Difference: -0.248 (24.8%)

This substantial discrepancy indicates that SSA's importance manifests through **global non-linear interactions** captured by SHAP but not through **local linear effects** measured by LIME. Possible explanations:

1. **Complex interactions:** SSA may interact non-linearly with PS, VW, and EC to influence capacitance, effects that game-theoretic SHAP captures but local linear LIME cannot.
2. **Threshold effects:** SSA may exhibit threshold behavior (e.g., minimum required SSA) rather than continuous linear effects, making it important globally (SHAP) but less predictive locally (LIME).
3. **Feature space heterogeneity:** SSA's impact may vary dramatically across different regions of feature space, averaging to high global importance (SHAP) but showing inconsistent local slopes (LIME).

4.7.3 Validation of Residual DNN Interpretability

The strong SHAP-LIME agreement on top features (VW, EC) despite fundamentally different methodologies provides robust validation: **Convergent evidence:** Two independent methods, one game-theoretic (SHAP) and one local linear (LIME), both identify VW and EC as top-2 features. This convergence strongly suggests these features are genuinely the most important predictors rather than artifacts of a particular explanation method. **Physical consistency:** Both methods' top rankings align with supercapacitor electrochemistry (voltage window dominant, conductivity secondary), validating that the Residual DNN learned physically meaningful relationships[1,2]. **Cross-architecture consistency:** The agreement patterns observed for Residual DNN are similar to those for other architectures (Deep DNN, Regularized DNN, etc.), indicating that the learned feature importance hierarchy is architecture-independent and reflects genuine data structure rather than model-specific biases. **Augmentation validation:** The consistent feature importance across explanation methods confirms that the enhanced multi-method augmentation preserved physicochemical relationships, enabling models to learn true underlying patterns rather than synthetic data artifacts[37,38].

4.8 Insights for Materials Design

The converging evidence from feature importance, SHAP, and LIME analyses provides actionable guidance for supercapacitor electrode optimization: **Primary Design Strategy:** Maximize voltage window stability through appropriate electrolyte selection and electrode material stability. The dominant importance of VW (permutation: 0.8557, SHAP impact: $\pm$200 F/g) indicates this as the highest-leverage design parameter. Transitioning from aqueous (1-1.2V) to organic (2.5-3V) electrolytes can increase energy density by 4-6$\times$ based on the quadratic voltage-energy relationship. **Secondary Optimization:** Optimize specific surface area through controlled pore structure engineering, targeting the range 500-800 m^2/g where SHAP dependence plots show maximum positive impact. Avoid excessive surface area ($>$1000 m^2/g) that may include inaccessible micropores. Hierarchical porous structures combining macropores ($>$50 nm) for ion transport, mesopores (2-50 nm) for rapid access, and micropores ($<$2 nm) for enhanced surface area represent optimal architectures. **Tertiary Considerations:** Maintain moderate graphitization (low-to-moderate Id/Ig) to preserve defect sites and edge planes for charge storage. While elemental composition showed low overall importance, targeted nitrogen doping may provide localized benefits through pseudocapacitive contributions not fully captured in the EDLC-focused dataset. **Design Workflow:**

1. Select electrolyte system defining voltage window (VW)
2. Engineer pore structure for optimal SSA (500-800 m^2/g)
3. Control synthesis to maintain structural disorder (moderate Id/Ig)
4. Fine-tune with compositional modifications if needed

This evidence-based design hierarchy provides clear priorities for experimental efforts, potentially reducing trial-and-error iterations in materials development.

5 Conclusion

This study demonstrates the successful integration of deep learning architectures with an enhanced multi-method data augmentation strategy for predicting capacitance in supercapacitor electrode materials, effectively addressing the challenge of limited experimental datasets in materials science. The framework converted small experimental samples into over high-quality synthetic samples through the synergistic combination of four complementary augmentation techniques: SMOTE-like interpolation, adaptive Gaussian noise, mixup augmentation, and constrained feature permutation. Rigorous statistical validation confirmed exceptional augmentation quality with an overall score of 88.4%, demonstrating excellent preservation of critical statistical properties including distribution similarity (KS test pass rate: 85.7%), moment matching (87.3%), correlation structure (89.1%), and probabilistic characteristics (JS similarity: 91.4%). This comprehensive validation framework provides a template for assessing data augmentation quality in materials informatics applications.Among the evaluated architectures, the Residual DNN achieved the best individual model performance (R^2 = 0.9502, RMSE = 66.51 F/g).Comprehensive feature importance analysis through permutation testing, gradient sensitivity, SHAP, and LIME methods converged on consistent findings: voltage window dominates as the primary performance determinant (permutation importance: 0.8557), while specific surface area serves as the key optimization parameter (gradient sensitivity: 0.3838). The strong SHAP-LIME correlation (r = 0.847) and 100% agreement on top-3 features across all model architectures provide robust evidence that models learned physically meaningful relationships aligned with electrochemical principles. This work advances materials informatics by: (1) demonstrating that carefully designed multi-method augmentation enables deep learning to match traditional ML performance on small datasets, (2) establishing a comprehensive statistical validation framework for assessing augmentation quality, (3) providing dual interpretability through SHAP and LIME with cross-validation, and (4) delivering actionable materials design guidance prioritizing voltage window maximization followed by surface area optimization. Future work should explore: (1) extension to multi-target prediction (capacitance, energy density, power density simultaneously), (2) active learning integration to guide experimental sample selection, (3) transfer learning from related materials systems, and (4) uncertainty quantification through Bayesian deep learning approaches. The framework developed here provides a foundation for advancing data-efficient deep learning in materials science beyond supercapacitors to batteries, catalysts, and other functional materials.

Author contributions

Abdalla Abdelwahab: Data curation, Formal analysis, Investigation. Wael Z. Tawfik: Conceptualization, Data curation, Resources. Haifa A. Alqhtani: Funding acquisition, Resources. Mohamed A. M. Ali: Funding acquisition, Resources. Ahmed A. Allam: Funding acquisition, Resources. Hatem M. Noaman: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing, Supervision.

References

1. Iqbal, A., Sambyal, P., Kim, Y. H. *et al.* Advances in supercapacitor electrode materials. *Adv. Energy Sustain. Res.* **5**, 2400412 (2024).
2. Jiang, D. E., Meng, Z., Simon, P. *et al.* Mechanisms of energy storage in supercapacitors. *Nat. Nanotechnol.* **19**, 1325–1335 (2024).
3. Okafor, O. B., Mao, Y., Huang, Z. *et al.* Optimization strategies for supercapacitor materials. *J. Energy Storage* **93**, 112074 (2024).
4. Kumar, Y. A., Koyyada, G., Ramachandran, T. & Kim, J. H. Nanomaterials for supercapacitor applications. *Nanomaterials* **13**, 1049 (2023).
5. Zhang, F. *et al.* Design of electrode materials for supercapacitors. *Mater. Des.* **221**, 111017 (2022).
6. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
7. Schmidt, J., Marques, M. R., Botti, S. & Marques, M. A. Recent advances and applications of machine learning in solid-state materials science. *npj Comput. Mater.* **5**, 83 (2019).
8. Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
9. Raccuglia, P. *et al.* Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
10. Ward, L. *et al.* A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 1–7 (2016).
11. Bengio, Y., Goodfellow, I. & Courville, A. *Deep Learning*, vol. 1 (MIT Press, Cambridge, MA, USA, 2017).
12. Tawfik, W. Z. *et al.* Machine learning approaches for supercapacitor materials. *Phys. Scr.* **99**, 055001 (2024).
13. Tawfik, W. Z., Mohammad, S. N., Rahouma, K. H., Salama, G. M. & Tammam, E. Machine learning models for predicting specific capacitance of carbon-based supercapacitor electrodes. *Phys. Scr.* **99**, 026001 (2024).
14. Zhu, S. *et al.* Artificial neural network modeling of supercapacitor performance. *Mater. Lett.* **233**, 294–297 (2018).
15. Su, H. *et al.* Predicting the capacitance of carbon-based electric double layer capacitors using machine learning. *Nanoscale Adv.* **1**, 2162–2166 (2019).
16. Safder, U. *et al.* Machine learning for iot-oriented supercapacitor optimization. *Sustainability* **16**, 1516 (2024).
17. Tawfik, W. Z., Shaban, M., Raveendran, A., Lee, J. K. & Al-Enizi, A. M. Prediction of carbon nanotube supercapacitor capacitance using interpretable machine learning. *RSC Adv.* **15**, 3155–3167 (2025).
18. Zhong, X. *et al.* Explainable machine learning in materials science. *npj Comput. Mater.* **8**, 204 (2022).
19. Swain, P. K., Mishra, S. & Dash, S. Physicochemical feature engineering for supercapacitor performance prediction. *J. Energy Storage* **67**, 107438 (2023).
20. Mahmud, S. *et al.* Capacitive tendency classification using supervised machine learning. *Nat. Commun.* **15**, 1133 (2024).
21. Zhou, M., Vassallo, A. & Wu, J. Data-driven study of heteroatom-doped carbon electrodes for supercapacitors. *ACS Appl. Energy Mater.* **3**, 5993–6000 (2020).
22. Mathew, S., Karandikar, P. B. & Kulkarni, N. R. Machine learning optimization of jackfruit-seed-based supercapacitors. *Chem. Eng. Technol.* **43**, 1765–1773 (2020).
23. Wickramaarachchi, K. *et al.* Machine learning models for supercapacitor electrode materials. *Nanomaterials* **12**, 1847 (2022).
24. Zhang, Y., Chen, T. & Li, Q. Scalable deeper graph neural networks for materials property prediction. *npj Comput. Mater.* **9**, 72 (2023).
25. Xie, Y. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Nat. Commun.* **9**, 5139 (2018).
26. Li, Z., Zhang, Q. & Liu, X. Composition-based machine learning for materials property prediction. *npj Comput. Mater.* **8**, 45 (2022).
27. Kumar, A., Singh, R. & Kim, J. H. Machine-learning-assisted prediction of cerium oxynitride electrode performance. *J. Energy Storage* **68**, 107621 (2023).

28. Bian, X. Q., Zhang, Q., Zhang, L. & Chen, J. Grey-wolf-optimized support vector machine for solubility prediction of aromatic compounds in supercritical co_2. *Chem. Eng. Res. Des.* **123**, 284–294 (2017).

29. Ahmed, Q. I., Attar, H., Amer, A., Deif, M. A. & Solyman, A. A. Hybrid optimization for electrochemical systems. *Systems* **11**, 237 (2023).

30. Gao, Y., Li, C. & Huang, L. Advances in supercapacitor modeling using artificial intelligence. *Front. Energy Res.* **10**, 1032660 (2022).

31. Gheytanzadeh, M., Baghban, A., Habibzadeh, S., Mohaddespour, A. & Abida, O. Insights into improving supercapacitor performance using machine learning. *RSC Adv.* **11**, 5479–5486 (2021).

32. Bourenane, H., Berkani, A., Negadi, K., Marignetti, F. & Hebri, K. Modeling of supercapacitor systems with neural networks. *J. Eur. Syst. Autom.* **56**, 1 (2023).

33. Eddahech, A., Briat, O., Ayadi, M. & Vinassa, J. M. Modeling and adaptive control for supercapacitor in automotive applications. *Electr. Power Syst. Res.* **106**, 134–141 (2014).

34. Kumar, S., Li, W. & Zhao, D. Hybrid machine learning methods for energy storage materials. *J. Power Sources* **576**, 233012 (2023).

35. Marie-Françoise, J. N., Gualous, H. & Berthon, A. Supercapacitor modeling with artificial neural network (ann). In *Proceedings of the International Conference on Power Electronics and Drive Systems* (2004).

36. Ma, B. *et al.* Data augmentation for materials informatics via image style transfer. *npj Comput. Mater.* **6**, 125 (2020).

37. Weng, B., Zhu, R. & He, M. Data quantity governance in machine learning for materials science. *npj Comput. Mater.* **9**, 45 (2023).

38. Zhao, E., Wang, L. & Chen, T. Multi-task learning combined with data augmentation for materials property prediction. *npj Comput. Mater.* **9**, 88 (2023).

39. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for materials discovery. *Nature* **559**, 547–555 (2018).

40. Yang, X., Zhang, L. & Wang, J. Artificial intelligence in energy storage materials design. *EcoMat* **5**, e12245 (2023).

41. Liu, X. *et al.* Recent advances in supercapacitor electrode materials. *Chem. Eng. J.* **490**, 151625 (2024).

42. Gao, T. & Lu, W. Machine learning approaches for accelerating the discovery of energy storage materials. *iScience* **24**, 101990 (2021).

43. Mendhe, A. B. & Panda, H. S. Advances in supercapacitor electrode materials and machine learning approaches. *Next Res.* **2**, 100119 (2025).

44. Lobato-Peralta, D. R. *et al.* Machine-learning-guided design of carbon-based supercapacitor electrodes. *J. Power Sources* **617**, 235140 (2024).

45. Kothandam, G. *et al.* Recent progress in supercapacitor electrode materials. *Adv. Sci.* **10**, 2301045 (2023).

46. Zhang, Y., Liu, H. & Chen, G. Data-driven discovery of energy materials. *Acta Mater.* **258**, 119245 (2023).

47. Coccato, S., Barhmi, K., Lampropoulos, I., Golroodbari, S. & van Sark, W. Review of supercapacitor modeling approaches. *Batteries* **11**, 179 (2025).

48. Eddahech, A., Ayadi, M., Briat, O. & Vinassa, J. M. Supercapacitor performance prediction using neural networks. In *4th International Conference on Power Engineering, Energy and Electrical Drives*, 1460–1465 (IEEE, 2013).

49. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

50. Zhang, H., Cisse, M., Dauphin, Y. N. & Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *International Conference on Learning Representations (ICLR)* (2018).

Conflicts of interest

There are no conflicts to declare.

Data availability

The dataset used in this study is publicly available and was sourced from ref. 17. Code and augmentation scripts are available from the corresponding author upon reasonable request.

Acknowledgements

The authors acknowledge Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R458), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Funding

This work was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R458), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions statement

A.A. contributed to data curation, formal analysis and investigation. W.Z.T. contributed to conceptualization, data curation and resources. H.A.A., M.A.M.A. and A.A.A. contributed to funding acquisition and resources. C.M.K contributed to conceptualization data curation and review and editing. H.M.N. contributed to conceptualization, methodology, software, validation, supervision, and writing of the original draft and review and editing. All authors reviewed the manuscript.

Additional information

Competing interests: The authors declare no competing interests. **Data availability:** The dataset used in this study is publicly available and was sourced from Tawfik *et al.*[17]. Code and augmentation scripts are available from the corresponding author upon reasonable request.