Bridging Data-Rich and Data-Poor Domains on Lithium-Ion Battery via Scanning Electron Microscopic Data Through Convolutional Neural Network Transfer Learning

Haein Jeon
Department of Artificial Intelligence
Kyungpook National University
Daegu, South Korea
hjeon@knu.ac.kr

Donghun Lee
Department of Mathematics
Korea University
Seoul, South Korea
holy@korea.ac.kr

Bo-Yeong Kang *
Department of Smart Mobility Engineering
Kyungpook National University
Daegu, South Korea
kby09@knu.ac.kr

Jimin Oh *
Department of Smart Mobility Engineering
Kyungpook National University
Daegu, South Korea
ojmhiin@knu.ac.kr

Abstract

Scanning Electron Microscopy (SEM) imaging provides a powerful yet often under-utilized method for diagnosing the state of lithium-ion battery cathodes. However, deep learning models for SEM-based state prediction often struggle with limited training data and domain shifts, especially when functional electrolyte additives are added. In this work, we propose a two-stage transfer learning framework using an EfficientNet-B0 backbone to robustly classify cathode SEM images across nine classes defined by material composition (NCM333, NCM622, NCM811) and aging state (pristine, formation-aged, 100 cycles). Our method first pretrains the model on a data-rich source domain of additive-free samples, then fine-tunes it on a smaller target domain containing additive-induced variations. To address class imbalance, we compare targeted oversampling and weighted loss strategies. Experimental results show that our framework consistently outperforms pretraining-only and fine-tuning-only baselines, achieving over 0.98 accuracy and F1 scores for domain-shifted classes. Qualitative analysis with Grad-CAM shows that the model identifies important physical features, such as particle cracking and boundary degradation. These findings demonstrate the effectiveness of transfer learning in reducing data scarcity and domain shift in SEM-based battery diagnostics, providing a practical solution for automated analysis in the development of next-generation batteries. Our code is available at here.

1 Introduction

Lithium-ion batteries (LIBs) are widely used in electric vehicles and grid-level energy storage systems because of their high energy density and long cycle life. An accurate assessment of battery state, such as aging or functional degradation, is important for ensuring safety, reliability, and optimal

*Corresponding authors.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: AI for Accelerated Materials Design.

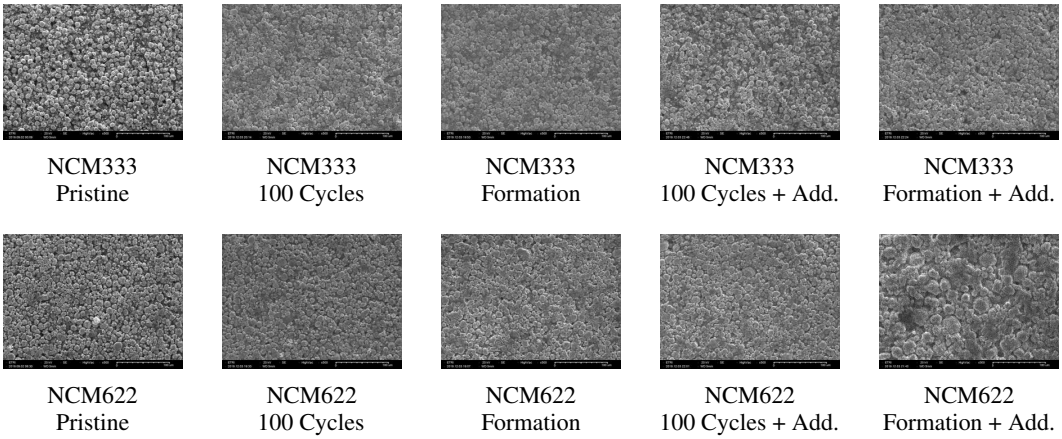

Figure 1: Representative SEM images of NCM cathode materials with varying nickel content (NCM333 and NCM622), cycling conditions (pristine, formation-aged, 100 cycles), and presence of functional additives.

performance [Zhang et al., 2025]. There are several factors that affect overall capacity and cyclability. Especially, a different nickel composition in lithium nickel cobalt manganese oxide cathode material affects the specific capacity and rate capability of the whole LIB. A typical LIB consists of a cathode, an anode, an electrolyte, and a separator, with the cathode influencing the overall behaviors of the battery. Various techniques have been used for battery state estimation, including electrochemical impedance spectroscopy, X-ray diffraction, and imaging-based methods. However, many of these approaches involve high-cost, complex equipment, or destructive testing procedures.

Among imaging techniques, scanning electron microscopy (SEM) is widely used for characterizing the morphology of cathode materials. SEM allows for direct visual inspection of microstructural changes linked to different electrochemical states, as shown in Figure 1. While SEM brightness can reflect variations in relative electron conductivity due to differences in electron density, it does not indicate changes in the elemental composition of the electrode. Particle morphology, however, can reveal whether a sample is pristine or cycled, since repeated charge-discharge cycles often cause particle fragmentation [Oh et al., 2024]. These features make SEM a convenient tool for assessing the chemical composition and degradation state of battery materials.

Furthermore, functional electrolyte additives are incorporated into positive electrode interfaces to enhance electrochemical performance, such as improving interfacial stability and suppressing side reactions [Pham et al., 2020]. Identifying these additive effects in the surface structure is important for understanding how they affect degradation behavior and performance changes. However, interpreting SEM images still requires manual analysis with considerable domain expertise [Wang et al., 2018, Sulzer et al., 2021].

To address this challenge, recent studies have begun to apply deep learning and machine learning techniques to automate SEM image analysis. These methods enable data-driven interpretation, reduce the reliance on expert heuristics, and can handle large-scale datasets more efficiently. For example, Oh et al. [2024] trained a CNN-based model using cathode SEM images to predict the composition of the material and the charge-discharge state, reporting high precision in the former but relatively limited performance in the latter. Despite these advances, labeled SEM data are still limited, especially for electrodes under diverse cycling conditions and functional electrolyte additives. This scarcity of data creates a problem for deep classifiers because they struggle to learn underrepresented classes.

We focus on the data-rich and data-poor situation. Transfer learning offers a solution to deal with data scarcity, using models that were pretrained on similar tasks and applying them to areas with scarce data, like medical imaging, remote sensing, and materials science. [Dip et al., 2024, Ma et al., 2024, Chen et al., 2024]. Models pretrained on data-rich source domains are adapted to target domains with limited labeled samples. However, its application to SEM-based battery state prediction remains unexplored. Most existing studies focus on training models from scratch, without leveraging transferable representations across experimental conditions or compositions.

In this work, we explore how transfer learning can be used for SEM-based classification of LIB cathodes. Specifically, our goal is to transfer knowledge learned from additive-free cathode images to improve classification performance on samples containing functional additives under various cycling conditions. We focus on NCM (Nickel-Cobalt-Manganese) cathodes, which are widely used in electric vehicles and grid-scale energy storage, with varying nickel content and cycling stages (e.g., pristine, formation-aged, and 100-cycle samples). By comparing models trained with and without transfer learning, we assess generalization performance and explore the effectiveness of pretrained representations in this domain.

The main contributions of this paper are as follows:

- We propose a transfer learning framework to address the data scarcity and class imbalance issues in SEM-based classification of LIB cathodes.
- We conduct a systematic evaluation across varying experimental conditions, including composition, additive presence, and cycle stages.
- We analyze the interpretability of the learned models using Grad-CAM visualization, providing insight into how morphological features contribute to prediction.

2 Related Work

Battery Life Prediction Recent studies have utilized machine learning to predict battery life and understand material degradation. In Severson et al. [2019]'s work, using features from the initial 5–100 charge/discharge cycles, a supervised machine learning model was trained to predict remaining useful life, achieving predictions within 9.1% error after only 100 initial cycles. While effective for early screening of cells for longevity, their model was limited to a single chemistry and aging protocol, so its generalizability to other cell types or conditions is limited. Zhang et al. [2025] introduced BatLiNet, a deep learning framework trained in a broad spectrum of aging conditions, such as varying charge protocols, temperatures, and chemistries, which achieved significantly lower error up to 40% than prior CNN approaches. However, BatLiNet struggled under highly mixed datasets, highlighting challenges in diverse conditions. Li et al. [2024] proposed a machine learning approach combining classification and regression to predict aluminum-ion battery cathode performance. They trained on known MXene cathode types and fine-tuned with a small amount of data from a new material, successfully predicting improved capacity and cycle life. This approach assumes the training materials' performance trends apply to the new material, and if the new material's behavior deviates radically, the model may need retraining. Yardimci and Ersoy [2025] sought to optimize deep learning architectures for classifying SEM image data categorized into three classes: non-defective, slightly defective, and defective. They examined how applying dimensionality reduction would impact the accuracy, speed, and resource usage of various convolutional neural network (CNN) architectures.

In addition, Oh et al. [2024] explored cathode composition and cycle state classification using SEM images. Their EfficientNet-based classifier [Tan and Le, 2019] achieved 99.6% accuracy on test SEM data, but performance degraded on unseen manufacturer samples with different additives, underscoring limited cross-domain robustness.

Transfer Learning under Data Scarcity Transfer learning has been widely adopted to address small-data challenges across various domains. Raghu et al. [2019] showed that the benefit of transfer learning from ImageNet to medical imaging often stems from weight scaling rather than feature reuse, and its advantage diminishes when sufficient labeled data are available. To overcome label scarcity, Chae and Kim [2023] proposed tailored transfer strategies, such as similar-image pretraining and RoI-focused fine-tuning, which significantly improved performance in cervical cancer and skin lesion analysis. Similarly, Song et al. [2024] reported that ImageNet-pretrained CNNs, combined with extensive augmentation and weighted cross-entropy loss, effectively mitigated overfitting and class imbalance in small medical datasets. In the materials domain, Semitela et al. [2025] applied transfer learning using ResNet-50 and Inception V3 for SEM-based surface defect classification, achieving improved performance under data scarcity. Liu et al. [2023] further enhanced transfer learning for imbalanced datasets through active sampling of minority classes. Beyond these applications, Maxwell et al. [2024] found that ImageNet-based transfer learning provided moderate gains in remote sensing tasks, suggesting that unsupervised or semi-supervised methods may be preferable when labeled data are extremely limited.

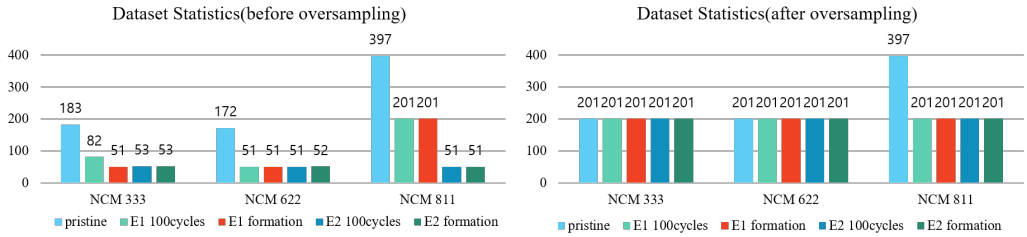

Figure 2: Class distribution before and after oversampling. Top: original sample counts per class. Bottom: balanced dataset with oversampling to 201 samples per class.

Summary of Gaps Overall, battery life prediction has advanced through early-cycle feature learning and inter-cell deep networks, but faces generalization issues under different chemistries and additives. Our work bridges these gaps by employing transfer learning strategies for SEM-based battery data under domain shift, leveraging pretraining on additive-free samples and fine-tuning on additive-containing data.

3 Proposed Method

We formulate the SEM image classification task as a multi-class supervised learning problem. Given a grayscale SEM image $x \in \mathbb{R}^{H \times W}$, the goal is to predict its corresponding class label $y \in \{1, \ldots, K\}$, where the class represents a combination of cathode composition (NCM333, NCM622, NCM811) and cycled state (pristine, formation-aged, 100 cycles). This task is challenged by limited labeled samples and class imbalance, particularly for additive-containing samples.

3.1 Model Architecture

We adopt EfficientNet-B0 as the backbone architecture, which has 5.3 million parameters and provides a good trade-off between efficiency and accuracy [Tan and Le, 2019]. The network is initialized with ImageNet-pretrained weights and fine-tuned on our dataset after replacing the final classification layer to match K classes. Pretraining on ImageNet, despite domain differences, can help improving convergence and generalization [Yosinski et al., 2014].

3.2 Data Preprocessing and Augmentation

Each SEM image is resized to 224×224. The raw dataset contains 1,700 images before augmentation. To enhance generalization, we apply random flips, rotations (±20°), contrast adjustment, and scaling during training. Resizing and contrast adjustment can potentially introduce artifacts; however, these operations were carefully applied to emulate variations that naturally occur under different SEM acquisition conditions, thereby improving the model's robustness in both data-rich and data-poor scenarios. To address class imbalance, we oversample the training set by replicating samples for each class until it matches the size of the second-largest class (201 samples). The resulting balanced class distribution is shown in Figure 2.

3.3 Framework Overview

We propose a two-stage transfer learning framework to improve classification performance under data scarcity and imbalance. The framework includes (1) pretraining on an additive-free source dataset, and (2) fine-tuning on a smaller target dataset containing additives.

Formally, let

$$\begin{aligned}
\mathcal{D}_{\text{src}} &= \{(x_i^{(s)}, y_i^{(s)})\}_{i=1}^{N_s}, && \text{(source, additive-free)} \\
\mathcal{D}_{\text{tgt}} &= \{(x_j^{(t)}, y_j^{(t)})\}_{j=1}^{N_t}, && \text{(target, with additives)} \\
f_\theta &: \mathbb{R}^{H \times W} \to \mathbb{R}^K, && \text{(CNN classifier)} \\
p &= \text{softmax}(f_\theta(x)), && \text{(predicted probabilities).}
\end{aligned}$$

After resizing, the grayscale SEM image is passed through EfficientNet-B0 [Tan and Le, 2019], which begins with a series of depthwise separable convolutions and squeeze-and-excitation blocks to extract low- to mid-level features. The backbone progressively encodes hierarchical morphological patterns while maintaining parameter efficiency through compound scaling.

The resulting high-dimensional feature map is then fed into a global average pooling layer, reducing spatial dimensions while preserving semantic content. Finally, a fully connected classification layer (replaced to match K target classes) produces a logit vector, which is converted to class probabilities via softmax activation.

This pipeline enables the model to capture fine surface features related to cathode composition and cycling state, supporting reliable SEM-based battery classification.

Stage 1: Pretraining on Source Dataset

The model is first trained on $\mathcal{D}_{\text{src}}$ using standard cross-entropy loss:

$$\mathcal{L}_{\text{src}}(\theta) = -\frac{1}{N_s} \sum_{i=1}^{N_s} \sum_{k=1}^{K} \mathbb{1}(y_i^{(s)} = k) \cdot \log p_k^{(i)}. \tag{1}$$

Stage 2: Fine-tuning on Target Dataset

The pretrained weights θ_{src} are used to initialize the model for fine-tuning on $\mathcal{D}_{\text{tgt}}$. We compare two fine-tuning approaches:

(a) Standard Cross-Entropy Loss

$$\mathcal{L}_{\text{tgt}}(\theta) = -\frac{1}{N_t} \sum_{j=1}^{N_t} \sum_{k=1}^{K} \mathbb{1}(y_j^{(t)} = k) \cdot \log p_k^{(j)}. \tag{2}$$

Oversampling is applied to balance the class distribution in this setting.

(b) Weighted Cross-Entropy Loss To address class imbalance without oversampling, we assign class weights $w_k = 1/n_k$ with n_k the number of samples in class k, and minimize

$$\mathcal{L}_{\text{wce}}(\theta) = -\tfrac{1}{N_t} \sum_{j=1}^{N_t} \sum_{k=1}^{K} w_k \, \mathbb{1}(y_j^{(t)} = k) \, \log p_k^{(j)}.$$

4 Experiments

4.1 Dataset

We performed SEM imaging on various NCM cathode samples with four different nickel compositions and three cycle states. The SEM images were captured at 500× magnification and an operating voltage of 20 kV using secondary electrons, following the methodology described in Oh et al. [2024]. The initial dataset contained 1,700 images before augmentation, covering both additive-free (E1) and additive-containing (E2) samples. For fair evaluation, the test set was separated from the experimental data used for training, and an 80:20 train–test split was applied. On the training portion, we performed 5-fold cross-validation to assess model stability and generalization.

To address class imbalance, all classes were oversampled to match the sample count of the second most frequent class, and data augmentation techniques—including random cropping, flipping, rotation, and their composites—were applied. This process expanded the dataset to a total of 3,211 images.

We plan to make the dataset available upon reasonable request for further research in SEM-based battery diagnostics.

4.2 Evaluation Metric

We evaluate our models using class-wise accuracy and F1 score.

$$\mathrm{Acc}_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c + \mathrm{FN}_c},\ \mathrm{F1}_c = 2 \times \frac{P_c R_c}{P_c + R_c},\ P_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FP}_c},\ R_c = \frac{\mathrm{TP}_c}{\mathrm{TP}_c + \mathrm{FN}_c}. \tag{3}$$

where TP_c is the number of true positives, FP_c is the number of false positives, and FN_c is the number of false negatives for class c. This metric measures how well the model predicts each individual class. The F1 score for class c is the harmonic mean of precision and recall, providing a balanced measure with our class-imbalanced dataset.

4.3 Baselines

We conducted experiments and statistical comparisons against prior work. We evaluated four different training strategies:

- **Pre+FT(OS)**: Pretraining on E1 data, followed by finetuning on E2 data with oversampling, and testing on E2.
- **Pre+FT(WL)**: Same as above, but using weighted cross-entropy loss instead of oversampling during finetuning.
- **FT(OS)**: Direct training on E2 data with oversampling, without E1 pretraining.
- **Pre Only**: Pretraining on E1 data and directly testing on E2 without additional finetuning.

We examine the impact of pretraining on additive-free data and subsequent finetuning on additive data, as well as to compare oversampling versus weighted loss for addressing class imbalance.

For qualitative evaluation, we use GradCAM [Selvaraju et al., 2019] to visualize and interpret the spatial regions of the SEM images that the model considers when making predictions. We analyze whether the learned features correspond to meaningful physical structures relevant to NCM composition or cycle state.

Additionally, we perform an ablation study, training different backbone models to compare their training performance. We report training time, overall accuracy, and F1 score to assess if there are trade-offs between computational efficiency and predictive performance.

4.4 Models

For our ablation study, we employed four different backbone architectures: EfficientNet-B0 [Tan and Le, 2019], ResNet-18 [He et al., 2015], MobileNet-v2 [Sandler et al., 2019], and DenseNet-201 [Huang et al., 2017]. These models were selected to cover a range of parameter sizes and computational complexities. Specifically, EfficientNet-B0 and MobileNet-v2 serve as lightweight models, while DenseNet-201 represents deeper architectures with larger parameter counts.

4.5 Experimental Setup

All experiments were performed with MATLAB R2024b. We installed additional toolboxes and add-ons to load pretrained networks and perform deep learning operations. The computations were executed on a device equipped with an Intel i5-12400 processor, 12GB RAM, and an NVIDIA RTX 3060 GPU.

We performed oversampling by replicating samples for each class up to 201 instances. Input images were resized to 224×224. Training used a mini-batch size of 16, the Adam optimizer with a learning rate of 0.001, and was run for a single epoch per stage. We performed 5-fold cross-validation on the training set, and for all models, repeated training ten times to report averaged results.

4.6 Experimental Results and Analysis

Table 1 compares per-class classification metrics under four training strategies. The proposed Pretraining + Finetuning approach consistently outperforms both the finetuning-only baseline and

Method	Pre+FT(OS)		Pre+FT(WL)		FT(OS)		Pre only	
Class	**Accuracy**	**F1**	**Accuracy**	**F1**	**Accuracy**	**F1**	**Accuracy**	**F1**
ncm333 pristine*	0.97	0.96	0.91	0.93	0.83	0.91	0.92	0.90
ncm333 100cycles	0.98	0.99	0.93	0.97	0.91	0.94	0.59	0.57
ncm333 formation	1.00	0.98	0.98	0.98	0.87	0.90	0.17	0.27
ncm622 pristine*	0.98	0.99	0.90	0.95	0.99	0.96	0.99	0.95
ncm622 100cycles	0.96	0.98	0.96	0.97	0.84	0.91	0.10	0.15
ncm622 formation	0.99	0.99	0.85	0.91	0.84	0.91	0.93	0.64
ncm811 pristine*	0.94	0.95	0.93	0.99	0.96	0.97	0.99	0.99
ncm811 100cycles	0.98	0.98	0.85	0.89	0.92	0.89	0.85	0.89
ncm811 formation	0.99	0.98	0.86	0.92	0.88	0.94	0.16	0.18

Table 1: Comparison of class-wise Accuracy and F1 Scores between Pre+FT(OS), Pre+FT(WL), FT(OS), and Pre only. * indicates pristine classes without an additive, thus no domain shift.

the pretraining-only method across nearly all 9 classes. In particular, Pre+FT with oversampling (Pre+FT(OS)) achieves the highest accuracy and F1 scores, often near 0.98–0.99, whereas the baseline FT(OS) model lags behind (typically in the 0.85–0.95 F1 range). The prior Pre only model performs worst on most classes, with some F1 scores falling below 0.3. This result indicates the effect of combining the pretraining stage of SEM data not containing additive with target-domain finetuning.

Our approach uses the features learned from pretraining while adapting to the new data distribution. For example, in the NCM333 formation class, Pre+FT(OS) achieves an F1 score of $\approx$0.98, whereas FT(OS) and Pre only yield $\sim$0.90 and 0.27, respectively, showing the effectiveness of our approach. Similar trends hold across other categories, showing that our Pre+FT method offers a substantial gain over the finetuning-only baseline and the previous pretraining-only model in this battery SEM image classification task.

Class-Specific Benefits of Finetuning The results also reveal that certain classes benefit more from finetuning than others, depending on their susceptibility to domain shift. Classes representing cycled or formation states (i.e. aged electrodes) see greater performance gains with finetuning. These classes exhibit significant domain-specific features (e.g. particle cracking) that were not fully captured by the pretraining alone. Consequently, without finetuning, the pretraining-only model struggled on these classes (e.g., for some 100-cycle and formation categories). Once fine-tuned on target data, however, the model's F1 for these classes jumps by 0.4–0.7 points in many cases, reaching $\sim$0.98 F1. This indicates that adaptation is effective for domain-shifted classes in the SEM analysis task. The Pristine classes (unaged cathodes) exhibit high performance even before finetuning because these images appear domain-invariant. As a result, finetuning yields relatively modest or no gains on pristine classes, since the model already generalizes well to them.

Impact of Oversampling vs. Weighted Loss Table 1 also contrasts two data imbalance mitigation strategies: oversampling (OS) vs. class-weighted loss (WL) within the Pre+FT framework. Oversampling is slightly more effective overall in our experiments. The Pre+FT(OS) variant achieves equal or higher F1 scores than Pre+FT(WL) for most classes, particularly for minority classes that had a few training examples. For example, in the NCM811 100-cycles class, Pre+FT with oversampling reaches $\sim$0.98 F1, compared to $\sim$0.89 F1 with weighted loss. Similar advantages are seen in other underrepresented classes (e.g., NCM811 formation). This suggests that duplicating minority class examples (with appropriate augmentation) gave the model more opportunities to learn their distinctive features, leading to better recognition of those classes than using a weighted loss alone. The weighted-loss approach still improved performance on rare classes relative to no balancing, but it did not fully close the gap – likely because simply increasing loss weight does not provide additional diverse exposures to minority-class patterns. In summary, oversampling produced superior or on-par performance across the board, with no observed overfitting detriment in this case, making it the preferred imbalance handling strategy for our task.

Generalization Limits of the Pretraining-Only Model The pretraining-only (Pre only) scenario highlights significant generalization limitations. In this setting, a model trained on a source SEM image dataset is applied directly to the target classification task without finetuning. While this model

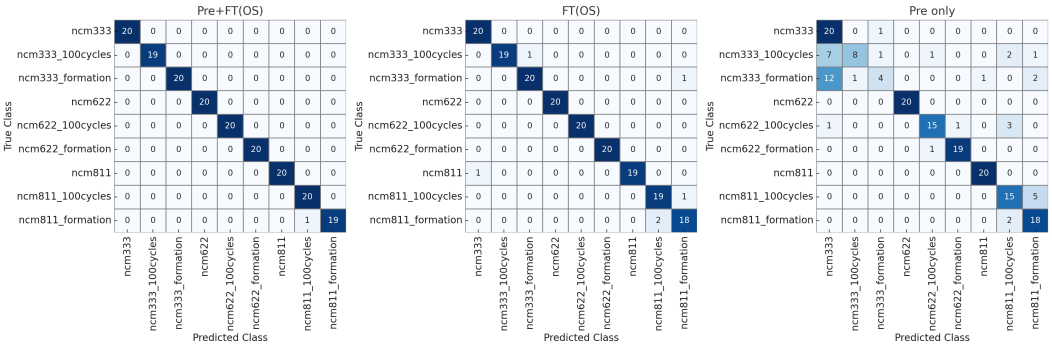

Figure 3: Confusion matrices for three training settings: (Left) Pre+FT(OS) — proposed two-stage training with oversampling; (Middle) FT(OS) — direct training on the target data with oversampling; and (Right) Pre only — pretrained model without fine-tuning. Each matrix shows predictions across nine classes.

(from a prior study of Oh et al. [2024]) achieved very high accuracy on its original training domain, it transfers poorly to our new dataset for several classes. As noted above, Pre only performance falls on classes involving different cycle states. For multiple 100 cycles and formation categories, it scored below 20–30% F1 in Table 1. These failures indicate that pretraining alone is insufficient for robust cross-domain generalization when the target data distribution shifts. Even for compositions seen during pretraining, changes in state (aging) can introduce new textures that the frozen pretrained model cannot interpret, resulting in misclassifications. This emphasizes why our proposed combined approach is needed: pretraining must be coupled with target-data finetuning to achieve high accuracy on new SEM image domains.

Model	TrainTime	Accuracy	F1
EfficientNet-B0	43.38	0.94	0.94
ResNet-18	26.62	0.66	0.59
MobileNet-v2	39.34	0.91	0.91
DenseNet-201	297.34	0.82	0.8

Table 2: Comparison of CNN backbones under the two-stage transfer learning setting. EfficientNet-B0 achieved the best trade-off between accuracy, F1 score, and training time.

Comparative Performance Evaluation of Models with fine-tuning We compared four CNN backbones: EfficientNet-B0, ResNet-18, MobileNet-v2, and DenseNet-201. All models were under the same two-stage transfer learning setup, pretrained on ImageNet and fine-tuned on our SEM dataset using the Pre+FT(OS) strategy.

As summarized in Table 2, EfficientNet-B0 achieved the best balance between accuracy (0.94), F1 score (0.94), and training time (43 s). MobileNet-v2 also performed well (accuracy 0.91, F1 0.91) with lower computational cost, suggesting that lightweight models can be effective for SEM transfer learning. In contrast, ResNet-18 showed limited accuracy (0.66) and F1 score (0.59), likely due to its shallower architecture, while DenseNet-201 incurred high training cost (297 s) without performance gains.

Overall, EfficientNet-B0 offered the best trade-off between performance and efficiency, indicating that its compound scaling is well-suited for capturing multi-scale morphological features in SEM images.

Qualitative Analysis To interpret the model's decision-making behavior and evaluate its alignment with domain-specific knowledge in battery materials, we conducted qualitative analyses using the confusion matrix and Grad-CAM visualizations.

The confusion matrix in Figure 3 reveals strong class-wise performance across nine categories, with especially high accuracy in pristine classes and well-separated compositions such as NCM811

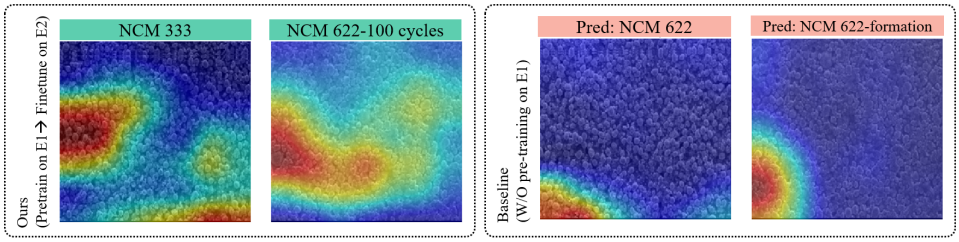

Figure 4: GradCAM comparison

pristine and NCM622 100 cycles. Misclassifications primarily occur between classes with similar morphological degradation patterns—for example, between NCM333 formation-aged and 100-cycle samples—highlighting the difficulty in visually distinguishing intermediate aging states based solely on SEM features. These observations align with known electrochemical behavior, where early-stage degradation often manifests subtly in particle boundaries, crack initiation, and surface texturing.

To enhance interpretability, we applied Grad-CAM to visualize the spatial attention maps of the trained model. As shown in Figure 4, the model primarily attends to regions with physical significance: smooth surfaces in pristine samples, and microcracks, voids, or fragmented grains in aged or additive-containing samples. This indicates that the model utilizes morphological cues relevant to battery state and composition rather than spurious features.

5 Conclusion

In this study, we addressed SEM-based state prediction for lithium-ion battery cathodes under data scarcity and domain shift caused by electrolyte additives. We proposed a two-stage transfer learning framework using EfficientNet-B0, which learns general features from additive-free SEM images and adapts to additive-containing samples through fine-tuning. To address class imbalance, targeted oversampling was applied and compared with weighted loss strategies.

Experiments across nine cathode classes—covering compositions (NCM333, NCM622, NCM811) and cycling states (pristine, formation, 100 cycles)—showed that our approach consistently outperforms finetuning-only and pretraining-only baselines. The model achieved over 0.98 accuracy and F1 scores even for domain-shifted classes such as NCM333 formation. Grad-CAM analysis confirmed that it captures physically relevant features, including surface cracking and particle boundaries.

This study reflects a common challenge in battery research: the tension between the need for data-driven modeling and the practical limitations of collecting large, well-controlled datasets. The proposed framework alleviates this issue by enabling robust classification and state prediction in data-scarce conditions.

Although demonstrated with SEM images, the method is readily applicable to other characterization modalities such as OM, XRD, XPS, and TEM. These can be integrated into a multimodal learning framework to further enhance predictive insights into electrode structure and degradation.

Overall, this work demonstrates that transfer learning can substantially improve SEM-based state prediction in practical, data-limited environments, paving the way toward scalable and automated diagnostics for advanced battery materials.

Acknowledgments

This present research has been conducted by 2025 Dongil Culture Foundation, Korea (202520400000). This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-02217276). This research was supported by the Technology Innovation Program (RS-2025-19532970) funded By the Ministry of Trade, Industry and Energy (MOTIE, Korea).

References

Jinyeong Chae and Jihie Kim. An investigation of transfer learning approaches to overcome limited labeled data in medical image analysis. *Applied Sciences*, 13(15):8671, 2023.

An Chen, Zhilong Wang, Karl Luigi Loza Vidaurre, Yanqiang Han, Simin Ye, Kehao Tao, Shiwei Wang, Jing Gao, and Jinjin Li. Knowledge-reused transfer learning for molecular and materials science. *Journal of Energy Chemistry*, 98:149–168, 2024.

Sajib Acharjee Dip, Kazi Hasan Ibn Arif, Uddip Acharjee Shuvo, Ishtiaque Ahmed Khan, and Na Meng. Equitable skin disease prediction using transfer learning and domain adaptation. In *Proceedings of the AAAI Symposium Series*, volume 4, pages 259–266, 2024.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL `https://arxiv.org/abs/1512.03385`.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

Jiahui Li, Shengkun Xi, Tongxing Lei, Ruijun Yao, Fanshuai Zeng, Junwei Wu, Shaobo Tu, and Xingjun Liu. Machine learning assisted prediction in the discharge capacities of novel mxene cathodes for aluminum ion batteries. *Journal of Energy Storage*, 82:110196, 2024.

Yang Liu, Guoping Yang, Shaojie Qiao, Meiqi Liu, Lulu Qu, Nan Han, Tao Wu, Guan Yuan, and Yuzhong Peng. Imbalanced data classification: Using transfer learning and active sampling. *Engineering Applications of Artificial Intelligence*, 117:105621, 2023.

Yuchi Ma, Shuo Chen, Stefano Ermon, and David B Lobell. Transfer learning in environmental remote sensing. *Remote Sensing of Environment*, 301:113924, 2024.

Aaron E Maxwell, Sarah Farhadpour, and Muhammad Ali. Exploring transfer learning for anthropogenic geomorphic feature extraction from land surface parameters using unet. *Remote Sensing*, 16(24):4670, 2024.

Jimin Oh, Jiwon Yeom, Benediktus Madika, Kwang Man Kim, Chi Hao Liow, Joshua C Agar, and Seungbum Hong. Composition and state prediction of lithium-ion cathode via convolutional neural network trained on scanning electron microscopy images. *npj computational materials*, 10(1):88, 2024.

Hieu Quang Pham, Marta Mirolo, Mohamed Tarik, Mario El Kazzi, and Sigita Trabesinger. Multifunctional electrolyte additive for improved interfacial stability in ni-rich layered oxide full-cells. *Energy Storage Materials*, 33:216–229, 2020.

Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems*, 32, 2019.

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019. URL `https://arxiv.org/abs/1801.04381`.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, 128(2):336–359, October 2019. ISSN 1573-1405. doi: 10.1007/s11263-019-01228-7. URL `http://dx.doi.org/10.1007/s11263-019-01228-7`.

Ângela Semitela, Miguel Pereira, António Completo, Nuno Lau, and José P Santos. Improving industrial quality control: A transfer learning approach to surface defect detection. *Sensors*, 25(2): 527, 2025.

Kristen A Severson, Peter M Attia, Norman Jin, Nicholas Perkins, Benben Jiang, Zi Yang, Michael H Chen, Muratahan Aykol, Patrick K Herring, Dimitrios Fraggedakis, et al. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy*, 4(5):383–391, 2019.

Zhenyu Song, Zhanling Shi, Xuemei Yan, Bin Zhang, Shuangbao Song, and Cheng Tang. An improved weighted cross-entropy-based convolutional neural network for auxiliary diagnosis of pneumonia. *Electronics (2079-9292)*, 13(15), 2024.

Valentin Sulzer, Peyman Mohtat, Antti Aitio, Suhak Lee, Yen T Yeh, Frank Steinbacher, Muhammad Umer Khan, Jang Woo Lee, Jason B Siegel, Anna G Stefanopoulou, et al. The challenge and opportunity of battery lifetime prediction from field data. *Joule*, 5(8):1934–1955, 2021.

Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019.

Xuefeng Wang, Yejing Li, and Ying Shirley Meng. Cryogenic electron microscopy for characterizing and diagnosing batteries. *Joule*, 2(11):2225–2234, 2018.

Cagri Yardimci and Mevlut Ersoy. Optimizing deep learning architectures for sem image classification using advanced dimensionality reduction techniques. *Research on Engineering Structures and Materials*, 11, 2025.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Han Zhang, Yuqi Li, Shun Zheng, Ziheng Lu, Xiaofan Gui, Wei Xu, and Jiang Bian. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nature Machine Intelligence*, pages 1–8, 2025.