

# Classification of Phase-Resolved Partial Discharge Images from Electrical Generators using Convolutional Neural Networks

E. Esenov, S. Sridhar  
Chalmers University of Technology  
[{esenov,sris}@chalmers.se](mailto:{esenov,sris}@chalmers.se)

**Abstract**—This work investigates the use of convolutional neural networks (CNNs) for classifying phase-resolved partial discharge (PRPD) images from electrical generators. Using the public dataset introduced by Zorrilla Henao *et al.* [1], which contains labeled PRPD patterns corresponding to internal, surface, and corona discharges, we first establish a compact baseline CNN model and evaluate common architectural and training modifications through a series of ablation studies. Batch normalization and increased network depth yielded the most consistent performance gains, improving both accuracy and  $F_1$ -score. Finally, transfer learning with a fine-tuned EfficientNetV2-B0 model pretrained on ImageNet achieved near-perfect classification across all discharge types. While these results demonstrate the potential of transfer learning for partial discharge analysis, they may also reflect the limited diversity and heavy augmentation of the examined dataset. Future work should assess generalization on larger and more challenging real-world data to validate the practical applicability of such models for insulation diagnostics.

## I. INTRODUCTION

Partial discharges (PDs) are localized electrical breakdown phenomena that occur within the insulation systems of electrically stressed equipment. Each discharge results from ionization within a confined region of the dielectric that does not span the entire insulation thickness. Although individual PD events are typically non-destructive, their cumulative effect can progressively deteriorate the insulation and eventually cause failure if the initiating defect is not detected and mitigated [2].

Conventional PD diagnosis relies heavily on expert interpretation of patterns obtained from time-, frequency-, or phase-resolved measurements [3]. In traditional workflows, diagnostic features, such as discharge magnitude distributions, phase occurrence, or statistical descriptors, are manually engineered to characterize the insulation condition. Recent advances in deep learning (DL) [4] have enabled the automated extraction of such diagnostic features, allowing neural networks to learn discriminative representations directly from raw or image-based data. Among these data formats, *phase-resolved partial discharge* (PRPD) images are particularly informative, as they represent the distribution of discharge pulses across the AC cycle, yielding characteristic patterns linked to specific defect mechanisms.

Code and trained models available at: <https://github.com/emiresenov/PRPD-Classification>

To support research on PD classification, Zorrilla Henao *et al.* [1] released a dataset of PRPD images obtained from both field measurements on operating power generators and controlled laboratory experiments. The dataset captures characteristic PD patterns associated with internal, surface, and corona defects, and offers a useful resource for developing and evaluating machine learning methods in insulation diagnostics.

In this paper, we investigate the performance of modern DL methods for PD classification on the Zorrilla Henao *et al.* dataset. We begin the study by establishing a baseline model with a compact convolutional neural network (CNN) [5] trained and evaluated on the dataset. Subsequently, a series of ablation studies is conducted to analyze the effect of common architectural and optimization choices, including batch normalization, dropout, learning rate scheduling, and network depth. Finally, we examine the applicability of transfer learning [6] by adapting a pretrained vision model to the PD classification task, thereby assessing whether feature representations learned from large-scale natural image datasets can enhance diagnostic accuracy for insulation systems.

## II. METHODOLOGY

### A. Dataset

The experiments were conducted using the public PRPD image dataset introduced by Zorrilla Henao *et al.* [1]. The dataset contains 945 labeled RGB images representing three discharge types: internal (321), surface (316), and corona (308). Each image has a resolution of  $570 \times 440$  pixels and was preprocessed to remove background regions unrelated to the discharge activity.

The collection includes field measurements from more than 40 power generators across Colombia, as well as laboratory simulations performed using an OMICRON PD testing system. To improve class balance, data quantity, and visual diversity, the authors applied controlled data augmentation involving random rotations, cropping, zooming, and horizontal flipping, with probabilities carefully chosen to preserve the diagnostic features of the PD patterns. Each discharge type contains between 9 and 24 combined real and simulated samples, complemented by approximately 300 augmented images. All images were released in a structured form suitable for direct use in machine learning experiments.

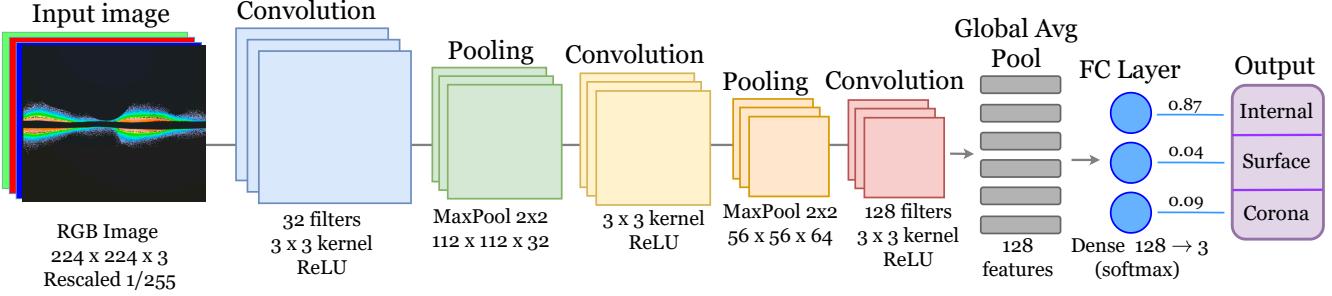


Fig. 1. Architecture of the baseline CNN used for PRPD image classification. The model consists of three convolutional-pooling blocks followed by global average pooling, dropout, and a softmax output layer producing class probabilities for internal, surface, and corona discharges.

### B. Preprocessing and Splits

In this study, all images were resized to  $224 \times 224$  pixels to reduce computational cost and accelerate training while maintaining sufficient spatial detail for reliable classification. Pixel intensities were normalized to the  $[0, 1]$  range to ensure consistent numerical scaling across inputs and improve the stability of gradient-based optimization. To enable balanced and reproducible evaluation, the dataset was randomly divided into 70% training, 15% validation, and 15% test subsets, stratified by class. No additional data augmentation was applied during training, as the original dataset already incorporated extensive augmentation.

### C. Model and Training Procedure

All experiments were implemented in Python using TensorFlow and the Keras API. Data management and preprocessing were performed with NumPy and `tf.data` pipelines. Model training was carried out in Google Colab using an NVIDIA T4 GPU-accelerated environment to reduce computational time.

The baseline model followed a compact CNN architecture that consisted of three convolutional layers with  $3 \times 3$  kernels and ReLU activations, using 32, 64, and 128 filters, respectively. Each convolutional layer was followed by a max-pooling layer for spatial downsampling. Input images were normalized to the  $[0, 1]$  range. The resulting feature maps were aggregated via global average pooling and passed through a dropout layer with a rate of 0.4 to reduce overfitting. The pooled features were then fed into a fully connected softmax layer for classification across the three target classes. The network was trained for 50 epochs using the Adam optimizer with a learning rate of  $10^{-3}$  and a sparse categorical cross-entropy loss, with a batch size of 32.

To assess the impact of common architectural and training modifications on overall model convergence and generalization, a series of ablation studies was conducted. Each ablation altered a single configuration of the baseline model while keeping all others fixed. Specifically, the following variants were examined: (i) the addition of batch normalization layers after each convolution; (ii) replacement of the Adam optimizer with AdamW including a weight decay term of  $10^{-4}$ ; (iii) the application of label smoothing with  $\varepsilon = 0.1$ ; and (iv) an

increased-depth model incorporating one additional convolutional block with 256 filters.

Finally, we explored transfer learning using the EfficientNetV2B0 [7] architecture pretrained on ImageNet. The pre-trained classification head was replaced by a global average pooling layer, dropout ( $p = 0.4$ ), and a new three-class softmax layer for PD classification. In the first training phase, the convolutional base was frozen and optimized with AdamW (learning rate =  $10^{-3}$ , weight decay =  $10^{-4}$ ) for up to 50 epochs with early stopping. In the second phase, the last 20 layers of the base were unfrozen and fine-tuned using a reduced learning rate of  $10^{-5}$  for up to 20 epochs. This two-stage training scheme aimed to adapt general visual representations from ImageNet to the PRPD domain while avoiding overfitting to the relatively small dataset.

### D. Evaluation

Model performance was quantitatively assessed using standard multi-class classification metrics, including accuracy, precision, recall, and  $F_1$ -score.

Let  $y_i \in \{1, \dots, C\}$  denote the true label and  $\hat{y}_i$  the predicted label for the  $i$ -th sample, with  $N$  total samples and  $C$  classes. Overall accuracy measures the proportion of correctly classified samples and is defined as

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\hat{y}_i = y_i), \quad (1)$$

where  $\mathbb{1}(\cdot)$  is the indicator function that equals 1 if the argument is true and 0 otherwise. For each class  $c$ , precision and recall are computed as

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c}, \quad \text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c}, \quad (2)$$

where  $\text{TP}_c$ ,  $\text{FP}_c$ , and  $\text{FN}_c$  denote the numbers of true positives, false positives, and false negatives for class  $c$ , respectively. The  $F_1$ -score combines precision and recall as their harmonic mean, reflecting the balance between false positives and false negatives:

$$F_{1,c} = 2 \cdot \frac{\text{Precision}_c \cdot \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}. \quad (3)$$

Finally, a confusion matrix has the structure  $\mathbf{M} \in \mathbb{N}^{C \times C}$ , where each entry  $M_{ij}$  represents the number of samples

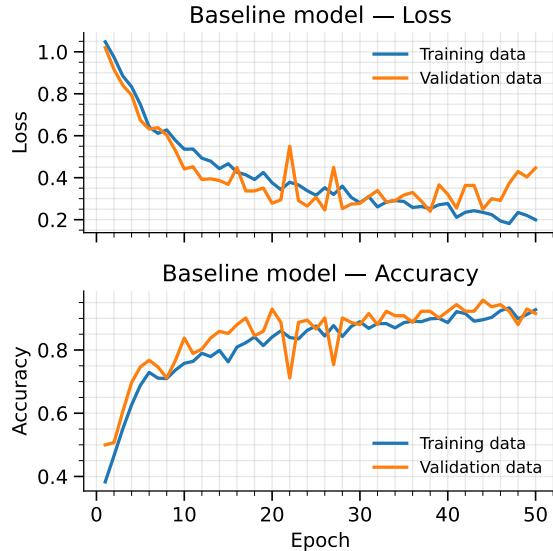


Fig. 2. Training and validation curves for the baseline CNN model. The plots show gradual convergence of both loss and accuracy over 50 epochs, with minor oscillations indicating training instability and potential overfitting.

belonging to class  $i$  that were predicted as class  $j$ . Diagonal elements correspond to correctly classified instances, while off-diagonal elements indicate misclassifications. All metrics were computed using the `scikit-learn` library.

### III. RESULTS AND DISCUSSION

#### A. Baseline CNN model

The baseline CNN showed overall convergence across training and validation datasets, as illustrated in Fig. 2, with both loss and accuracy improving over 50 epochs. However, the training process displayed signs of instability, evident in the pronounced oscillations in both metrics and the partial divergence between training and validation loss. Such fluctuations may result from a relatively high learning rate, sensitivity to batch variability, or limited regularization.

Quantitatively, the baseline model reached a validation accuracy of 0.9155 and an  $F_1$ -score of 0.9143. On the held-out test dataset, the model achieved an accuracy of 0.8451, precision of 0.8788, recall of 0.8472, and an  $F_1$ -score of 0.8359 (Table I). These results suggest that the network successfully captured key spatial-phase features of PRPD patterns.

#### B. Ablation study results

Ablation experiments examined the impact of architectural and training modifications relative to the baseline (Table I). Among the tested variants, batch normalization yielded some of the most consistent improvements, increasing test accuracy from 0.8451 to 0.9859 and  $F_1$ -score to 0.9860. This suggests that normalization effectively stabilized gradient updates and mitigated internal covariate shift. The AdamW optimizer also improved performance, achieving a test accuracy of 0.9437, indicating that explicit weight decay provided a beneficial form of regularization.

In contrast, label smoothing offered marginal gains on the validation set and slightly lower test performance. Increasing the network depth with one additional convolutional block produced the highest validation accuracy (0.9789) but did not translate to further improvements on the test set. Overall, these results show that normalization and depth were the most influential factors in enhancing performance, while other modifications offered less significant returns on this dataset.

#### C. Fine-tuned transfer learning model

The fine-tuned EfficientNetV2-B0 model converged rapidly within the first ten epochs, achieving close alignment between training and validation curves (Fig. 4). Loss decreased smoothly and stabilized close to zero, while accuracy approached unity, reflecting both stable optimization and strong generalization. The absence of the oscillations observed in the baseline model indicates more stable optimization, likely aided by the pretrained initialization and the two-stage fine-tuning strategy.

Evaluation on the test set confirmed this behavior, with perfect class separation observed in the confusion matrix (Fig. 5). All samples across the three defect types were correctly classified. Representative examples in Fig. 3 illustrate consistent predictions across diverse PRPD patterns. Overall, the fine-tuned model substantially outperformed the baseline and ablated CNNs, demonstrating the effectiveness of transfer learning for leveraging prior visual knowledge in this PRPD classification task.

### IV. CONCLUSION

This study evaluated CNN-based models for classifying PRPD images from electrical generators. Beginning with a

TABLE I  
VALIDATION AND TEST PERFORMANCE OF THE CNN ABLATION STUDY FOR PRPD IMAGE CLASSIFICATION.

Method	Validation dataset				Test dataset			
	Accuracy	F1	Precision	Recall	Accuracy	F1	Precision	Recall
Baseline	0.9155	0.9143	0.9276	0.9165	0.8451	0.8359	0.8788	0.8472
BatchNorm	0.9648	0.9644	0.9655	0.9647	<b>0.9859</b>	<b>0.9860</b>	<b>0.9860</b>	<b>0.9861</b>
AdamW (wd=1e-4)	0.9578	0.9576	0.9600	0.9582	0.9437	0.9429	0.9471	0.9441
Label smoothing ( $\epsilon=0.1$ )	0.9155	0.9161	0.9264	0.9159	0.8944	0.8928	0.9047	0.8952
Deeper (+1 block)	<b>0.9789</b>	<b>0.9789</b>	<b>0.9793</b>	<b>0.9790</b>	0.9789	0.9789	0.9792	0.9792

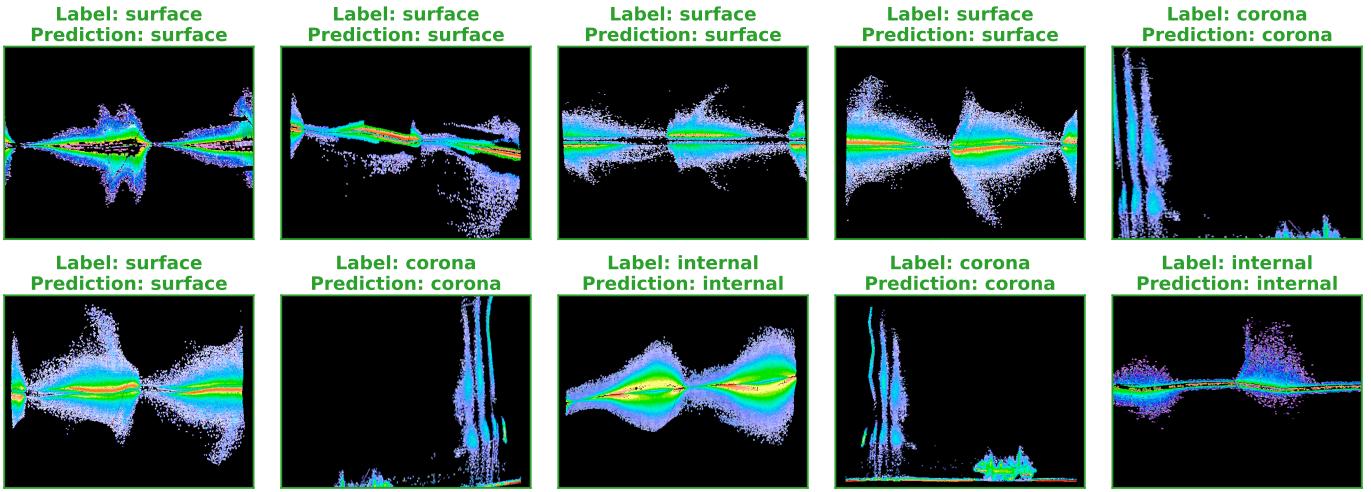


Fig. 3. Representative PRPD images with corresponding predicted and true labels using the fine-tuned EfficientNetV2-B0 model. The model achieves correct classification across diverse discharge patterns for corona, surface, and internal defects.

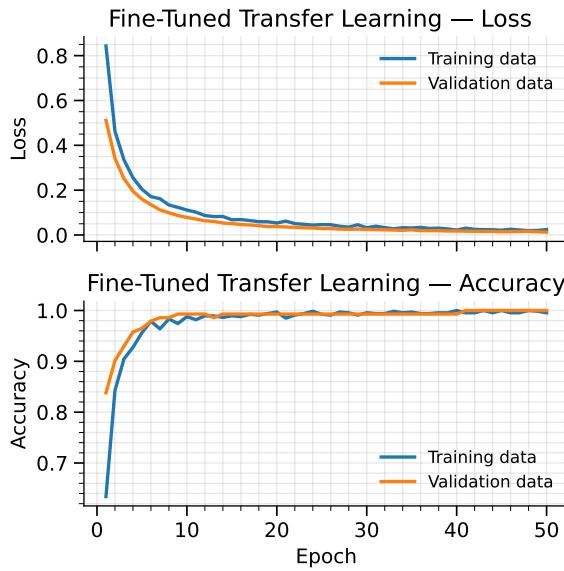


Fig. 4. Training and validation curves for the fine-tuned transfer learning model. The curves show rapid and stable convergence, with nearly identical performance on training and validation data, indicating strong generalization.

compact baseline model, a series of ablation studies assessed the influence of normalization, regularization, and network depth. Batch normalization and increased depth produced the most consistent gains, highlighting the importance of stable feature scaling and representational capacity. Fine-tuning a pretrained EfficientNetV2-B0 further improved performance, achieving near-perfect classification accuracy across all defect types.

Although these results demonstrate the strong potential of transfer learning for PD pattern recognition, the exceptionally high accuracy may partly reflect the ease of class separation or unintended data overlap caused by extensive augmentation and limited sample diversity in the examined dataset. Future work should therefore focus on evaluating model robustness on

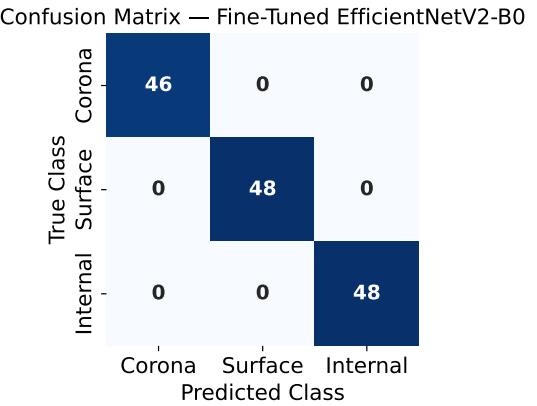


Fig. 5. Confusion matrix for the fine-tuned EfficientNetV2-B0 model on the test dataset. Perfect class separation is achieved, with all samples correctly classified into their respective discharge categories.

larger and more heterogeneous datasets, as well as exploring domain-specific pretraining and interpretability methods to better understand the learned diagnostic features.

## REFERENCES

- [1] J. D. Z. Henao, A. Segura, A. Tenorio, H. J. Diaz, and A. Paz, "Dataset of phase-resolved images of internal, corona, and surface partial discharges in electrical generators," *Data Brief*, vol. 52, p. 109992, 2024.
- [2] R. Bartnikas, "Partial discharges: Their mechanism, detection and measurement," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 9, no. 5, pp. 763–786, 2002.
- [3] N. C. Sahoo, M. M. A. Salama, and R. Bartnikas, "Trends in partial discharge pattern classification: A survey," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 12, pp. 248–264, Apr. 2005.
- [4] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.
- [5] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.
- [6] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, p. 9, 2016.
- [7] M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," 2021.