A Deep Learning Approach for Discrimination of Single- and Multi-Source Corona Discharges

Moein Borghei, *Graduate Student Member, IEEE*, and Mona Ghassemi, *Senior Member, IEEE*

***Abstract*— Insulation system health is crucial for reliable, lifelong operation of almost any electrical apparatus. While many studies have focused on the testing, modeling, and analysis of insulation aging mechanisms, research is needed to overcome new challenges in electric power systems. Fortunately, the progress in data analytics methods has opened up new opportunities to extract information from datasets. This study aims to make use of deep learning algorithms to lay the foundation for an online condition monitoring system that is capable of discriminating single- and multi-source corona discharges. In this article, we report the results of experimental testing and conversion of the data into phase-resolved partial discharge images, which we fed into deep neural networks. We begin by reviewing some of the most successful image recognition models including AlexNet, Inception-V3, residual network (ResNet), and DenseNet. Thereafter, we develop and optimize a ResNet model to achieve the highest accuracy model with the lowest computational cost.**

***Index Terms*— Convolutional neural network (CNN), corona discharge, deep learning, partial discharge (PD).**

I. Introduction

FOR more than a century, high-voltage engineering has dealt with the challenges that occur when electrical equipment is exposed to high-voltage levels. An inseparable part of any piece of electrical equipment is its insulation system. Dielectrics (insulating materials) keep conductive parts electrically apart and prevent short circuits between them. The behavior of dielectrics is highly dependent on the electric field distribution, which can become extremely high in a high-voltage apparatus. The understanding of high-voltage and insulation engineering has significantly improved with the development of numerical models and powerful computers.

The proper functioning of equipment is strongly dependent on the health of its insulation system. Dielectrics can be in the form of a gas, liquid, or solid, and the insulating material may undergo various electrical, mechanical, thermal, chemical, and radiation tensions during operation. Aggregation of these factors can cause irreversible degradation of the insulating properties and reduce life expectancy.

Manuscript received June 3, 2021; accepted July 31, 2021. Date of publication August 11, 2021; date of current version September 15, 2021. This work was supported in part by the Office of Naval Research (ONR) under Award N00014-19-1-2343 and in part by the National Science Foundation (NSF) under Award 1942540. The review of this article was arranged by Senior Editor S. J. Gitomer. *(Corresponding author: Mona Ghassemi.)*

The authors are with The Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: moeinrb@vt.edu; monag@vt.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPS.2021.3102115.

Digital Object Identifier 10.1109/TPS.2021.3102115

One of the main aging mechanisms in dielectrics is partial discharge (PD). According to the International Electrotechnical Commission (IEC) standard 60270 [1], "Partial discharge (PD) is a localized electrical discharge that only partially bridges the insulation between conductors and which may or may not occur adjacent to a conductor." Three types of PD can happen depending on the medium, the electrode type, and the electric field distribution: internal discharge, surface discharge, and corona discharge.

Some of the principal factors that influence PDs and other such discharges are the gaseous environment pressure, gas composition, magnitude and frequency of the applied voltage, electrode arrangement and geometry, and the properties, condition, and age of the insulation [2].

Measurement of PDs has been carried out since the 1940s and has provided us with an extensive amount of data [3]. Traditionally, experimental data provide useful information regarding the type of discharge, intensity, and frequency of discharges, as well as the phase-resolved PD (PRPD) patterns used to adjust simulation models. However, with the rapid advancement of data science, these invaluable data resources can now be used to provide novel insights that could not be obtained through traditional methods.

Machine learning (ML) provides a widely accepted tool for the extraction of useful information from datasets. However, to feed the ML models with experimental data, the datasets should first be turned into vectors of features. These features can be obtained through statistical, fractal, temporal, or frequency feature extraction methods. Below are summaries of some of the works that study PD using ML.

In [4], the PD measurements of different PD types (corona, internal, and surface discharges) were preprocessed using the wavelet decomposition process. The preprocessed dataset was then normalized and, after selecting the Gaussian radial basis function as the kernel, the support vector machine (SVM) was used to identify PD sources. Liao *et al.* [5] proposed a data mining approach to recognize different PD patterns in substations based on the shape of PD signals obtained from sensors. In this study, principal component analysis (PCA) was used to reduce the dimensionality and extract features. Then, the normalized autocorrelation functions (NACFs) were used to choose features for clusters. Finally, unsupervised clustering algorithms such as K-means and hierarchical clustering were used to form classes. Song *et al.* [6] performed PD defect-type classification based on 3-D phase-resolved pulse sequence (PRPS) graphs of the ultrahigh-frequency (UHF) detection

system of a gas-insulated substation (GIS). First, a set of features were extracted from these graphs using the histogram of oriented gradient (HOG) technique. Then, a naïve Bayes classifier was adopted to classify defect types to achieve 99.32% accuracy.

In addition, in [7], extensive experimental data were obtained for eight types of defects in a 220 kV GIS, which were classified through a random forest sequential forward selection method based on analysis of variance (RF-VA). In [8], electromagnetic interference (EMI) signals were used to distinguish between various PD and noise types using adaptive local iterative filtering (ALIF). In this study, an SVM was used for classification. Morette *et al.* [9] used K-means and transudative SVM learning methods to discriminate surface PD from noise signals. The features for classification were the mean, variance, skewness, and kurtosis of wavelet detail coefficient distribution at five levels of decomposition. In [10], real-time PD occurrence detection was used after transforming pulse sequence data into a feature vector stream and using an anomaly detection method based on the difference against normal data vectors. Also, Ardila-Rey *et al.* [11] gave an overview of the advances in clustering techniques used for discrimination of multiple PD sources and electrical noise.

Conventionally, the features used for PD classification are engineered not learned. Statistical parameters, image processing features, frequency/time features, and fractal features are among those used for this purpose. However, a set of features that successfully represent one dataset may not suit another dataset.

Recently, deep learning, a subfield of ML, has received a tremendous amount of attention due to its ability to automatically extract features and its relative independence from human intervention. Deep neural networks (DNNs) can make informed decisions based on the input data without manual feature engineering. Significant success has been reported using DNN in image classification, speech recognition, and language processing. A convolutional neural network (CNN) is a class of DNN that uses convolution (instead of matrix multiplication) in at least one of its layers. It has been observed that DNNs offer higher accuracy than the conventional ML methods as they are faster and do not require human intervention.

PD researchers recently targeted the use of deep learning models due to their promising performance. Recent progress in the application of deep learning for automated PD identification was reviewed in [12]. In [13], a deep CNN was developed to use UHF signals of a 120-kV GIS for the identification of PD sources. In this model, through a time-frequency analysis, UHF signals were represented as 2-D spectral frames using short-time Fourier transform (STFT). Wan *et al.* [14] developed a 1-D CNN to classify five types of PD defects in GIS based on the images of unstructured time-domain UHF waveforms. A CNN was used in [15] to automatically extract features from 3-D PRPD images and to classify four types of defects in GIS. In [16], the MobileNets CNN (MCNN), developed by Google in 2018, was used to recognize PD patterns and classify faults in a GIS with four defect types. In [17], the method of deep convolutional generative adversarial networks (DCGANs) was used to artificially produce PD image data. This can be helpful when there are not sufficient PD data, such as in industrial places where noise interference does not allow one to obtain valid measurements. Wang *et al.* [18] developed a generative adversarial network (GAN)-based method to augment the UHF PD dataset. This method also automatically generates deep features and performs classification of PD types.

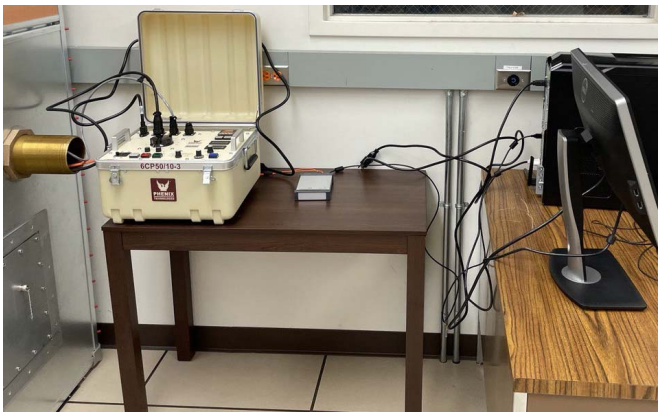
The deep belief network (DBN) method was developed in [19] by stacking restricted Boltzmann machines for both feature extraction and classification of PRPD patterns based on PD type (corona, internal, and surface discharges). In [20], four defect types of GIS PD were classified based on a combination of UHF and simulation data. A denoising-stacked autoencoder was used for classification, which achieved 82.2% accuracy. In [21], a recurrent neural network (RNN) structure with two long short-term memory (LSTM) layers classified artificially generated PD sources in a 345-kV GIS. This method achieved 96.62% accuracy, which was higher than that obtained by SVM and conventional artificial neural network (ANN). In [22], a stacked sparse autoencoder (SSAE) model was used for PRPD patterns based on UHF PD signals to classify four levels of PD severity from normal to dangerous. This method had an average of 90% accuracy, which is 10% higher than that of the SVM method. Balouji *et al.* [23] targeted the recognition of internal PD location and size when a bipolar voltage was applied. For this purpose, bagging trees and LSTM RNNs presented the best performance by achieving 95.5% and 98.3% accuracy, respectively. Instead of using signal measurements, five features were used as input for the RNN method, which was acquired as sequences of information correlating a PD to other PDs in the same cycle.

In this study, we aimed to use PD measurements to discriminate single- and double-source corona discharges. For this, PRPD patterns were used for identifying the PD source, which was advantageous as it represented the evolution of the signal over time. However, PRPD patterns become harder to interpret in the presence of high electrical noise or when multiple PD sources exist in an object. For this purpose, ML-based methods were used on the spectral domain to discriminate PD sources and electrical noise.

In real-world cases, it is common to have multiple sources of PD working at the same time. It is not a trivial task to detect multiple sources of PD from the patterns within a short period of time. In this study, we address this research gap by testing single and double sources of corona discharge and provide a deep learning framework to discriminate these different sources. Hence, the main contributions of this study are.

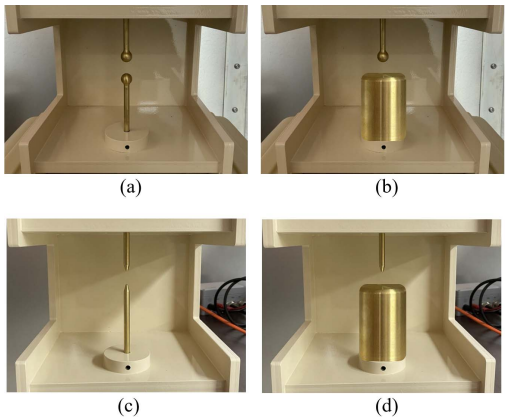
1) Experimentation on coronas under strongly inhomogeneous electric fields. This allowed us to simulate real-world cases where insulation systems undergo this type of electric stress. We then transformed the raw data into clean, normalized PRPDs patterns.
2) A review of some of the most successful DNN models in the recent decade is introduced and briefly discussed. These models include AlexNet (2012),

(a)



(b)

Fig. 1. Experimental setup. (a) Inside shielded room. (b) Outside shielded room.



(a) (b) (c) (d)

Fig. 2. Experiment system. (a) Test setup. (b) Sphere–sphere. (c) Sphere–plane. (d) Needle–needle. (e) Needle–plane electrode configuration.

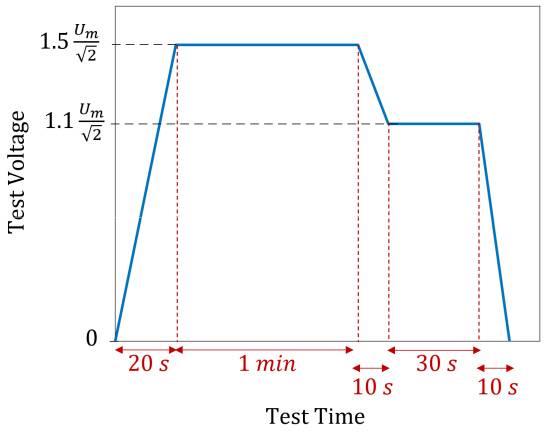


Fig. 3. Test procedure based on IEC standard 61287 [24].

Inception-V3 (2015), Residual Network (ResNet) (2015), and DenseNet (2016).

3) After the application of different variations in these DNN concepts, the ResNet model was optimized to capture the highest amount of useful information from PRPD images.

To the best of the authors' knowledge, the use of very DNNs for the identification of corona discharge sources has not been addressed before. This approach demonstrates a great potential for being used in online monitoring systems of electrical equipment with highly exposed insulation systems. Specifically, in the case of electric ships or aircraft, this approach can ensure the safety of electric machinery and prevent disastrous consequences due to insulation failure.

II. Test Setup

The test setup was assembled following the IEC 60270 standard and is shown in Fig. 1. To perform tests in a radio-silent environment and prevent EMI, a shielded room made of two layers of copper was used.

The test samples were combinations of four fundamental configurations in terms of the electric field homogeneity. The configurations considered in this work were: 1) sphere–plane; 2) sphere–sphere; 3) needle–plane; and 4) needle–needle (see Fig. 2). Among these four configurations, the first two led to weakly nonuniform electric field distributions and the other two produced extremely nonuniform fields.

Moreover, three double-source corona discharges, as follows, were tested.

1) Needle–needle and sphere–sphere.
2) Needle–plane and sphere–sphere.
3) Sphere–plane and needle–needle.

A standardized test procedure was used for all these scenarios to generate comparative datasets. The test procedure was inspired by IEC 61287-1 [24] (Fig. 3).

III. Overview of Deep Learning Models

ML approaches have shown groundbreaking performance in object recognition and classification. Specifically, CNNs have revolutionized computer vision approaches. Using large datasets and developing models that are resistant to overfitting, neural networks can achieve the lowest loss. The great success achieved by this deeper network has resulted in the move from LeNet architecture (1989) with five layers to VGG with 19 layers and, later, to residual neural networks, which have up to 1000 layers. However, due to the vanishing gradient problem, the deeper CNNs suffer from their long pathway from input to output to retrieve useful information. As we see in the following sections, different data science research

Fig. 4. AlexNet architecture.

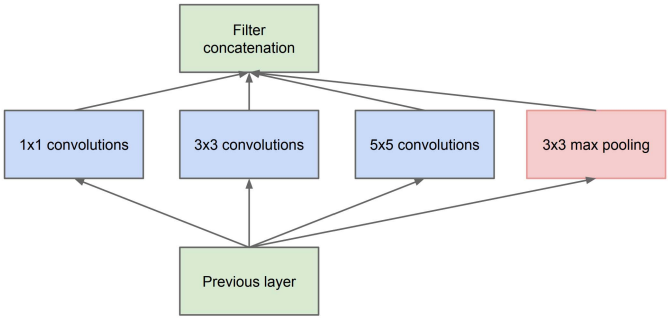


Fig. 5. Primitive inception block [27].

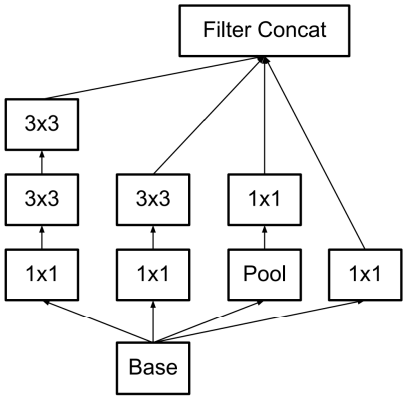


Fig. 6. Sample inception block [26].

groups have come up with different strategies to tackle this issue.

A. AlexNet

AlexNet is a CNN named after its creator Alex, which was described in collaboratory work published in 2012 [25]. AlexNet architecture includes eight weighted layers, five of which are convolutional and the rest are fully connected layers as shown in Fig. 4. The output of the last fully connected layer goes to an n-way softmax activation function to generate seven classes. The input to the first convolutional layer includes images of $32 \times 32 \times 1$ with 96 kernels of size (11, 11) with a stride of 4 pixels. The second convolutional layer has 256 filters with kernel size (5, 5). The next three convolutional layers also have (3, 3) kernels and 384, 384, and 256 filters, respectively. All the convolutional layers are followed by a batch normalization layer and a rectified linear unit (ReLU) activation function. The first two and the last convolutional layers have a maximum pooling layer with a pool size of (2, 2) and a stride of 2 pixels. The number of neurons in fully connected layers is 4096.

B. Inception-V3

Led by a group of researchers from Google, Inception-V3 was proposed in 2015. Inception-V3 architecture is based on the idea of approximating an optimal local sparse structure in a convolutional network [26]. The concept that helps achieve this goal is the inception module. Instead of limiting to a single filter size, an inception module allows using several types of filter sizes for a single image block and concatenating their outputs into a single vector. This helps capture more information from the input images in deep weighted layers.

A basic module is shown in Fig. 5 where three convolutional layers with filter sizes of 1×1, 3×3, and 5×5 are used in a single block followed by a pooling layer, which is key to the success of convolutional networks. The modules stack upon each other and as the number of layers increases, more abstract features are discovered by the model [27].

However, an issue arises from the expensive computational burden imposed by 5×5 convolutions even when there is only a small number of them. In Inception-V3, the authors deployed two ideas to tackle this challenge. First, they performed a dimension reduction using 1×1 convolutions before the troublesome 3×3 and 5×5 convolutions. Also, additional ReLU activation functions add to the beneficial effect. Szegedy *et al.* [26] proposed the replacement of the factorization of larger convolutions, such as 5×5, with two 3×3 convolutions on top of each other. In this way, 25 parameters in a 5×5 filter are reduced to 18 parameters in two 3×3 convolutions. This results in the inception block shown in Fig. 6. Later, a 3×3 convolution was replaced by two asymmetric 3×1 and 1×3 convolutions. Also, some auxiliary classifiers were placed between the layers to regularize neurons. The architecture of the entire model is shown in Fig. 7.

C. ResNet

In 2015, a group of Microsoft researchers proposed a deep residual learning scheme to overcome two well-known

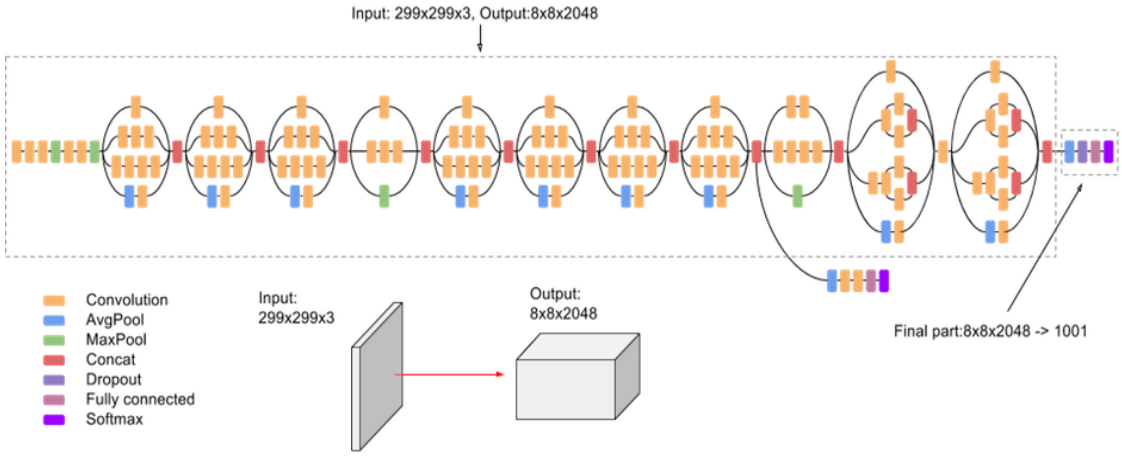


Fig. 7. Inception-V3 architecture.

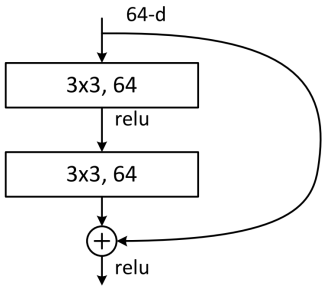


Fig. 8. Sample residual block [28].

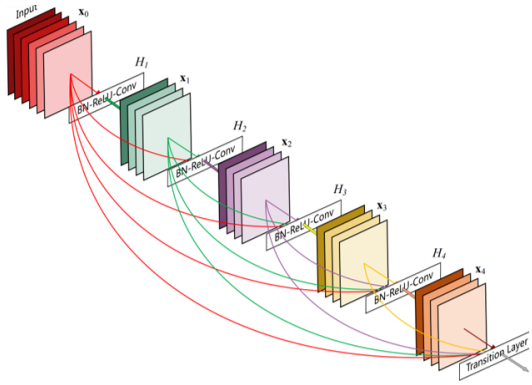


Fig. 9. Basic DenseNet architecture [29].

problems of DNNs: high computation time and overfitting [28]. In this method, the ResNet fits a residual mapping using the block shown in Fig. 8 that shows a feed-forward neural network. In this figure, identity mapping occurs after two layers, but this number can vary depending on the problem and, as seen in the next sections, we use this parameter as a tool to optimize ResNet architecture. Compared with CNNs, ResNets have a lower number of filters and are relatively less complex in training.

D. DenseNet

In 2016, this notion was proposed "to ensure maximum information flow between layers in the network, we connect all layers directly to each other" [29]. In a feed-forward network, each layer receives inputs from all preceding layers, which provides input for all the subsequent layers. In contrast to ResNet, in this approach, the features are concatenated instead of aggregated. Therefore, the number of connections in an L-layer network is $L(L+1)/2$ instead of L as in conventional networks. The dense connection of layers in DenseNet leads to a lower number of parameters needed for the convolutional networks due to decreased redundancy. The basic architecture of DenseNet is shown in Fig. 9.

IV. Data Preprocessing

The measured corona signals were used to train deep network models introduced in the previous section. The inputs were in the form of PRPD patterns. However, PRPD images must be preprocessed before use in the model. The preprocessing stage was implemented in MATLAB and followed these steps.

1) Importing raw data to MATLAB.
2) Filtering discharges that are too small in magnitude.
3) Defining proper time intervals for generating PRPD patterns.

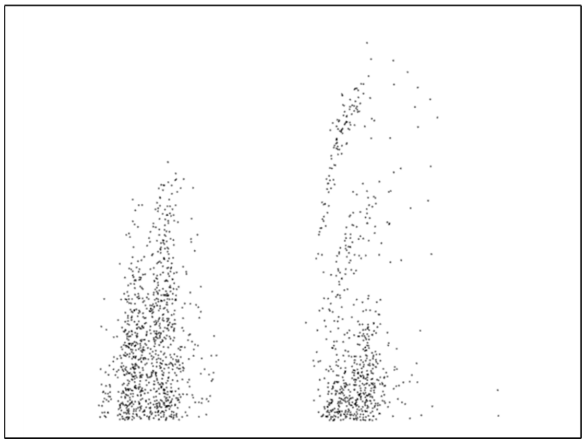


Fig. 10. Sample PRPD image fed to DNNs.

TABLE I
LEARNING RESULTS FOR THE STANDARD NEURAL NETWORK MODELS

Model	Layers	Training Loss	Validation Loss	Training Acc.	Validation Acc.	Time/epoch
AlexNet	8	0.003	0.560	1	0.869	4.8 s
Inception-V3	48	0.208	0.569	0.924	0.807	14.3 s
DenseNet121	121	0.037	1.446	0.988	0.748	9.7 s
DenseNet 169	169	0.028	1.424	0.992	0.789	12.3 s
DenseNet201	201	0.028	1.376	0.991	0.761	15.3 s
ResNet50	50	0.052	1.190	0.983	0.760	9 s
ResNet101	101	0.028	1.172	0.991	0.780	16.3 s
ResNet152	152	0.064	1.078	0.978	0.747	21.4 s

4) Generating PRPD patterns of the desired size and under a unified x and y range.

For PD measurement, the aforementioned process was applied to the 60-s period in Fig. 3 where the voltage amplitude was $1.5U_m/\sqrt{2}$ since this was the interval over which the most intense PDs occurred. Fig. 10 shows a sample PRPD pattern. Note that since the range of the x- and y-axes is both unified, we have omitted the x- and y-axes to prevent exhaustion of the ResNet model with useless visual information. The y-axis extends from 0.5 to 25 nC, which was the highest observed PD magnitude in all cases. The sampling rate was ten cycles (60 Hz) per PRPD plot, meaning that each image stands for 167 ms of data. The PRPD images were produced at a rate of 20 patterns per second. This gives about 1200 images per electrode configuration. Therefore, we had a dataset containing 8400 images.

The generated images were all in greyscale and 75×75 pixels. However, depending on the model, the images could be reshaped. The PRPD pixels were scaled and normalized. Also, the classes/labels (which are the type of discharge) were converted into a one-hot vector for adaptability to the deep learning models. Each of the above electrode configurations received a numerical value to ease the labeling process. The label numbers started from 0 for the needle–needle configuration and continued until 6 for the double-source sphere–plane and needle–needle configurations. About 70% of the 8400 images were allocated for training and the remaining 30% was used for validation of the performance of the neural network.

For this stage, eight standard neural networks, based on the four deep learning concepts discussed earlier, were used to discriminate seven corona discharge sources.

1) AlexNet standard model.
2) Inception-V3 standard model.
3) ResNet50.
4) ResNet101.
5) ResNet152.
6) DenseNet121.
7) DenseNet169.
8) DenseNet201.

The numbers at the end of the variations in the ResNet and DenseNet models correspond to the number of weighted layers used in these networks. For all cases, the batch size and the number of epochs were 128 and 120 m respectively. All the deep learning networks were implemented in Python with the aid of the Keras (Tensorflow), Numpy, Pandas, and SciKit Learn libraries. The computer server used for simulation had two 16-core Intel Xeon Gold processing units, 192 GB of RAM, and 5 GB of dedicated GPU memory that is used for speeding up the training process.

V. PRELIMINARY RESULTS

After training the explained models with the corona discharge data, examination of the accuracy level, loss, and training time provided insightful comparisons. A summary of these quantities, as well as the number of layers, is shown in Table I.

Based on these results, the use of deeper models did not significantly improve the accuracy of discrimination. AlexNet, which has the lowest depth, gave the best performance. But that did not necessarily mean that as the model goes deeper, the level of loss increases. There is agreement among data scientists that deeper models require larger training datasets. Therefore, for our 8400-image dataset, we expected that a deeper model would not lead to better performance.

In Fig. 11, the variations in loss and accuracy for different models over the 120 iterations are shown at both the training and validation levels. These figures provide more insight, in terms of monotonicity, stability of models, and how fast these models reach their best performance level. The accuracy figure for AlexNet shows the severe fluctuations during the first 100 iterations. In fact, for about 90 iterations, the model had less than 40% accuracy while it exceeded 80% about 20 iterations later. Therefore, it takes at least 100 epochs (each takes 4.8 s) to reach a plateau.

On the other hand, the fluctuation degree was considerably lower for deeper models, especially the ResNet and DenseNet models. Still, the deepest models of ResNet and DenseNet (i.e., ResNet152 and DenseNet201) presented inferior performance compared with those models with less depth. Inception-V3's performance, in contrast, fluctuated somewhere between AlexNet's and DenseNet's. A severe accuracy drop of 60% occurred in the 110th iteration; however, it returned to the normal 80% accuracy soon afterward.

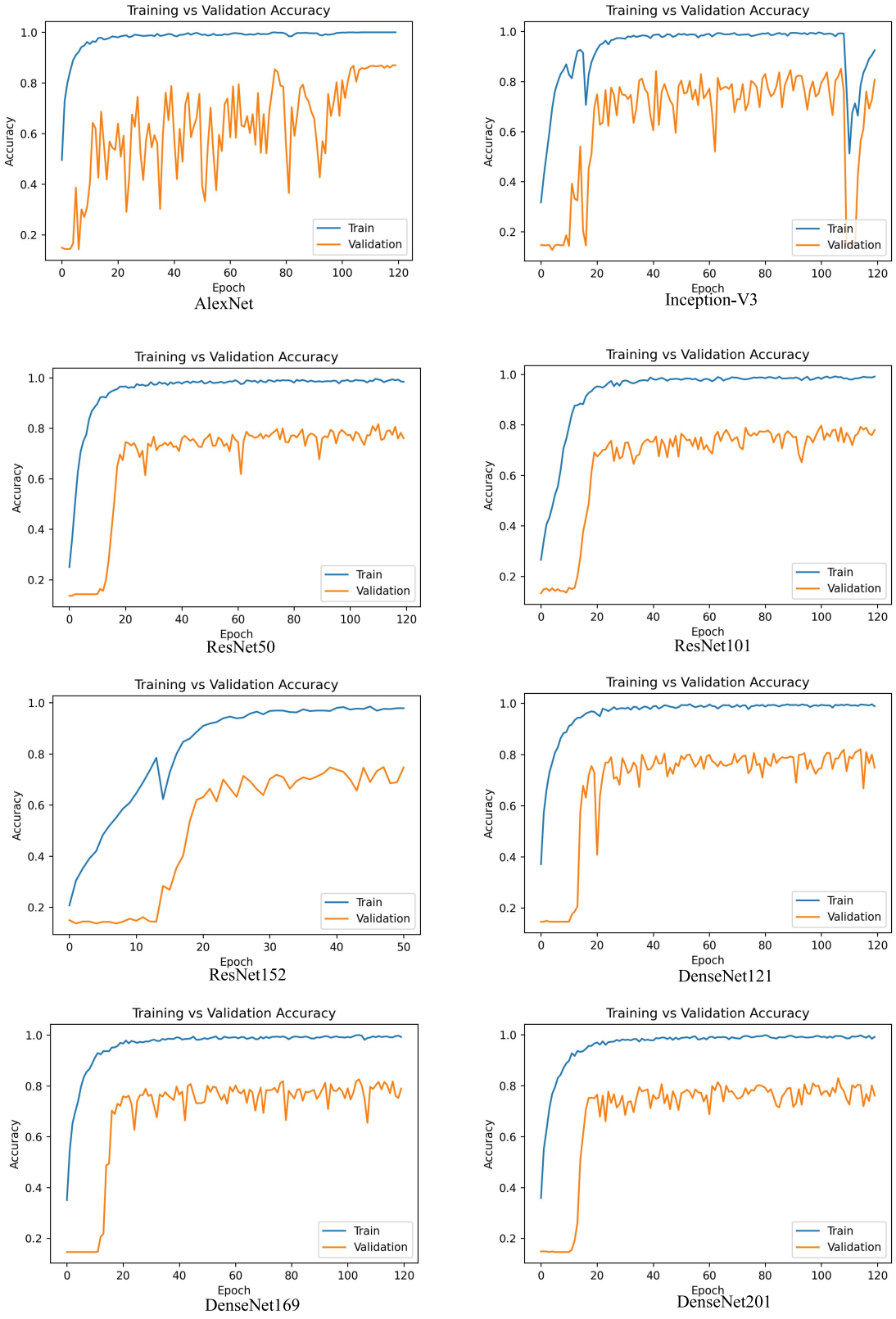


Fig. 11. Comparison of training and validation accuracy level among different models.

Therefore, we conclude that even though AlexNet presents the highest ultimate accuracy, it has broad swings at a very high level and requires at least 100 iterations for mid-level stability. On the other hand, the ResNet and DenseNet models reach a stable level of high accuracy within 20 iterations. As mentioned in the literature, based on the dataset, there is

TABLE II
LEARNING RESULTS FOR THE NEURAL NETWORKS WITHIN OPTIMIZATION FRAMEWORK

Model No.	Block Size	Initial Filter	Layers	Tr. Loss	Val. Loss	Tr. Acc.	Val. Acc.	Time/epoch
1	2	16	16	0.010	0.762	1.000	0.830	3.7 s
2	2	32	16	0.010	0.618	1.000	0.854	5.1 s
3	2	64	16	0.010	0.576	1.000	0.876	9.2 s
4	2	128	16	0.010	0.522	1.000	0.881	23.5 s
5	3	16	22	0.010	0.898	1.000	0.820	4.9 s
6	3	32	22	0.010	0.732	1.000	0.852	7.15 s
7	3	64	22	0.022	1.517	0.996	0.727	13.3 s
8	4	16	26	0.010	0.868	1.000	0.831	5.1 s
9	4	32	26	0.010	0.716	1.000	0.863	9.2 s
10	5	64	30	0.010	0.693	1.000	0.876	21.4 s
11	7	16	38	0.010	0.911	1.000	0.836	8.2 s
12	9	16	46	0.010	0.886	1.000	0.835	10.2 s
13	2	16	16	0.010	0.762	1.000	0.830	3.7 s

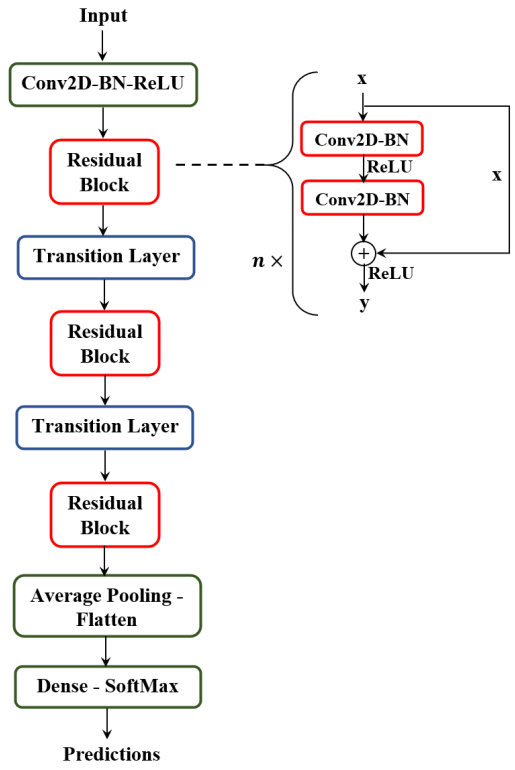


Fig. 12. Base ResNet model for optimization.

an optimal depth for the best model performance, which is not necessarily the lowest or highest depth. In searching for this optimal depth, we built a customized ResNet architecture that had the high accuracy of the AlexNet model while benefitting from the stability of deep ResNets.

VI. OPTIMIZATION OF RESNET

As discussed in the previous section, deeper neural networks provided more stable results and converged quickly. However, a standardized approach did not suit our database. In this regard, we built a customized model based on the concept of residual neural networks. In Fig. 12, the base ResNet model that we used for optimization is shown. The residual block concept was used, which used network layers to fit a residual mapping, instead of directly trying to fit a desired underlying map (see the residual block). In the design of convolutional layers, the layers had the same number of filters as the output feature size.

In addition, to maintain the computational complexity relatively constant, the number of filters had an opposite relationship with the feature map size. This meant that the number of filters doubled if the feature map size was cut in half. Therefore, the number of filters was doubled and each dimension was halved periodically.

Moreover, there were transition layers between residual blocks (a 1×1 2-D convolutional layer with a stride of 2) to shrink the number of feature maps and make the model more compact. After each convolutional layer, the batch of data was normalized to reduce the internal covariant shift. This layer was followed by a ReLU activation function.

At the end of the residual blocks, we conducted down sampling by having convolutional layers that had a stride of 2. The network concluded with a global average pooling layer and a seven-class fully connected layer with softmax. Two parameters were used as degrees of freedom to optimize the ResNet model.

1) The repetition rate of residual double convolutional layers in each residual block (n).
2) The initial number of filters (in the first layer).

The performance of this proposed framework, using different values for these two parameters, is reported in Table II. A summary of the models in Table II shows that the optimized network provided a highly accurate model for this study. Almost all the models have 100% training accuracy and more than 80% validation accuracy.

The previous section showed that going deeper than 50 weighted layers did not improve the overall performance of the model (for this dataset). The best performance was obtained when the number of filters in the convolutional layers was the highest. In the fourth model, the initial number of filters was 128, which doubled after each residual block, reaching 512 filters in the last convolutional layer. Although this model offered slightly higher accuracy compared with the other models, the higher number of filters translated into a longer iteration time. The third model was the second most accurate model and only had 16 layers and 64 initial filters. Note that the training time in each iteration of this model was 60% lower than the fourth model. Therefore, we chose model 3 as the optimal model for our dataset. Fig. 13 shows

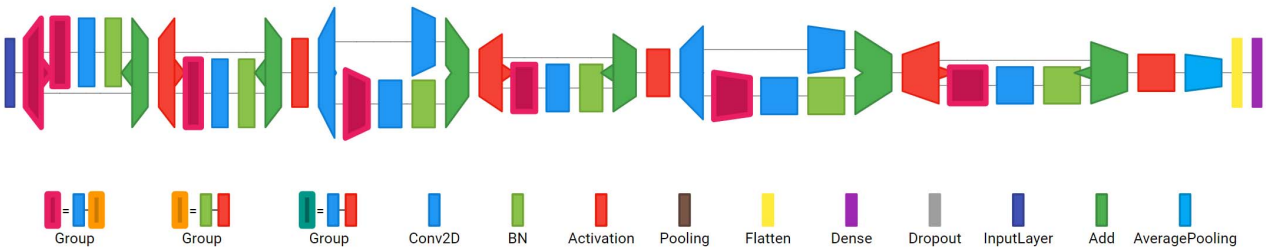


Fig. 13. Architecture of the third model.

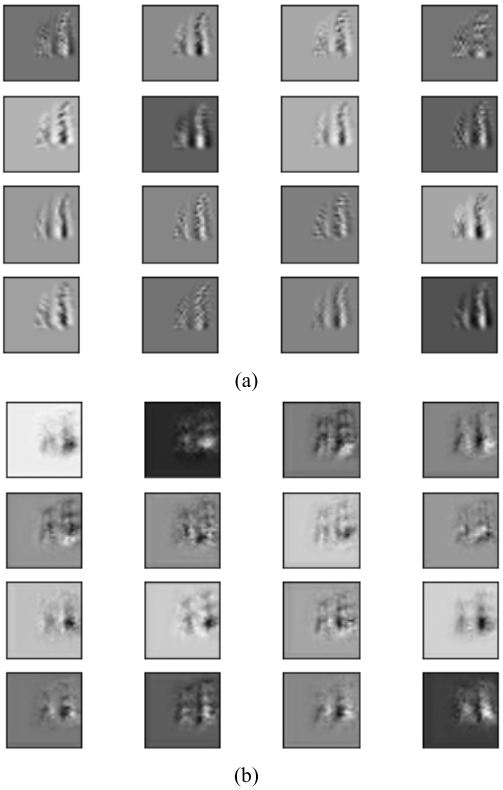


Fig. 14. Visualization of features in the first residual block after (a) first convolutional layer and (b) last convolutional layer.

the architecture of this model and Fig. 14 demonstrates the 16 filters' outputs of two convolutional layers on a test image (shown in Fig. 10) in the first residual block. Fig. 14(a) shows the first convolutional layer output characterizing 16 features from the test image. Fig. 14(b) shows the 16 relatively more sophisticated features extracted by the last convolutional layer in the first residual block.

To further investigate the performance of this model, Fig. 15 shows the evolution of training and validation accuracy/loss over the iterations. These figures imply that this model reaches its peak performance within 40 iterations and remains almost intact thereafter. Comparing the convergence time of this model with AlexNet, we determined that this optimized model reached a completely stable result after 368 s, while AlexNet required 528 s to reach a rough plateau. Therefore, our proposed model offers superior performance compared with AlexNet. Finally, one should consider that for a much larger dataset, the proposed ResNet can be easily adjusted to achieve high performance.

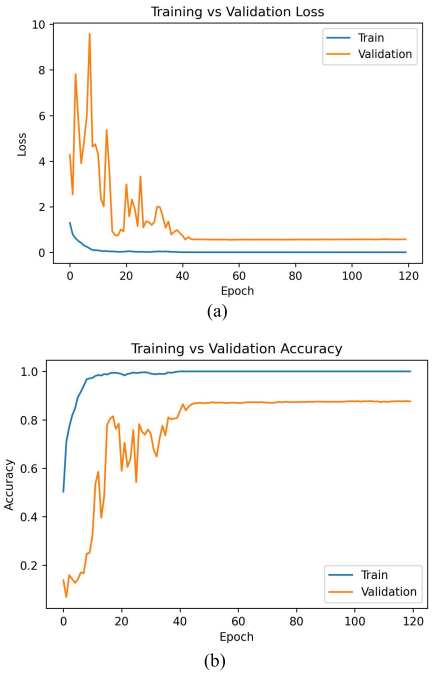


Fig. 15. Training and validation accuracy and loss of the third model. (a) Accuracy. (b) Loss.

VII. Conclusion

In this study, we investigated one of the most detrimental phenomena in high0voltage engineering: PDs. Even though PDs have been studied for more than a century, new technologies and applications have introduced new challenges and threats to insulation systems. These new technologies have expanded the domain of PD activity and, in cases such as more electric aircraft, low-pressure environments lead to unprecedentedly vulnerable dielectric performance against the PD phenomenon.

To this end, this study aimed to experiment with single- and double-source corona discharges and use deep learning to develop a PD source discrimination model. The breakthrough of DNNs has furnished us with opportunities to learn useful information from large datasets. In this regard, we evaluated the performance of some of the most successful DNNs in identifying PD sources, specifically, corona discharge. Standardized measurement data were used to generate PRPD patterns and we then used these figures to feed neural networks. The initial examination of DNNs showed that there was no need to go deeper than 50 layers to achieve accurate results. However, deeper networks, such as ResNets, were found to provide faster convergence and more stability. Thus, to use these merits, we

developed a customized residual neural network and optimized it to meet our needs. The optimized ResNet offered 100% training accuracy and more than 87% validation accuracy after 40 iterations. In future work, we will cover a variety of PD types and expand our dataset to improve the deep learning model for fast and accurate discharge source discrimination.

REFERENCES

[1] *High-Voltage Test Techniques—Partial Discharge Measurements*, Standard IEC 60270, 2000.

[2] D. G. Kasten, X. Liu, S. A. Sebo, D. F. Grosjean, and D. L. Schweickart, "Partial discharge measurements in air and argon at low pressures with and without a dielectric barrier," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 12, no. 2, pp. 362–373, Apr. 2005.

[3] D. A. Nattrass, "Partial discharge. XVII. The early history of partial discharge research," *IEEE Elect. Insul. Mag.*, vol. 9, no. 4, pp. 27–31, Jul./Aug. 1993.

[4] L. Hao, P. L. Lewin, and S. G. Swingler, "Use of machine learning for partial discharge discrimination," in *Proc. 11th Int. Electr. Insul. Conf.*, 2009, pp. 1–6.

[5] R. Liao, Y. Fernandess, K. Tavernier, and M. R. Irving, "Recognition of partial discharge patterns," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2012, pp. 1–8.

[6] S. Song, Y. Qian, H. Wang, Y. Zang, G. Sheng, and X. Jiang, "Partial discharge pattern recognition based on 3D graphs of phase resolved pulse sequence," *Energies*, vol. 13, no. 16, p. 4103, Aug. 2020.

[7] R. Yao, J. Li, M. Hui, L. Bai, and Q. Wu, "Feature selection based on random forest for partial discharges characteristic set," *IEEE Access*, vol. 8, pp. 159151–159161, 2020.

[8] I. Mitiche, G. Morison, M. Hughes-Narborough, A. Nesbitt, P. Boreham, and B. G. Stewart, "Classification of partial discharge signals by combining adaptive local iterative filtering and entropy features," in *Proc. IEEE Conf. Electr. Insul. Dielectr. Phenomenon (CEIDP)*, Oct. 2017, pp. 335–338.

[9] N. Morette, L. C. C. Heredia, T. Ditchi, A. R. Mor, and Y. Oussar, "Partial discharges and noise classification under HVDC using unsupervised and semi-supervised learning," *Int. J. Electr. Power Energy Syst.*, vol. 121, Oct. 2020, Art. no. 106129.

[10] J. Kim and C. H. Park, "Partial discharge detection based on anomaly pattern detection," *Energies*, vol. 13, no. 20, p. 5444, Oct. 2020.

[11] J. A. Ardila-Rey, M. P. Cerda-Luna, R. A. Rozas-Valderrama, B. A. de Castro, A. L. Andreoli, and F. Muhammad-Sukki, "Separation techniques of partial discharges and electrical noise sources: A review of recent progress," *IEEE Access*, vol. 8, pp. 199449–199461, 2020.

[12] S. Barrios, D. Buldain, M. P. Comech, I. Gilbert, and I. Orue, "Partial discharge classification using deep learning methods—Survey of recent progress," *Energies*, vol. 12, no. 13, p. 2485, Jun. 2019.

[13] G. Li, M. Rong, X. Wang, X. Li, and Y. Li, "Partial discharge patterns recognition with deep convolutional neural networks," in *Proc. Int. Conf. Condition Monitor. Diagnosis (CMD)*, Sep. 2016, pp. 324–327.

[14] X. Wan, H. Song, L. Luo, Z. Li, G. Sheng, and X. Jiang, "Pattern recognition of partial discharge image based on one-dimensional convolutional neural network," in *Proc. Condition Monitor. Diagnosis (CMD)*, Sep. 2018, pp. 1–4.

[15] F. C. Gu, "Identification of partial discharge defects in gas-insulated switchgears by using a deep learning method," *IEEE Access*, vol. 8, pp. 163894–163902, 2020.

[16] Y. Wang, J. Yan, Q. Sun, J. Li, and Z. Yang, "A mobilenets convolutional neural network for gis partial discharge pattern recognition in the ubiquitous power Internet of Things context: Optimization, comparison, and application," *IEEE Access*, vol. 7, pp. 150226–150236, 2019.

[17] J. A. Ardila-Rey, J. E. Ortiz, W. Creixell, F. Muhammad-Sukki, and N. A. Bani, "Artificial generation of partial discharge sources through an algorithm based on deep convolutional generative adversarial networks," *IEEE Access*, vol. 8, pp. 24561–24575, 2020.

[18] X. Wang, H. Huang, Y. Hu, and Y. Yang, "Partial discharge pattern recognition with data augmentation based on generative adversarial networks," in *Proc. Condition Monitor. Diagnosis (CMD)*, Sep. 2018, pp. 1–4.

[19] M. Karimi, M. Majidi, H. MirSaeedi, M. M. Arefi, and M. Oskuoee, "A novel application of deep belief networks in learning partial discharge patterns for classifying corona, surface, and internal discharges," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3277–3287, Apr. 2020.

[20] Y. Zhao, J. Yan, Y. Wang, T. Liu, and J. Jiang, "Partial discharge patterns recognition of gis with denoising-stacked autoencoder networks," in *Proc. 5th Asia Conf. Power Electr. Eng. (ACPEE)*, Jun. 2020, pp. 1815–1818.

[21] M.-T. Nguyen, V.-H. Nguyen, S.-J. Yun, and Y.-H. Kim, "Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 11, no. 5, p. 1202, May 2018.

[22] J. Tang, M. Jin, F. Zeng, X. Zhang, and R. Huang, "Assessment of PD severity in gas-insulated switchgear with an SSAE," *IET Sci., Meas. Technol.*, vol. 11, no. 4, pp. 423–430, Jul. 2017.

[23] E. Balouji, T. Hammarstrom, and T. McKelvey, "Partial discharge classification in power electronics applications using machine learning," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.

[24] *Railway Applications—Power Converters Installed on Board Rolling Stock—-Part 1: Characteristics and Test Methods, Edition 3.0*, Standard IEC 61287-1, 2014.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[27] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

Moein Borghei (Graduate Student Member, IEEE) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2018, and the M.Sc. degree in electrical engineering from Virginia Tech, Blacksburg, VA, USA, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering.

His research interests include high-voltage engineering, dielectrics, electrical insulation, multiphysics modeling, and transmission line design.

Mona Ghassemi (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees (Hons.) in electrical engineering from the University of Tehran, Tehran, Iran, in 2007 and 2012, respectively.

From 2013 to 2015, she was a Post-Doctoral Fellow with the High Voltage Laboratory of NSERC/Hydro-Quebec/UQAC, as the Industrial Chair on Atmospheric Icing of Power Network Equipment, and the Canada Research Chair on Power Network Atmospheric Icing Engineering, University of Quebec, Chicoutimi, QC, Canada. She was a Post-Doctoral Fellow with the Electrical Insulation Research Center, Institute of Materials Science, University of Connecticut, Storrs, CT, USA, from 2015 to 2017. In 2017, she joined the Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA, as an Assistant Professor, and in 2021, she was appointed as the Steven O. Lane Junior Faculty Fellow and the College of Engineering Faculty Fellow. She has been a registered Professional Engineer in the province of Ontario, Canada, since 2015. Her research interests include electrical insulation materials and systems, high-voltage/field technology, multiphysics modeling, electromagnetic transients in power systems, and power system analysis and modeling.

Dr. Ghassemi is an At-Large Member of the Administrative Committee of the IEEE Dielectrics and Electrical Insulation Society from 2020 to 2023, a Corresponding Member of the IEEE Conference Publication Committee of the IEEE Power and Energy Society, an Active Member of several CIGRE working groups and IEEE Task Forces, and a Member of the Education Committee of the IEEE DEIS and PES. She was a recipient of the 2020 Contribution Award from the *High Voltage* journal, the 2020 National Science Foundation (NSF) CAREER Award, the 2020 Air Force Office of Scientific Research (AFOSR) Young Investigator Research Program (YIP) Award, and the 2021 Department of Energy Early Career Research Program Award. She is an Associate Editor of the IEEE TRANSACTIONS ON DIELECTRICS AND ELECTRICAL INSULATION, the IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, *IET High Voltage*, the *International Journal of Electrical Engineering Education*, and *Power Electronic Devices and Components* (Elsevier), and a Guest Editor for *Energies*.