A Foundational Framework for Substation Fault Diagnosis from Imbalanced Thermal Data Using Transfer Learning

Felipe Nascimento[1*], Herman Lepikson[1,2]
[1] *Universidade Federal da Bahia, PPGM, Salvador, Bahia, Brazil*
[2] *Universidade SENAI CIMATEC, Management and Technology, Salvador, Bahia, Brazil*
**Corresponding author: felipern@ufba.br*

Abstract: A key component of predictive maintenance for vital electrical substation equipment is thermographic inspection, which makes it possible to identify thermal abnormalities early on before they cause failures. Maintenance workflows suffer by the subjective, time-consuming, and human error-prone manual interpretation of the resulting thermal images. Although deep learning offers an efficient automation solution, a significant real-world obstacle to its widespread use is the extreme scarcity and class imbalance of available fault data. In order to close this gap, a comprehensive methodological framework for creating a reliable baseline diagnostic model is proposed and described in this paper. The strategy is based on transfer learning, which involves optimizing strong, pre-trained Convolutional Neural Network (CNN) architectures to take advantage of their acquired features and lessen the impact of sparse data. From data preparation and aggressive minority fault class augmentation to the use of a weighted loss function during training, the framework describes the complete pipeline. The experimental pipeline was validated after this methodology was put into practice. The model's anticipated initial bias towards the prevalent "Normal" class—a direct result of the data imbalance—was validated by preliminary observations. The main result of this work is this procedural validation, which provides a strong, repeatable basis for further investigation. While conclusive performance analysis and the investigation of explainability techniques remain open for future work, this study offers a practical route for creating trustworthy diagnostic tools under practical data constraints.
Keywords: Artificial Intelligence. Thermography. Deep Learning. Fault Detection. Electrical Substations.
Abbreviations: CNN, Convolutional Neural Network. AI, Artificial Intelligence. MLP, Multilayer Perceptron. SVM, Support Vector Machine. GANs, Generative Adversarial Networks. GAP, Global Average Pooling. XAI, Explainable.

1. Introduction

Reliability of power systems is something modern society takes for granted but is reliant on flawless working of vital infrastructure, most notably electrical substations. These units are head of the grid, and collapse of one component of it (a transformer, a circuit breaker, or simply a connection) can result in mass power cuts with dire economic as well as social consequences. To prevent this, utilities rely heavily on predictive patterns of maintenance, and of available tools, infrared-based thermography has been useful. It offers a non-intrusive, in-real-time means of "seeing" heat, being thus extremely good at detecting the tell-tale signs of a potential failure, such as an overheated joint, method well-tested in industry standards [1].

Despite thermography, power is ultimately limited by the human eye and brain. Each of these thermal images must be manually inspected by a skilled technician, which can be a slow and subjective process. What in one expert's eyes is an essential fault, in another's is nothing but solar reflection or normal variability in operations. An acute risk and a challenge of scalability are present. A central consideration guiding this effort is the need to automate the inspection process to ensure it is quick, objective, and scalable.

The easy solution lies in Artificial Intelligence (AI), with deep learning structures like Convolutional Neural Networks (CNNs), which have significantly impacted the field of image processing [2]. It is not news to train a model to recognize thermal fault patterns. Most research works, however, overlooks the most crucial in-the-field limitation: data. Faults in a well-maintenance substation happen rarely. This creates a familiar machine learning problem: a small, highly imbalanced set, making it extremely difficult to train a trustworthy model from scratch. This work targets just such problem at first principles. It provides an elaborate methodological framework towards the first stage of an automated diagnosis program. The contribution of this work is designing, in fine detail, a baseline experiment using transfer learning to leverage ideas learned under large datasets as well as data augmentation to virtually supplement a sparse set of fault images. This work clearly defines architecture, data processing strategies, as well as evaluation measures, of a model, once in working, will serve as a baseline reference point to justify as well as guide subsequent research in next-level-data generation as well as explainability strategies.

2. Related work

This project is founded on two pillars of research: applying thermography in predictive maintenance as well as using deep learning in image classification.

The infrared thermography theory and application to electrical equipment are well-documented. [3] defines in detail heat transfer physics, emissivity, as well as reading thermal patterns of likely failure, as defined in standards by [1, 4]. These patterns, most of them being hotspots or unknown thermal gradients, are precisely those on which was trained the AI model herein proposed.

Early efforts at automating this activity used traditional machine learning strategies. For instance, [5] built a predictive maintenance framework with a Multilayer Perceptron (MLP), a classical neural network structure. It relied on feature extraction of eleven hand-designed statistical features (e.g., mean, variance, skewness) in thermal images to differentiate equipment as “defect" or "non-defect" with 84% accuracy, with subsequent graph-cut-based refinement. This work demonstrated possible automation of diagnostics but identified manual feature engineering as an essential condition.

Subsequent research involved using CNNs to automatically learn image features through deep learning. [6] provided a novel method of using a special CNN with a Support Vector Machine (SVM) classification method. A key part of their research involved complex preprocessing to locate and extract temperature values from the

scale bar of the thermal image, using extracted characters to build a training set.

More recently, focus has moved on to current state-of-the-art object detection architectures. [7] used a YOLOv4-based approach to fault diagnosis of some substation equipment, including insulators, cables, and transformers. It locates equipment along with any abnormal heating areas, calculates spatial overlap between them to determine a fault status. Performance reported average precision of 92.2% and confirmed the potential of current object detection architectures for this task.

But the weakness of most of this presented research is the assumption of a large, balanced dataset. In most cases, this most severe of challenges, the scarcity of data, is handled superficially. Transfer learning has been an extremely useful means of making up for this. As illustrated by [8] and [9], learned features such as edges, textures, and shapes of a model pre-trained with a large dataset such as ImageNet [10], can be transferred in a new domain, such as images of electrical equipment under thermal imaging, with much less data as a fine-tuning set. This approach has been successful in some medical and industrial imaging applications with limited data. This research is placed firmly within the context of pragmatic studies, with a focus on an exacting use of transfer learning as being the most realistic first step towards this specific problem, as it opens towards further advanced solutions such as Generative Adversarial Networks (GANs) towards synthetic data generation in due course [11, 12].

3. Proposed methodology

This section details the proposed framework for developing and evaluating the baseline fault diagnosis model.

3.1. Dataset and preprocessing

This work's first dataset is composed of thermal images acquired using active substation equipment. It is composed of diverse equipment, including bushings, connections, transformers, circuit breakers, relay and surge arresters, with varied operational as well as environmental conditions. In this first task, images will be grouped into two main groups: “Normal” and “Anomaly” as the base of a binary classification problem.

Before being presented to the model, they go through a standardized preprocessing pipeline:

1. Resizing: All images were resized with a fixed size of (224×224 pixels) to meet the input condition of our adopted pre-trained CNN architecture.
2. Normalization: The values of pixel intensity were normalized such that they remain in the interval [0, 1], inviting

stability and acceleration of training in models.

3.2. Data augmentation

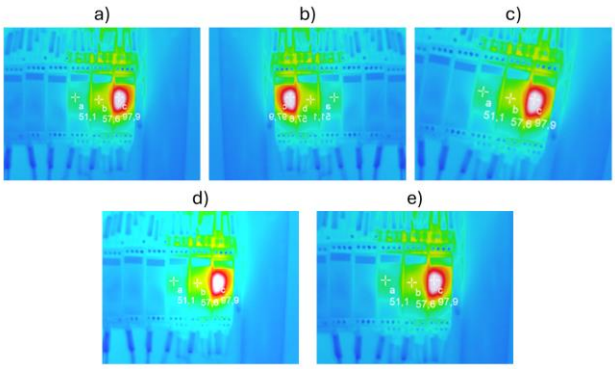
Given the anticipated class imbalance, where "Anomaly" images are expected to be significantly fewer than "Normal" ones, data augmentation plays a critical role. To enhance the representation and variability of the minority class, a series of geometric and photometric transformations will be selectively applied to the "Anomaly" images. This strategy helps reduce model bias toward the majority class and improves generalization.

The augmentation techniques applied were:

- Horizontal flips.
- Random rotations (within a range of -15 to +15 degrees).
- Random adjustments to brightness and contrast.
- Random scaling or zoom (90-110% of the original size).

Figure 1 illustrates the effect of these transformations.

Figure 1. An example of original thermal image of a fault, and 4 augmented versions.



a) Original. b) Flipped. c) Rotated. d) Brightness adjusted. e) Scaled.

3.3. Transfer learning model architecture

This framework utilizes a transfer learning approach with the architecture of ResNet50 [13], a highly regarded and potent CNN pre-trained with the ImageNet dataset. The reasoning is to build upon the rich low-level features learned in ResNet50, rather than attempting to learn them from a limited dataset.

The execution of the training was as follows:

1. Load Pre-trained Base: The ResNet50 was initialized with its ImageNet weights, while ignoring its final classification layer.
2. Freeze the Base Layers: The weights of the base convolutional layers were "frozen," meaning they were not updated during the initial training phase. This helped preserve the valuable features already learned.

3. Add a Custom Classifier Head: A fresh stack of layers was added atop the "frozen" base model. This head included a Global Average Pooling (GAP) layer, followed by a Dropout layer for regularization purposes. Ultimately, there was a dense layer featuring a Sigmoid activation function to generate a probability score for binary classification.
4. Train and Fine-Tune: The model underwent two training stages. Initially, only the new classifier head was trained. After it converged, some upper layers of ResNet50 were unfrozen for fine-tuning with a low learning rate, allowing the pre-trained features to better adapt to the thermal imagery.

3.4. Validation of the Framework and Behavioral Analysis

The evaluation in this early phase of research is not concerned with optimizing performance metrics, but rather with two main goals: (1) verifying the overall experimental framework's functional integrity, and (2) conducting a preliminary behavioral analysis of the model when it is exposed to the highly imbalanced dataset.

To make sure that the evaluation is objective, the dataset will be split into standard training, validation, and testing sets. The main sign that the basic model is learning will be that the training loss function is going down over the first few epochs. This shows that the model can extract features from the data.

The preliminary analysis will be based on a qualitative assessment of the predictive effectiveness of the model. The main goal is to explain why the model initially favors the "Normal" class, which is a common and expected result from the data distribution. The goal of this analysis is not to determine precise performance scores. Rather, it examines how the model's predictions evolve over time by testing them on the validation set. Making qualitative observations and verifying procedures are among the objectives of this study's first section. The behavioral analysis will yield preliminary findings that will be used to empirically support the next steps in the research process, which will center on systematic hyperparameter tuning and the application of more sophisticated techniques to address the observed class imbalance.

4. Experimental setup and preliminary process validation

This section presents the implementation of the framework outlined in Section 3. The primary aim of this initial stage was not to produce a fully optimized model, but rather to construct and validate the entire experimental pipeline, ensuring the robustness of the data processing, model training, and evaluation process.

4.1. Execution of the proposed methodology

The workflow was carried out exactly as the methodology specified. A typical 70-15-15 split was used to divide the small dataset into training, validation, and testing sets. Using a custom data loader, the data augmentation techniques, such as random rotations, flips, and brightness adjustments, were successfully applied in real-time to the training batch. The experiment's main component involved loading a pre-trained ResNet50 model, freezing its convolutional base layers, and swapping out the final classification layer for a new head designed specifically for the binary classification task ("Normal" vs. "Anomaly").

The training script was developed to execute the two-stage training process: initial training of only the new classifier head, followed by a fine-tuning phase with a low learning rate. The process was monitored using standard metrics, such as training and validation loss and accuracy, which were logged after each epoch.

4.2. Preliminary observations

The initial training epochs were primarily used to confirm that the experimental pipeline was functioning as planned. A distinct and illuminating performance trend emerged from these early runs: the model performed significantly better at correctly classifying images in the "Normal" class while struggling to identify images in the "Anomaly" class.

The inherent structure of the training dataset is the direct cause of this behavior. The "Normal" class gives the model a rich and varied collection of examples by supplying a large number of thermal images from a wide range of equipment types. Because of the data's richness, the network can learn a reliable and broadly applicable feature representation of what constitutes typical equipment operation in various settings. As a result, the model gains the ability to recognize new, undetectable examples of equipment that is healthy.

The "Anomaly" class, on the other hand, is incredibly sparse, covering only a few types of equipment and making up a very small portion of the entire dataset. This gives the model a very small feature space to learn from. Insufficient variation significantly impairs the model's capacity to generalize the appearance of an "anomaly," which results in a high rate of misclassification for this crucial class. The model rapidly discovers that creating a strong bias towards the majority "Normal" class is the best way to reduce the overall error.

This observation, while predictable, is a crucial preliminary finding. It verifies empirically that the main obstacles to overcoming are the stark class disparity and the lack of diversity in the fault data. The pipeline's successful deployment, together with this diagnostic of the model's

initial behavior, confirms the research direction and emphasizes the need for the suggested methodological focus on new strategies to generate synthetic data and weighted loss functions that are intended to address this imbalance.

5. Conclusion and future work

A thorough and practical methodological framework for creating a fundamental AI model for substation fault diagnosis using thermal images was provided in this paper. The suggested method makes use of transfer learning and data augmentation to establish a strong baseline while acknowledging the practical limitation of data scarcity. The creation and procedural validation of this framework, which verifies that every step is operational and appropriately configured, constitutes the main contribution of this study.

The first crucial step in a larger research agenda is the effective application of this framework. To fully develop the diagnostic capabilities of the model, future work will involve the thorough execution of the training and hyperparameter tuning process. The final performance outcomes from that thorough process will be saved for a later publication that will include a detailed evaluation of the model's efficacy as well as research into explainability strategies to increase confidence and real-world adoption.

The future research will proceed along three general paths:

1. Improvement of the transfer learning technique: Evaluating the use of thermal image datasets from similar equipment to augment the current dataset, as well as testing other pre-trained networks.
2. Generation of Advanced Data: Based on the baseline output, studies and implementation of advanced artificial data generation schemes will be under consideration, using likely Generative Adversarial Networks (GANs), aimed at narrowing down and refining the dataset with rare fault examples as well as severe ones.
3. Explainable AI (XAI): Explainability models, such as Grad-CAM, will be integrated to have graphical evidence of the model's predictions. That is necessary to build confidence and change the model from being a “black box” to an explainable and reliable maintenance engineer tool.

Acknowledgement
This work was carried out with the support of CTG Brasil, to whom the authors are grateful for providing the inspection data used in this project. The data were part of the ANEEL Research and Development (R&D) project PD-10381-0121/2021, titled "Autonomous Robot for Inspection in Galleries and Substations,"

developed by SENAI CIMATEC in partnership with CTG Brasil and Pollux Automation.

References

[1] National Fire Protection Association. *NFPA 70B: Standard for Electrical Equipment Maintenance*, 2023.

[2] Goodfellow, I., Bengio, Y., & Courville, A. *Deep Learning*. MIT Press, 2016. ISBN: 9780262035613.

[3] Usamentiaga, R., Venegas, P., Guerediaga, J., Vega, L., Molleda, J., & Bulnes, F. G. Infrared thermography for temperature measurement and non-destructive testing. *Sensors*. 2014;14(7):12305-12348. DOI:10.3390/s140712305.

[4] NETA MTS-2019, *Standard for Maintenance Testing Specifications for Electrical Power Equipment and Systems*, 2019.

[5] Ullah, I., Yang, F., Khan, R., Liu, L., Yang, H., Gao, B., & Sun, K. Predictive Maintenance of Power Substation Equipment by Infrared Thermography Using a Machine-Learning Approach. *Energies*. 2017;10(12):1987. DOI: 10.3390/en10121987.

[6] Wang, K., Zhang, J., Ni, H., & Ren, F. (2021). Thermal Defect Detection for Substation Equipment Based on Infrared Image Using Convolutional Neural Network. *Electronics*. 2021;10(16):1986. DOI: 10.3390/electronics10161986.

[7] Liu, T., Li, G., & Gao, Y. Fault diagnosis method of substation equipment based on You Only Look Once algorithm and infrared imaging. *Energy Reports*. 2022;8:171-180. DOI: 10.1016/j.egyr.2022.05.074.

[8] Mahmoud, K.A.A., Badr, M.M., Elmalhy, N.A., Hamdy, R.A., Ahmed, S., Mordia, A.A. Transfer learning by fine-tuning pre-trained convolutional neural network architectures for switchgear fault detection using thermal imaging. *Alexandria Engineering Journal*. 2024;103:327-342. DOI: 10.1016/j.aej.2024.05.102.

[9] Elgohary, A.A., Badr, M.M., Elmalhy, N.A., Hamdy, R.A., Ahmed, S., Mordia, A.A. Transfer of learning in convolutional neural networks for thermal image classification in Electrical Transformer Rooms. *Alexandria Engineering Journal*. 2024;105:423-436. DOI: 10.1016/j.aej.2024.07.077.

[10] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. ImageNet: A large-scale hierarchical image database. *IEEE conference on computer vision and pattern recognition*. 2009:248-255. DOI: 10.1109/CVPR.2009.5206848.

[11] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems*. 2014:2672-2680. DOI: 10.48550/arXiv.1406.2661.

[12] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. Self-attention generative adversarial networks. *International conference on machine learning*. 2019:7354-7363). DOI: 10.48550/arXiv.1805.08318.

[13] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *IEEE conference on computer vision and pattern recognition*. 2016:770-778. DOI: 10.1109/CVPR.2016.90