2024 International Nuclear Atlantic Conference – INAC 2024
Escola de Guerra Naval, Rio de Janeiro, RJ, May 6 - 10, 2024

ISBN: 978-65-594-1256-3

Explainable Artificial Intelligence Applied to Images of Two-Phase Flow

S. M. C. Schott[1] and R. N. de Mesquita[2]

Instituto de Pesquisas Energéticas e Nucleares (IPEN / CNEN - SP) Av. Professor Lineu Prestes 2242

[1] *sandro.schott@usp.br*
[2] *rnavarro@ipen.br*

1. Introduction

The interest in research within the field of Artificial Intelligence (AI) has been growing due to significant advances that have emerged in the area over the last decade. An increasing number of applications of significant social and economic impact have emerged . A substantial part of this progress is due to a specific domain within AI, namely the methods of deep neural networks, or deep learning. These methods have significantly elevated the state of the art in tasks such as computer vision (e.g., image classification, localization, and segmentation), speech recognition, natural language processing, drug discovery, and genomics [1].

However, this progress has mainly concentrated on enhancing task performance, highlighting a key aspect of neural networks: their nature as 'black boxes', meaning their decision-making processes are not understood by humans.. For many applications, it is important to have reliable, transparent and explainable models. Therefore, the field of eXplainable Artificial Intelligence (XAI) is becoming more important.[2].

The classification of two-phase flow patterns is important as they are used in models predicting heat transfer and pressure drop [3]. Among various approaches, a technique that has been studied involves classifying flow patterns based solely on images in the visible spectrum. These studies have followed the traditional approach of image classification with a first step of feature extraction and a second step of classification [3,4]. Recently, deep learning has been applied to this domain [5,6].

The present work applies XAI techniques to two-phase flow images connecting performance with transparency. Two XAI techniques are employed: filter visualization, visualization of trained filters, and layer output visualization. This application reveals aspects of the previously occult mechanisms used byCNNs to perform classification, helping to improve its performance. Moreover, biases in this classification due to experimental setup and other factors can be identified [6] where the CNN exclusively uses the edges of the channel for classification. Therefore, this work contributes to a more transparent and reliable application of AI.

This study applies XAI techniques to CNNs applied to images experimentally acquired from the three stages of chugging instability [7] under specific natural circulation conditions. The utilized database contained approximately 1500 images. A series of training sessions were conducted with simplified models referred to as "Toy" for comparison with more complex models, namely "Conv" and "VGG-16," along with a baseline. The dataset is split into training and testing sets, with the former used for parameter adjustment and the latter for performance evaluation.

2. Methodology

2.1 Convolutional Neural Networks

Traditional the use of neural networks for image classification involve two distinct modules. The first module manually extracts features from the image, which are chosen by experts where the variables are suited to the classification task. The second module differentiates classes through training. However, Convolutional Neural Networks (CNNs) have introduced an innovative approach by eliminating the need for manual feature extraction, allowing the network to automatically identify aspects relevant to the problem[1].

CNNs are designed to efficiently handle n-dimensional inputs, such as images, by exploiting the correlation between neighboring pixels. This capability enables effective classification of small areas into a limited number of categories, something often overlooked in conventional neural network approaches to image processing [1].

2.2 image acquisition

This study employs a dataset that was previously established and utilized in earlier research [3,8] . Detailed information regarding the formation of this dataset can be found in [3]. The images were acquired using an experimental setup known as the Natural Circulation Circuit (NCC).

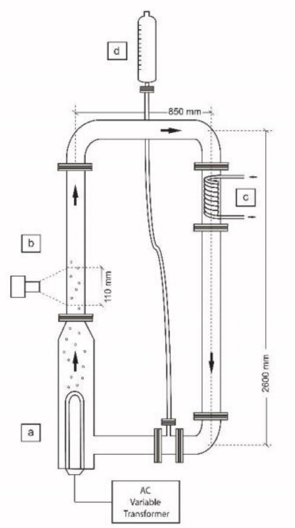

Figure 1: Natural Circulation Circuit (NCC).

3. Results and Discussion

In this section, three sets of results will be presented, one of performance and two applying two XAI techniques: training with different architectures on the dataset, an output from a second layer model with 8 filters, and the filters from a second layer model with 8 filters. The second and third results are from simplified models with few filters; this is done to enhance the visualization of XAI techniques. The results of all training sessions are displayed in Table I. The models are named based on the first number representing the filter size and the subsequent numbers indicating the number of filters per layer. For instance, the model "Toy_conv_32_4x8" has a filter size of 32, with 4 and 8 filters in the first and second layers, respectively.

Table I: The performance on the test subset of the Toy architectures is as follows

	Precision	**Recall**	**F1-Score**	**Accuracy**
Linear	0,932	0,889	0,907	0,947
Toy_conv_32_4x8	0,923	0,901	0,910	0,952
Toy_conv_3_4x8	0,956	0,948	0,952	0,973
Conv_3_64x128	0,980	0,944	0,960	0,973
VGG-16	0,995	0,986	0,990	0,995

Through Table I, it is observed that the linear model can achieve an already high F1-Score result of 0.907, and the addition of convolutional layers with size 32 practically does not change this value. Despite this, such an architecture allows for better visualization of filters and technique studies. On the other hand, changing the filter size to 3 increases the F1-Score to 0.952, a substantial improvement, indicating that this is the ideal size when aiming for performance. There is also a slight performance increase when significantly increasing the number of filters in the Conv_3_64x128 architecture. Finally, a considerable gain is noted with the use of VGG-16, achieving 0.990 in F1-Score.

Applying filters (the first XAI technique employed in this study) to the images yields outputs, or feature maps, or outputs of layers, for each layer. For brevity, only the results for the Incubation class are presented.

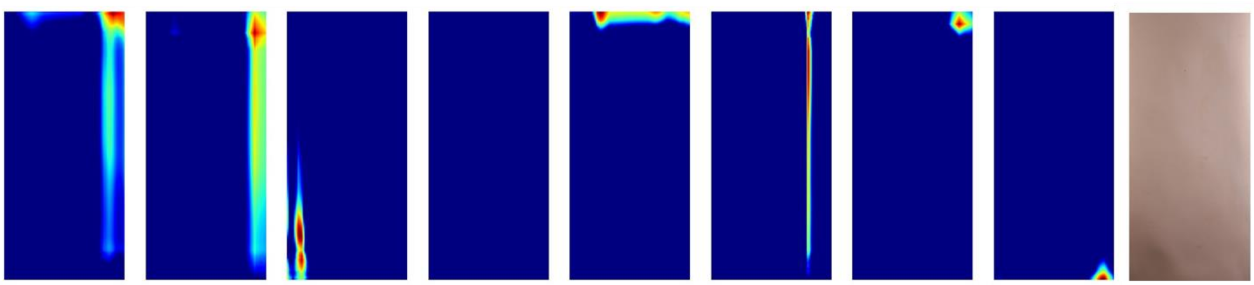

Figura 2: Features maps , second layer, Incubation class, Toy_4x8_32 model

In Figure 2, it is observed that significant activation occurs only in the first feature map. This aligns with the fact that expressive pattern formations occur only in the first filter. Another noteworthy point is the presence of a setup bias, as for the incubation class, activation primarily occurs along the edges, neglecting the central patterns of the images.

Another XAI technique employed is the visualization of filters. To achieve this, a CNN with filters of size 32 was utilized. The output was printed in grayscale for enhanced visibility. The results are shown in figure 3.

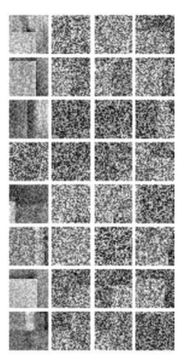

Figure 3: Filters from the second layer after training the Toy_4x8_32 model.

Through Figure 3, it is evident that the patterns in the second layer are indeed more complex and concentrated in the first slice, represented in first column in Figure 3, created by the first filter in that model. Results indicate the presence of square patterns in both filters and activations, as expected, given the observation in Figure 2, given that the patterns in the filters create the patterns in the feature maps

4. Conclusions

This experiment demonstrates that the visualization of filters and layer outputs is not productive as it works well only for simple models. Firstly, in more complex models, the filter size is typically set to 3, resulting in limited explanatory power in their visualization. Secondly, complex models may involve up to 500 filters, compromising explicability. Thus, it becomes necessary to explore more compelling eXplainable Artificial Intelligence (XAI) techniques as Integrated Gradients.

Acknowledgements

This research was supported by Comissão Nacional de Energia Nuclear (CNEN).

References

[1] LECUN, Y.; BENGIO, Y.; HINTON, G.. “Deep learning”. *nature*, (2015).

[2] GUNNING, D.; AHA, D. “DARPA’s explainable artificial intelligence (XAI) program”. *AI magazine*, (2019).

[3] DE MESQUITA, R. N. et al. “Classification of natural circulation two-phase flow patterns using fuzzy inference on image analysis”. *Nuclear Engineering and Design*, (2012).

[4] GHANBARZADEH, S.; HANAFIZADEH, P.; HASSAN SAIDI, M. “Intelligent Image-Based Gas-Liquid TwoPhase Flow Regime Recognition”. *Journal of Fluids Engineering*, (2012).

[5] JUNIOR, L. H. S.; BARBOSA JR, J. R.; DA SILVA, A. K. “Multi-parameter classification and quantification of R-134a condensation using machine learning”. *Applied Thermal Engineering* (2023).

[6] SCHOTT, S. M, C. “Classificação de padrões de escoamento bifásico por meio de redes neurais convolucionais”. Universidade de São Paulo, (2022).

[7] BOURE, J. A.; BERGLES, A. E.; TONG, L. S. “Review of two-phase flow instability”. *Nuclear Engineering and Design*, (1973)

[8] DE MESQUITA, R. N. et al. “Classification of natural circulation two-phase flow image patterns based on selforganizing maps of full frame DCT coefficients”. Nuclear Engineering and Design, (2018).