Received September 18, 2020, accepted October 13, 2020, date of publication October 16, 2020, date of current version October 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3031683

Applying Convolutional Neural Networks to Detect Natural Gas Leaks in Wellhead Images

ROBERLÂNIO OLIVEIRA MELO, M. G. F. COSTA, (Member, IEEE), AND CICERO F. F. COSTA FILHO, (Member, IEEE)

Department of Electronic and Computation, Federal University of Amazonas, Manaus 69080-900, Brazil

Corresponding author: Roberlânio Oliveira Melo (roberlaniomelo@gmail.com)

This work, according to Article 48 of Decree n° 6.008/2006, was supported in part by the Samsung Electronics of Amazonia Ltd., through the terms of Federal Law n° 8.387/1991, signed with the Center for Research and Development in Electronics and Information from the Federal University of Amazonas-CETELI/UFAM under Agreement 004; and in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, under Grant 001.

ABSTRACT Detecting natural gas leaks is one of the most important measures in the oil industry for preventing accidents. The literature provides different techniques for detecting natural gas leaks. However, except for previous studies by the authors on this topic, there remains a gap in the literature on leak detection of natural gas using digital images, without the need for sensors or special cameras calibrated for the spectrum of methane molecules. These previous studies used image-processing techniques associated with a novelty filter classifier to detect the presence or absence of visible cloud of hydrocarbon vapors, that is, a natural gas plume in Closed Circuit Television (CCTV) frames installed in onshore wellheads. In this article, we present a new method for detecting natural gas leaks in oil facilities that enhances the results obtained previously, along with the Gradient-weighted Class Activation Mapping Algorithm (Grad-CAM) to identify natural gas leaks. In this new method, convolutional neural networks (CNN) are applied to classify images (CCTV frames) as belonging to classes with or without natural gas leaks in onshore wellheads. Experimental results showed that the best performance model presented an accuracy of 99.78% and false negative rate of 0.00%.

INDEX TERMS Gas leak, gas detection, convolutional neural network, onshore wellhead image, gas and oil industry.

I. INTRODUCTION

Natural Gas (NG), extracted from oil wells, is composed of methane, ethane, propane, butane, pentane and other hydrocarbons, in addition to nonhydrocarbon compounds (water, carbon dioxide, etc.), but 70 to 98% of the NG is composed of methane [1], being a major contributor to the greenhouse effect [2]. In addition to the environmental impacts, the effects of natural gas leak accidents are commonly lethal to humans and capable of destroying oil and gas processing plants [3].

Considering industrial processes involving natural gas, whether in production, processing, transmission and distribution of this fuel, the sources of methane emission to the atmosphere are due to well completions, blowdowns, workovers, reciprocating compressor rod packing, pipe leaks, compression station leaks, and oil wellhead leaks [4]. Therefore, to mitigate the effects of natural gas leaks on the greenhouse effect, it is mandatory to develop efficient prevention systems, such as natural gas leak detection systems.

Natural gas leaks, besides contributing to global climate change, have catastrophic consequences in industrial facilities, especially when such leaks are followed by an explosion [3], causing significant loss of human lives and financial losses [5]. To mitigate events involving methane leaks and explosions, the options range from area classification to monitoring and leak detection of this gas in industrial processing plants.

From the underground rocks, wellheads extract oil, a gaseous mixture of hydrocarbons and other non-hydrocarbons under high pressure. Leaking gas under high pressure will produce a visible vapor cloud emanating from a wellhead leak site. This cloud is the result of hydrocarbon and water vapors that condense the humidity of the surrounding air [6], [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenhua Guo.

In the literature, several methodologies have been published regarding leak detection. The biological monitoring methods involve employing people and dogs [8]. The hardware-based methods require specialized sensors for this purpose [9], such as acoustic, optical, electrochemical, and ultrasonic sensors [8]–[10]. Software-based methods, using intelligent computational algorithms [8], analyze, and classify sensor signals that monitor variables, such as flow, pressure, and temperature [9].

Other studies suggest the detection of leakage of combustible gases using image processing, however, the images of these works are thermal images acquired by high-cost hardware, which are infrared cameras specifically for detecting the infrared spectrum of the methane molecule [11]–[14].

Only two previously works, that we published in literature [15], [16], used visible image processing to detect gas leakage. These works employed the novelty filter classifier.

In this article, we also intend to detect gas leakage using visible images. Motivated by the excellent performance of CNNs in image classification tasks and given the lack of studies applying CNN to image classification regarding the presence or not of natural gas leakage (two-class problem), using visible images, we aim at investigating this problem. Therefore, this article proposes a new model for natural gas leak detection, using visible images, based on Convolutional Neural Networks.

The great difference between our work and others that use the infrared spectrum is that in our study we only use images obtained by simple cameras of CCTV systems to detect visible gas clouds.

Another objective of the work is to provide a location map that highlights the most important regions in the image that contributed to the detection of natural gas leakage, that is, this work will generate visual explanations of the leak, which would help untrained users to discern the leak location. For this we used the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [17].

To compare the results of this new method with those obtained by previous studies that used the novelty classifier [15], [16], this work uses the same image dataset of these studies.

To minimize the overfitting, we enrich the data set, through a data augmentation technique and also used the Transfer Learning (TL) technique. In TL, the following pre-trained deep networks were evaluated: AlexNet [18], DenseNet-201 [19], GoogLeNet [20], MobileNet-v2 [21], ResNet-18 [22] and VGG-16 [23].

Therefore, the main objectives of this article are as follows:

- Propose and evaluate the performance of different CNN architectures and choose the one that provides the best performance in natural gas leak detection.
- Compare the best performance model with the results obtained with TL, using AlexNet, DenseNet-201, GoogLeNet, MobileNet-v2, ResNet-18 and VGG-16.
- Compare the performance of the CNN based model with the previous novelty filter based model [15], [16].

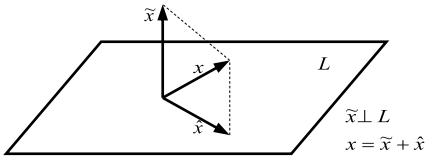

FIGURE 1. Illustration of novelty filter concept in R^3 space. Concept using Gram-Schmidt Orthogonalization Method [16].

- Provide a visual explanation of gas leakage, illustrating the regions of the image that were relevant to the ''with leak'' class label.

The paper is organized as follows: section II introduces the related; section III provides details of the image datasets used in this work and presents the proposed CNN models and performance evaluation metrics; section IV presents and discusses the results and the last section provides a conclusion.

II. RELATED WORKS

In [15], [16] the authors described a first natural gas leak detection system, using production head images, based on a surveillance system (CCTV). This system did not need installation of specific sensors. The proposed method was based on the use of the RGB color space component R and the novelty filter concept [24]. The novelty filter classifier obtained an AUC (Area Under Curve ROC - Receiver Operating Characteristic) of 98.53%, with an average error of 0.37%.

The novelty filter concept was described by Kohonen and Oja [24], based on the classic Gram-Schmidt orthogonalization method. Let $\{x_1, x_2, \ldots x_m\}$ be a set of n dimensional Euclidian vectors which span an m dimensional subspace $L \subset R^n$, with $m < n$. Considering the subspace L, an arbitrary vector $x \in R^n$ can be divided into two components, $\hat{x}$ and $\tilde{x}$ and Figure 1 illustrates these two components in R^3 space, where: $\hat{x} \in L$ is the projection of x on L and $\tilde{x} \perp L$ is the projection of x perpendicular to L. Vector $\hat{x} \in L$ is the component of x that is ''known'' by the subspace L and can be defined as a linear combination of $\{x_1, x_2, \ldots x_m\}$. Vector $\tilde{x} \perp L$ is the new information, that is unknown by the subspace L and cannot be represented as a linear combination of $\{x_1, x_2, \ldots x_m\}$ [15].

In this method, the magnitude of the vector $\tilde{x}$, $\|\tilde{x}\|$, is considered a measure of dissimilarity, that is, the higher the value $\|\tilde{x}\|$, the greater the probability that an image X is not part of subspace L. Therefore, considering subspace L, $\tilde{x}$ is named novelty and the system that extracts this component from x and shows it as an output can be named the Novelty Filter. In terms of pattern recognition, the novelty filter novelty rule can be applied to solve problems in which it is desired to classify a pattern as belonging or not to a class, subspace L.

Convolutional Neural Networks (CNN) [25] have been delivering excellent results in solving problems of different image processing areas that have a texture similar to the natural gas clouds used in this work. The first one work is this of cloud classification [26] and other is fire and smoke detection [27].

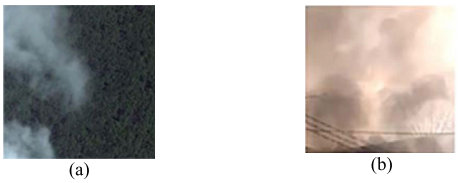

FIGURE 2. (a) image used by [26] to detect clouds and (b) image used by [27] to detect smoke.

We highlight the studies of cloud classification [26] and fire and smoke classification [27], as their images have a similar texture to the natural gas leak clouds. Figures 2(a) and 2(b), show clouds and smoke images, respectively, while Figures 4(b) and 4(d) show gas leak clouds.

To detect and distinguish thin clouds from thick clouds, Xie *et al.* [26] combine the superpixel approach with deep learning. The experiment dataset consists of 81 satellite images. From these images, patches were extracted with thick clouds, thin clouds and cloudless, totaling 26,000, 22,000 and 156,000 patches, respectively. They obtained a hit rate of 0.9454 and an error rate of 0.0330.

Namozov and Cho [27] applied the CNN concepts to fire and smoke detectors, but instead of using traditional rectified linear units or tangent functions, they used an adaptive piecewise linear unit as activation function of the hidden layers. The original dataset of this study is composed of 2,440 images, divided into 1,220 smoke images and 1,220 fire images. To overcome the overfitting problem, they used a combination of traditional image transformation methods with Generative Adversarial Networks (GANs) to create a large dataset. An accuracy and a false alarm rate of 0.9485 and 0.63, respectively, were obtained.

Wang *et al*. [14] proposed the detection of methane leaks using a database composed of infrared images obtained by a special camera, FLIR GF-320 infrared camera, whose imaging sensor is specifically sensitized to detect methane molecules. These authors applied CNNs for leak detection. The overall detection accuracy was close to 95%.

Shi *et al*. [13] also used the FLIR-320 infrared camera, the same hardware used by Wang *et al*. [14], to detect methane leaks. The authors used the Faster Region-Convolutional Neural Network technique to analyze the videos, achieving an average accuracy of 98%.

III. MATERIALS AND METHODS

The block diagram of Figure 3 presents the main steps of the methodology employed in this work to detect natural gas leak. Initially, at the data preparation block, the dataset images go through a preprocessing stage. After, we used a data augmentation technique. This augmented dataset is divided into three subsets: training, validation and testing. Section III.A presents the materials of this work and the steps just described.

In the model building block, three CNN architectures are proposed. Section III.B presents these architectures. In training these architectures, to obtain a network with better performance, we evaluated different optimization algorithms and regularization techniques. The training, validation and testing of the proposed architectures are detailed in section III.C. Once the best performance network and its parameters have been chosen, pre-trained networks are implemented through TL to perform the classification task. TL networks are presented in section III.D. Finally, we used the best performance network to visualize the class activation map. This theme is presented in section III.E.

A. DATASET

The original image dataset used in previous studies with the novelty filter is comprised of 2,810 images [16]. The size of these images is 184 × 160 pixels. However, in a recent study [28], this set was supplemented with images from the same wellhead, totaling 3,060 samples.

In this work, to reduce possible overfitting [29], we used the largest dataset. In addition, images with similar content were excluded. These excluded images were obtained from consecutive frames, in which the time between two consecutive images was 1/24 seconds. In this period the shape of the vapor cloud did not change. After exclusion, the total number of images was 2,980. This dataset is divided into 2,000 images of the ''without leak'' class and 980 of the ''with leak'' class. Figure 4 illustrates some of these images.

To obtain leak images, natural gas leak simulations were performed at the wellhead, by opening a needle valve intentionally and in a controlled way. To prevent accidents or needless damage to the environment, the volume of gas released during the leak simulations was the minimum necessary to collect the images to be used in this work.

Regarding the dimensions of the dataset, Aggarwal [25] states that CNNs should work with square images. Because the images have the dimensions of 184 × 160 pixels, we chose to resize all images to a standard size of 186 × 186 pixels. Aiming to prevent overfitting, we used the technique of on-the-fly data augmentation, also known as online augmentation [30]. In our proposal, the applied transformations were the following: rotated ± 20°, horizontal and vertical translation of ± 3 pixels, and a reduction of the image size with a factor between 0.5 and 1.0 [31]. These operations resulted in an augmented dataset with 10,132 images. Figure 5 illustrates such transformations

B. CNN ARCHITECTURE AND TRAINING PARAMETERS

According with Aggarwal [25], the increase in the number of hidden units in CNN increases the complexity of the modeled function, which can be useful for modeling difficult problems. However, this can cause overfitting if the problem is not so complex or when the dataset is small. A way to decrease complexity is to increase the length of the network by decreasing the number of units hidden in each layer [25]. This procedure was used by Miyagawa *et al.* [32]. The first CNN architecture proposed in this study for leak detection of natural gas, referred to as Long CNN (LCNN), is a serial architecture with 18 convolutional layers, based on the work of Miyagawa *et al*. [32]. Figure 6(a) illustrates this architecture. The width x length and depth of the first convolution

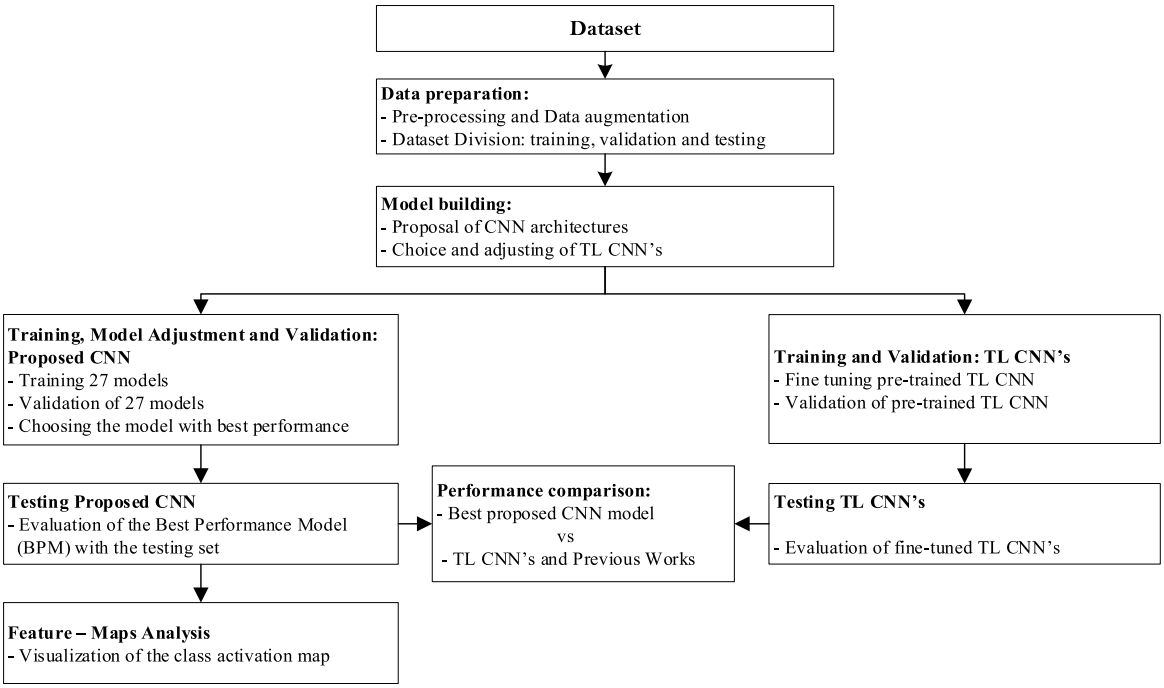

FIGURE 3. Overview of the proposed approach.

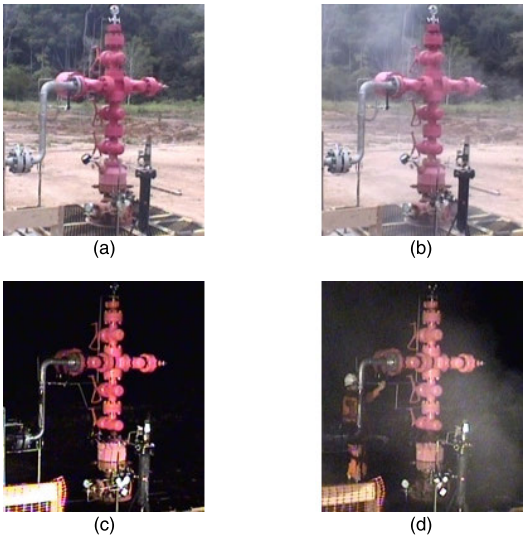

FIGURE 4. Sample images from the image dataset: (a) Daytime image without gas leak (without the hydrocarbon cloud), (b) Daytime image with gas leak (with the hydrocarbon cloud), (c) Nighttime image without gas leak (without the hydrocarbon cloud), (d) Nighttime image with gas leak (with the hydrocarbon cloud).

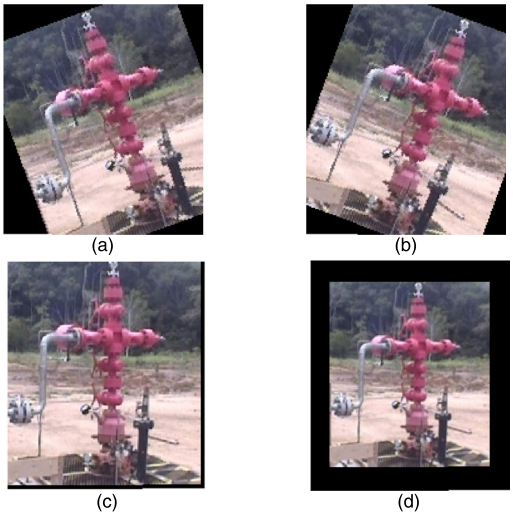

FIGURE 5. Visualization of examples of random transformations applied to four images of the dataset: (a) rotation of + 20°; (b) rotation of −20 °; (c) translation of − 3 by − 3 pixels; (d) image resizing with factor 0.8.

layer has the following dimensions: 186 ×186 ×16. Depth convolution layers have shorter width and length. The last convolution layer has a feature map with dimension of 11 ×11 ×128.

To investigate whether the LCNN was complex enough to solve the problem, with possible occurrence of overfitting, even after this tapering of units in the hidden layers, we proposed two other architectures with a lower number of convolution layers. These new architectures, also shown in Figure 6, are the following: Medium architecture (MCNN) with 6 convolution layers. And Short architecture (SCNN) with 2 convolution layers.

Each convolution layer is followed by a batch normalization layer [33], as a method to address the vanishing and exploding gradient problems [25], and a ReLu layer as the activation function. As shown in Figure 6, to reduce the spatial size of the feature map in depth layers, we applied max-pooling operations with 2 × 2 kernels, with stride 1 ×1. Each architecture has a classification step composed of the following layers: fully connected and softmax layer, which outputs a probability distribution of two classes, and a final classification layer.

To optimize the performance of the architectures, we evaluated the application of the following technique for improving generalization: Dropout [34], which randomly eliminates neurons of a given CNN layer, at a rate of 30%,

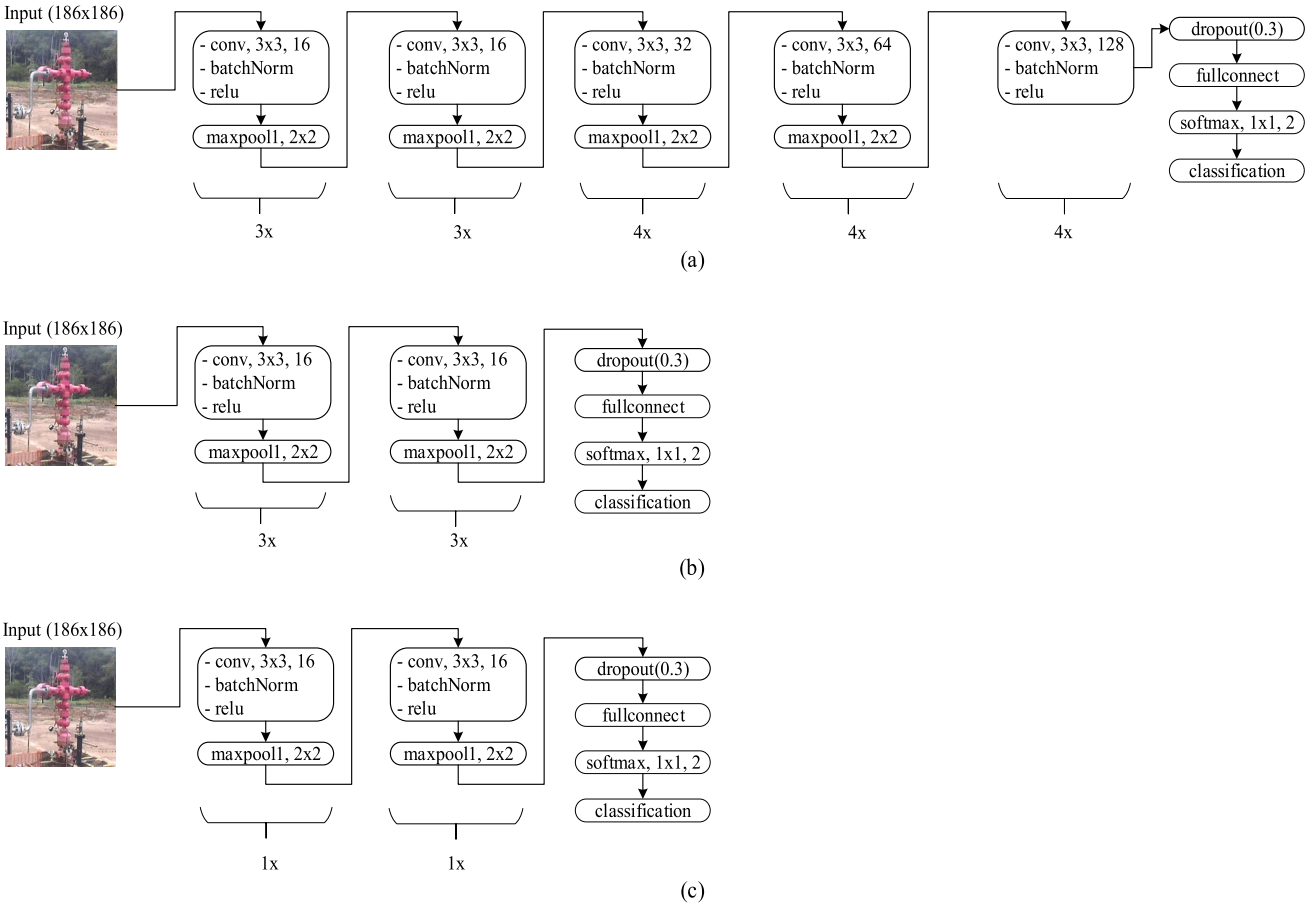

FIGURE 6. Proposed network architectures: a) LCNN, with 18 convolution layers; b) MCNN, with 6 convolution layers; c) SCNN, with 2 convolution layers.

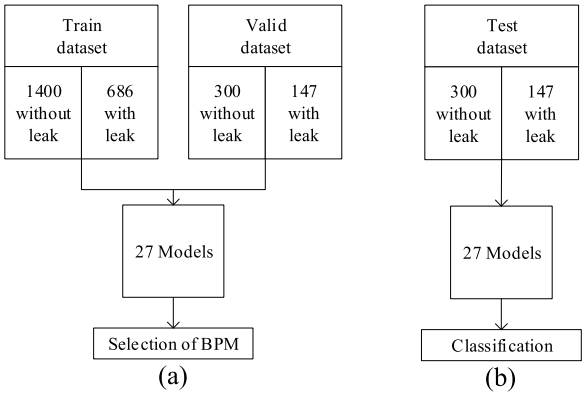

FIGURE 7. Block diagram f or training, evaluation and testing: (a) CNN training and evaluation to obtain the best performance model (BPM model); (b) BPM test with test dataset.

and L2 regularization [35], that regularizes the weights in the loss function with a smoothing coefficient of 0.0001.

The following optimization algorithms were evaluated: Adaptive Moment Estimation (ADAM) with decay rates β_1 of 0.9 and β_2 of 0.9 [36]; Root Mean Square Propagation (RMSProp) with decay rates β_2 of 0.9 [37], and Stochastic Gradient Descent with Momentum (SGDM), with momentum of 0.9 [38].

Other training hyperparameters are fixed as follows: maximum number of epochs is 15, mini-batch size is 16 and

TABLE 1. Contingency table used to calculate χ^2.

CNN	Novelty filter
Number of correct classifications	Number of correct classifications
Number of wrong classifications	Number of wrong classifications

learning is 0.001. Finally, for training stop, we used early stop with a validation patience of 4. This means that the training is interrupted when the loss in the validation set increases in four consecutive mini-batches [39], [40].

The combination of three architectures, three optimization algorithms, and two techniques for improving generalization resulted in twenty-seven distinct network models being analyzed.

C. TRAINING, VALIDATION AND TESTING

To evaluate the 27 models, the dataset is divided into the following proportions: 70% training (*train* dataset), 15% validation (*valid* dataset), and 15% testing (*test* dataset).

The methodology used for training, validation and testing is shown in Figure 7. As shown in Figure 7(a), the *valid* dataset is used to choose the Best Performance Model (BPM). After choosing the BPM, it is tested with the *test* dataset as shown in Figure 7(b).

The following metrics were used in evaluating the performance of the CNN models: accuracy, false negative rate,

TABLE 2. Performance of the 27 CNN models with the valid dataset showing the metrics accuracy (ACC), false negative rate (FNR) precision, recall, F1-score and chi-square (χ^2) test performed comparing the results of the CNN with the results of [16].

Architecture	Optimization	Regularization	Model	ACC (%)	FNR (%)	Precision (%)	Recall (%)	F1-Score (%)	χ^2
LCNN	ADAM	Dropout	1	99.33	2.04	100.0	97.96	98.97	25.57
		L_2	2	99.55	1.36	100.0	98.64	99.32	27.70
		L_2 + Dropout	3	99.55	1.36	100.0	98.64	99.32	27.70
	RMSProp	Dropout	4	76.06	72.79	100.0	27.21	42.78	104.35
		L_2	5	99.55	1.36	100.0	98.64	99.32	27.70
		L_2 + Dropout	6	99.55	1.36	100.0	98.64	99.32	27.70
	SGDM	Dropout	**7**	**99.78**	**0.00**	99.32	**100.0**	**99.66**	**29.95**
		L_2	8	99.55	1.36	100.0	98.64	99.32	27.70
		L_2 + Dropout	9	99.55	1.36	100.0	98.64	99.32	27.70
MCNN	ADAM	Dropout	10	97.32	8.16	100.0	91.84	95.74	10.75
		L_2	11	99.55	1.36	100.0	98.64	99.32	27.70
		L_2 + Dropout	12	97.32	8.16	100.0	91.84	95.74	10.75
	RMSProp	Dropout	13	95.53	13.61	100.0	86.39	92.70	3.28
		L_2	14	97.32	8.16	100.0	91.84	95.74	10.75
		L_2 + Dropout	15	92.17	23.81	100.0	76.19	86.49	0.51
	SGDM	Dropout	16	99.55	1.36	100.0	98.64	99.32	27.70
		L_2	17	97.32	8.16	100.0	91.84	95.74	10.75
		L_2 + Dropout	18	95.08	14.97	100.0	85.03	91.91	2.12
SCNN	ADAM	Dropout	19	97.32	8.16	100.0	91.84	95.74	10.75
		L_2	20	97.32	8.16	100.0	91.84	95.74	10.75
		L_2 + Dropout	21	98.88	3.40	100.0	96.60	98.27	21.61
	RMSProp	Dropout	22	99.55	1.36	100.0	98.64	99.32	27.70
		L_2	23	99.33	2.04	100.0	97.96	98.97	25.57
		L_2 + Dropout	24	98.66	4.08	100.0	95.92	97.92	19.78
	SGDM	Dropout	25	94.86	15.65	100.0	84.35	91.51	1.65
		L_2	26	99.55	1.36	100.0	98.64	99.32	27.70
		L_2 + Dropout	27	99.55	1.36	100.0	98.64	99.32	27.70

precision, recall and F1 score [41]. To present these metrics, we need the following definitions: True-Positives (TP): images with natural gas leak that CNN has classified as such; True-Negatives (TN): images without natural gas leak that CNN has classified as such; False-Positives (FP): images without natural gas leak that CNN classified as with natural gas leak and False-Negatives (FN): images with natural gas leak that CNN has classified as without natural gas leak.

The accuracy (ACC), shown in Equation (1), was used to define the BPM from the 27 proposed models.

In many machine learning applications, it is crucial to minimize the False Negatives Rate (FNR), as is the case with medical diagnoses, where a late or mistaken diagnosis can reduce the effectiveness of treatment [42]. Also, in autonomous vehicles, a small failure could cause a momentary malfunction in the car, rendering it a hazard [43]. This also applies to the oil industry: a false negative, when an actual leakage is classified as non-leakage, can lead to serious accidents. FNR is defined according to Equation (2). Other metrics are also used in this work to evaluate model performance, such as: precision [41], which determines the rate of positive samples that were correctly classified, that is, the ability of the model to avoid false positive; *recall* or sensitivity [41], which determines the rate of true positives that the classifier can provide and F1-*Score* [44] with values between 0 and 1, which represents a harmonic mean of precision and sensitivity. These metrics are shown in Equations (3)-(5).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$FNR = \frac{FN}{FN + TP} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{5}$$

In this work, we may be interested in answering whether the performance of the BPM is better than the performance obtained in the previous work of Melo and Costa [16]. To answer this question, we formulate a null hypothesis that the results are similar. To evaluate this null hypotheses we employed the Chi-Square Test [45], using the contingency table shown in Table 1. The values of Table 1 were collected in the *valid* dataset.

To confirm the results presented in this work, we adopted a significance level of 99% ($\alpha = 0.01$). The null hypothesis (H_0) proposes that no difference exists between the two methods. Considering that the dataset used for comparison between the two methods has N = 447 samples (*valid* dataset) and with a degree of freedom of 1, we have the value of $\chi^2_{critical} = 6.63$. If the value of $\chi^2_{calculated} > \chi^2_{critical}$, the null hypothesis is rejected.

The hardware used in this work has the following configuration: Intel^{TM}i7 processor 8th generation and GeForce MX150 GPU of 2 GB and software used was Matlab^{TM} version 2019b.

D. TRANSFER LEARNING

As stated by Yosinski *et al*. [46], when the dataset is limited, transfer learning can be a powerful tool to allow training the network without generating overfitting. In this work, six transfer learning models were fine-tuned: AlexNet [18], DenseNet-201 [19], GoogLeNet [20], MobileNet-v2 [21], ResNet-18 [22] and VGG-16 [23]. Transfer learning consists of using a pre-trained model and adapting it to a new dataset [47]. In this work, these models are shallow tuned (only the parameters of the last fully connected layer are tuned) to identify one of two classes: with or without a natural gas leak.

For training stop, both the proposed and transfer learning architectures used the early-stop method, with a patience of 4. A maximum number of 15 epochs was used.

E. VISUAL EXPLANATION

The gas leak clouds have slight nuances that are difficult for operators to perceive, leaving them uncertain that they are facing a real leak. Therefore, just identifying an image with a gas leak and informing the operator of its presence may not resolve all doubts. To improve the quality of the information provided to the operator, we employed the Gradient-weighted Class Activation Mapping (Grad-CAM) [17]. This tool provides a visual explanation, showing which region of an input image was relevant for the classifier to make the decision that there was a leak.

Grad-CAM produces heatmaps that highlight specific regions of the image that contributed to CNN prediction. This heat map is obtained by calculating gradients between the last convolution layer and the output of the desired class.

IV. RESULTS AND DISCUSSION

Table 2 presents results of the twenty-seven CNN models, showing the metrics presented in the last section. These results were obtained using the *valid* dataset. Among the 27 CNN models, model 7 showed the highest accuracy. This model, which uses the LCNN architecture, the SGDM optimization algorithm and L_2 regularization, obtained an accuracy of 99.78% and an FNR of 0%. Therefore, model 7 is our Best Performing Model (BPM).

In Table 2, models 2, 3, 5, 6, 8, 9, 11,16, 22, 26, and 27 provided ACC = 99.55%, lower than 99.78% of BPM. However, despite the high accuracy obtained (99.55%), these listed models presented FN classifications. This type of error can put an installation at risk, as classifying a positive image as negative for natural leakage means neglecting the potential risk of an accident. The BPM, besides presenting the highest accuracy (99.78%) among all models, is the only one free of FNs, that is, FNR = 0%.

TABLE 3. Performance comparison between the classifier BPM and novelty filter classifier [16].

Method	ACC (%)	FNR (%)	Precision (%)	Recall (%)	F1 Score (%)
Novelty Filter	93.15	8.78	91.22	91.22	91.22
CNN (BPM)	99.78	0.00	99.32	100.00	99.66

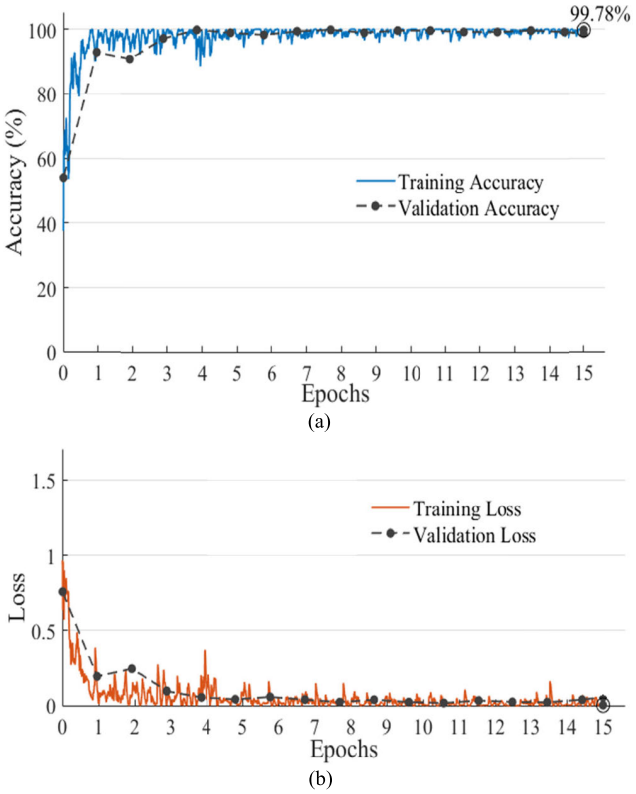

FIGURE 8. Training and validation of BPM: (a) accuracy during training and validation; (b) loss during training and validation.

Most of the models shown in Table 2 presented a higher performance compared with the previous method using the novelty filter [16], with a statistically significant difference ($\chi^2_{calculated} > 6.63$). The method using the novelty filter presented an accuracy of 93.15%. In this work, as shown in Table 2, BPM achieved an accuracy of 99.78%.

As shown in Table 2, the value of $\chi^2_{calculated} = 29.95$ for the BPM is greater than the $\chi^2_{critical}$. Thus, the null hypothesis is rejected. Therefore, CNN performed better than the novelty filter classifier, and this difference is statistically significant. Table 3 presents a comparison between the metrics obtained with the BPM model and the metrics obtained with the novelty filter technique.

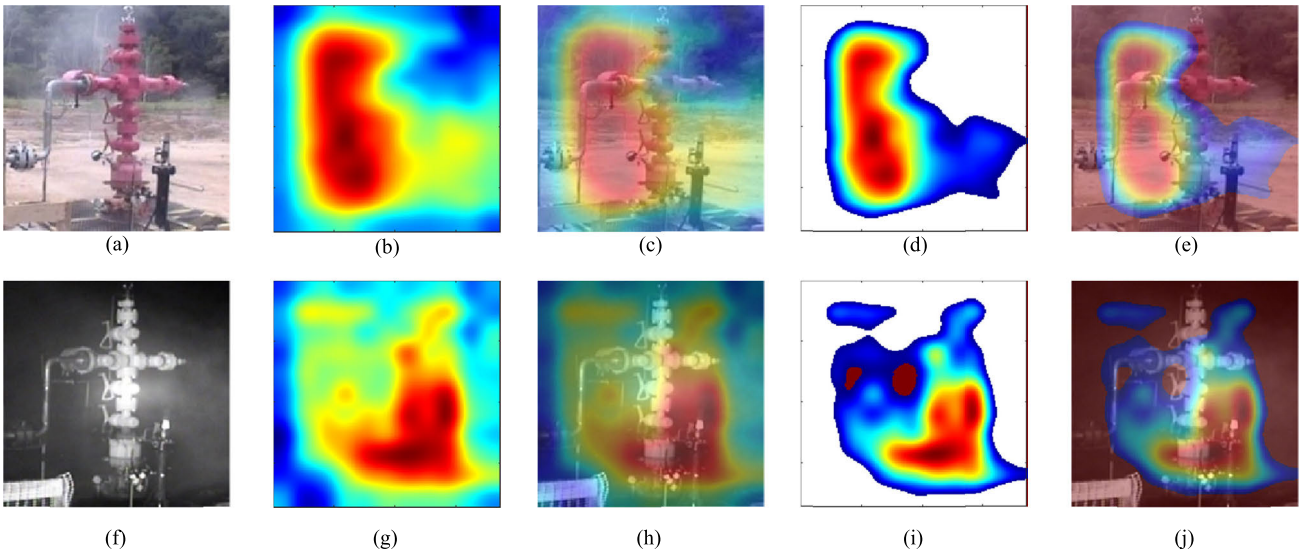

FIGURE 9. (a) and (f) examples of original images "with leak"; (b) and (g) heatmap of the original images; (c) and (h) locating the class discriminatory regions through Grad-CAM; (d) and (i) new heatmap with pre-processed magnitudes; (e) and (j) identification of the pixels that contributed to the "with leak" prediction. The pixels with red color contribute strongly to the score of the "with leak" class.

The accuracy and loss graphs during training and validation of the BPM model are shown in Figure 8(a) and Figure 8(b) respectively. The training accuracy was 100% and the validation accuracy was 99.78%. Classification of the *test* dataset (not used during training), also resulted in an accuracy of 99.78%, but with one FP.

A. PRE-TRAINED CNNS PERFORMANCE IN RELATION TO BPM

Among the pre-trained CNNs, DenseNet-201 and AlexNet provided the best results. As shown in Table 4, the accuracies obtained with pre-trained networks were $ACC_{DenseNet} = 99.55\%$ and $ACC_{AlexNet} = 99.33\%$. Therefore, the performance of the two pre-trained networks are lower than BPM, which presented an accuracy of 99.78%. This demonstrates the superiority of the BPM model compared with DenseNet-201 and AlexNet, and over other pre-trained networks.

According to Table 4, MobileNet-v2 presented the worst value for recall, 91.84%. The lower accuracy of MobileNet-v2 compared to BPM is due to the number of FNs. The accuracy of DenseNet-201 is 99.55%, close to the BPM, but the DenseNet-201 presented an FNR of 1.36%. This type of failure can be catastrophic in the oil industry, delaying or obstructing decisions that could prevent the explosion of a natural gas cloud.

With AlexNet, using the *test* dataset, an FNR of 2.04% was obtained, with three FNs and zero FP. With DenseNet-201, an FNR of 1.36% was obtained, with two FNs and zero FP.

B. VISUAL EXPLANATIONS

After defining the BPM, it was used to evaluate images with a gas leak using the Grad-CAM tool, aiming to identify the gas cloud. Two original images with a gas leak are shown in Figure 9(a) and 9(f). With Grad-CAM a heatmap is obtained that shows, in the original images, the pixels that have more influence in the CNN prediction of the "with leak" class.

TABLE 4. Performance comparison between the classifier BPM with the pre-trained networks: AlexNet, DenseNet-201, GoogLeNet, MobileNet-v2, ResNet-18 and VGG-16.

Method	ACC (%)	FNR (%)	Precision (%)	Recall (%)	F1 Score (%)	Time training and validation (min:sec)	Average time test (sec)
BPM	**99.78**	**0.00**	**99.32**	**100.00**	**99.66**	**11:46**	**0.0055**
DenseNet-201	99.55	1.36	100.00	98.64	99.32	397:29	0.0484
AlexNet	99.33	2.04	100.00	97.96	98.97	11:22	0.0049
GoogLeNet	98.66	4.08	100.00	95.92	97.92	21:27	0.0096
ResNet-18	98.43	4.76	100.00	95.24	97.56	16:04	0.0082
VGG-16	97.76	6.80	100.00	93.20	96.48	73:19	0.0228
MobileNet-v2	97.32	8.16	100.00	91.84	95.74	52:26	0.0180

Figure 9(b) and 9(g) show the heatmaps corresponding to original images shown in Figure 9(a) and 9(f), respectively. In these heatmaps, Grad-CAM shows that pixels in red color have a strong correlation with the "with leak" class. In Figure 9(c) and 9(h) we superimposed these red regions over the original images. As shown in these figures, there is a high coincidence between the red pixels and the gas cloud. To better highlight the regions strongly correlated with the prediction of the class "with gas", we calculated the average value of this heatmap and excluded activations with scores below this average value, resulting in the new heatmap shown in Figure 9(d) and 9(i).

Finally, in Figure 9(e) and 9(f) we superimposed the heatmaps of Figure 9(d) and 9(j) over the input images, providing better visualization of the pixels of the input image that contribute to the prediction of gas leakage

In Figure 10(a) we show an original image with gas leak that was incorrectly classified as with no gas leak in a previous work [15] and correct classified as with gas leak

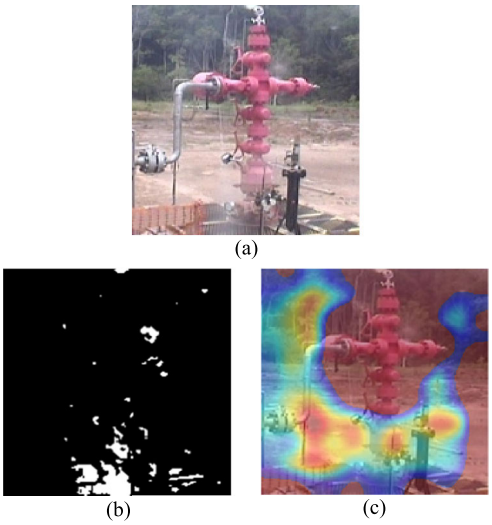

FIGURE 10. **(a) Image classified as FN in previous work [15] and properly classified (TP) by BPM.; (b) novelty region obtained by the technique proposed in. [15]; (c) heatmap of the original image in this work.**

in present work. Figure 10(b) shows the region detected as a novelty in [15], while Figure 10(c) shows the heatmap obtained with the BPM of the present work. Visually, the novelty region obtained by the novelty filter is lower than the region of high magnitude scores provided by the Grad-CAM heat map, thus suggesting that our technique has greater sensitivity than the previous one [15].

V. CONCLUSION

This study proposed a methodology using deep learning tools, more specifically convolutional neural networks, for digital image classification of a wellhead into two classes, ''with natural gas leak'' and ''without natural gas leak''. It is noteworthy that these images were obtained through CCTV surveillance systems already existing in the wellhead site, with no demand for additional facilities. To achieve our goal, we evaluated the performance of 27 different CNN models. The best performing model, BPM, had the following characteristics: 18 convolution layer architecture, SGDM optimization algorithm and dropout regularization technique. The accuracy obtained in the test set, 99.78%, is higher than similar studies found in the literature, which used the novelty classifier technique [16]. After applying a Chi-square (χ^2) hypothesis testing, it became clear that this higher performance is significant at a 99% significance level. The aforementioned accuracy obtained in the *test* dataset is close to the accuracy obtained in the *train* set, 100%, suggesting that the generalization techniques employed avoid overfitting. We also showed that the BPM outperforms pre-trained networks AlexNet, DenseNet-201, GoogLeNet, MobileNet-v2, ResNet-18 and VGG-16. Additionally, these pre-trained networks result in FNs, while BPM presented no FNs. The Grad-CAM tool proved to be of great importance for energy facility operators, as it shows, with heatmaps, the location of the gas cloud in original images. For future works we aim at evaluating the generalization of the model on unseen images of different types of wellheads.

ACKNOWLEDGMENT

Academic English Solutions (https://www.academicenglishsolutions. com) revised this article.

REFERENCES

[1] W. D. McCain, *Properties of Petroleum Fluids*. Tulsa, OK, USA: Pennwell Books, 2017.
[2] C. MAgee, ''Survey of natural gas leakage abatement best practices,'' California Public Utilities Commission, San Francisco, CA, USA, Tech. Rep. R.15-01-008 CEK/ek4, 2015.
[3] Y. F. Khalil, ''A probabilistic visual-flowcharting-based model for consequence assessment of fire and explosion events involving leaks of flammable gases,'' *J. Loss Prevention Process Industries*, vol. 50, pp. 190–204, Nov. 2017.
[4] M. Fulton, N. Mellquist, S. Kitasei, and J. Bluestein, ''Comparing life-cycle greenhouse gas emissions from natural gas and coal,'' Deutsche Bank Group, DB Climate Change Advisors, Frankfurt, Germany, Tech. Rep., 2011.
[5] M. T. Dröge and R. Kenter, ''Gas pipeline incidents: 10th report of the european gas pipeline incident data group,'' Tech. Rep. VA 17.R.0395, Mar. 2018, no. 10, pp. 1–50.
[6] G. Atkinson, J. Hall, and A. McGillivray, ''Review of vapour cloud explosion incidents,'' Health Saf. Executive, Harpur Hill, U.K., RR1113 Res. Rep., 2017.
[7] Board and B. M. I. Investigation, ''Buncefield major incident investigation: Initial report to the health and safety commission and the environment agency of the investigation into the explosions and fires at the buncefield oil storage and transfer depot, Hemel Hempstead, 11 December 2005,'' Health Safety Commission, Hertfordshire, U.K., Tech. Rep., 2006.
[8] M. A. Adegboye, W.-K. Fung, and A. Karnik, ''Recent advances in pipeline monitoring and oil leakage detection technologies: Principles and approaches,'' *Sensors*, vol. 19, no. 11, p. 2548, Jun. 2019.
[9] K. Sachedina and A. Mohany, ''A review of pipeline monitoring and periodic inspection methods,'' *Pipeline Sci. Technol.*, vol. 2, no. 3, pp. 187–201, Sep. 2018.
[10] P.-S. Murvay and I. Silea, ''A survey on gas leak detection and localization techniques,'' *J. Loss Prevention Process Industries*, vol. 25, no. 6, pp. 966–973, Nov. 2012.
[11] M. S. Jadin and K. H. Ghazali, ''Gas leakage detection using thermal imaging technique,'' in *Proc. UKSim-AMSS 16th Int. Conf. Comput. Modeling Simulation*, Mar. 2014, pp. 302–306.
[12] B. Liu, H. Ma, X. Zheng, L. Peng, and A. Xiao, ''Monitoring and detection of combustible gas leakage by using infrared imaging,'' in *Proc. IEEE Int. Conf. Imag. Syst. Techn. (IST)*, Oct. 2018, pp. 1–6.
[13] J. Shi, Y. Chang, C. Xu, F. Khan, G. Chen, and C. Li, ''Real-time leak detection using an infrared camera and faster R-CNN technique,'' *Comput. Chem. Eng.*, vol. 135, Apr. 2020, Art. no. 106780.
[14] J. Wang, L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell, D. Zimmerle, S. Savarese, and A. R. Brandt, ''Machine vision for natural gas methane emissions detection using an infrared camera,'' *Appl. Energy*, vol. 257, Jan. 2020, Art. no. 113998.
[15] C. F. F. C. Filho, R. de O. Melo, and M. G. F. Costa, ''Detecting natural gas leaks using digital images and novelty filters,'' in *Proc. 3rd Int. Conf. AIS*, vol. 7326 LNAI. London, U.K.: Springer, 2012, pp. 242–249.
[16] R. O. de Melo, M. G. F. Costa, and C. F. F. C. Filho, ''Using digital image processing and a novelty classifier for detecting natural gas leaks,'' in *Studies in Computational Intelligence*, vol. 542, L. Chen, S. Kapoor, and R. Bhatia, Eds. Cham, Switzerland: Springer, 2014, pp. 409–422.
[17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, ''Grad-CAM: Visual explanations from deep networks via gradient-based localization,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''Imagenet classification with deep convolutional neural networks,'' in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst.*, vol. 2, Dec. 2012, pp. 1097–1105.
[19] 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.
[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, ''Going deeper with convolutions,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[21] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[22] 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.
[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015, pp. 1–14.
[24] T. Kohonen and E. Oja, "Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements," *Biol. Cybern.*, vol. 21, no. 2, pp. 85–95, Jun. 1976.
[25] C. C. Aggarwal, *Neural Networks and Deep Learning*. Cham, Switzerland: Springer, 2018.
[26] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3631–3640, Aug. 2017.
[27] A. Namozov and Y. I. Cho, "An efficient deep learning algorithm for fire and smoke detection with limited data," *Adv. Electr. Comput. Eng.*, vol. 18, no. 4, pp. 121–128, 2018.
[28] R. O. Melo, C. F. F. C. Filho, and M. G. F. Costa, "Leak detection of natural gas with base on the components of color spaces RGB and HSI using novelty filter," *IEEE Latin Amer. Trans.*, vol. 12, no. 8, pp. 1560–1565, Dec. 2014.
[29] 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.
[30] P. Molchanov, S. Gupta, K. Kim, and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 1–7.
[31] J. Talukdar, A. Biswas, and S. Gupta, "Data augmentation on synthetic images for transfer learning using deep CNNs," in *Proc. 5th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2018, pp. 215–219.
[32] M. Miyagawa, M. G. F. Costa, M. A. Gutierrez, J. P. G. F. Costa, and C. F. F. C. Filho, "Detecting vascular bifurcation in IVOCT images using convolutional neural networks with transfer learning," *IEEE Access*, vol. 7, pp. 66167–66175, 2019.
[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 1, 2015, pp. 448–456.
[34] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[35] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, CA, USA, 2014. [Online]. Available: https://arxiv.org/abs/1412.6980
[37] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.
[38] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural Networks: Tricks of the Trade*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., 2nd ed. Berlin, Germany: Springer, 2012, pp. 599–619.
[39] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, "Identifying quantum phase transitions using artificial neural networks on experimental data," *Nature Phys.*, vol. 15, no. 9, pp. 917–920, Sep. 2019.
[40] C. Kofler, R. Muhr, and G. Spöck, "Classifying image stacks of specular silicon wafer back surface regions: Performance comparison of CNNs and SVMs," *Sensors*, vol. 19, no. 9, p. 2056, May 2019.
[41] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009.
[42] D. S. Kermany, M. Golfbaum, W. Cai, C. C. S. Valentim, H. Liang, S. L. Baxter, A. McKeown, and G. Yang, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, Feb. 2018.
[43] H. M. Song, J. Woo, and H. K. Kim, "In-vehicle network intrusion detection using deep convolutional neural network," *Veh. Commun.*, vol. 21, Jan. 2020, Art. no. 100198.
[44] C. J. Van Rijsbergen, "Foundation of evaluation," *J. Documentation*, vol. 30, no. 4, pp. 365–373, Apr. 1974.
[45] D. C. Howell, *Stastical Methods for Psychology*, 7th ed. Belmont, CA, USA: Cengage Wadsworth, 2010.
[46] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 2, 2014, pp. 3320–3328.
[47] A. Darwish, D. Ezzat, and A. E. Hassanien, "An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis," *Swarm Evol. Comput.*, vol. 52, Feb. 2020, Art. no. 100616.

ROBERLÂNIO OLIVEIRA MELO received the B.Sc. degree from Paulista University, in 2009, and the master's degree from the Electrical Engineering from the Federal University of Amazonas, in 2012, where he is currently pursuing the Ph.D. degree. His current research interests include neural networks and artificial intelligence, computer vision, and image processing.

M. G. F. COSTA (Member, IEEE) received the Electrical Engineering degree from the Federal University of Amazonas, Brazil, in 1983, and the M.Sc. and D.Sc. degrees in electrical engineering from the State University of Campinas, Brazil, in 1988 and 1996, respectively. She has been with the Federal University of Amazonas, since 1990, where she is currently a Full Professor with the Electrical Engineering Program, Federal University of Amazonas. Her research interests include biomedical image and artificial.

CICERO F. F. COSTA FILHO (Member, IEEE) received the Electrical Engineering degree from the Federal University of Pernambuco, Brazil, in 1982, and the M.Sc. and D.Sc. degrees in electrical engineering from the State University of Campinas, Brazil, in 1985 and 1996, respectively. He is currently a Full Professor with the Electrical Engineering Program, Federal University of Amazonas. His research interests include biomedical image and artificial intelligence.