Scientific Paper / Artículo Científico

UNIVERSIDAD POLITÉCNICA SALESIANA ECUADOR
INGENIUS
Revista de Ciencia y Tecnología
https://doi.org/10.17163/ings.n26.2021.07
pISSN: 1390-650X / eISSN: 1390-860X

Transfer Learning for Binary Classification of Thermal Images

Transfer learning en la clasificación binaria de imágenes térmicas

Daniel Pérez-Aguilar[1,*], Redy Risco-Ramos[1,*], Luis Casaverde-Pacherrez[1]

Received: 13-11-2020, Received after review: 17-12-2020, Accepted: 22-02-2021, Published: 01-07-2021

Abstract

The classification of thermal images is a key aspect in the industrial sector, since it is usually the starting point for the detection of faults in electrical equipment. In some cases, this task is automated through the use of traditional artificial intelligence techniques, while in others, it is performed manually, which can lead to high rates of human error. This paper presents a comparative analysis between eleven *transfer learning* architectures (AlexNet, VGG16, VGG19, ResNet, DenseNet, MobileNet v2, GoogLeNet, ResNeXt, Wide ResNet, MNASNet and ShuffleNet) through the use of fine-tuning, in order to perform a binary classification of thermal images in an electrical distribution network. For this, a database with 815 images is available, divided using the 60-20-20 hold-out technique and cross-validation with 5-Folds, to finally analyze their performance using Friedman test. After the experiments, satisfactory results were obtained with accuracies above 85 % in 10 of the previously trained architectures. However, the architecture that was not previously trained had low accuracy; with this, it is concluded that the application of *transfer learning* through the use of previously trained architectures is a proper mechanism in the classification of this type of images, and represents a reliable alternative to traditional artificial intelligence techniques.

Keywords: fine-tuning, Friedman test, pre-training, thermal images, transfer learning

Resumen

La clasificación de imágenes térmicas es un aspecto clave en el sector industrial, debido a que suele ser el punto de partida en la detección de fallos en equipos eléctricos. En algunos casos, esta tarea se automatiza mediante el uso de técnicas tradicionales de inteligencia artificial, mientras que en otros, es realizada de manera manual, lo cual puede traer consigo altas tasas de error humano. Este artículo presenta un análisis comparativo entre once arquitecturas de *transfer learning* (AlexNet, VGG16, VGG19, ResNet, DenseNet, MobileNet v2, GoogLeNet, ResNeXt, Wide ResNet, MNASNet y ShuffleNet) mediante el uso de fine-tuning, con la finalidad de realizar una clasificación binaria de imágenes térmicas en una red de distribución eléctrica. Para ello, se dispone de una base de datos con 815 imágenes, divididas mediante la técnica tipo hold-out 60-20-20 y validación cruzada con 5-folds, para finalmente analizar su rendimiento mediante el test de Friedman. Luego de los experimentos, se obtuvieron resultados satisfactorios con exactitudes superiores a 85 % en diez de las arquitecturas previamente entrenadas. Sin embargo, la arquitectura que no se entrenó previamente presentó una exactitud baja; concluyéndose que la aplicación de *transfer learning* mediante el uso de arquitecturas previamente entrenadas es un mecanismo adecuado en la clasificación de este tipo de imágenes, y representa una alternativa confiable frente a técnicas tradicionales de inteligencia artificial.

Palabras clave: imágenes térmicas, fine-tuning, preentrenamiento, test de Friedman, transfer learning

[1,*]Laboratorio de Sistemas Automáticos de Control, Universidad de Piura, Perú.
Autor para correspondencia: ✉: daniel.perez.a@alum.udep.edu.pe

Suggested citation: Pérez-Aguilar, D.; Risco-Ramos, R. and Casaverde-Pacherrez, L. (2021). «Transfer Learning for Binary Classification of Thermal Images». Ingenius. N.° 26, (july-december). pp. 71-86. doi: https://doi.org/10.17163/ings.n26.2021.07.

1. Introduction

At present, artificial intelligence (AI) is a booming discipline that has redefined many of the processes carried out in industry, showing very diverse applications, which include object recognition through artificial vision, voice recognition and synthesis, reading comprehension, translation systems, language comprehension, etc. [1]. AI is not a new term, since it has existed for many years; however, what has changed in recent times is the computational power, which can be used to compute much more complex models in shorter time [2]. AI is defined as a set of algorithms whose purpose is creating machines that emulate the capabilities of human beings. Seen in another way, it is a software that may be trained for recognizing patterns and performing predictions, in some cases more accurately than human beings [3]. Terms such as Machine Learning (ML) and Deep Learning are found within AI [4].

Machine Learning (ML) or automatic learning, is a branch of AI that seeks generalizing behaviors of a set of input data, i.e., their objective is to predict future behaviors based on finding patterns within big data sets [5]. Likewise, Deep Learning (DL) is a part of Machine Learning [6], whose objective is that systems automatically mimic the behavior and reasoning of people; in other words, that humans are involved as little as possible in the process. This objective is based on the use of artificial neural networks (ANN), which simulate the synapsis of the human brain [7]. Figure 1 displays the traditional structure of an artificial neural network, which includes the input layer, the hidden layers and the output layer.

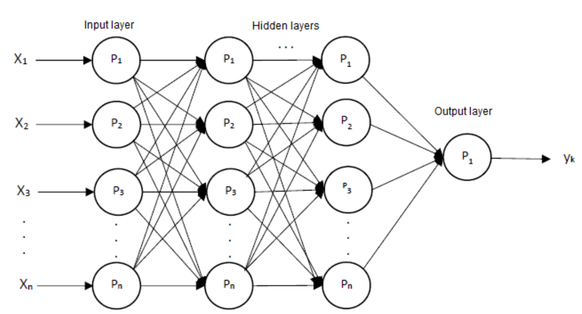

Figure 1. Structure of an artificial neural network [8]

In addition, it should be pointed out that within ANNs there are the convolutional neural networks (CNNs), which are an advanced and high potential type of the classical artificial neural network model, designed to address more complex problems, and generally used in image classification [9].

Regarding data separation, the holdout technique [10] is one of the most commonly used, and consists of dividing the data in three subsets: 60 % for training, 20 % for validation and the remaining 20 % for testing the model, as observed in Figure 2. However, this type of technique cannot be considered enough to evaluate the performance of the models, and consequently literature suggests to apply a k-folds cross-validation [11] by randomly dividing the data set in k subsets, of which k-1 are used to train the model, and one to validate it. This mechanism must be repeated k times in each iteration, using different validation subsets, as seen in Figure 3. Finally, it is recommended to perform a statistical comparison of the results of each model [12], by means of parametric techniques such as ANOVA, or non-parametric ones such as Friedman Test [13].

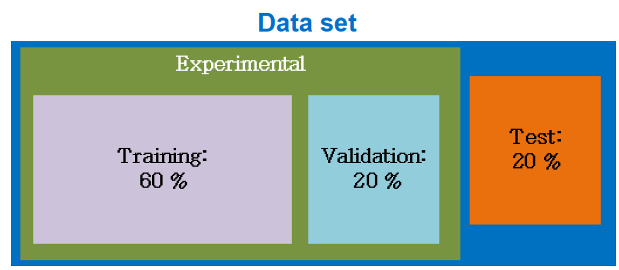

Figure 2. Holdout technique for separation

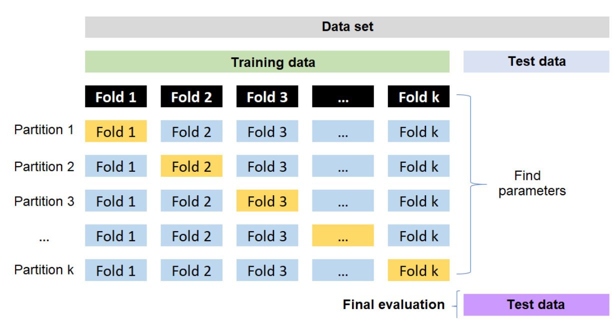

Figure 3. K-folds cross-validation

Some research works show that the application of artificial intelligence is useful in the electricity sector through the use of thermal images to automate their classification [14]. This type of images is captured by means of infrared cameras, thus involving another relevant term: thermography, a technique that enables capturing infrared radiation of the electromagnetic spectrum, whose main advantage is not requiring physical contact with the object or piece under study, with which machinery operation does not stop [15]. Figure 4 shows an example of the thermal image of a high-voltage equipment captured from the ground.

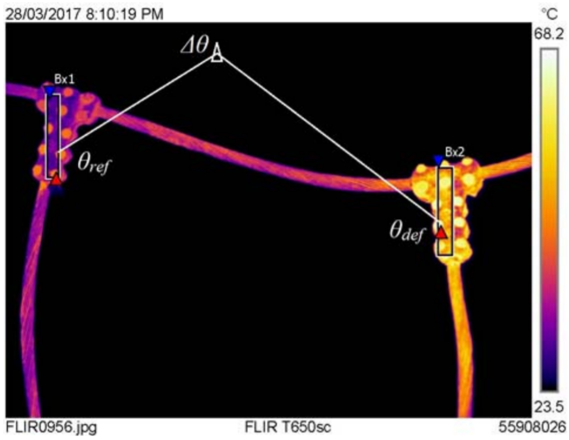

Figure 4. Thermal image of a high-voltage equipment [15]

Hereafter, the most relevant research works about the use of traditional artificial intelligence techniques for classification of thermal images are described. A clear example is the proposal of an automatic recognition system for classifying thermographic images of an electric power distribution network [16], where it was implemented a CNN and the JSEG or J segmentation algorithm, which consists of reducing the number of colors and their fusion based on the similarity of the regions of the image [17]. Similarly, a research work carried out in the Chongqing Technology Department, China [18], addresses computer vision using infrared thermal images captured without disturbing the operation of electrical substations. For this purpose, they trained a multi-layer perceptron (MLP), which is a type of artificial neural network constituted by various layers of intermediate or hidden neurons, used for solving problems that cannot be linearly separated [19].

A semiautomatic approach is proposed in [20] to evaluate the thermal condition of the electrical installations of a building through the analysis of infrared images, using a multilayer perceptron (MLP) and principal components analysis (PCA); the latter is a statistical technique whose purpose is simplifying the complexity of the sample through the selection or extraction of the most representative features of the input data [21]; whereas, an intelligent diagnosis method is described in [22], for classifying different conditions of electrical equipment using data obtained from infrared images using the K-means algorithm, which is in charge of grouping the images of electrical equipment to determine and classify clusters or groups with similar features [23].

The aforementioned research works are focused on training a model from scratch or in the traditional way for a specific scenario; however, at present there are techniques that simplify this process, such as transfer learning (TF), which is part of deep learning and consists of using a pretrained network, i.e., reusing the architecture and weights of a model trained with great amounts of input data, and apply them to different scenarios with other data sets, seeking to carry out classifications more rapidly and using lower computational load [24]. An example of the databases used to train these models is ImageNet, which contains more than fourteen millions of images [25].

One of the paradigms of transfer learning is fine-tuning the model, which seeks to adapt it to a new application domain [26]; for this purpose, it is taken the pretrained model and some parameters such as the learning rate are varied, with the objective of achieving significant improvements in the predictions [27].

The literature review reveals that there are different transfer learning applications in sectors such as (i) health, through the classification of pathologies in neurological images [28], detection of objects such as guns or knives in X-ray images [29] or cervical [30], among others. (ii) In the agroindustry, as observed in [31], which presents a comparison of the ResNet, Googlenet, VGG16, Alexnet and DenseNet transfer learning architectures, with the purpose of classifying a data set that contains images of flowers, demonstrating that the pretrained VGG16 architecture obtains accuracy levels larger than the others. (iii) In the food sector, as described in [32], where it is indicated that CNNs are the most frequently used image classification techniques; this research is focused in classifying food with the purpose of obtaining a healthier lifestyle, for which they use a database with 500 images, in addition to the pretrained architectures VGG16, VGG19, ResNet and InceptionV3, with the latter achieving the best results.

However, when studying applications of thermal images and transfer learning techniques in the electric sector, it is found a scarcity of them. One of the most representative is the case of [33], where it is proposed a mechanism for classifying thermal images of rotor bearing systems; for this purpose, they modify a convolutional neural network with the use of transfer learning, nevertheless they do not specify the TF architecture used. On the other hand, in [34] it is sought to automate the supervision of the state of industrial machinery through the use of thermal images and a CNN, indicating that a drawback of the latter is the need for great amounts of data for training; as a consequence, it is proposed the use of the VGG16 architecture as a method for reusing layers of the neural network.

Based on what has been pointed out in previous paragraphs, there is an evidence that the advantages of transfer learning are not being fully exploited in the electrical sector, since there is no study that applies different architectures with the same data set, and thus the objective of the present study is to propose an alternative to the traditional use of artificial intelligence techniques through the analysis of eleven transfer learning architectures and the self-tuning paradigm, applied to the binary classification of thermal images in an electric power distribution network.

2. Materials and methods

Figure 5 represents the methodology followed in this research, which starts with the collection of field data through the capture of thermal images; afterwards, it is carried out the design of a base architecture which includes different pretrained transfer learning architectures, each of which is trained and tested to finally compare the results obtained. All this process was conducted with the help of the Google Collaboratory online service, through the use of Jupyter Notebooks [35].

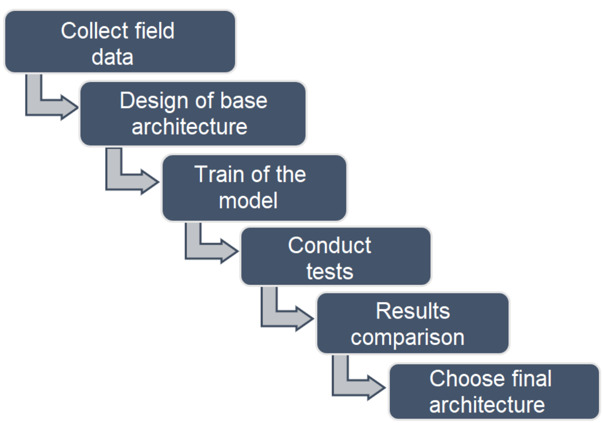

Figure 5. Methodology

2.1. Database

The database used in this work corresponds to 815 thermal images, belonging to a Peruvian electric power distribution company, classified in two categories, electric substations and transmission lines. This data set is larger than the ones used in [18], [20], [22], [36] and [37]. The images were captured using the TP8S infrared camera, whose specifications are seen in Table 1.

Table 1. Technical specifications of the TP8S camera [38]

Feature	Description
Type of detector	FPA (384 × 288 pixels, 35 m)
Spectral range	8-14 m
Thermal sensitivity	0.08 °C a 130 °C
Field of vision	22° × 16° / 35 mm
Electronic focus	Automatic or motorized
Zoom	Continuous from ×1 a ×10

The procedure that the electric distribution company has been executing includes five stages which are described in detail in the following and are shown in Figure 6. (i) First, an external company is hired for the capture of thermal images, specifically transmission lines and electric substations, which is carried out weekly or monthly. (ii) The service company delivers all images to a specialist certified in thermal image analysis. (iii) The specialist should manually classify the images and divide them between transmission lines or equipment in electric substations, since they require different types of analyses. (iv) Then the specialist proceeds to analyze each image to determine, based on knowledge and experience, if there is an evidence of failures due to the detection of a hotspot. (v) Finally, if the specialist detects a hotspot, he/she prepares the corresponding report and takes the appropriate corrective actions.

The present study focused in automating stage 3 of the process using transfer learning techniques.

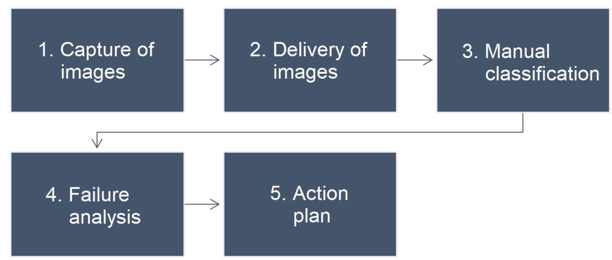

Figure 6. Process for classifying the thermal images

The images have a resolution of 384 × 288 pixels. Figure 7 shows some examples of typical images corresponding to transmission lines equipment, and Figure 8 shows some images corresponding to electrical substations, which represent the two classes of the model.

Figure 7. Typical image of transmission lines equipment

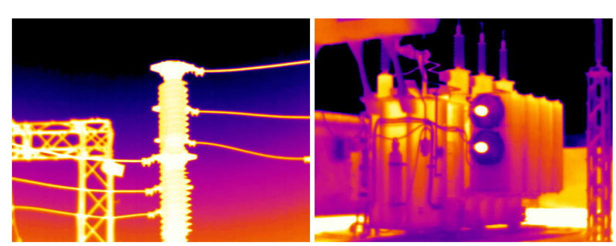

Figure 8. Typical image of electrical substations equipment

2.2. Data set

A total of 815 thermal images were used for carrying out the experiments; these images were divided in three data sets with a split of 60-20-20, known as holdout separation (See Figure 2). The 60 % of the images were considered as training data and 20 % as validation data, while the remaining 20 % includes the data for testing the model (see Table 2), i.e., that the model is trained and validated in parallel and finally tested with new images that have not been previously considered. It is indicated in [39] that the models that generalize appropriately show similar accuracy and loss metrics in training and validation, thus preventing overfitting.

Table 2. Distribution of the data sets

Class	Train	Validation	Test
Line	206	68	68
Substation	283	95	95
TOTAL	489	163	163

2.3. Architectures

Eleven architectures of previously trained models were considered in this study, through the use of the TorchVision package, which is part of PyTorch, an open-source automatic learning library; as indicated in [40], the PyTorch models are faster and easier to implement and train. The architectures used were:

2.3.1. AlexNet

Convolutional neural network constituted by eight layers [41], five of which are max-pooling, and the three remaining are fully connected. This architecture was trained with the ReLU (Rectified Linear Units) activation function and the ImageNet database. As it is observed in Figure 9, the input to the network are images of 224 × 224 pixels, which are transformed in each of the layers up to the obtaining the output, the classification of one thousand categories.

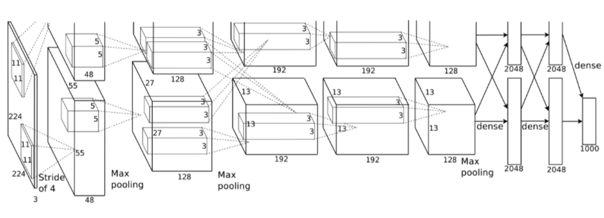

Figure 9. AlexNet architecture [41]

2.3.2. VGG16

Network constituted by 16 layers, which was also trained with the ImageNet database [42], assuming improvements compared to the AlexNet architecture since it replaces the large kernels filters by a set of 3 × 3 filters. The VGG16 architecture is shown in Figure 10.

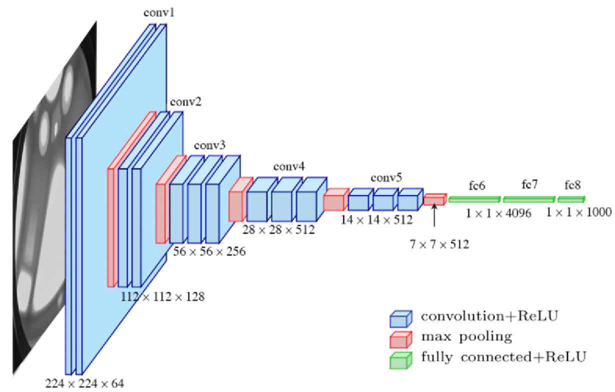

Figure 10. VGG16 architecture [42]

2.3.3. VGG19

Convolutional neural network constituted by 16 convolutional layers [43], three fully-connected, five MaxPool and one SoftMax, with an approximate of 143 million of parameters. The VGG19 architecture is shown in Figure 11.

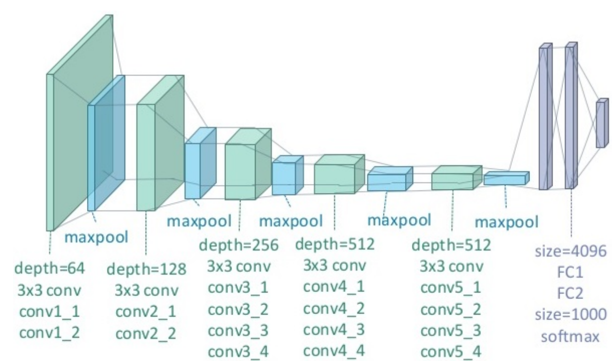

Figure 11. VGG19 architecture [43]

2.3.4. ResNet

Architecture that seeks that the increase of layers is performed different to the traditional way [44], and thus it adds a residual connection with an identity layer that is directly passed to the next layer, considerably improving the training of the model. A traditional block of the ResNet architecture is shown In Figure 12.

2.3.5. DenseNet

CNN in which every layer obtains additional inputs from all previous layers and passes its own feature maps to all further layers [45], i.e., that all outputs from previous layers are concatenated with further layers, seeking to have a smaller number of parameters and an accuracy greater than the one achieved with networks such as ResNet. The DenseNet architecture is shown in Figure 13.

2.3.6. GoogLeNet

Neural network developed by Google with the purpose of classifying images. This CNN is based on the Inception architecture [46], and thus it uses modules that give the possibility of choosing among different sizes of convolutional filter in each of the blocks. An example of the Inception module is shown in Figure 14.

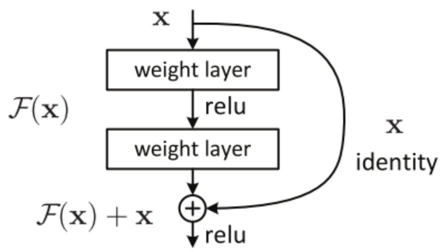

Figure 12. Block of the ResNet architecture [44]

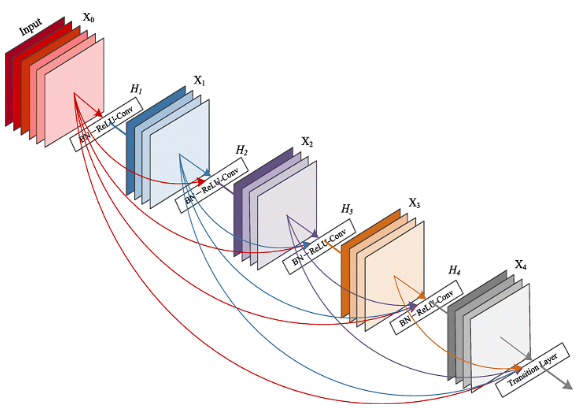

Figure 13. Block of the DenseNet architecture [45]

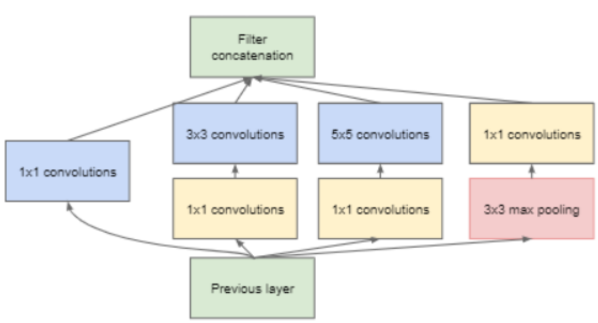

Figure 14. Inception module with reduced dimension [47]

2.3.7. MobileNet v2

is based on the use of depthwise separable convolutions and uses an inverted residual structure [48], where the input and output of the residual block are thin bottleneck layers opposed to traditional residual models that use expanded representations in the input, as it is shown with detail in Figure 15.

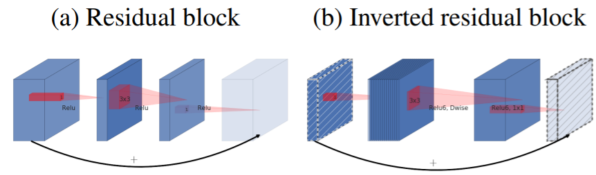

Figure 15. Blocks of the MobileNet v2 architecture [48]

2.3.8. ResNeXt

A variant of ResNet that seeks to increase the number of paths or routes parallel to the residual connection [49], i.e., that ResNeXt is a CNN with multiple branches, as seen in Figure 16, with shows a block with a cardinality of 32.

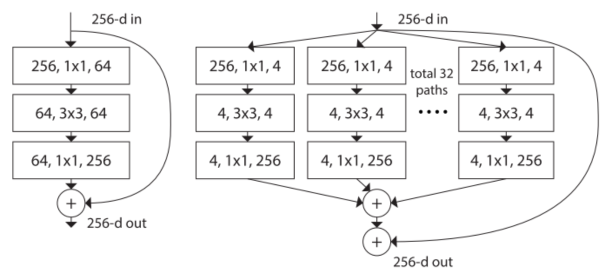

Figure 16. Block of the ResNeXt architecture [49]

2.3.9. Wide ResNet

A neural network that represents a variation to the traditional ResNet architecture [50], reducing the depth of the model and increasing the width of the residual networks. The characteristic blocks of this CNN are: basic, bottleneck, basic-wide and wide-dropout. Figure 17 shows the details of the latter.

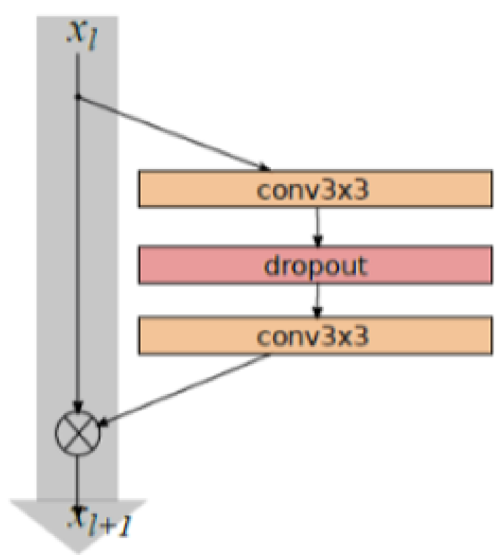

Figure 17. Wide-dropout block [50]

2.3.10. MNASNet

Is a convolutional neural network that, similar to the MobileNet [51], is designed and optimized for mobile devices and seeks that the model obtains an equilibrium between latency and accuracy. Figure 18 shows an example of the design of a convolutional layer with a 5 × 5 kernel.

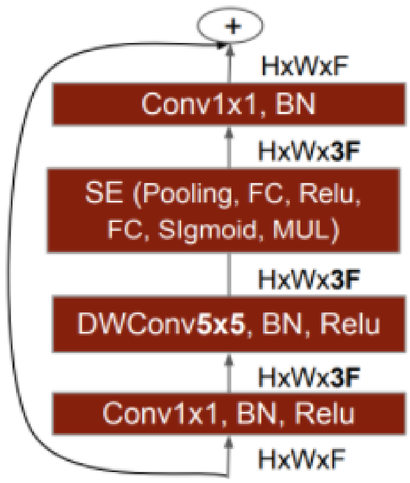

Figure 18. Convolutional layer of MNASNet [51]

2.3.11. ShuffleNet

A CNN whose main component is a new channel reorganization operation [52], seeking that the information flows more easily in them. Figure 19 shows a ShuffleNet unit, that is a central element within this architecture.

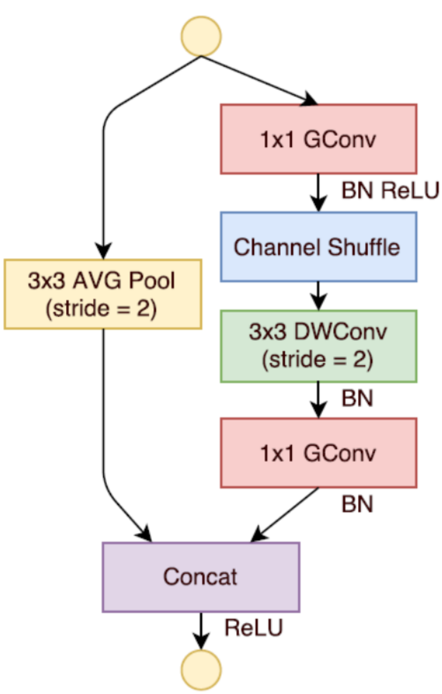

Figure 19. ShuffleNet unit [52]

At last, Table 3 shows information provided in the TorchVision [53] web site; it summarizes the features of these architectures related with the size in megabytes and the number of parameters used in training the model, highlighting that ShuffleNet is the only architecture that currently does not allow the use of the pretraining configuration parameter, i.e., its size is zero megabytes.

Table 3. Features of the architectures used

N°	Arquitecture	Size (mb)	Parameters (millions)
1	AlexNet	233	61.1
2	VGG16	528	138.36
3	VGG19	548	143.67
4	ResNet	230	60.19
5	DenseNet	77.4	20.01
6	GoogLeNet	49.7	13
7	MobileNet v2	13.6	3.5
8	ResNeXt	340	44.55
9	Wide ResNetx	243	126.89
10	MNASNet	16.9	4.38
11	ShuffleNet	0*	7.39

2.4. Model

Figure 20 shows the design of the base architecture to be used. The first section corresponds to the input layer, in which images of 328 × 288 pixels were included. Afterwards, the eleven architectures were added in the «Transfer learning model architecture» section, highlighting that the classification layer of each of them was edited to perform a binary classification, since they were originally designed to classify an approximate of one thousand images. At last, there is the output layer that corresponds to the predictions of the model.

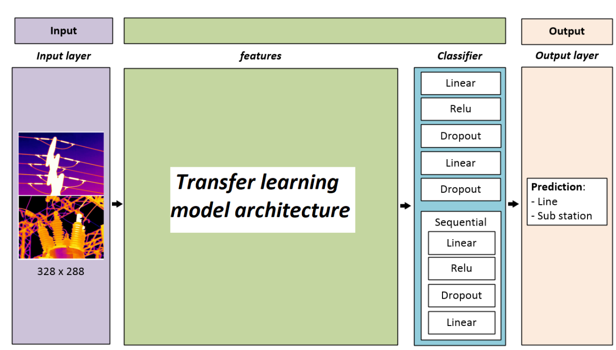

Figure 20. Base architecture

Since no research works were found that compare this number of architectures applied to thermography, and with the purpose of carrying out a fair comparison between the models, the same hyperparameters were considered in all tests conducted. They were selected based on empirical experiments, as proposed in [54–56], based on the values in Table 4. According to the results of the initial experiments, there is a breaking point at approximately epoch 20, in which the training and validation curves show a separation trend, evidencing problems of overfitting (see Figure 21). It is concluded that the models obtain better results according to the hyperparameters shown in Table 5, and these data were used to execute the remaining tests.

Table 4. Empirical experiments with hyperparameters

Hyperparameter	Value
Learning rate	10^{-2}, 10^{-3}, 10^{-4} y 10^{-5}
Images per Batch	16, 32, 64 y 128
Number of epochs	5, 10, 20, 30, 50 y 100
Cost function	Cross Entropy, Multi Margin Loss y MSE
Optimizer	Adagrad, Adam, Adamax, RMSprop y SGD

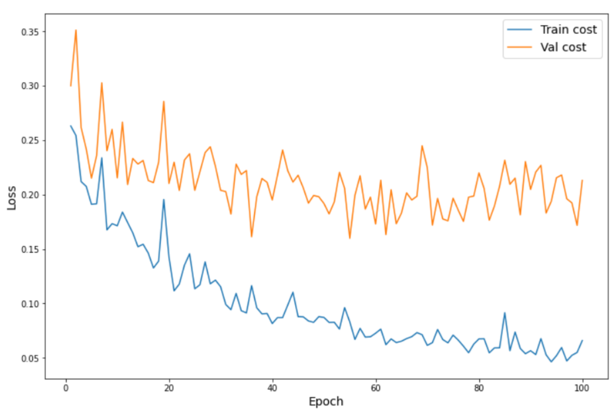

Figure 21. Experiment with 100 epochs

Table 5. Final configuration of hyperparameters

Hyperparameter	Value
Learning rate	0.0001
Images per Batch	32
Number of epochs	20
Cost function	Multi Margin Loss
Optimizer	Adam

Models improved their performance when the input data is normalized calculating the standard deviation and the mean of the data sets. In addition, the data set (815 images) is larger than in research works such as [18], [20], [22], [36] and [37], in which the maximum number of images used is 500. The literature suggests as a good practice the application of techniques that contribute to improve the quality of the training, and for this reason it was used Data Augmentation [57, 58] through the random horizontal flip, random vertical flip and random rotation transformations. Figure 22 shows some results of the transformations used.

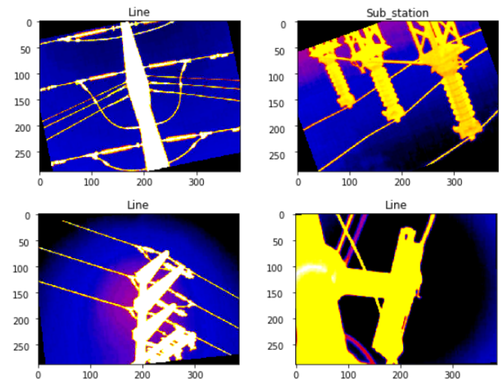

Figure 22. Images with data augmentation

3. Results and discussion

The eleven architectures were trained using the values of Table 5, obtaining the accuracy results shown in Table 6. On the other hand, Table 7 shows the loss rate for each architecture. Based on the results, it is seen that DenseNet yields a higher accuracy, while VGG16 shows the lowest loss rate. An additional point to be considered is that ShuffleNet yields the worst results since it was the only architecture without pre-training, evidencing that the pretrained architectures yield better results.

Table 6. Accuracy of the architectures

Arquitecture	Train	Validation	Test
DenseNet	96.52	92.02	98.15
VGG19	93.66	90.18	96.31
Wide ResNetx	94.68	90.18	96.31
MobileNet v2	94.68	88.95	95.70
VGG16	95.91	91.41	95.09
ResNeXt	94.06	92.02	95.09
ResNet	93.66	84.66	94.47
AlexNet	95.50	91.41	93.86
GoogLeNet	95.09	88.95	93.86
MNASNet	71.41	69.93	79.14
ShuffleNet	62.78	68.09	76.68

Table 7. Loss rate of the architectures

Arquitecture	Train	Validation	Test
VGG16	0.130571	0.195825	0.106733
VGG19	0.175259	0.257767	0.124051
MobileNet v2	0.179914	0.283927	0.124497
DenseNet	0.134446	0.198757	0.126488
AlexNet	0.127838	0.226890	0.134585
GoogLeNet	0.205683	0.242438	0.164706
ResNet	0.220600	0.295843	0.175979
ResNeXt	0.215153	0.241967	0.179627
Wide ResNetx	0.246464	0.271462	0.19046
MNASNet	0.507049	0.529690	0.395779
ShuffleNet	0.647315	0.579723	0.524782

Research works such as [59], point out that the accuracy and loss rate metrics exhibit a high degree of subjectivity, and consequently it is proposed the use of statistical techniques to evaluate the results of the architectures, specifically F1-score, whose calculation mechanism is observed in (1). Precision and recall are obtained from (2) and (3), respectively, where TP represents true positives, FP false positives and FN the false negatives. The results are shown in detail in Table 8, demonstrating that the VGG16 architecture achieves the first place with the highest F1-score, corresponding to 95.11 %.

$$F1_{score} = 2 \times \frac{precision \times recall}{precision + recall} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

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

Table 8. Comparison between architectures: F1-score

Arquitecture	Precision	Recall	F1-score
VGG16	96.12	94.12	95.11
ResNeXt	92.86	93.79	93.32
MobileNet v2	93.62	92.54	93.08
ResNet	92.36	93.48	92.92
VGG19	92.25	92.85	92.55
DenseNet	93.48	91.59	92.53
Wide ResNetx	93.13	91.8	92.46
AlexNet	93.03	89.91	91.44
GoogLeNet	92.11	89.39	90.73
MNASNet	85.71	86.11	85.91
ShuffleNet	29.14	50.00	36.82

The holdout separation mechanism may not be enough when comparing different models, and with the purpose of eliminating this source of variability other experiments were conducted by means of 5-folds cross-validation (see Figure 23) as suggested in [60], followed by a statistical comparison of the results of each model, as it is performed in [12]. The results of the cross-validation may be observed in Table 9.

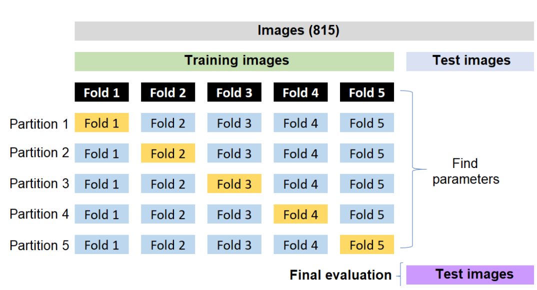

Figure 23. 5-folds cross validation

Table 9. Results – 5-folds cross validation

Arquitecture	Fold 1	...	Fold 5	Average
VGG16	93.87	...	98.16	96.81
VGG19	93.87	...	98.77	95.83
AlexNet	90.18	...	99.38	95.34
ResNeXt	94.48	...	98.16	94.6
DenseNet	88.34	...	98.16	94.36
MobileNet v2	90.18	...	96.32	93.99
ResNet	93.87	...	95.09	93.74
GoogLeNet	88.96	...	93.86	93.5
MNASNet	86.51	...	98.15	92.52
Wide ResNetx	84.66	...	94.48	89.82
ShuffleNet	68.09	...	72.39	70.06

Regarding the statistical tests to evaluate the performance of the architectures, first every model was executed 30 times, as it was done in [12]. Afterwards, residue and normal probability analyses were carried out; according to the former, it is evidenced that it might be applied a parametric test since the residues exhibit a similar dispersion (see Figure 24). However, when analyzing the normal probability plot to verify that the residues approximately fit a normal distribution, it is observed that there are data outside the confidence interval, with a Shapiro-Wilk coefficient equal to 0.932994 and a p-value equal to 0 (see Figure 25). Similarly, it was carried out a data transformation using the square root, but these are still outside the confidence interval.

Since populations do not fit a normal distribution, it cannot be applied a parametric test; for this reason, it is necessary to use a non-parametric test, specifically Friedman test, in which it is not required to meet the normality or homoestacity (equality of variances) condition. By means of this analysis shown in Table 10, it is obtained a p-value equal to zero, i.e., there is a difference between the populations, showing that VGG16 is better than the other architectures.

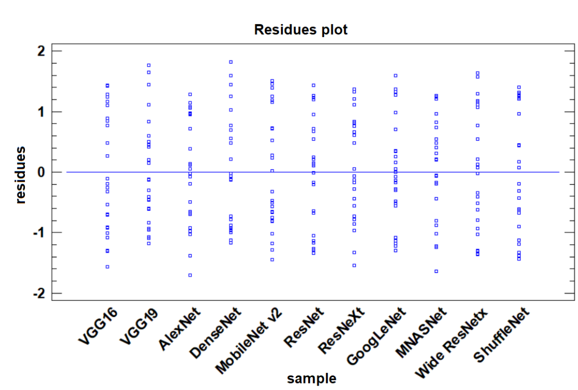

Figure 24. Residues plot

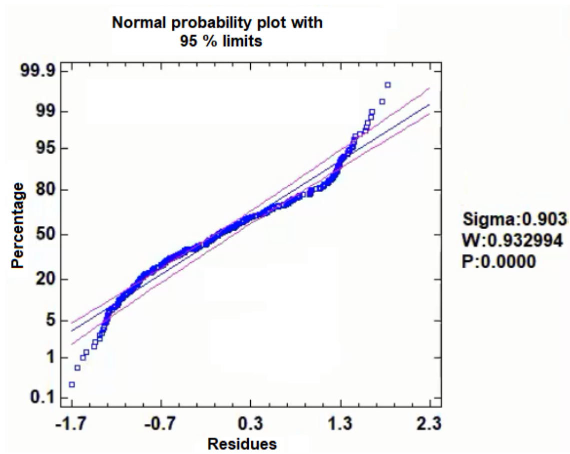

Figure 25. Normal probability plot

Table 10. Statistical results –Friedman test

Ranking	Arquitecture	Average range
1	VGG16	10.68
2	AlexNet	9.28
3	VGG19	9.28
4	DenseNet	6.48
5	MobileNet v2	6.32
6	ResNeXt	6.06
7	ResNet	5.78
8	GoogLeNet	5.16
9	MNASNet	3.96
10	Wide ResNetx	2.00
11	ShuffleNet	1.00

This same result is evidenced in the analysis of means, in which it is seen that the confidence interval that has been constructed both with the Fisher LSD (Figure 26) and Tukey HSD (Figure 27) methods, is better for the case of VGG16 since it does not intersect the corresponding to other architectures. In the case of the analysis of medians it is observed that the VGG16 architecture is better than most of the others, but it shows a small intersection with the AlexNet and VGG19 architectures (see Figure 28).

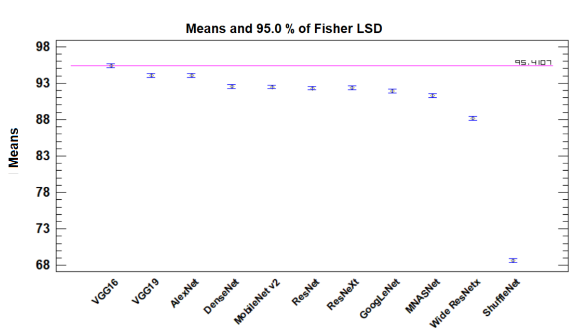

Figure 26. Plot of means – Fisher LSD

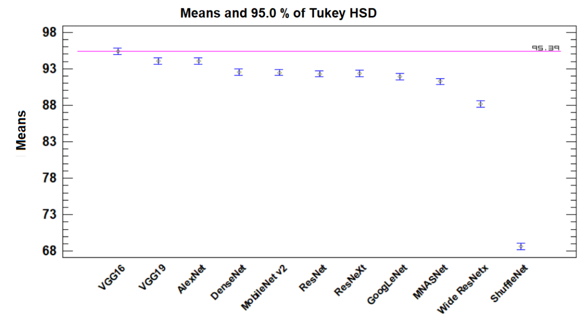

Figure 27. Plot of means – Tukey HSD

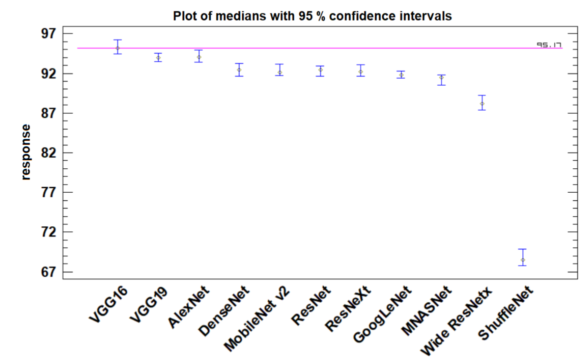

Figure 28. Plot of medians

In addition, it was performed the multiple comparison analysis by means of the multiple range test (see Table 11), where if the "X" of homogeneous groups are in the same column the architectures behave similarly, observing that VGG16 is better and different than the other architectures.

Table 11. Multiple range test

Arquitecture	Mean	Homogeneous groups
ShuffleNet	686.560	X
Wide ResNetx	881.820	X
MNASNet	912.540	X
GoogLeNet	919.068	XX
ResNet	923.080	X
ResNeXt	923.704	X
MobileNet v2	924.816	X
DenseNet	925.336	X
AlexNet	940.500	X
VGG19	940.652	X
VGG16	953.756	X

The present paper is not intended to obtain a «better» model but presenting an alternative mechanism in front of traditional artificial intelligence techniques. However, the results corresponding to the model with better performance, i.e., VGG16, are presented for academic purposes; the original architecture of this model is seen in Figure 10, and the resulting final architecture is shown with details in Figure 29.

It is seen in this architecture that the input is represented by images of 328 × 288 pixels, which enter the VGG16 pretrained convolutional neural network, constituted by thirteen convolutional layers followed by three fully connected layers, the first two of which have 4096 channels and the last one 1000 channels; therefore, this was edited to be able to perform a binary classification (2 channels). The hidden layers use the ReLU activation function and besides different 3 × 3 kernels. Finally, the output corresponds to the classification between both classes, i.e., lines and substations.

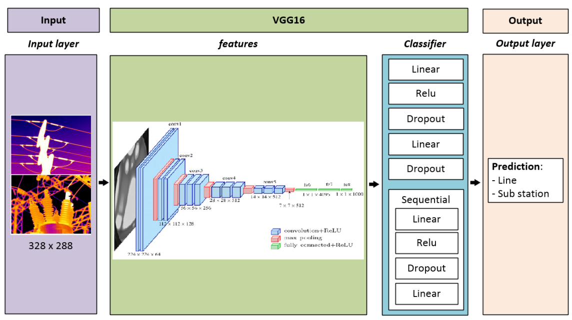

Figure 29. Proposed transfer learning architecture for classifying thermal images

Precisions of 95.91 % and 91.41 % in training and validation, respectively, were obtained with this model (Figure 30). This architecture was tested with new images belonging to the test data set, obtaining an accuracy of 94.43 % for the category lines, and 92.81 % for substations. This may be seen in the confusion matrix shown in Figure 31.

The accuracy of the model was graphically represented through the receiver operating characteristic (ROC) curve, whose area under the curve (AUC) shows a value of 94 %, which indicates a high performance of the proposed architecture in the classification of thermal images (see Figure 32).

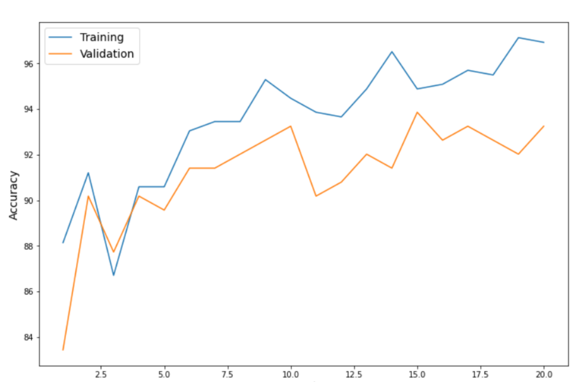

Figure 30. Accuracy: training and validation of the model

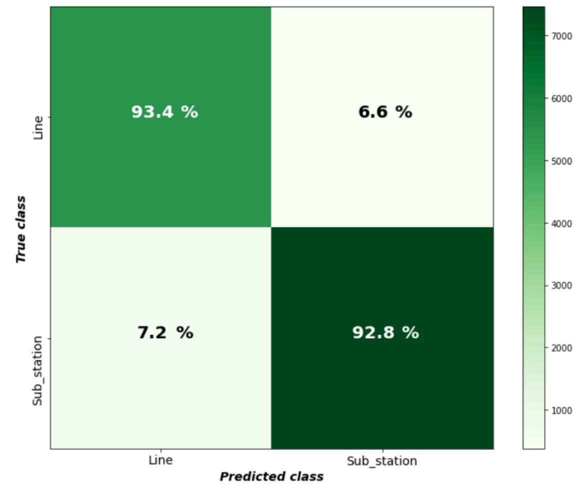

Figure 31. Confusion matrix

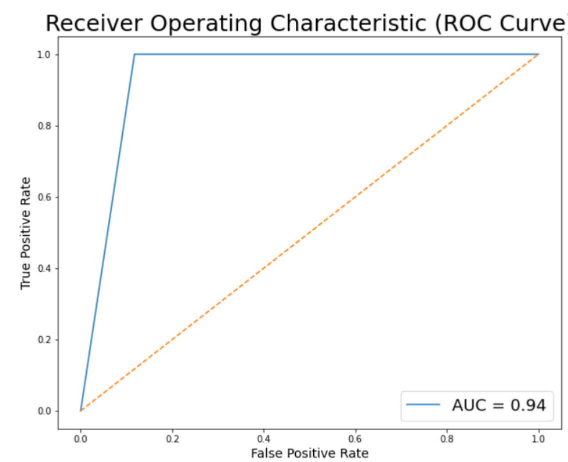

Figure 32. ROC curve

At last, Figure 33 shows some examples of the predictions by the model. The real classification is represented in the left side of the title of each image, and in the right side the one obtained by the model.

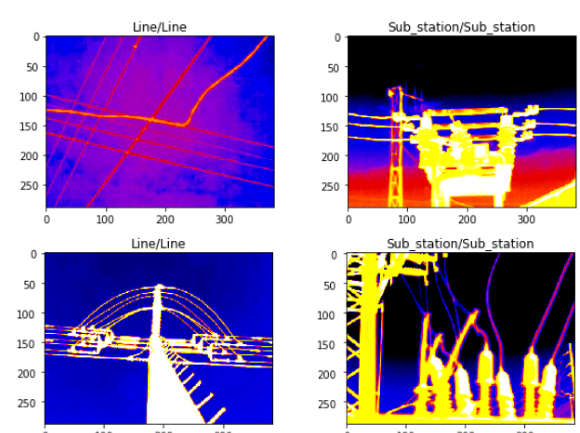

Figure 33. Predictions by the model

4. Conclusions

In this work, it is analyzed the performance of eleven pretrained networks that use the transfer learning paradigm based in fine-tuning the model, for binary classification of thermal images. The final objective is not finding a «better» model, but presenting alternatives to traditional artificial intelligence techniques, seeking to save computational time and load.

The models yield accuracies between 79.14 % and 98.15 %, and values of F1-score between 85.91 % and 95.11 % in the pretrained architectures; these results are an indication that the use of transfer learning techniques represents a reliable alternative as a mechanism for classifying thermal images in the electric sector; however, it is recommended to perform a specific analysis in each particular case of application.

The use of data augmentation, transformations and normalization of the images, are important aspects to improve the performance of the model; whereas the division of the data set in the training, validation and test subsets using the holdout technique helped to prevent overfitting, generalize the model and, therefore, carry out more accurate predictions. However, with the purpose of performing a fairer comparison, the study also included applying a 5-folds cross-validation, and moreover a statistical analysis by means of Friedman Test.

Acknowledgement

Daniel Pérez-Aguilar acknowledges the financial support of the CONCYTEC-World Bank Project, through its executive unit, the Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (FONDECYT), within the call E033-2018-01-BM under Contract Nº 06-2018-FONDECYT/BM, for the research work Transfer learning in binary classification of thermal images, executed as part of the doctoral program in Engineering with focus on Process Automation, Control and Optimization, carried out in the Automatic Control Systems Laboratory of the Universidad de Piura, Peru.

References

[1] M. Haenlein and A. Kaplan, “A brief history of artificial intelligence: On the past, present, and future of artificial intelligence,” *California Management Review*, vol. 61, no. 4, pp. 5–14, 2019. [Online]. Available: https://doi.org/10.1177/0008125619864925

[2] M. Flasiński, *Introduction to artificial intelligence*. Springer International Publishing, 2016. [Online]. Available: http://doi.org/10.1007/978-3-319-40022-8

[3] M.-H. Huang and R. T. Rust, “Artificial intelligence in service,” *Journal of Service Research*, vol. 21, no. 2, pp. 155–172, 2018. [Online]. Available: https://doi.org/10.1177/1094670517752459

[4] Z. Aung, I. S. Mikhaylov, and Y. T. Aung, “Artificial intelligence methods application in oil industry,” in *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, 2020, pp. 563–567. [Online]. Available: https://doi.org/10.1109/EIConRus49466.2020.9039330

[5] T. P. Carvalho, F. A. A. M. N. Soares, R. Vita, R. da P. Francisco, J. ao P. Basto, and S. G. S. Alcalá, “A systematic literature review of machine learning methods applied to predictive maintenance,” *Computers & Industrial Engineering*, vol. 137, p. 106024, 2019. [Online]. Available: https://doi.org/10.1016/j.cie.2019.106024

[6] S. Wan, L. Qi, X. Xu, C. Tong, and Z. Gu, “Deep learning models for real-time human activity recognition with smartphones,” *Mobile Networks and Applications*, vol. 25, no. 2, pp. 743–755, Apr. 2020. [Online]. Available: https://doi.org/10.1007/s11036-019-01445-x

[7] V. Golodov, A. Zavei-Boroda, S. Ivanov, and K. Nikolskaya, “Development of a deep learning neural network for human movements analysis,” in *2017 Second Russia and Pacific Conference on Computer Technology and Applications (RPC)*, 2017, pp. 72–74. [Online]. Available: https://doi.org/10.1109/RPC.2017.8168071

[8] E. A. Galindo, J. A. Perdomo, and J. C. Figueroa-García, “Estudio comparativo entre máquinas de soporte vectorial multiclase, redes neuronales artificiales y sistema de inferencia neuro-difuso autoorganizado para problemas de clasificación,” *Información tecnológica*, vol. 31, pp. 273–286, 02 2020. [Online]. Available: http://dx.doi.org/10.4067/S0718-07642020000100273

[9] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey,” *Neurocomputing*, vol. 300, pp. 17–33, 2018. [Online]. Available: https://doi.org/10.1016/j.neucom.2018.01.092

[10] I. Yildiz, P. Tian, J. Dy, D. Erdoğmuş, J. Brown, J. Kalpathy-Cramer, S. Ostmo, J. Peter Campbell, M. F. Chiang, and S. Ioannidis, “Classification and comparison via neural networks,” *Neural Networks*, vol. 118, pp. 65–80, 2019. [Online]. Available: https://doi.org/10.1016/j.neunet.2019.06.004

[11] Y. Jung, "Multiple predicting k-fold cross-validation for model selection," *Journal of Nonparametric Statistics*, vol. 30, no. 1, pp. 197–215, 2018. [Online]. Available: https://doi.org/10.1080/10485252.2017.1404598

[12] F. Pacheco, J. Valente de Oliveira, R.-V. Sénchez, M. Cerrada, D. Cabrera, C. Li, G. Zurita, and M. Artés, "A statistical comparison of neuroclassifiers and feature selection methods for gearbox fault diagnosis under realistic conditions," *Neurocomputing*, vol. 194, pp. 192–206, 2016. [Online]. Available: https://doi.org/10.1016/j.neucom.2016.02.028

[13] D. W. Zimmerman and B. D. Zumbo, "Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks," *The Journal of Experimental Education*, vol. 62, no. 1, pp. 75–86, 1993. [Online]. Available: https://doi.org/10.1080/00220973.1993.9943832

[14] C. Lile and L. Yiqun, "Anomaly detection in thermal images using deep neural networks," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 2299–2303. [Online]. Available: https://doi.org/10.1109/ICIP.2017.8296692

[15] A. Dragomir, M. Adam, M. Andruççâ, A. Munteanu, and E. Boghiu, "Considerations regarding infrared thermal stresses monitoring of electrical equipment," in *2017 International Conference on Electromechanical and Power Systems (SIELMEN)*, 2017, pp. 100–103. [Online]. Available: https://doi.org/10.1109/SIELMEN.2017.8123307

[16] F. Fambrini, Y. Iano, D. G. Caetano, A. A. D. Rodríguez, C. Moya, E. Carrara, R. Arthur, F. C. Cabello, J. V. Zubem, L. M. Del Val Cura, J. a. B. Destro Filho, J. R. Campos, and J. H. Saito, "Gpu cuda jseg segmentation algorithm associated with deep learning classifier for electrical network images identification," *Procedia Computer Science*, vol. 126, pp. 557–565, 2018, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia. [Online]. Available: https://doi.org/10.1016/j.procs.2018.07.290

[17] X. W. X. L. Z. J. Wenzhen Yang, Jiali Luo and Z. Pan, "Image tactile perception with an improved jseg algorithm," *International Journal of Performability Engineering*, vol. 14, no. 1, p. 77, 2018. [Online]. Available: https://doi.org/10.23940/ijpe.18.01.p9.7788

[18] I. Ullah, F. Yang, R. Khan, L. Liu, H. Yang, B. Gao, and K. Sun, "Predictive maintenance of power substation equipment by infrared thermography using a machine-learning approach," *Energies*, vol. 10, no. 12, 2017. [Online]. Available: https://doi.org/10.3390/en10121987

[19] H. Ramchoun, Y. Ghanou, M. Ettaouil, and M. A. J. Idrissi, "Multilayer perceptron: Architecture optimization and training," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, 2016. [Online]. Available: http://doi.org/10.9781/ijimai.2016.415

[20] A. S. Nazmul Huda, S. Taib, M. S. Jadin, and D. Ishak, "A semi-automatic approach for thermographic inspection of electrical installations within buildings," *Energy and Buildings*, vol. 55, pp. 585–591, 2012. [Online]. Available: https://doi.org/10.1016/j.enbuild.2012.09.014

[21] C. Yuan, X. Sun, and R. Lv, "Fingerprint liveness detection based on multi-scale 0PQ and PCA," *China Communications*, vol. 13, no. 7, pp. 60–65, 2016. [Online]. Available: https://doi.org/10.1109/CC.2016.7559076

[22] H. Zou and F. Huang, "A novel intelligent fault diagnosis method for electrical equipment using infrared thermography," *Infrared Physics & Technology*, vol. 73, pp. 29–35, 2015. [Online]. Available: https://doi.org/10.1016/j.infrared.2015.08.019

[23] S.-S. Yu, S.-W. Chu, C.-M. Wang, Y.-K. Chan, and T.-C. Chang, "Two improved k-means algorithms," *Applied Soft Computing*, vol. 68, pp. 747–755, 2018. [Online]. Available: https://doi.org/10.1016/j.asoc.2017.08.032

[24] T. V. Phan, S. Sultana, T. G. Nguyen, and T. Bauschert, "Q - transfer: A novel framework for efficient deep transfer learning in networking," in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, 2020, pp. 146–151. [Online]. Available: https://doi.org/10.1109/ICAIIC48513.2020.9065240

[25] M. Ebrahim, M. Al-Ayyoub, and M. A. Alsmirat, "Will transfer learning enhance imagenet classification accuracy using imagenet-pretrained models?" in *2019 10th International Conference on Information and Communication Systems (ICICS)*, 2019, pp. 211–216. [Online]. Available: https://doi.org/10.1109/IACS.2019.8809114

[26] T. Alshalali and D. Josyula, "Fine-tuning of pre-trained deep learning models with extreme learning machine," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2018, pp. 469–473. [Online]. Available: https://doi.org/10.1109/CSCI46756.2018.00096

[27] G. Vrbançiç and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196 197–196 211, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.3034343

[28] T. Kaur and T. K. Gandhi, "Automated brain image classification based on vgg-16 and transfer learning," in *2019 International Conference on Information Technology (ICIT)*, 2019, pp. 94–98. [Online]. Available: https://doi.org/10.1109/ICIT48102.2019.00023

[29] R. L. Gálvez, E. P. Dadios, A. A. Bandala, and R. R. P. Vicerra, "Threat object classification in X-ray images using transfer learning," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*, 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/HNICEM.2018.8666344

[30] D. Xue, X. Zhou, C. Li, Y. Yao, M. M. Rahaman, J. Zhang, H. Chen, J. Zhang, S. Qi, and H. Sun, "An application of transfer learning and ensemble learning techniques for cervical histopathology image classification," *IEEE Access*, vol. 8, pp. 104 603–104 618, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2999816

[31] E. Cengil and A. Çinar, "Multiple classification of flower images using transfer learning," in *2019 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2019, pp. 1–6. [Online]. Available: https://doi.org/10.1109/IDAP.2019.8875953

[32] J. R. Rajayogi, G. Manjunath, and G. Shobha, "Indian food image classification with transfer learning," in *2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS)*, vol. 4, 2019, pp. 1–4. [Online]. Available: https://doi.org/10.1109/CSITSS47250.2019.9031051

[33] H. Shao, M. Xia, G. Han, Y. Zhang, and J. Wan, "Intelligent fault diagnosis of rotor-bearing system under varying working conditions with modified transfer convolutional neural network and thermal images," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3488–3496, 2021. [Online]. Available: https://doi.org/10.1109/TII.2020.3005965

[34] O. Janssens, R. Van de Walle, M. Loccufier, and S. Van Hoecke, "Deep learning for infrared thermal image based machine health monitoring," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 151–159, 2018. [Online]. Available: https://doi.org/10.1109/TMECH.2017.2722479

[35] T. Carneiro, R. V. Medeiros Da NóBrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance analysis of google colaboratory as a tool for accelerating deep learning applications," *IEEE Access*, vol. 6, pp. 61 677–61 685, 2018. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2874767

[36] A. S. N. Huda and S. Taib, "Suitable features selection for monitoring thermal condition of electrical equipment using infrared thermography," *Infrared Physics & Technology*, vol. 61, pp. 184–191, 2013. [Online]. Available: https://doi.org/10.1016/j.infrared.2013.04.012

[37] M. S. Jadin, S. Taib, and K. H. Ghazali, "Feature extraction and classification for detecting the thermal faults in electrical installations," *Measurement*, vol. 57, pp. 15–24, 2014. [Online]. Available: https://doi.org/10.1016/j.measurement.2014.07.010

[38] W. I. Technology, *ThermoProTP8S™ IR Thermal Camera. User Manual.* Wuhan Guide Infrared Technology Co., Ltd., 2007. [Online]. Available: https://bit.ly/3bVGd0u

[39] L. Sandjakoska and F. Stojanovska, "How initialization is related to deep neural networks generalization capability: Experimental study," in *2020 55th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, 2020, pp. 163–166. [Online]. Available: https://doi.org/10.1109/ICEST49890.2020.9232882

[40] C. Heghedus, A. Chakravorty, and C. Rong, "Neural network frameworks. comparison on public transportation prediction," in *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, 2019, pp. 842–849. [Online]. Available: https://doi.org/10.1109/IPDPSW.2019.00138

[41] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing alexnet deep transfer learning for ear recognition," in *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, 2018, pp. 1–5. [Online]. Available: https://doi.org/10.1109/INFRKM.2018.8464769

[42] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 730–734. [Online]. Available: https://doi.org/10.1109/ACPR.2015.7486599

[43] J. Xiao, J. Wang, S. Cao, and B. Li, "Application of a novel and improved VGG-19 network in the detection of workers wearing masks," *Journal of Physics: Conference Series*, vol. 1518, p. 012041, apr 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1518/1/012041

[44] A. Budhiman, S. Suyanto, and A. Arifianto, "Melanoma cancer classification using resnet with data augmentation," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2019, pp. 17–20. [Online]. Available: https://doi.org/10.1109/ISRITI48646.2019.9034624

[45] K. Zhang, Y. Guo, X. Wang, J. Yuan, and Q. Ding, "Multiple feature reweight densenet for image classification," *IEEE Access*, vol. 7, pp. 9872–9880, 2019. [Online]. Available: https://doi.org/10.1109/ACCESS.2018.2890127

[46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9. [Online]. Available: https://doi.org/10.1109/CVPR.2015.7298594

[47] GeeksforGeeks. (2020) Understanding googlenet model - CNN architecture. [Online]. Available: https://bit.ly/2RLmiuc

[48] K. Fu, L. Sun, X. Kang, and F. Ren, "Text detection for natural scene based on mobilenet V2 and U-net," in *2019 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2019, pp. 1560–1564. [Online]. Available: https://doi.org/10.1109/ICMA.2019.8816384

[49] C. Qiu, M. Schmitt, H. Taubenböck, and X. X. Zhu, "Mapping human settlements with multi-seasonal sentinel-2 imagery and attention-based resnext," in *2019 Joint Urban Remote Sensing Event (JURSE)*, 2019, pp. 1–4. [Online]. Available: https://doi.org/10.1109/JURSE.2019.8809009

[50] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proceedings of the British Machine Vision Conference (BMVC)*, E. R. H. Richard C. Wilson and W. A. P. Smith, Eds. BMVA Press, September 2016, pp. 87.1–87.12. [Online]. Available: https://dx.doi.org/10.5244/C.30.87

[51] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "Mnasnet: Platform-aware neural architecture search for mobile," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 2815–2823. [Online]. Available: https://doi.org/10.1109/CVPR.2019.00293

[52] Y. Li and C. Lv, "Ss-yolo: An object detection algorithm based on YOLOv3 and shufflenet," in *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol. 1, 2020, pp. 769–772. [Online]. Available: https://doi.org/10.1109/ITNEC48623.2020.9085091

[53] PyTorch. (2019) TORCHVISION.MODELS. [Online]. Available: https://bit.ly/2QSClGe

[54] X. Song, Y. Du, and J. Jackson, "An empirical study on hyperparameters and their interdependence for RL generalization," *arXiv preprint arXiv*, vol. abs/1906.00431, 2019. [Online]. Available: https://bit.ly/3ulY3zZ

[55] J. N. van Rijn and F. Hutter, "Hyperparameter importance across datasets," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 2367–2376. [Online]. Available: https://doi.org/10.1145/3219819.3220058

[56] A. Aravkin, J. V. Burke, A. Chiuso, and G. Pillonetto, "On the estimation of hyperparameters for empirical bayes estimators: Maximum marginal likelihood vs minimum MSE," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 125–130, 2012, 16th IFAC Symposium on System Identification. [Online]. Available: https://doi.org/10.3182/20120711-3-BE-2027.00353

[57] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop (IIPhDW)*, 2018, pp. 117–122. [Online]. Available: https://doi.org/10.1109/IIPHDW.2018.8388338

[58] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 2019. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0

[59] D. Avola, L. Cinque, G. L. Foresti, F. Lamacchia, M. R. Marini, L. Perini, K. Qorraj, and G. Telesca, "A shape comparison reinforcement method based on feature extractors and f1-score," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2019, pp. 2155–2159. [Online]. Available: https://doi.org/10.1109/SMC.2019.8914601

[60] J. Amat Rodrigo. (2020) Validación de modelos predictivos: Cross-validation, oneleaveout, bootstraping. [Online]. Available: https://bit.ly/3bYgPHk