Power Distribution Insulators Classification Using Image Hybrid Deep Learning

Eduardo F. Simas Filho*, Ricardo M. Prates*‡, Rodrigo P. Ramos‡, Jaime S. Cardoso†
*Digital Systems Laboratory, PPGEE, Federal University of Bahia, Salvador, Brazil
Emails: ricardo.prates@univasf.edu.br, eduardo.simas@ufba.br
†INESC TEC, Portugal
Faculty of Engineering, University of Porto, Portugal
Email: jaime.s.cardoso@inesctec.pt
‡Federal University of Vale do São Francisco, Petrolina, Brazil
Email: rodrigo.ramos@univasf.edu.br

***Abstract*—The Overhead Power Distribution Lines present a wide range of insulator components, which have different shapes and types of building materials. These components are usually exposed to weather and operational conditions that may cause deviations in their shapes, colors or textures. These changes might hinder the development of automatic systems for visual inspection. In this perspective, this work presents a robust methodology for image classification, which aims at the efficient distribution insulator class identification, regardless of its degradation level. This work can be characterized by the following steps: implementation of Convolutional Neural Network (CNN); transfer learning; attribute vector acquisition and design of hybrid classifier architectures to improve the discrimination efficiency. In summary, a previously trained CNN goes through a fine tuning stage for later use as a feature extractor for training a new set of classifiers. A comparative study was conducted to identify which classifier architecture obtained the best discrimination performance for non-conforming components. The proposed methodology showed a significant improvement in classification performance, obtaining 95% overall accuracy in the identification of non-conforming component classes.**

***Index Terms*—distribution insulators, convolutional neural network (CNN), transfer learning, hybrid classifiers**

I. Introduction

The conformity in electrical power supply is closely related to the quality of the preventive and corrective maintenance services, provided by power utilities companies, in the Medium Voltage Power Lines (MVPL). As a prerequisite for performing these activities, the first step to be carried out is inspection. It consists in the search and identification of non-conformities present in the MVPL components, such as defective, polluted and badly positioned elements. From the determination of the type of component and its respective defect, it becomes possible to act preventively on the electrical system [1]. These activities provide improvement of the supplied quality (e.g. electric voltage in permanent regime) and reduce the probability of power outages.

Nowadays, visual inspection activities in the Overhead Power Distribution Lines (OPDLs) are performed by professionals trained by the utility company, in the modality of visual inspection by "walking patrol" or assisted by helicopter [2], [3]. These activities are based on empirical methods, where these professionals point out components classes and presence of defects in a visual and intuitive way. Most of the time, they are performed with live circuits, which can bring risks to these professionals safety [4], [5]. These activities are also susceptible to subjective and heuristic interpretations, which can lead to inaccurate or incorrect identification/diagnosis on the inspection process, resulting in registration failures and inadequate maintenance planning.

Therefore, it can be seen that there is still a latent need for improvement in techniques for both component identification and defect analysis in OPDLs. To this end, one approach looks promising: development and implementation of automated/intelligent data acquisition and information analysis techniques. The computer vision techniques can be explored to reduce subjectivities inherent to current OPDL inspection methods. As far as the authors' knowledge, the use of intelligent techniques for power systems inspection is a current but a not yet very explored area. A recent literature review on this topic can be found in [6].

In this work, the development of a methodology was initiated intended to distribution insulators classification, through the use of deep learning techniques and hybrid classifiers. To this end, a sequence of activities was performed, namely: insulators image acquisition; convolutional neural network (CNN) [7], [8] fine-tuning from ImageNet database [9]; extraction of 2,048 parameters belonging to ResNet-50 flat layer; and second training with a new set of algorithms - Suport Vector Machine (SVM) [10], k-Nearest Neighbors (kNN), Random Forest [11] and Bayes Net. In the experimental stage, the models were trained using datasets with intact components and their respective performances were evaluated with a database composed only of defective ones. At the end, a set of performance metrics were evaluated for the proposed techniques.

This research can be understood as a necessary step to support the development of intelligent visual inspection systems for OPDLs, embedded in drones. At this early stage, it becomes necessary to develop robust models that can discriminate insulator classes, regardless of the non-conformity type and degradation level. In a future step, the verification of the

978-9-0827-9703-9/19/$31.00

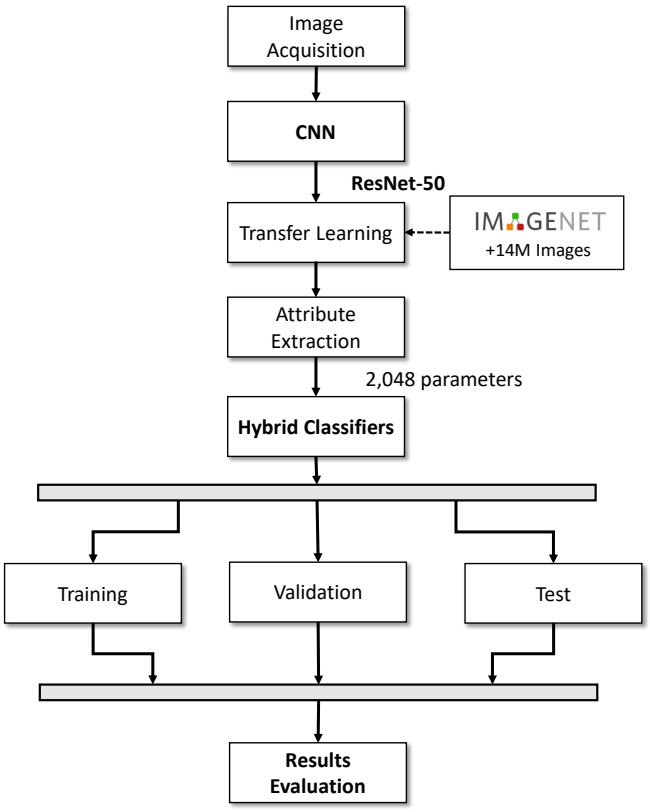

Fig. 1. Diagram of the proposed methodology for insulator classification.

robustness of these techniques in real-world environments will be performed.

II. Methodology

The main steps of a pattern recognition system are data acquisition, pre-processing, attribute extraction and pattern classification [12], [13]. The different experimental approaches used in the stages of the classification process will be described. Figure 1 presents the flowchart of the procedures proposed in this work.

A. Image acquisition

At the data collection stage, samples were obtained from four types of insulators commonly used in 13.8 kV OPDLs (see Figure 2). These components have different shapes and building materials, listed as follows:

- Ceramic Insulators - Pin (CPI) and bi-color (CBI),
- Polymeric Insulator - Grey type (PGI),
- Glass Insulator - Green type (GGI).

For defective components, insulators samples were collected containing different types of non-conformities, such as breakage, fracture, deformation and color deviation caused by electrical discharges. Figure 3 illustrates some examples. In a second moment, a set of photos of the mentioned components was collected in several angles and positions, totaling 1600 photos of intact components (400 of each class) and 800 photos of defective ones (200 per class). This step was performed in a research lab, with controlled lighting and background color. We called DATA-INT the dataset formed by intact components and DATA-DEF the dataset of defective ones (see Figures 2 and 3). Finally, for computational complexity reduction, the images were resized to 224×224 pixels.

(a) PGI (b) CPI (c) GGI (d) CBI

Fig. 2. Images samples of intact components (DATA-INT).

(a) Deformed (b) Fractured (c) Stained (d) Broken

Fig. 3. Image samples of defective components (DATA-DEF).

B. Convolutional neural network architecture

The Residual Neural Network (ResNet) was introduced by Kaiming [7]. These CNNs have Residual Units which facilitate the training of very deep convolutional networks. A variation of this model, with 152 layers, was the winner of the 2015 ImageNet Large Scale Visual Recognition Competition.

In this architecture, indicating the desired underlying mapping as $H(x)$, instead of learning in a direct way, with few stacked layers, these layers are permitted to fit into a residual mapping. In other words, stacked nonlinear layers are allowed to fit into another mapping of $F(x) = H(x) - x$. Then, a new type of mapping is created, which can be expressed by $H(x) = F(x) + x$, where $F(x)$ and x represent the stacked non-linear layers and the identity function, respectively. A current review about this topic is present in reference [8].

ResNet-50 architecture was chosen as the base model for the simulations. Figure 4 illustrates, in a simplified way, the network architecture along with its Residual Units. Also, the last layer of this CNN was replaced to represent the classes belonging to the studied datasets. The activation function chosen for the referred layer was the softmax type [14].

C. Transfer Learning

Transfer learning techniques are commonly used to reduce the computational cost and increase classification efficiency [15], [16]. The chosen CNN was adjusted with learned weights from the ImageNet competition dataset [9]. During the fine-tuning of the network, the last dense layer was retrained and the rest of the network weights were frozen. The Adam optimization method [17] was used with a low learning rate (10^{-4}) for adjustments to the datasets used.

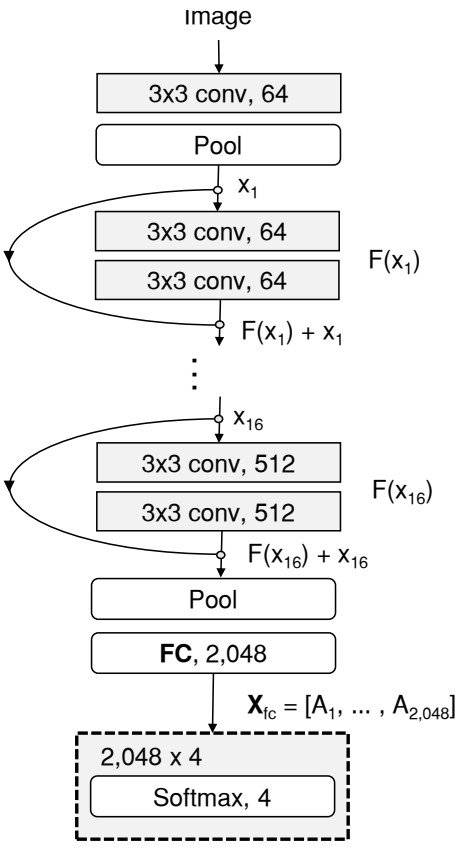

Fig. 4. Partial representation of the CNN ResNet-50 architecture.

D. Attribute Extraction

For this stage, the ResNet-50's $\mathbf{X}_{fc}$ vector, illustrated in Figure 4, will serve as an attribute generator for a new training database. Each image inserted in the CNN will compose a new set of 2048 parameters, used as input to the hybrid classifiers.

E. Hybrid classifiers

The use of hybrid classifiers was focused on the potential performance gain provided by the combination of two or more machine learning techniques [18]–[20]. In this research, deep learning techniques were combined with traditional classifiers. The pre-training CNN served as a first learning stage, providing attributes for training SVM, kNN, Bayes Net and Random Forest algorithms. These techniques are briefly described as follows.

a) Support Vector Machine (SVM): Support vector machine is a machine learning technique introduced in 1995 [10] that has been broadly and successfully applied in several research fields [21], [22]. It aims at defining an optimal decision boundary that best separates the classes under study. This is achieved by determining the best separation hyperplane through the minimization of the error probability of a given training set. The SVM transforms the feature space into a higher dimensional space through the use of kernel functions.

To be more specific, it is considered a training set defined as $\mathbf{z}_i$ which represents an n_f-dimensional feature vector, with n_s samples, and defined $y_i \in \{-1, +1\}$ as a target value indicating an i-th class. The solution is obtained finding the Lagrange multipliers $\{\alpha_i\}_{i=1}^{n_s}$ that minimize the objective function

$$Q(\alpha) = \frac{1}{2} - \sum_{i=1}^{n_s} \alpha_i + \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \alpha_i \alpha_j y_i y_j K(\mathbf{z}_i, \mathbf{z}_j), \quad (1)$$

subject to

$$\sum_{i=1}^{n_s} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C \quad \text{for } i = 1, 2, \ldots, n_s, \quad (2)$$

where C is a positive constant given by the user, and $K(\cdot, \cdot)$ is a partially defined positive kernel function. If $\{\alpha_i\}_{i=1}^{n_s}$ is an optimal solution of Equations (1) and (2), the decision function is obtained as

$$\beta_{SVM}(\mathbf{z}) = \sum_{i=1}^{n_s} \alpha_i y_i K(\mathbf{z}_i, \mathbf{z}) + b, \quad (3)$$

where b is a bias parameter. A new data point is classified as one of the classes according to the value of $\beta_{SVM}(\mathbf{z})$.

b) k-nearest neighbors (kNN): As an instance-based learning model, kNN does not have a training step, since the only parameter that is used is k, which is provided by the user. Given a new observation $\mathbf{x}$, the training set is reordered by relative distance, $Z = \{\|z_1 - x\|, \|z_2 - x\|, ..., \|z_{n_s} - x\|\}$, and this set is truncated to be of size k. The respective $y_1, y_2, \ldots, y_{n_s}$ are then aggregated into a frequency vector $\mathbf{f}_k$ for each class k, and the final prediction is produced by majority voting, $\hat{y}_i = \max_k f_k$. Usually, Euclidean distance (L2-norm) is used as the distance function. A successful application of kNN for nondestructive evaluation is presented in [23].

c) BayesNet (BN): BayesNet produces the probability that a new observation $\mathbf{x}$ belongs to each class k, $P(Y = k \,|\mathbf{x})$. This *posterior* probability is computed from the Bayes Theorem,

$$P(Y = k \,|\mathbf{x}) = \frac{P(Y = k) P(\mathbf{x} \,|\, Y = k)}{P(\mathbf{x})}. \quad (4)$$

All these terms are learned from the training set $\mathbf{z}$. The denominator is just a scaling factor. $P(Y = k)$ represents the *prior* probability that the hypothesis is true based on what is known from the population and $P(\mathbf{x} \,|\, Y = k)$ is the *likelihood* which represents how compatible the evidence is with the given hypothesis.

d) Random Forest (RF): Random forest is defined as a classifier built from a collection of classification trees. It is a concept of regression trees, induced by a bootstrap sampling method of a training data set, using random descriptors selected in the tree induction process. It is based on the idea of bagging used to average noise and improve the variance reduction by reducing the correlation between trees [11].

The main characteristic is that for the k-th tree in a collection of trees, a random vector Θ_k is generated, with statistical independence from other trees previous vectors, $\Theta_1, \Theta_2, \ldots, \Theta_{k-1}$, and with equal probability distribution. A

tree is grown with the training set and the vector Θ_k, resulting in a specific classifier for a given input vector. After a defined number of trees is generated in this way, they vote for the most popular class for a given input.

III. Experiments

At the simulation stage, we organized the datasets into four distinct classes:

- **Class 1** - Polymeric Grey Insulator (PGI);
- **Class 2** - Ceramic Pin Insulator (CPI);
- **Class 3** - Glass Green Insulator (GGI);
- **Class 4** - Ceramic Bi-color Insulator (CBI);

Five computational models were proposed for image classification. The first one was the pre-trained CNN ResNet-50 (RN) and the remaining four were combinations of RN with the algorithms kNN, SVM, BN and RF. The database percentage split for the training, validation and testing steps were arbitrated as follows:

- **Training** - 50% DATA-INT (800 samples);
- **Validation** - 50% DATA-DEF (400 samples);
- **Tests** - 50% DATA-INT and 100% DATA-DEF.

A. Performance metrics

For performance evaluation, two analytical tools were used: the Receiver Operating Characteristic (ROC) curve and the Confusion Matrix (CM). The CM is a table that allows visualization of the learning algorithm performance. The matrix columns represent the instances of a predicted class, while the rows represent the cases of the actual (real) class. The main equations used for CM estimation were presented in [24] and are listed below:

- True Positive Rate (*TPR*):

$$TPR \approx \frac{\sum True\ Positive}{\sum Condition\ positive} \tag{5}$$

- False Positive Rate (*FPR*):

$$FPR \approx \frac{\sum False\ Positive}{\sum Condition\ Negative} \tag{6}$$

- Precision (*Pr*):

$$Pr = \frac{\sum True\ Positive}{\sum Predicted\ Condition\ Positive} \tag{7}$$

- Accuracy (*Acc*):

$$Acc = \frac{\sum True\ Positive + \sum True\ Negative}{Total\ Population} \tag{8}$$

The ROC curve is a two-dimensional graph where TPR is plotted on the Y axis and FPR is plotted on the X axis. The ROC chart describes the relationship between the benefits (true positives) and the costs (false positives). Areas under the ROC curve (AUC) close to one indicate high discrimination performances. When the area is one the curve is flattened at the top of the graph, corresponding to 100% sensitivity (TPR) and 100% specificity ($1 - FPR$).

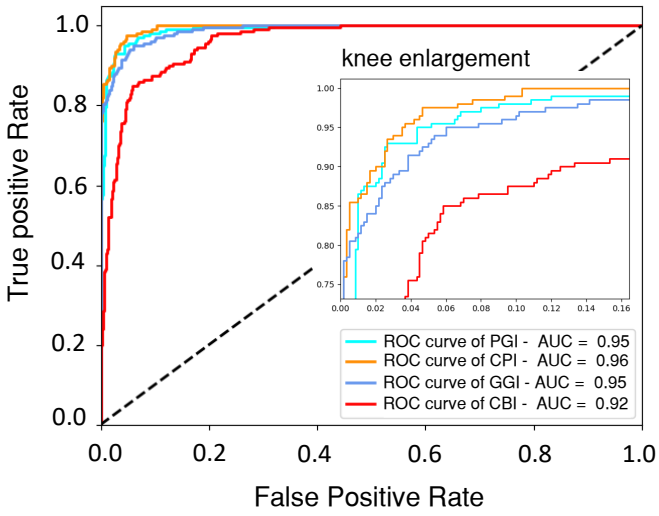

Fig. 5. ROC Curves for ResNet-50 DATA-DEF classification.

TABLE I
CLASS PRECISION FOR DEFECTIVE COMPONENTS.

CLASS	RN	RN + BN	RN + kNN	RN + SVM	RN + RF
PGI	96%	97%	83%	95%	94%
CPI	83%	88%	95%	91%	91%
GGI	96%	100%	97%	100%	99%
CBI	69%	77%	96%	95%	88%

B. Results and discussion

The ROC curves presented in Figure 5 show the performance of the different RN computational approaches, that reached values close to the top of the graph and AUC above 0.9.

The overall classification accuracy of all models are presented in Figure 6. The RN + SVM configuration achieved the best performance, with 95% Acc in the DATA-DEF classification. For the precision values, shown in Table I, it is possible to affirm that the hybrid configurations provided substantial improvements. CBI's precision reached 96%.

Table II presents the confusion matrix of the RN + SVM model. The precision and sensitivity of the classes were greater than or equal to 90%. However, 7.5% of the components belonging to the CBI class had misclassification for CPI. We can say that very aggressive defects in CBI (see Figure 7) can generate classification errors from these two components. This way, there is still room for classifiers optimization.

TABLE II
CM OF THE SVM MODEL FOR DEFECT INSULATOR CLASSIFICATION.

		Predicted				
		PGI	**CPI**	**GGI**	**CBI**	
Actual	**PGI**	**185**	5	0	10	92.50%
	CPI	5	**195**	0	0	97.50%
	GGI	0	0	**200**	0	100%
	CBI	5	15	0	**180**	90.00%
		94.87%	90.70%	100%	94.74%	**95.00%**

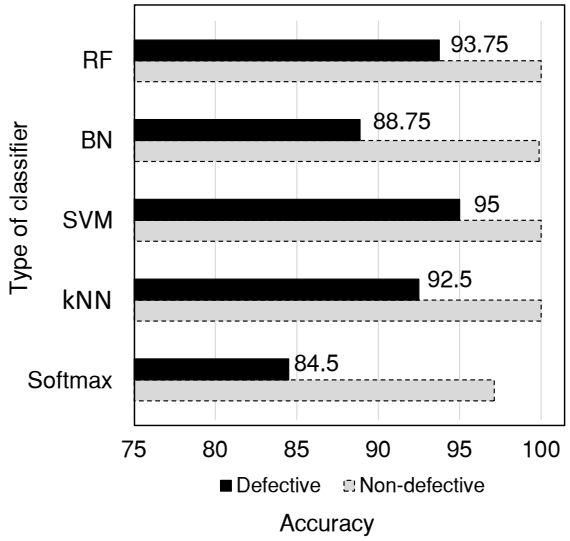

Fig. 6. Overall accuracy (in %) for each configuration model.

Fig. 7. CBI component samples that lead to misclassifications.

IV. Conclusions

This study presented a methodology for automatic classification of ceramic, polymeric and vitreous power distribution insulators through the use of a CNN ResNet-50 combined with other intelligent algorithms. Insulator images were used as inputs for classification. The **ResNet-50 + SVM** architecture showed the best classification performance for the tested hybrid configurations, achieving 95% overall accuracy in classification of the defective component type and approximately 100% for the non-defective ones. Another point of interest in this work was the use of the ResNet-50 as an attribute extractor. The 2,048 values extracted from the second-to-last flattened layer of this CNN were able to provide useful information for all the test classifiers. The CBI, for example, showed an increase in precision of 27 percent points due to the use of hybrid techniques. In future studies, we intend to expand the types of insulators, inserting, for instance, Pillar and Suspension types. It is also planned to evaluate the efficiency of the proposed techniques for insulator images with background interference from real Distribution Networks.

Acknowledgements

The authors would like to thank FAPESB, CAPES and CNPq for the financial support.

References

[1] V. Tajnsek, J. Pihler, and M. Roser, "Advanced logistical systems for the maintenance of overhead distribution lines through DCC with the use of laser monitoring," *IEEE Transactions On Power Delivery*, vol. 26, no. 3, pp. 1337–1343, 2011.

[2] Miller, Abbasi, and Mohammadpour, "Power line robotic device for overhead line inspection and maintenance," *Industrial Robot: An International Journal*, vol. 44, pp. 75–84, 2017.

[3] Katrasnik, Pernus, and Likar, "A survey of mobile robots for distribution power line inspection," *IEEE Transactions on Power Delivery*, vol. 25, pp. 485–493, 2010.

[4] Tompkins, "Look up and live: Developing an effective overhead powerline safety program," *IEEE Transactions on Industry Applications*, vol. 50, pp. 4323–4328, 2014.

[5] J. Koustellis, S. Anagnostatos, C. Halevidis, F. Karagrigoriou, A. Polykrati, and P. Bourkas, "Contact of heavy vehicles with overhead power lines," *Safety Science*, vol. 26, pp. 951–955, 2011.

[6] P. S. Prasad and B. P. Rao, "Review on machine vision based insulator inspection systems for power distribution system," *Journal of Engineering Science and Technology Review*, vol. 9, no. 5, pp. 135–141, 2016.

[7] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 770–778.

[8] Z. Wu, C. Shen, and A. Hengel, "Wider or deeper: Revisiting the resnet model for visual recognition," *Pattern Recognition*, 2019.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classifcation with deep convolutional neural networks," *Pattern Recognition*, vol. 60, no. 6, pp. 84–90, 2017.

[10] V. N. Vapnik, "An overview of statistical learning theory," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 988–999, 1999.

[11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] C. Solomon and T. Breckon, *Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab*. Wiley Blackwell, 2011. [Online]. Available: https://www.wiley.com/en-us/Fundamentals+of+Digital+Image+Processing

[13] R. C. Gonzales and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson Education Limited, 2018. [Online]. Available: https://www.bookdepository.com/Digital-Image-Processing-Global-Edition-Rafael-C.-Gonzalez

[14] J. L. McClelland, "Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review," *Frontiers in Psychology*, vol. 4, 2013.

[15] Z. Kolar, H. Chen, and X. Luo, "Transfer learning and deep convolutional neural networks for safety guardrail detection in 2d images," *Automation in Construction*, vol. 89, pp. 58–70, 2018.

[16] H. Shin, H. R. Roth, and et al, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1285–1297, 2016.

[17] D. P. Kingma and J. L. Ba, "Adam : A method for stochastic optimization," *ICLR*, 2014.

[18] X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, pp. 1318–1325, 2012.

[19] G. Cao, S. Wang, B. Wei, Y. Yin, and G. Yang, "A hybrid CNN–RF method for electron microscopy images segmentation," *Biomimetics Biomaterials and Tissue Engineering*, vol. 18, pp. 1–6, 2013.

[20] P. Roy, R. Tennakoon, K. Cao, S. Sedai, D. Mahapatra, S. Maetschke, and R. Garnavi, "A novel hybrid approach for severity assessment of diabetic retinopathy in colour fundus images," *IEEE*, vol. 26, no. 3, pp. 1078–1082, 2017.

[21] R. K. Fukuchi, B. M. Eskofier, M. Duarte, and R. Ferber, "Support vector machines for detecting age-related changes in running kinematics," *Journal of Biomechanics*, vol. 44, no. 3, pp. 540–542, 2011.

[22] R. Lopes, A. Ayache, N. Makni, P. Puech, A. Villers, S. Mordon, and N. Betrouni, "Prostate cancer characterization on mr images using fractal features," *Medical Physics*, vol. 38, no. 1, pp. 83–95, 2011.

[23] L. F. Rodrigues, F. C. Cruz, M. A. Oliveira, E. F. Simas Filho, M. C. Albuquerque, I. C. Silva, and C. T. Farias, "Carburization level identification in industrial hp pipes using ultrasonic evaluation and machine learning," *Ultrasonics*, vol. 94, pp. 145 – 151, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0041624X18302889

[24] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.