UNIVERSIDADE FEDERAL DO MARANHÃO - UFMA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE ELETRICIDADE/CCET
MESTRADO EM ENGENHARIA DE ELETRICIDADE

LUIS FERNANDO MARIN SEPULVEDA

META-LEARNING APPLICATIONS IN DIGITAL IMAGE PROCESSING

SÃO LUÍS - MA
2019

LUIS FERNANDO MARIN SEPULVEDA

META-LEARNING APPLICATIONS IN DIGITAL IMAGE PROCESSING

Dissertação apresentada ao programa de pós-graduação em engenharia de eletricidade da UFMA como requisito parcial para obtenção do grau de Mestre em Engenharia de Eletricidade.

Orientador: Prof. Dr. Aristófanes Corrêa Silva

SÃO LUÍS - MA

2019

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Núcleo Integrado de Bibliotecas/UFMA

Marin Sepulveda, Luis Fernando.
Meta-Learning Applications in Digital Image Processing / Luis Fernando Marin Sepulveda. - 2019.
111 f.

Orientador(a): Aristófanes Corrêa Silva.
Dissertação (Mestrado) - Programa de Pós-graduação em Engenharia de Eletricidade/ccet, Universidade Federal do Maranhão, Núcleo de Computação Aplicada CCET / UFMA, 2019.

1. CNN. 2. Feature Selection. 3. Image Processing. 4. Machine-Learning. 5. Meta-Learning. I. Silva, Aristófanes Corrêa. II. Título.

LUIS FERNANDO MARIN SEPULVEDA

META-LEARNING APPLICATIONS IN DIGITAL IMAGE PROCESSING

Dissertação apresentada ao programa de pós-graduação em engenharia de eletricidade da UFMA como requisito parcial para obtenção do grau de Mestre em Engenharia de Eletricidade.

Dissertação aprovada em _____ de ____________________ de ___________.

Prof. Dr. Aristófanes Corrêa Silva
Orientador

Prof. Dr. Aura Conci
Membro da Banca Examinadora

Prof. Dr. Denivaldo Cicero Pavao Lopes
Membro da Banca Examinadora

Prof. Dr. João Dallyson Sousa De Almeida
Membro da Banca Examinadora

SÃO LUÍS - MA
2019

To my beloved wife, parents, brother and especially my grandmother

ACKNOWLEDGEMENTS

To my family for their support, to my wife for her patience and dedication, to my brother for his unconditional support and in memory of my beloved grandmother.

To counselor Aristófanes Silva for his trust, support and guidance at all times. To the professors Vicente Leonardo Paucar and Denivaldo Lopes, who received me and helped me on my arrival in Brazil and UFMA.

To my laboratory colleagues, who have always supported me and made me grow as a person and as a professional. Especially to Dênes, Antonino and João for always helping me in the realization of this work.

To the OEA and UFMA for giving me the opportunity and funding to conduct studies in Brazil.

All thanks

"When a man does not sacrifice himself for his ideas, or they are worthless, or man is worth nothing"

Plato

ABSTRACT

In recent decades, advances in capture devices and increase of available digital image data have stimulated the creation of methodologies for data processing that produce various forms of valuable models, such as descriptors, classifiers, approximations and visualizations. These models are often developed in the field of machine learning, which is characterized by a large number of available algorithms, these algorithms often do not have guidelines to identify the most appropriate one based on specific data to which they will be applied and nature of problem under analysis. There is a knowledge that allows to relate the features of the algorithms and data that present a good performance to fulfill a specific task, known as Meta-Knowledge, which can include information on algorithms, evaluation metrics to calculate similarity of datasets or relation of tasks. Being Meta-Learning the study of methods based on principles that explore the Meta-Knowledge to obtain efficient models and solutions, adapting the processes of Machine Learning and Data Mining. The research carried out in this work analyzes the applications and advantages offered by Meta-Learning in field of digital image processing. To carry out this task, different types of images, characterizers, and feature analysis techniques are used; in addition, multiple Machine Learning techniques are applied. The results obtained show that methodology based on Meta-Learning is efficient when applied in processing of digital images for identification and storage of experience generated by developing methodologies for classification of different types of images, obtaining a high performance with respect to an evaluation metrics. This statement means that Meta-Learning allows recommending the most appropriate methodology to perform the processing of a specific type of image based on features of dataset under analysis and the type of specific task to be performed.

Keywords: Meta-Learning, Image Processing, Machine-Learning, Feature Selection, Meta-Data, CNN.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AECLBP	Adjacent Evaluation Completed Local Binary Patterns
AUC	Area Under Curve
AvTD	Mean Taxonomic Distinction
BN	Bayesian Networks
CFS	Correlation-based Feature Selection
CNN	Convolutional Neural Network
CT	Computed Tomography
DC	Ductal Carcinoma
DDSM	Digital Database for Screening Mammography
DFM	Average Phylogenetic Diversity
EQE	Extensive Quadratic Entropy
EQI	Intensive Quadratic Entropy
HMM	Hidden Markov Model
JDT	J48 Decision Tree
K-NN	K-nearest neighbor
LIDC-IDRI	Lung Image Database Consortium
MLP	Multilayer Perceptron
MNND	Mean Nearest Neighbor Distance
MRI	Magnetic Resonance Imaging
NB	Naive Bayes
RC	Random Committee
ReLU	Rectified Linear Unit

RF	Random Forest
RR	Rules Roughset
RT	Random Tree
RTD	Taxonomic Diversity Index
PCA	Principal Component Analysis
q	Index of Taxonomic Distinction
SBP	Basic Sum of Normalized Weights
SVM	Support Vector Machines
TTD	Total Taxonomic Distinction
WMD	Pure Diversity Index

CONTENTS

1 INTRODUCTION

Recent decades have brought a large amount of data, as a result of technological advances and increased data capture, they are eligible for automated analyses that could result in descriptions, classifiers, approximations, visualizations or other forms of valuable models. These models are often developed and used in field of Data Mining and Machine Learning, being characterized by having a large number of algorithms and are applied to perform multiple tasks, such as credit rating, medical diagnosis and classification of images (BRAZDIL; GIRAUD-CARRIER, 2018),(WEISS; KHOSHGOFTAAR; WANG, 2016). Specifically, image processing has been stimulated by the large-scale increase in image datasets (XIA et al., 2016), which leads to creation of new methods that perform specific processing tasks, nevertheless, these algorithms often do not have guidelines that select the correct method to perform a task, based on nature of problem under analysis (BRAZDIL; GIRAUD-CARRIER, 2018),(SMITH et al., 2014).

However, there is a knowledge that identifies the similarities between the algorithms and data that work best to fulfill a specific task, known as Meta-Knowledge, which can include information about the algorithms, metrics available to calculate the similarity of dataset or relationship of tasks. Being the Meta-Learning the study of methods based on principles that explore the Meta-Knowledge to obtain efficient models and solutions, adapting the processes of Machine Learning and Data Mining (RAHMAN; BHATTACHARYA, 2017).

Consider a scenario that requires identify the best methodology for a task but has a large set of possibilities, without a recommendation strategy, one option is test each methodology independently, with the disadvantage of repeat the tests for each new dataset under study (DE MORAIS; MIRANDA; SILVA, 2017),(DE MELO; PRUDÊNCIO, 2014), this happens because there is no learning from experience and each dataset has inherent features that are used by methodologies to provide results (PIMENTEL; CARVALHO, 2019).

One recommendation strategy has the advantage of avoiding reprocessing by establishing the relationship between features of dataset and performance of methodology, this relationship is known as Meta-Knowledge and represents the experience. According to the task, Meta-Knowledge can include parameters configuration, performance evaluation metrics and analysis of the most representative features (CUNHA; SOARES; CARVALHO, 2018),(RAHMAN; BHATTACHARYA, 2017).

One of the fields that Meta-Learning can act on is the image processing, used as a tool for selection of classification, segmentation and diagnosis methods, based on Meta-Knowledge extraction and features of datasets (CAMPOS; BARBON; MANTOVANI, 2017),(CHEPLYGINA et al., 2017),(AMIRI et al., 2014). However, there are few studies found.

Based on the statements presented, this work aims to explore different Meta-Learning applications in field of image processing, validating performance when applied to several tasks, datasets, techniques and parameter settings.

1.1 Objective

General aim of research is to explore different applications of Meta-Learning in the field of image processing.

Having as specific objectives:

1. Study and use classification and characterization techniques based on pattern recognition and applied to image processing.

2. Investigate the concepts and applications of Meta-learning, its techniques, development and implementation in images.

3. Construction of a methodology for selection of the most suitable methodology for image classification based on Meta-Learning.

4. Validation of proposed methodology through experiments, using multiple datasets with different classification aims.

1.2 Work Organization

This work presents several investigations of Meta-Learning applications in image processing, for this reason, in addition to chapters introduction, related works, theoretical foundation, discussion and conclusion, each chapter presents a new application and behaves as an independent investigation.

- Chapter 2 presents a summary of related works found in state of the art, related to Meta-Learning and image processing.
- Chapter 3 presents the theoretical foundation used for construction of this research. Addressing Meta-Learning, classification techniques based on Machine Leaning and Deep Leaning, image characterization techniques, feature analysis techniques and performance evaluation metrics.
- Chapter 4 discusses the construction of Meta-Data through the creation of a new methodology for classification of medical images in order to identify Ductal Carcinoma.
- Chapter 5 establishes the construction of Meta-Model, for classification of different types of medical images with different classification tasks.
- Chapter 6 proposes the extraction of experience from the methodologies proposed in literature for construction of a methodology recommendation system for different types of medical images using CNN.
- Chapter 7 proposes the extraction of experience from the methodologies proposed in literature for construction of a methodology recommendation system for different types of industrial images.
- Chapter 8 proposes the use of Meta-Learning for identification of CNN configuration parameters that will be applied for image classification.
- Chapter 9 presents discussions about the lessons learned in research.
- Chapter 10 presents conclusions and scientific productions.

2 RELATED WORKS

In this chapter, works found in literature about Meta-Learning and its application in images are presented. Similarly, image processing works are presented, since this field is the basis on which the proposed methodologies are developed, and are considered sources of experience on which Meta-Leaning is based.

2.1 Meta-Learning in Image Processing

Meta-Learning has been used by Maicas et al. (2018) as a method of automatic classification of Dynamic Contrast Enhanced Magnetic Resonance Images (DCE-MRI) of breasts. Meta-Data is obtained from the training applied to a series of individual tasks, based on incremental learning performed by radiologists, this learning consists of simple classification problems with a small training set. The Breast DCE-MRI dataset used contains 117 patients, where the final label of the classification process can be "not found", "malignant" or "benign"; the area under the ROC curve (AUC) is used as evaluation parameter, the best result obtained is 90%. Maicas et al. (2018) employs Meta-Learning as a DCE-MRI classification method, using the results of a set of simple classification tasks such as experience or learning data, which involves dividing the main classification aim into multiple tasks.

The method proposed by Maicas et al. (2018) is limited to a single specific aim of classification on a single type of image, methodology proposed in this dissertation has the advantage of allowing the use of multiple types of images with different classification aims, in addition, proposed methodology differs in the aim of classification and content of Meta-Data, being the aim, to identify the best classification methodology for different types of images, and Meta-Data contains the results of tests that identify the most appropriate classification methodology for each type of image, including features that best represent the image.

Campos, Barbon e Mantovani (2017) presents a Meta-Learning approach for recommendation of image segmentation algorithms, four Meta-Bases were used to evaluate three segmentation methods. A set of 44 features were extracted based on color, frequency domain, histogram, texture, contrast and quality to represent the images. The database used correspond to different images with different segmentation objectives, the first one has the objective of segmenting Chicken breast, the second requires the identification of wounds, the third has the objective of segmenting clouds, the last data set is made up of all the previous images. The segmentation methods to be evaluated are: Otsu's thresholding, K-means, and vector support machine (SVM), and the tested Meta-Learning methods are: a linear classifier (LogReg), C4.5 Decision Tree (used through the J48 implementation), Naïve Bayes (NB), K-Nearest Neighbors (K-NN), Neural Networks Using Model Averaging (avNNet), Random Forest (RF) and SVM. The evaluation criteria used for the selection of the Meta-Model classification method were AUC, average, F-measure, average of true positive rate (TPRate) and average of true negative rate (TNRate). The results show that the Random forest method presents a better performance as a classifier of the best segmentation method for the database under study, obtaining average of 88.7%, AUC of 93.7%, F-Measure of 81.4%, TPRate of 81.3% and TNRate of 88.5%. In this article Campos, Barbon e Mantovani (2017) employs Meta-Learning for the recommendation of image segmentation algorithms, Meta-Data is obtained from the extracted features and the results of testing three segmentation methods in each dataset. Identifying Random Forest as the most suitable for the Meta-Model, finally the importance of each feature is extracted using the Mean Decrease Gini value.

The approach proposed by Campos, Barbon e Mantovani (2017) performs the features analysis only for RF classifier, which limits the result analysis. The methodology proposed in this dissertation has the advantage of performing analysis of features extracted for different classification techniques, except CNN. There are also differences in the objective of the techniques under study, being the classification and not the segmentation of images, in the same way, there are differences in the characterization techniques used and in the type of dataset under study.

Santos et al. (2017) conducted a study concerning the combination of multiple classification approaches for the problem of interpretation of remote sensed hyperspectral images. Expanding the traditional weighted linear combination algorithm optimized by genetic algorithms (WKC-GA), to fine-tune the class probabilities within the combination process. The comparison of the performance of the traditional algorithm and its extension is done with a more complex non-linear Meta-Learning strategy with a radial kernel SVM classifier. The tests were conducted using the datasets of the Indian Pines and the Pavia University, and show as a result that the proposed method exceeds the predictions of the traditional approach. The overall accuracy achieved by the Meta-Learning method is

91.4% for Indian Pines data and 98.7% for Pavia University data. In this article, Santos et al. (2017) use Meta-Learning as a classification method based on Stacked Generalization, which uses different methods to calculate the probability of a sample belonging to a class. The Meta-Data is constructed by concatenating the probabilities obtained by each method of the stack for each class.

The approach used by Santos et al. (2017) is limited to a single specific type of image, methodology proposed in this dissertation has the advantage of allowing the use of multiple types of images. The proposed methodology differs in the aim task of classification techniques, in characterization of images and the type of dataset.

2.2 Meta-Learning in Data Processing

Meta-Learning is used by Tanfilev, Filchenkov e Smetannikov (2018) to choose the feature selection algorithm from a preselected set to produce feature rankings. Selection of final features is made based on Aggregated Extended Adjusted Ratio of Ratios (AEARR) metric using fivefold cross-validation, in addition, Bayesian Networks were used to measure the efficiency of feature selection algorithm applied to 75 dataset of different domains, research verifies that use of rank aggregation methods show a better result than basic algorithms, obtaining an accuracy of 99.4%.

The approach proposed by Tanfilev, Filchenkov e Smetannikov (2018) limits the evaluation of performance of methods to dataset using the AEARR metric. Methodology proposed in this dissertation has the advantage of measuring performance of feature subset selection techniques in relation to particular aim task of each dataset.

Pimentel e Carvalho (2019) uses Meta-Learning in a set of Meta-Features to predict the performance of different clustering algorithms, a new set of Meta-Features was proposed based on correlation and dissimilarity measures, using 219 datasets from Open Machine Learning (OpenML) like study object. K-NN and RF techniques were used to validate the proposed method, results showed that grouping algorithm recommended obtained a better grouping quality (affirmed by the Adjusted Rand), likewise, some Meta-Features proved to be more suitable than others for recommendation of grouping algorithms. In this article, Pimentel e Carvalho (2019) shows the effectiveness of methods based on Meta-Learning for recommendation of algorithms, besides, it is possible to identify features that are more significant for the process.

The approach proposed by Pimentel e Carvalho (2019) has the limitation of having the same aim for each dataset, given that aim of study is clustering algorithm recommendation. Methodology proposed in this dissertation presents a different application of Meta-Learning and differs form of characterization of datasets, besides it has the advantage of allowing multiple aims in objects of study.

Mantovani et al. (2015) uses Meta-Learning to recommend value of operational parameters for SVM technique, based on features of the dataset to be applied. For the experiments, 145 test datasets and 21 training datasets were used, obtained from University of California Irvine Machine Learning Repository and OpenML. Six classification methods for Meta-Model were tested: J48 Decision Tree (J48), NB, K-NN with k = 3, Multilayer Perceptron (MLP), RF and SVM. Results show that use of new set of operational parameter values produced significantly better models than the default values suggested by Machine Learning tools. In this article, Mantovani et al. (2015) shows how Meta-Learning can be used for selection of operational parameters of SVM technique, based on features of datasets to be classified.

Methodology proposed in this dissertation presents a different technique aim, using CNN as object of study for recommendation of configuration parameters.

2.3 Image Processing

Araujo et al. (2017) proposed a method based in Convolutional Neural Network (CNN) for classification of histologist images of breast cancer using hematoxylin and eosin staining (H&E), with purpose of retrieving information at different scales, including nuclei and general organization of tissue. Images are classified in four classes, normal tissue, benign lesion, in situ carcinoma and invasive carcinoma, and in two additional classes, carcinoma and noncarcinoma. The network is trained in an increased patch dataset and tested on a separate set of images. The extracted features are also used to train an SVM classifier. Both CNN and SVM classifiers achieve comparable results. A high sensitivity of method was obtained for carcinoma cases of 95.6%, and accuracy of 77.8% for four classes and 83.3% for carcinoma / noncarcinoma. In this article, Araujo et al. (2017) evidences how classification of histological images of breast cancer with H&E can be executed with CNN and SVM, results show a high sensitivity for detection of cases with carcinoma.

Methodology proposed in this dissertation presents a different form of preprocessing of images that improves the performance of methodologies based on CNN.

Spanhol et al. (2016) presented a set of histopathological images of breast cancer, called BreaKHis, this is made up of 7909 images acquired in 82 patients. The dataset currently contains four histological types of benign breast tumors: adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma; and four malignant tumors (breast cancer): ductal carcinoma (DC), lobular carcinoma, mucinous carcinoma, and papillary carcinoma. Six methods of feature extraction were used: local binary patterns (LBP), completed LBP, local phase quantization, gray-level co-occurrence matrix, threshold adjacency statistics, and one keypoint descriptor, named ORB. Four different classifiers were used to assess the feature sets: a 1-nearest neighbor (1-NN), quadratic linear analysis, SVMs, and random forests of decision trees. The article makes the classification of two classes, benign and malignant breast images, obtaining a precision that oscillates between 80% and 85%. In this article, Spanhol et al. (2016) presents the BreaKHist dataset and introduces its processing through the extraction of features and use of machine learning methods, results are focused on accuracy for classification of dataset in benign and malignant cases.

Methodology proposed in this dissertation compare different classification methodologies with different ways of data characterization and add a new dataset that increase the accuracy.

In Mercan et al. (2018), breast images were segmented into eight tissue types for diagnosis using two different methods, an SVM-based approach that uses color and texture features to classify superpixels to produce tissue labeling, and a CNN-based approach that use raw images. Then, frequency of tissue labels and co-occurrence of histograms based on superpixel segmentation to classify images into diagnostic categories is calculated. Two schemes of classification were compared, SVM to classify images into four diagnostic categories, and a series of SVM to diagnose one class at a time. Results present a accuracy of 83% in critical atypia vs the ductal carcinoma in situ threshold. In the article, Mercan et al. (2018) shows a development that classifies images of breast tissue in benign, atypia, DC in situ and invasive DC, using the SVM and CNN methods, achieving as best result identification of DC in situ versus atypia.

Methodology proposed in this dissertation improves the accuracy for DC classification.

3 THEORETICAL FOUNDATION

This chapter presents the theoretical foundations used by the proposed methodologies, explaining Meta-Learning and its functionality, presenting different techniques of classification and characterization of images, as well as the methods of relevance analysis used in the extracted features.

3.1 Meta-Learning

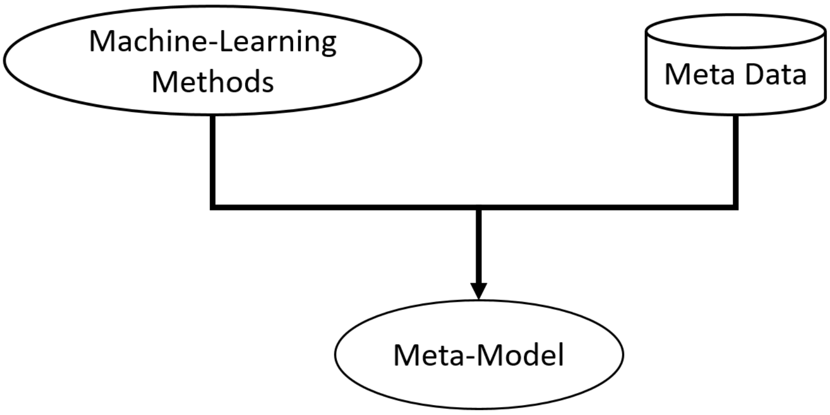
Initially, the term Meta-Learning was used in the area of educational psychology, where it is interpreted as an understanding and adaptation of the learning process itself, beyond the acquisition of knowledge, which allows to evaluate the learning approach and adjust it according to a specific task (LEMKE; BUDKA; GABRYS, 2015).

The application of Meta-Learning in the context of Machine Learning and Data Mining is similar, since it focuses on the learning process for adaptation according to a specific task. Data Mining and Machine Learning have a large number of available algorithms, but lack guidance to select the appropriate method according to the nature of the problem under analysis. Since the implementation and testing of these algorithms represent a significant consumption of resources, professionals and researchers often use only a set of algorithms, hoping that these methods fit with data features. This practice has stimulated research that seeks to determine if the application of Data Mining and Machine Learning depends only on the search for a good algorithm that adjusts the data, or if there are several operational layers that can be exploited to produce an increase in performance in comparison with time. The last option implies that it is possible to learn from the process that allows identifying the best algorithm for a specific type of data, determining the relationship between the performance of an algorithm and features of the data, this relationship and the extracted features are considered experience and It is possible to use them as a learning source for the system, which allows generating additional knowledge that simplifies the automatic selection of models for new datasets (BRAZDIL et al., 2008).

Meta-Learning offers two particular advantages, the first one is related to the methods and their possible combinations to perform specific tasks, this particularity is presented in an already developed system, which gives the user access to predictive models (BRAZDIL et al., 2008). The second particularity refers to taking advantage of the repetitive use of a predictive model in similar tasks and its adaptation to new needs, using the experience acquired in previous executions, avoiding to start learning the model again from the beginning, this task is also known as learning to learn . Meta-Learning helps to control the process of exploitation of the accumulated experience by searching for patterns in the data used for a specific task (BRAZDIL et al., 2008).

The learning source of Meta-Learning is composed of experience acquired when developing methodologies to fulfill specific tasks using defined data types, this is done with the aim of constructing a model that allows to recommend the appropriate methodology to perform a task based on features of data (GRABCZEWSKI, 2014). Making a recommendation requires that model employ a classification technique, usually using machine learning methods. Experience used for training is known as Meta-Data, while the methodology that includes Machine Learning is known as Meta-Model. Figure 1 presents the basic components of a system based on Meta-Learning.

Figure 1 – Basic Components of Meta-Learning.



Source: Prepared by author

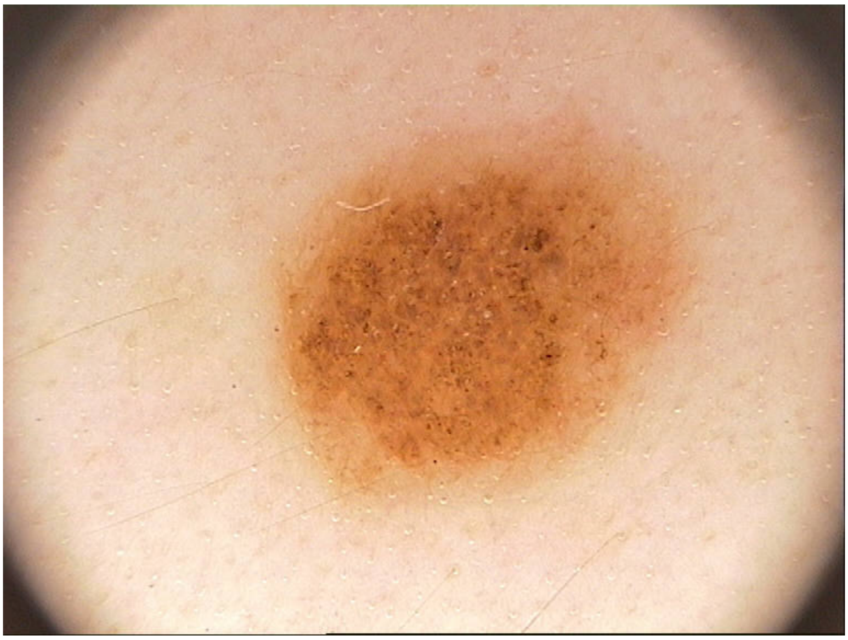
3.1.1 Meta-Data

Meta-Data contains the experience on which Meta-Learning is based to carry out the recommendation process, it is also known as Meta-knowledge. The type of data contained depends on the task to be performed, e.g., feature, images, methods and configuration parameters. Conceptually, Meta-Data consists of two components, the first one allows identifying the features of the data on which experience has been obtained. The second component refers to the methodology used in the data to perform the task, including the methods used, its parameter settings and preprocessing (GIRAUD-CARRIER, 2008).

Since the Meta-Data is the basis on which learning occurs, it must contain at least two components, the first one allows to identify the data in which a processing has been performed, this identification is carried out by extracting features, also called Meta-features. The second component refers to the performance that different techniques have obtained when applied to the data. The union of these two components, allows to make the recommendation of which technique is the one that presents a better result, for future data without the need to perform the experiments again (BRAZDIL et al., 2008).

To illustrate the idea of Meta-Data, consider Figure 2, for which a binarization process is required, in order to make the region of interest appear in black and background in white, for this task it is used a function that, based on the level of pixel intensity and a defined threshold, performs the binarid process, Figure 3 show the results achieved.

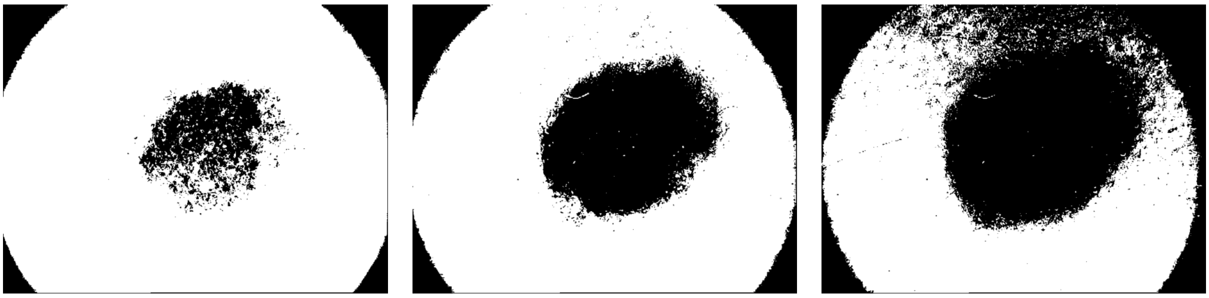
Figure 2 – Image to Binarize.



Source: PH^2 Dataset (MENDONCA et al., 2013)

To extract the first component of Meta-Data it is necessary to consider the classification technique that will be used for Meta-Model, for example, if a CNN is used, the image is taken as a features. The second element of Meta-Data is the performance achieved by each technique, for this example the same function was used, modifying only the threshold, for this reason the second component would be formed by the methodology

Figure 3 – Binarized image.



(a) *Threshold: 134, Accuracy:33%.*

(b) *Threshold: 170, Accuracy:84%.*

(c) *Threshold: 170, Accuracy:70%.*

Source: Prepared by author

that presents a better result, in this case the experience is formed by: binarization function, threshold value of 170 and accuracy of 84%.

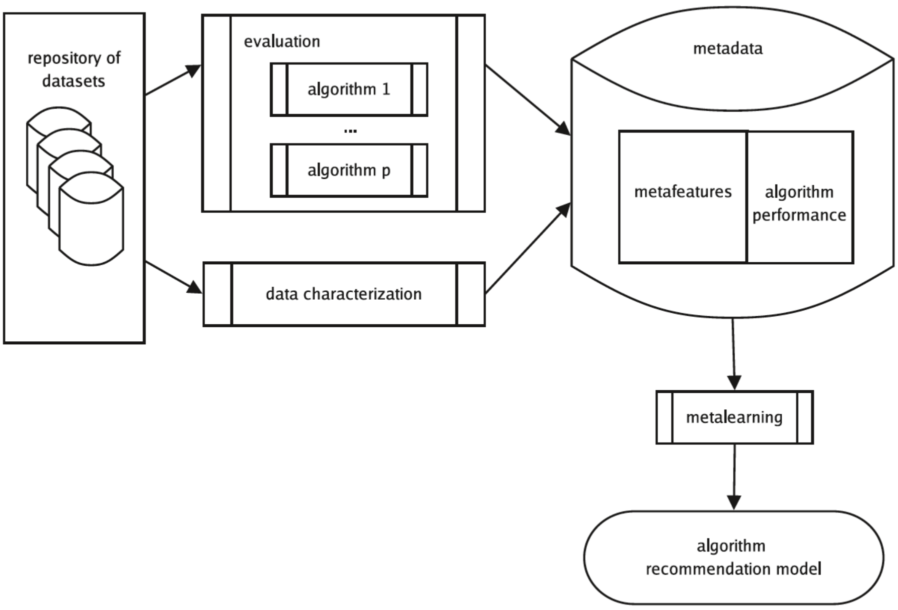
3.1.2 Meta-Model

When the recommendation system has already been developed, the Meta-Model is responsible for processing the new data and depending on its features, recommending the most appropriate methodology to fulfill a specific task (CUNHA; SOARES; CARVALHO, 2018). The objective or Meta-Target function refers to form of recommendation that is provided to user, one form used is the classification of base algorithms. The type of Meta-Target determines the type of Meta-Algorithm, that is, the Machine Learning methods that can be used (BRAZDIL et al., 2008).

As presented in Figure 1, the Meta-Model uses Meta-Data as a training source and a Machine Learning technique as a classification system. In the construction of Meta-Model, different classification techniques are tested taking the Meta-Features as input parameters, the aim of classification is that Meta-Model can identify for each Meta-Data element the methodology presented a better performance. Considering the example of Figure 2, the goal of Meta-Model is that for each image of this type it is identified as the best classification methodology, which uses a threshold of 170 and achieves an accuracy of 84%. Once the Meta-Model has been built, it is used as a recommendation system, for this task, it extracts the features of the new data and uses the experience gathered in the Meta-Data. This implies that the more experience of different types of data, the Meta-Data contains, the recommendation made by the Meta-Model will be more accurate.

To facilitate the understanding of Meta-Learning, consider the scenario presented in Figure 4, in which an algorithm recommendation model is developed. The construction of the model begins with the repositories of data sets, these contain the data for which it is sought to determine the most appropriate algorithm to fulfill a task, each repository is tested with each of the evaluation algorithms, storing its performance in Meta-Data, in the same way that the characterization of the data is performed, in the end, the Meta-Data contains a representation of each repository, formed by characteristics or Meta-characteristics that point to an algorithm and its performance. In the next step, Meta-Learning is applied and based on Meta-Target, the Machine-Learning method that shows better performance using Meta-Data is selected, the result is an algorithm recommendation model or Meta-Model.

Figure 4 – Meta-Learning to obtain Meta-Data for algorithm selection.



Source: Adapted from (BRAZDIL et al., 2008)

3.2 Classification Techniques

This section presents different classification techniques that are tested to be part of the Meta-Model, since the data under study consist of images, the techniques are selected using as criteria their applicability in classification of images, likewise, it is taken into consideration the techniques present in the state of the art.

3.2.1 Bayesian Networks (BN)

BN is an acyclic directed graph where the nodes represent variables and the arcs relations, each variable possesses a limited set of excluding states. Bayesian Networks are based on the probability of occurrence of each node and based on the path of the graph, the output of the traveled nodes is presented as a result (KHAKZAD, 2015).

3.2.2 Convolutional Neural Network (CNN)

CNN is a neural network that uses the convolution process in some of its layers, the convolution has the objective of extracting features from the original data by applying filters known as a convolution filter, in CNN each neuron can apply a different filter, at the end, the data is taken to the next layer in which a sub-sampling process is performed, producing new information, before or after the sub-sampling it is possible to apply an activation function in order to discard unnecessary data. The process of convolution and sub-sampling can be repeated, producing diverse results with each configuration of the layers. Finally, it has a full-connected layers as a Multilayer Perceptron Network (LECUN; BENGIO; HINTON, 2015).

3.2.3 Gaussian Mixture Model

The Gaussian mixture model is a parametric probability density function that represents the density of sub populations that have a Gaussian distribution, represented in Equation 3.1 (REYNOLDS, 2015).

$$P(x|\theta) = \sum_{i=1}^{k} w_i g(x|u_i, \sum i) \tag{3.1}$$

where x is a D-dimensional continuous-valued data vector, w_i are the mixture weights and $g(x|u_i, \sum i)$ are the component Gaussian densities, which is a $D-$variate Gaussian function Equation 3.2.

$$g(x|u_i, \sum i) = \frac{1}{(2\pi)^{\frac{D}{2}} |\sum_i|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu_i)' \sum_i^{-1}(x-\mu_i\} \tag{3.2}$$

with mean vector μ_i and covariance matrix $\sum_i$.

3.2.4 Hidden Markov Model (HMM)

HMM is a stochastic process based on two levels of uncertainty: a random observation process associated with each hidden state and a Markov chain that characterizes the probability relationship between the states. HMM is used for the prediction of states based on the information of the current state and the conditional probability of the hidden states, and is also used for pattern recognition (RABINER, 1989).

3.2.5 J48 Decision Tree (JDT)

The J48 algorithm consists of a decision tree based on the ID3 and C4.5 algorithms; where the internal nodes represent an operation, the branches the results and the leaves denote class labels (YADAV; CHANDEL, 2015),(QUINLAN, 1996). The tree is made up of decision nodes, random nodes and final nodes, these are often presented graphically by means of squares, circles and triangles. Equation 3.3 presents the entropy E (SUGUMARAN; MURALIDHARAN; RAMACHANDRAN, 2007) that is used to calculate the decision criteria that identify the relevant input variables.

$$E = -P * \log_2(P) - N * \log_2(N) \tag{3.3}$$

where P, N are proportion of positive and negative examples of the training dataset.

3.2.6 K-Nearest Neighbor (K-NN)

K-NN is a method used for classification based on proximity distance, let $\{K_{ij}, i = 1...n, j = 1...m\}$ be a test set where n is a number of class and m is the size of the feature vector and let $\{E_j, j = 1...m\}$ be a sample set. K-NN execution consists in the repetition of two steps until reaching a predetermined limit or until the k samples remain unchanged, the first step is to calculate the distance d_i of each element E with each element of K, different types of distances can be used, including the Euclidean distance represented in Equation 3.4.

$$d_i = \sqrt{\sum_{j=1}^{m}(E_j - K_{ij})^2}, \quad \forall i \in \{1...n\} \tag{3.4}$$

The second step is to assign a class label to each element of the sample set based on the proximity to the element k that represents a class, finally a new test k is chosen based on the class boundaries (DENG et al., 2016).

3.2.7 Linear Logistic Regression

Linear logistic regression models calculate the posterior class probabilities $Pr(G = j \mid X = x)$ shown in Equation 3.5, for class j through linear functions in x (SUMNER; FRANK; HALL, 2005).

$$Pr(G = j \mid X = x) = \frac{e^{F_j(x)}}{\sum_{k=1}^{J} e^{F_k(x)}}, \sum_{k=1}^{J} F_k(x) = 0, \tag{3.5}$$

where $F_j(x) = \beta_j^T.x$ are linear regression functions, β_j^T is the transpose of the parameter vector j and x is the input vector, it is usually fit by finding maximum likelihood estimates for the parameters β_j.

3.2.8 Multilayer Perceptron (MLP)

MLP consists of a network of neurons, separated into layers that transform the input data through simple functions or non-linear activation, the response of the first layer of neurons is normalized according to the connection weight and is used as input for a new layer (GARDNER; DORLING, 1998). The last layer of the network is known as the output layer and is responsible for delivering the result that identifies the class to which the input data belongs.

3.2.9 Naive Bayes (NB)

The NB technique is used in classification problems and is based on the Bayes theorem, it consists of a simple form of a Bayesian network and is widely used in real-world applications. Given a variable X represented by a vector $< a_1, a_2, ..., a_m >$ (JIANG et al., 2019), NB uses the Equation 3.6 to estimate the probability of class membership and Equation 3.7 to predict the class tag.

$$P(C \mid X) = \frac{P(C) \prod_{j=1}^{m} P(a_j \mid C)}{\sum_{c' \in C} \prod_{j=1}^{m} P(a_j \mid C\prime)} \tag{3.6}$$

$$C(X) = \arg\max_{c \in C} P(C \mid X) \tag{3.7}$$

where m is the number of attributes, a_j is the jth attribute value of X, C is the collection of all possible class labels c.

3.2.10 Random Committee (RC)

Uses a set of base classifiers, each classifier uses a subset of characteristics selected randomly. The final classification is determined using the direct average of the predictions generated by the base classifiers (LIRA et al., 2007).

3.2.11 Random Forest (RF)

RF consists of a combination of a set of prediction trees, each tree uses a subset of training samples randomly selected, which implies that the same sample can be selected by more than one tree, the labelling of the class is done by the average response given by the set of prediction trees (BELGIU; DRăGUţ, 2016).

3.2.12 Rules Roughset (RR)

Rules Roughset is based on the approximation of concept through two exact sets called upper and lower approximation of the concept that allows to identify the partial or total dependencies in the data and provide a null value approach (WEI; LIANG, 2019).

3.2.13 Random Tree (RT)

The operation of Random Tree is similar to Quinlan's C4.5 or CART algorithms, the inner nodes correspond to input features and the leaf node represents class labels; the difference is that RT selects a random subset of features in each division, the number of features used is a determined as input parameter, this work use Equation 3.8 to determine the number of feature to be selected (KARGER et al., 1997).

$$F = \log_2(z) + 1 \tag{3.8}$$

where z is the number of predictors.

3.2.14 Support Vector Machines (SVM)

The foundations of SVM were developed by Vapnik (1995) and consist in the construction of a hyperplane of equidistant separation from the closest examples of each class to achieve the maximum margin on each side of the hyperplane. SVM is a supervised Machine Learning algorithm that can be used for classification or regression. In this algorithm, each data element for training is represented as a point in the n-dimensional space (where n is the number of features). Then, the classification is made by finding the hyperplane that separates the sets of points of each class Equation 3.9 and makes use of Equation 3.10 to transforms the original data space.

$$min\frac{1}{2} \parallel w \parallel^2 \quad s.t. \quad y_i(w * x_i + b) \geq 1, \forall x_i \tag{3.9}$$

where w is a m-dimensional vector, b is a bias term and $y \in \{-1, 1\}$.

$$K(x_i, x_j) = \exp(\frac{-(x_i - x_j)^2}{2} * \sigma^2) \tag{3.10}$$

where σ^2 represents the variance.

3.3 Feature Extraction

The feature extraction process is carried out in order to identify attributes in the data that result in quantitative information of interest and allow differentiating one object class from another, this process is also known as characterization. The extraction of features in images seeks to obtain insensitivity to capture and lighting noise, in the same way it must be independent of certain variations such as translation, rotation, scale and transformations. Image characterization can be used in processes that require segmentation of elements that make up the image or can also be used to extract data that allows a classification of the entire image (GONZALEZ; WOODS, 2008).

The techniques are selected considering the ability to characterize a complete image for a classification process, the extracted features are used as input data for the Machine Learning techniques presented in Section 3.2, except for CNN, which uses as input the image itself. The process of characterization of images is used for construction of Meta-Data presented in Figure 1.

3.3.1 Phylogenetic Indexes

The feature extraction method used in this work, correspond to phylogenetic indexes based on the diversity of species, were selected due to the potential in the characterization of a region or image, which allows establishing kinship relationships between different individuals and species (NETO et al., 2018). Diversity is a term frequently used in the area of ecology, where an index of diversity describes the variety of species present in a community or region (MAGURRAN, 2004), phylogeny is a branch of biology responsible for the study of the evolutionary relationships between the species in order to determine possible common ancestors (BAXEVANIS; OUELLETTE, 2004). To use this approach in images, it is necessary to make an analogy between its properties, Table 1 shows the correspondence between the terms.

Table 1 – Matching Terms Proposed Between Biology and Image Processing.

Biology	**Image processing**
Community	Image
Species	Level of intensity
Individual	Pixel or voxel

The characterization of images is done by using thirteen phylogenetic indexes:

1. **Index of Taxonomic Distinction (q)**: Represents the taxonomic distance between two individuals, with the restriction that they belong to different species, represented in Equation 3.11 (CLARKE; WARWICK, 1998).

$$q = \frac{\sum \sum_{i<j} w_{ij} x_i x_j}{\sum \sum_{i<j} x_i x_j} \tag{3.11}$$

where $x_i \, (i = 0, ..., s)$ is the abundance (number of pixels) of the i-th species, $x_j \, (j = 0, ..., s)$ is the abundance of the j-th species, s indicates the number of species, and w_{ij} is the distance between the species i and j.

2. **Taxonomic Diversity Index (RTD)**: The index includes the abundances of the species and the taxonomic relationship between them. Expresses the average taxonomic distance between any two individuals, chosen randomly in a sample, shown in Equation 3.12 (CLARKE; WARWICK, 1998).

$$RTD = \frac{\sum \sum_{i<j} w_{ij} x_i x_j}{[n\,(n-1)\,/2]} \tag{3.12}$$

where $x_i \, (i = 0, ..., s)$ is the abundance (number of pixels) of the i-th species, $x_j \, (j = 0, ..., s)$ is the abundance of the j-th species, s indicates the number of species, w_{ij} is the distance between the species i and j, and n is the total number of individuals.

3. **Intensive Quadratic Entropy (EQI)**: When the abundance values are the same (hypothetical or formal), the EQI is a function that represents the number of species and their taxonomic relationships, represented in Equation 3.13 (IZSÁK; PAPP, 2000).

$$EQI = \frac{\sum w_{i,j}}{s^2} \tag{3.13}$$

where $w_{i,j}$ represents the distance between species i and j, and s represents the number of species.

4. **Extensive Quadratic Entropy (EQE)**: Represents the sum of the differences of the species, shown in Equation 3.14 (IZSÁK; PAPP, 2000).

$$EQE = \sum w_{i,j} \tag{3.14}$$

where $w_{i,j}$ represents the distance between species i and j.

5. **Average Taxonomic Distinction (AvTD)**: Indicates the average taxonomic distance between any two randomly chosen species, represented in Equation 3.15 (CLARKE; WARWICK, 1998).

$$AvTD = \frac{\sum \sum_{i<j} w_{ij}}{[s(s-1)/2]} \tag{3.15}$$

where w_{ij} is the distance between species i and j, s represents the number of species.

6. **Total Taxonomic Distinction (TTD)**: Represents the mean phylogenetic distinction summed along all species, shown in Equation 3.16 (CLARKE et al., 2007).

$$TTD = \sum_i \frac{\sum_{i \neq j} w_{ij}}{(s-1)} \tag{3.16}$$

where w_{ij} is the distance between species i and j, s is the number of species.

7. **Pure Diversity Index (WMD)**: It is responsible for calculating the value of the distance of a species for its closest neighbor, represented in Equation 3.17 (FAITH,

1994),(WEITZMAN, 1992).

$$WMD = \sum w_{imin} \tag{3.17}$$

where w_{imin} indicates the distance of the nearest neighbor of species i for all other species.

8. **Mean Nearest Neighbor Distance (MNND)**: Represent the average of the phylogenetic distance of the closest relative to all species, equivalent to the rates of species by gender. MNND is calculated from the weighted average of the phylogenetic distance of each neighbor closest to the species, with equal weights to the abundance of the species, presented in Equation 3.18 (BOOTS; GETIS, 1988).

$$MNND = \frac{\sum w_{imin}}{s} \tag{3.18}$$

where w_{imin} is the distance of the nearest neighbor of species i and s is the number of species.

9. **Basic Sum of Weights (SBP)**: It represents the sum of the nodes of the roads, starting from the root to all the species of the phylogenetic tree, shown in Equation 3.19 (KEITH et al., 2005),(POSADAS; ESQUIVEL; CRISCI, 2001),(VANE-WRIGHT; HUMPHRIES; WILLIAMS, 1991).

$$SBP = \sum SBP_i = \sum \frac{I}{I_i} \tag{3.19}$$

where I_i is the number of nodes between the root and the species i, and I is the sum of I_i.

10. **Basic Sum of Normalized Weights (SPN)**: Indicates the normalized weight for each species, represented in Equation 3.20 (KEITH et al., 2005),(POSADAS; ESQUIVEL; CRISCI, 2001),(VANE-WRIGHT; HUMPHRIES; WILLIAMS, 1991).

$$SPN = \sum SPN_i = \sum \frac{SBP_i}{SBP_{min}} \tag{3.20}$$

where SBP_{min} represents the quotient of the minimum path from the root to the species.

11. **PDNode**: The quantitative measure of phylogenetic diversity is defined as the minimum total length of all the phylogenetic branches needed to measure a taxon in a phylogenetic tree, presented in Equation 3.21 (FAITH, 1992).

$$PD_{Node} = \sum n_i \tag{3.21}$$

where n_i represents the number of nodes i in the minimum path of each species present in diversity.

12. **PDRoot**: The phylogenetic diversity including base branches indicate a number of nodes within the maximum rooted path, shown in Equation 3.22 (RODRIGUES; GASTON, 2002).

$$PD_{Root} = \sum n_{iRoot} \tag{3.22}$$

where n_{iRoot} is the number of nodes within the path.

13. **Average Phylogenetic Diversity (DFM)**: It is the average of the quantitative measure of phylogenetic diversity, presented in Equation 3.23 (FAITH, 1992).

$$DFM = \frac{PD_{Node}}{s} \tag{3.23}$$

where s indicates the total of species.

3.3.2 Adjacent Evaluation Completed Local Binary Patterns (AECLBP)

AECLBP is a method of characterization that decomposes the local differences of the image into two complementary components: the sign (S_p) and the magnitude (m_p), shown in Equation 3.24, then two operators AECLBP_S presented in Equation 3.25, and AECLBP_M shown in Equation 3.26 are used to perform the coding (SONG; YAN, 2013).

$$S_p = S(a_p - g_c),\ m_p = |a_p - g_c|, \tag{3.24}$$

where g_c is the value of the gray scale of the center point, a_p is the average value of the evaluation window pth excluding the value from the center of the window.

$$AECLBP\_S_{P,R} = \sum_{p=0}^{P-1} S(a_p - g_c)2^P, S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.25}$$

where P points spaced equidistantly around a circle of radius R.

$$AECLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c)2^P, t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases} \tag{3.26}$$

where c is the average value of m_p of the whole image.

3.4 Feature Analysis and Selection

The analysis of the characteristics is carried out to identify the most relevant ones, this allows to reduce the amount of data that will be stored as a training source, and according to the selected method, it is possible to reduce the computational cost.

The main objective of using feature analysis in this work is to reduce the amount of data that is required to store for the creation of Meta-Data, for this purpose, selected techniques have two different approaches. The first approach uses all the extracted features to create a new set, this implies that for the new data it is still necessary to extract the entire set of original features and then perform a processing. The second approach performs an analysis of correlation between features and class to which it belongs, discarding those features that have a high correlation with each other, this approach has two advantages: 1) it allows to identify the most relevant features to identify the class, and 2) reduce the computational cost, since the features that are discarded do not need to be extracted for new data.

3.4.1 Principal Component Analysis (PCA)

The PCA has the purpose of transforming a set of variables, calls of originals, into a new set of variables called principal components. The new variables are linear combinations and are constructed according to the order of importance in terms of the total variability that they collect from the sample (JOLLIFFE, 2002).

The concept of more information is related to that of greater variability or variance. The greater the variability of the data (variance), it is considered that there is more information. That is, the greater its variance, the greater the amount of information that this component has incorporated. For this reason, the one with the highest variance is selected as the first component, while the last component is the one with the lowest variance (JOLLIFFE, 2002).

3.4.2 Correlation-based Feature Selection (CFS)

CFS is a filter algorithm that classifies feature subsets based on correlation, using a heuristic evaluation function. The bias of the function is towards subsets that contain features that are highly correlated with the class and not with each other, Equation 3.27 has an evaluation function (HALL; SMITH, 1999).

$$M_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{3.27}$$

where M_s is the heuristic value of the subset feature S, which contains k features, $\overline{r_{cf}}$ is the mean feature-class correlation ($f \in S$), finally $\overline{r_{ff}}$ is the average feature-feature inter correlation.

3.5 Performance Metrics

Since the present work uses different types of images with different classification objectives, it is necessary to use more than one performance measurement technique, because for some developed applications of Meta-Learning it is necessary to measure the success rate only between two classes, while for others, it is necessary to measure several classes, in the same way that some sets are unbalanced. For this reason, different metrics are selected that allow evaluating performance based on features of image sets used.

3.5.1 Accuracy

Accuracy refers to the degree to which the predictions made by a model match the reality that is modeled, it is applied when the test data is labeled, it can be calculated as the number of classified objects correctly divided over the total number of objects.

Table 2 shows the confusion matrix, on which Equation 3.28 is based to calculate the accuracy (SAMMUT; WEBB, 2017).

Table 2 – Confusion Matrix.

		Predicted Class	
		Positive	Negative
True Class	Positive	TP	FN
	Negative	FP	TN

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.28}$$

where TP are the true positive cases, TN are the true negative cases, FP are the false positive cases and FN are the false negative cases.

3.5.2 Sensitivity

Sensitivity represents the estimator's ability to correctly identify cases that are considered true positives, represented in Equation 3.29 (SAMMUT; WEBB, 2017).

$$Sensitivity = \frac{TP}{TP + FN} \tag{3.29}$$

3.5.3 Specificity

Specificity represents the ability of the estimator to correctly identify the cases that are considered true negatives, represented in Equation 3.30 (SAMMUT; WEBB, 2017).

$$Sensitivity = \frac{TN}{TN + FP} \tag{3.30}$$

3.5.4 Area Under Curve (AUC)

AUC is a measure of classification performance based on the area under the Receiver Operating Characteristics (ROC) curve, which relates the sensitivity to specificity in a binary classification system. The AUC allows to determine the level of success of a classifier to distinguish the elements of the classes, it is expressed in a scale from 0 to 1, where values close to 1 indicate a better performance of the classifier (SAMMUT; WEBB, 2017).

3.5.5 Precision

Precision represents the ability of the predictor to correctly identify positive cases in relation to the total positive cases predicted by the model. Precision is defined in Equation 3.31 (SAMMUT; WEBB, 2017).

$$Precision = \frac{TP}{TP + FP} \tag{3.31}$$

3.5.6 F-measure

The F-measure metric takes into account FP and FN to calculate the weighted average of Precision and Sensitivity. F-measure is used when classes are not balanced, its interpretation indicates that high values mean greater classification Precision. F-measure is represented by Equation 3.32 (POWERS, 2011).

$$\text{F-measure} = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \tag{3.32}$$

4 META-DATA CONSTRUCTION

As presented in Section 3.1.1, Meta-Data stores the experience that Meta-Learning uses as a training base, in this context, Meta-Data is the fundamental component that allows the recommendation process, for this reason, identifying and transferring the experience created when developing a methodology towards Meta-Data is a process of great importance for a system based on Meta-Learning.

The aim of this chapter is to show the process of construction and testing of Meta-Data, to fulfill this purpose a classification methodology is developed that allows identifying images that present cases of ductal carcinoma (DC) in breast tissue biopsy slides images. To test obtained Meta-Data, a new classification system based on Meta-Learning is used for selection of the best DC identification methodology. Conceptually, this task consists of two stages:

First, a methodology is developed that allows identification of DC in breast tissue biopsy slides images, the stage that makes up the methodology, as well as the technique and parameters used are considered experience and form Meta-Data.

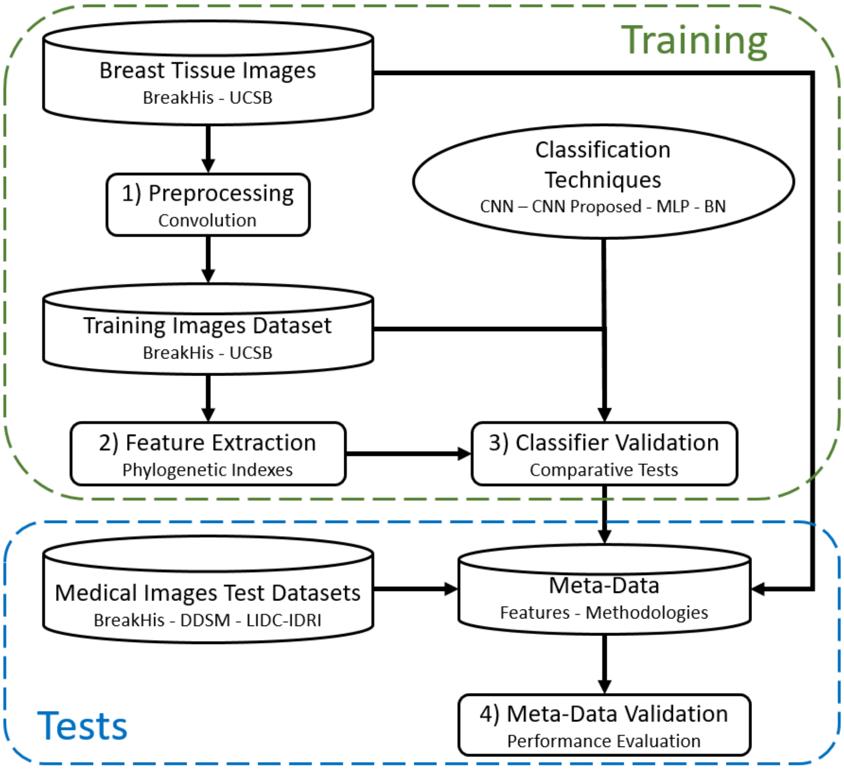
The second stage, demonstrates the effectiveness of Meta-Data in a classification system based on Meta-Learning. To perform this task, two new datasets with different types of images are added to Meta-Data created in first stage, the classification system is tested using a new set of test images. It is important that Meta-Data contain dataset information with different types of images and their respective processing methodologies to test Meta-Model; otherwise, it would not be possible to verify its performance. When executing the tests, images whose best methodology is already known are used, in this way it is possible to measure the performance of Meta-Model and Meta-Data.

The chapter is organized by sections, in Section 4.1 is presented the datasets used, in Section 4.2 the proposed methodology is presented, Section 4.3 shows the results, finally in Section 4.4 presents final considerations. Figure 5 presents the proposed methodology, including datasets used.

4.1 Datasets

In this chapter, four medical image datasets are used, the first two contain breast tissue biopsy images and are used as object of study for construction of a methodology

Figure 5 – Proposed Methodology for Meta-Data Construction.



Source: Prepared by author

to identify DC. The last two datasets contain different types of medical images and are used as a test object to determine the effectiveness of Meta-Data. Initially, the images are used maintaining the original parameters with respect to size, color channels, format and histogram, only the modifications described in the proposed methodology are made.

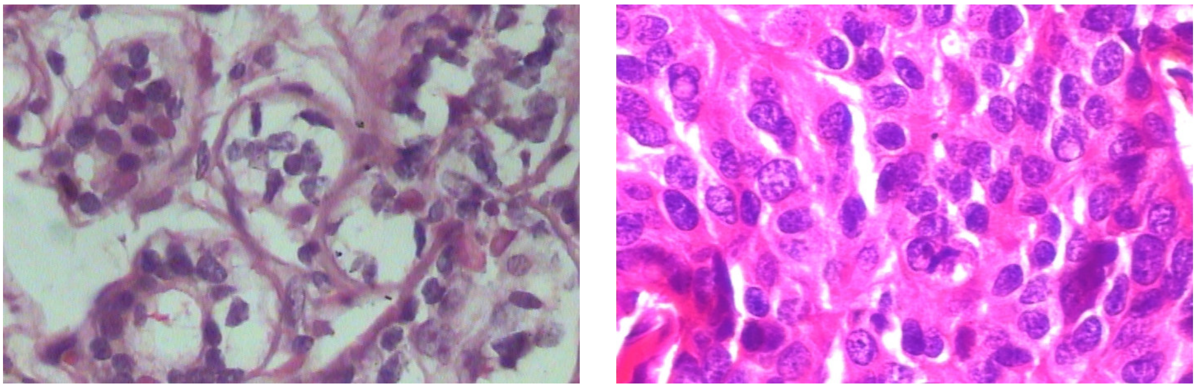
The **BreakHis dataset** consists of 7909 biopsy images of breast tissue divided into two groups, benign and malignant cases, the images were acquired using four magnification measures 40X, 100X, 200X and 400X. The dataset has eight classes that correspond to the histological subtypes. Histologically benign is a term referring to a lesion that does not match any criteria of malignancy, e.g., marked cellular atypia, mitosis, disruption of basement membranes, metastases. Normally, benign tumors present a slow growing and remains localized. Malignant tumor is a synonym for cancer, it is a lesion that can invade and destroy adjacent structures (locally invasive) and spread to distant sites (metastases) to cause death (SPANHOL et al., 2016).

The **UCSB Bio-Segmentation Benchmark dataset** consisted of 162 breast cancer slide images scanned at 40X. 277 524 patches of size 50x50 pixels were extracted where 78 786 presented cases of Invasive DC and 198 738 cases of negative Invasive DC.

The Invasive DC is the most common subtype of breast cancer. The pathologists usually focus on the areas affected by the Invasive DC to determine the degree of aggressiveness, which makes its identification and classification an important task (GELASCA et al., 2008).

Figure 6 present a sample of breast tissue datasets, which are used as object of study to construct the Meta-Data, in order to develop a methodology that allows the identification of DC in biopsy images of breast tissue. However, to verify the effectiveness of Meta-Data, it is necessary to have new datasets that contain different types of images, for this reason two new datasets are used.

Figure 6 – Breast Tissue Example.



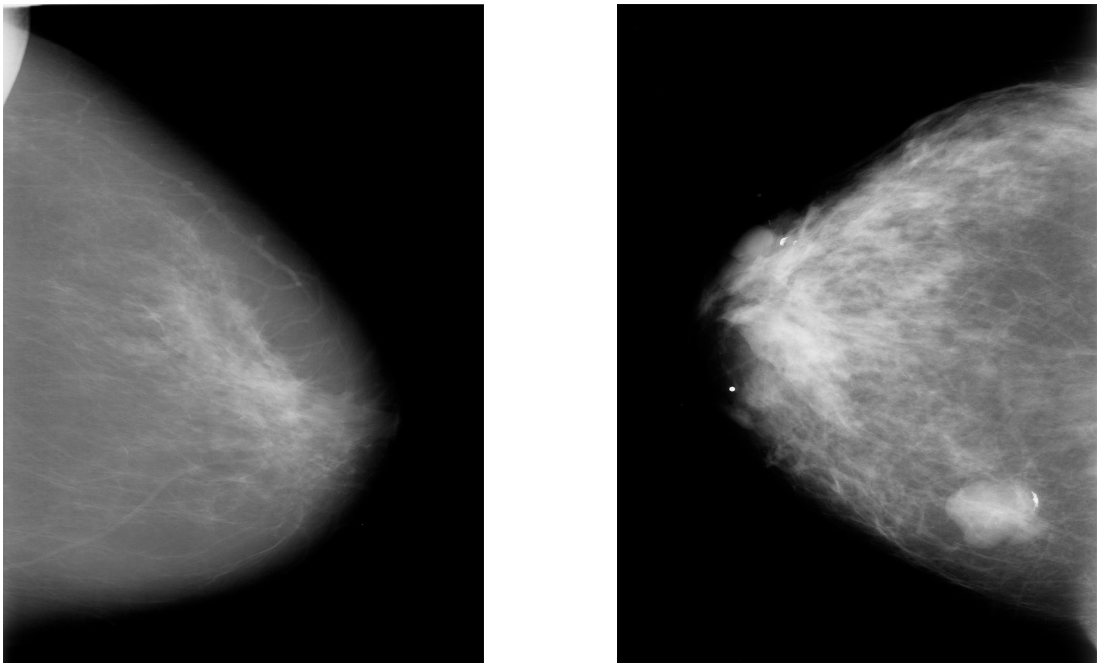
(a) *Benign Case* (b) *Malignant Case*

Source: BreakHis (SPANHOL et al., 2016)

The **Digital Database for Screening Mammography (DDSM)** consists of scanned film mammography studies with 2620 images, with diagnostic annotations of calcification or mass, and include segmentation of the regions of interest. The DDSM was developed through a grant from the DOD Breast Cancer Research Program, US Army Research and Material Command and comes from data collected at Massachusetts General Hospital, Washington University School of Medicine, Sacred Heart Hospital and Wake Forest University School of Medicine (LEE et al., 2017),(HEATH et al., 1998), Figure 7 shows an example of dataset cases.

The **Lung Image Database Consortium (LIDC-IDRI)** is made up of a set of computer tomography (CT) images that includes diagnostic label, segmenting and describing lesions for lung cancer detection. Computer assisted diagnostic methods can use it as a resource for development, evaluation and training. Initiated by the National Cancer Institute, most advanced by the Foundation for the National Institutes of Health

Figure 7 – Mammography Example.

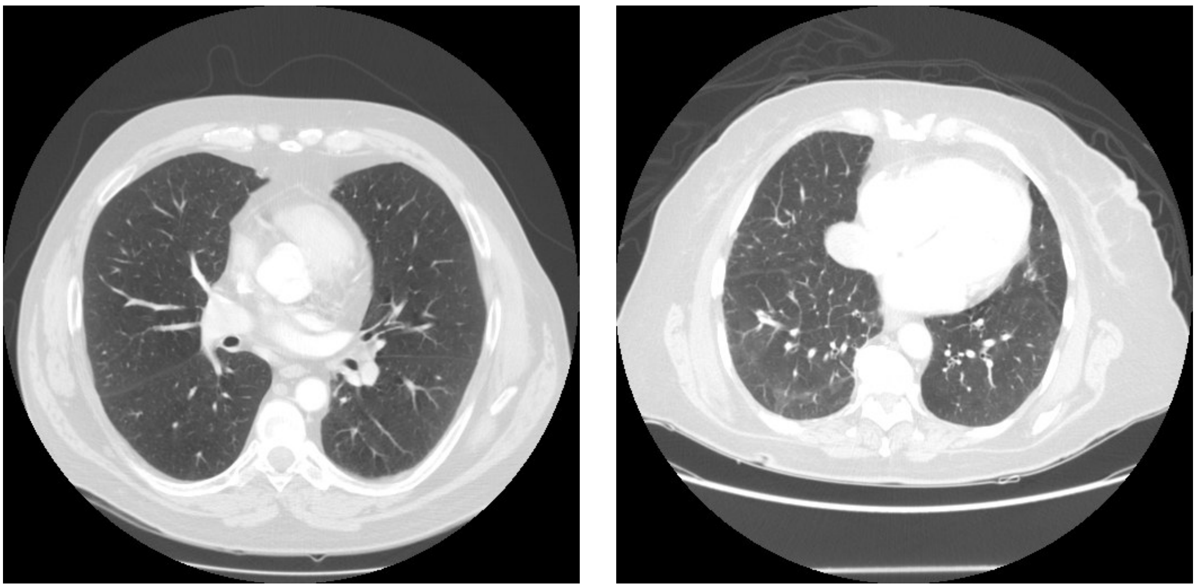


(a) *Calcification Case* (b) *Mass Case*

Source:DDSM (LEE et al., 2017),(HEATH et al., 1998)

and accompanied by the Food and Drug Administration. LIDC-IDRI consists of 1018 cases that includes XML files and images. The dataset contains 7371 lesions identified as nodule, the annotations were made by four thoracic radiologists (ARMATO et al., 2011). Figure 8 shows an example of CT images.

Figure 8 – Lung CT Example.



(a) *Non Malignant Case* (b) *Non Small Cell Lung Cancer*

Source: LIDC-IDRI (ARMATO et al., 2011)

4.2 Proposed Methodology

The proposed methodology consists of four steps: 1) **Preprocessing**, for elements contained in **Breast Tissue Image** dataset, a convolution process is applied to improve the classification, result of this step is **Training Images Dataset**. 2) **Feature Extraction**, phylogenetic indices are applied as image characterization method. 3) **Classifier Validation**, in this step, four **Classification Techniques** are executed and compared, using **Training Image Dataset** as input data for two CNN and features extracted in previous step for MLP and NB techniques. When comparing performance based on evaluation measures, the best technique is selected, along with its data characterization technique, this information is stored in **Meta-Data**. 4) **Meta-Data Validation**, a validation of Meta-Data used Meta-Learning is carried out, to identify the most suitable classification methodology for identification of DC in biopsy images of breast tissue slides. Before starting this step, **Medical Images Test Dataset** is included to **Meta-Data**, adding two different kinds of medical images. Figure 5 shows proposed methodology.

4.2.1 Preprocessing

The aim of this step is to highlight the features of each type of image to allow classification, for this task convolution technique is used to obtain different transformations of original images.

The convolution is selected as a transformation technique given its ability to perform different transformations of images by altering only the convolution Kernel used, filters applied are selected given their application in literature. Convolution process is performed on each image using the intensity information of each pixel, applying the kernel to each color channel.

The Kernel found constitute a valuable experience and are stored in **Meta-Data**.

4.2.2 Feature Extraction

Feature extraction is performed using the Phylogenetic Indexes q, RTD, EQI, EQE, AvTD, TTD, WMD and MNND, represented in Equation 3.11 to Equation 3.18 presented in Section 3.3.1.

The feature extraction process is used only for those techniques that do not use images themselves as input data, this means that characterization through phylogenetic indexes is not used for CNN.

4.2.3 Classifier Validation

In this step, four **classification techniques** are tested and compared to determine the most appropriate for the identification of DC. Validation requires the creation of training and testing sets. To build training set, 100% of the images from the BreakHis dataset were used, however, this dataset contains four different histological subtypes classified as malignant, but this study only requires cases with DC diagnosis, for this reason images of malignant diagnosis that do not correspond to DC are removed. Additionally, 70% of images from UCSB, randomly selected, are added. This initial training dataset contains a large number of images, which leads to a high processing time, to face this difficulty, tests are carried out to determine the amount of images necessary for training, using as evaluation criteria, the stabilization of the classification performance, which implies that the tests are carried out with an initial number of images and the classification performance is analyzed, the test is repeated increasing the number of images used and comparing the new performance, ending the test when the classification performance does not change. One hundred images are used for the initial test and the same amount is increased with each iteration.

The **classification techniques** were selected according to their use in the literature, in addition, a CNN is proposed to obtain information about the effect caused by the number of layers:

1. LeNet Network that is reported as the most famous convolutional network, this includes two layers of convolution, the first layer consisting of twenty neurons with a 5x5 core and a ReLU activation function, the second layer uses fifty neurons and configuration similar to the first, then, two subsampling operations that employ a 2x2 core. Finally, two internal product operations are performed before entering the classification layer (DRISS et al., 2017).

2. The proposed CNN is designed to determine if more than one convolution layer is necessary to perform the classification correctly, the proposed configuration consists of a convolution layer with 20 neurons and 3x3 kernel, a 2x2 subsampling layer with ReLU activation, finally a classification layer that uses a SoftMax function that calculates the probability of the image belonging to each class.

3. MLP presented in Section 3.2.8, the parameters used for processing are those defined as standard: one hundred instances, learningRate 0.3, momentum 0.2, two decimals, five hundred times, number of hidden layers equal to half the sum of the attributes and classes, limit of validation twenty and number of epochs to train through 3000.

4. BN presented in Section 3.2.1, the parameters used for processing are: one hundred instances, two decimals and without debugging.

At the end of this step, the methodology that presents the best classification performance is selected as experience and is stored in the **Meta-Data**, methodology includes preprocessing, classification and characterization techniques.

4.2.4 Meta-Data Validation

Validation of **Meta-Data** is a process that involves the creation of a new classification system, which implies a new classification technique and a new set of training and tests. This step demonstrates that **Meta-Data** created and the features extracted from all types of images used allow to identify for each type of image its best classification methodology.

As a Meta-Learning classification methodology, the CNN LeNet are selected given its ability to extract features from images and perform the classification process without the help of additional methods, using CNN implies that the **Meta-Data** must include training images, for this reason 70% of elements coming from the datasets with **Breast Tissue Images** are randomly selected. These images are linked to the **Meta-Data** obtained in the **Classifier Validation** step (Section 4.2.3).

Testing the **Meta-Data** of the developed methodology requires having **Meta-Data** for different types of images, otherwise it would be using a classification system trained with a single type of class, for this reason 70% of images, randomly selected, from DDSM and LIDC-IDRI datasets are loaded, which make up the **Medical Images Test Dataset**. Images belonging to DDSM dataset are associated with a classification methodology named as methodology 1 and the images from the LIDC-IDRI dataset are associated with a classification methodology named as methodology 2, this assignment is made because to determine the best classification methodologies for DDSM and LIDC-IDRI images is outside the scope of this chapter. All training images together with their classification methodology are added to **Meta-Data**, which finally consists of three different types of images that constitute three classes.

Validation of **Meta-Data** is done using a set of test images, made up of 30% of BreakHis, DDSM and LIDC-IDRI datasets, taking precautions so that no image of the test set is part of training set at the same time. The aim of this validation is that Meta-Learning can recommend as a classification methodology for images coming from BreakHis the methodology selected in Section 4.2.3.

4.3 Results

In this step, the performance of **classification techniques** is evaluated and compared. The processing of data and applications of methods are carried out using Waikato Environment for Knowledge Analysis (WEKA) (WITTEN et al., 2016). This tool was selected, since allows to use various Machine Learning and Deep Learning techniques, in the same way, it allows the analysis of the extracted features and the configuration of different test parameters, all these characteristics allow to perform the necessary tests, in addition, WEKA proved to be easy to install.

As described for the **Preprocessing** step in Section 4.2.1, different convolution kernels were used, creating multiple data sets that were tested separately with each classification technique described in Section 4.2.3. The best convolution kernel found is presented in Table 3, Figure 9 shows the result of applying the kernel in an example image.

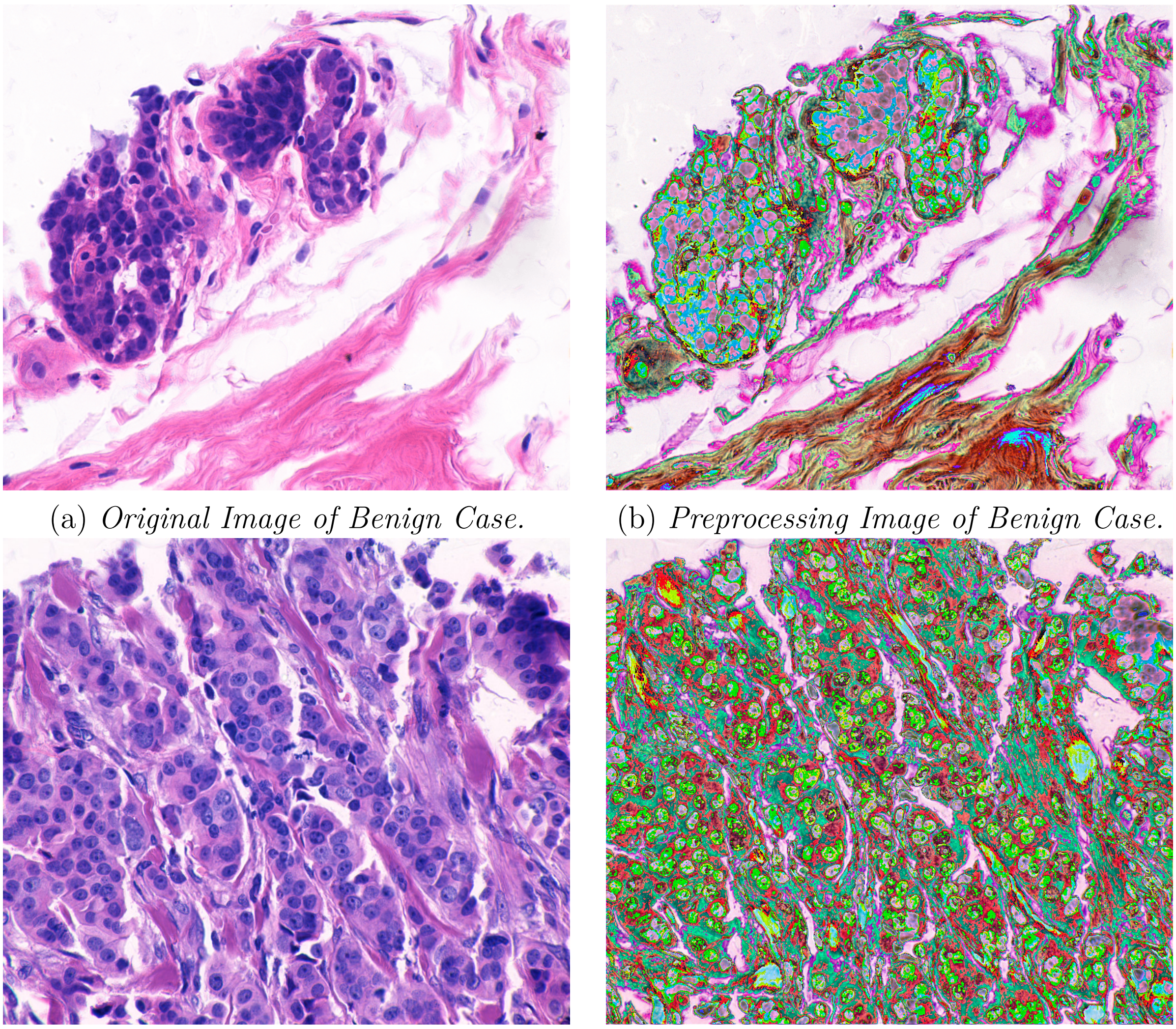
Table 3 – Kernel of Convolution.

-0.25	-0.25	-0.25
-0.25	3	-0.25
-0.25	-0.25	-0.25

The results of tests in Section 4.2.3 to determine the size of training set from the UCSB dataset, showed that it is necessary to use 0.4% of the images, threshold reached in tenth iteration. In total, training set consists of 6931 images, 2980 labeled as benign and 3951 as malignant.

Table 4 shows the average of five results of **Classifier Validation** step (Section 4.2.3). Accuracy, AUC, sensitivity, specificity and cost of processing time are used as performance evaluation measures for each classification technique.

Figure 9 – Preprocessing Example.



(a) *Original Image of Benign Case.* (b) *Preprocessing Image of Benign Case.*

(c) *Original Image of DC Case.* (d) *Preprocessing Image of DC Case.*

Source: Prepared by author

Table 4 – Average Results with Preprocessing.

Technique	Accuracy %	AUC %	Sensitivity %	Specificity %	Processing Time/sec
LeNet	**0.93**	0.97	0.90	**0.95**	2217.75
Proposal CNN	**0.93**	**0.98**	**0.98**	0.87	1307.78
MLP	0.72	0.75	0.68	0.71	33.14
BN	0.76	0.82	0.66	0.84	**1.33**

Analyzing the results of Table 4, the proposal technique presents a better performance with respect to AUC, sensitivity and processing time, nevertheless the LeNet presents greater specificity, which implies a lower rate of false positives.

When comparing the results with and without **Preprocessing**, the performance of the CNN-based methods increases when using **Preprocessing**, while the others techniques suffer variations without definite trend.

When comparing the processing times, it is evident that Machine Learning based techniques require less than 1% of the time used by CNN techniques, in the same way, it is proved that when increasing a convolution layer, the processing time increases by 58.9%.

The results show that two of analyzed technique have a better performance: LeNet and Proposal. It is possible to conclude that the use of a single convolution layer and a smaller Kernel increases AUC and sensitivity, and reduces processing time and specificity. Taking into account the relationship between sensitivity and specificity, LeNet presents a better balance, reducing false negatives, for this reason and considering the aim of this work, LeNet is identified as the best technique for identification of Ductal Carcinoma.

Table 5 shows the results of the state of the art and the proposed methodology. When making the comparison, it is possible to find:

Table 5 – Results of the State of Art and Proposed Methodology for DC Identification.

Methodology	**Accuracy %**	**AUC %**	**Sensitivity %**	**Specificity %**
Spanhol et al. (2016)	0.85	0.86	-	-
Araujo et al. (2017)	0.78	-	0.81	-
Mercan et al. (2018)	**0.94**	-	0.88	0.78
Selected (LeNet)	0.93	**0.97**	**0.90**	**0.95**

- The proposed methodology presents a better accuracy in comparison with Spanhol et al. (2016), for the identification of ductal carcinoma and benign cases.

- When comparing the results with Araujo et al. (2017), the proposed work obtains a better performance in sensitivity and accuracy in identification of two classes. it is not possible to compare the results for the identification of four classes, since it is outside the scope of this work.

- For Mercan et al. (2018), analysis of results must be done taking into account the accuracy, sensitivity and specificity, when comparing all results the proposal methodology presents a better balance, however, it is clarified that the present work does not perform segmentation of image in components and similarly does not perform classification by analysis of each component.

When determining the appropriate methodology for identification of DC, **Training** process presented in Figure 5 is completed, all collected experience is stored in **Meta-Data** and consists of three elements:

1. Information contained in Kernel of Convolution (Table 3).
2. CNN that presents a better performance according to Results with Preprocessing (Table 4) and its parameter settings shown in Section 4.2.3.
3. Given that selected method is based on CNN, the **Training Images Dataset** used are stored.

Meta-Data contains all experience gained in developing a methodology for DC detection. The next step is to validate **Meta-Data**, for this the **Medical Images Test Datasets** and their classification methodologies are added to **Meta-Data** as explained in **Meta-Data Validation** step (Section 4.2.4).

Validation consists in verifying the performance of a classifier based on Meta-Learning to identify the most appropriate classification methodology for each type of test image, based on information contained in **Meta-Data**. To perform this task, Meta-Data is used as training for the Meta-Leaning-based classifier, which uses CNN LeNet, this model is provided with the test base. Results present 99.6% accuracy, 99.9% AUC and 99.7% F-measure, these results demonstrate that **Meta-Data** constructed for identification of Ductal Carcinoma in breast tissue biopsy slides images are effective, and completing the **Tests** presented in Figure 5.

4.4 Final Considerations

In order to create a **Meta-Data** that contains the methodology for identification of DC in medical images, two public dataset were used, a **Preprocessing** process was carried out with each dataset and four **classification techniques** are evaluated, in the same way **Meta-Data** obtained was evaluated using two dataset with different types of medical images. Based on the results of this work, it is concluded:

1. The proposed methodology presents an improvement in identification of DC in images from breast tissue biopsy slides with respect to methods found in state of the art.

2. The use of convolution kernel for **Preprocessing** of biopsy images of breast tissue proves to be efficient for classification of malignant and benign cases.

3. The **Meta-Data** constructed for identification of DC is efficient to be used in recommendation systems based on Meta-Learning.

When comparing the performance of the different **classification techniques** tested in this work, it is possible to affirm that two CNN-based methods have a better performance, however it is necessary to consider the implications of the specificity and sensitivity results before selecting the best one.

The main contributions of methodology proposed in this chapter are:

1. Development of a methodology that presents a better performance than those found in state of the art for identification of DC in biopsy images of breast tissue.

2. Construction of Meta-Data for Meta-Learning that allows to determine the best methodology for identification of DC in biopsy images of breast tissue.

Information contained in this chapter is presented in article "Meta-Data Construction for Selection of Breast Tissue Biopsy Slides Image Classifier to Identify Ductal Carcinoma", accepted for Brazilian Conference on Intelligent Systems (BRACIS 2019); 15 a 18 de octubro de 2019; Salvador - Bahia - Brazil.

5 META-MODEL CONSTRUCTION

In Chapter 4 a methodology was developed for identification of a specific type of tumor in breast biopsy images in order to build Meta-Data, this chapter, in addition to building Meta-Data for different types of medical images with various classification objectives, shows the construction of Meta-Model, using techniques based on Machine Learning. The Meta-Model, as presented in Section 3.1.2, is responsible for making the methodology recommendation based on Meta-Data and feature of data under analysis.

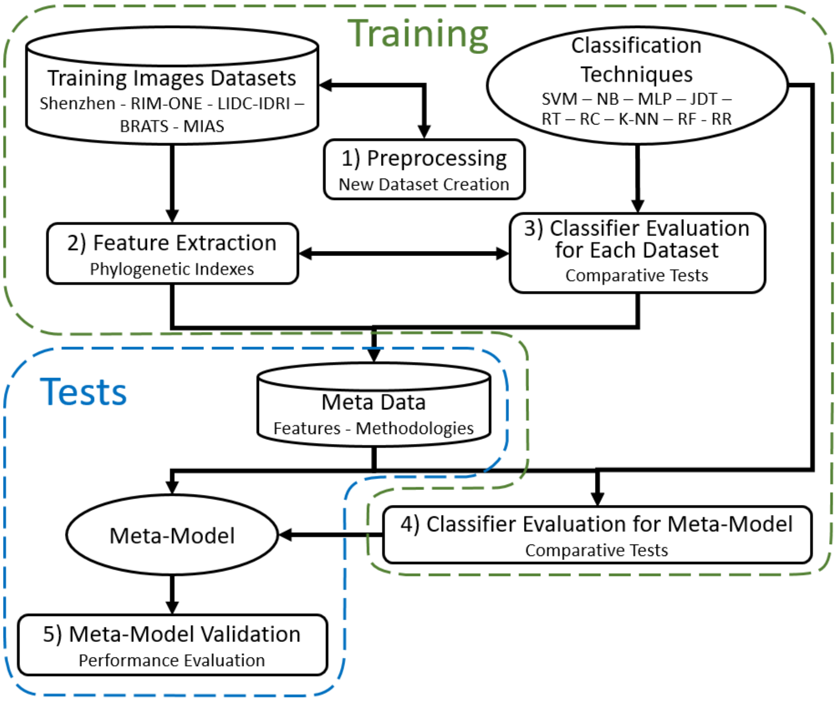
The aim of this chapter is to develop a Meta-Model that allows to identify the most adequate methodology of classification for different types of medical images, through the application of Machine Learning and Meta-Learning. Six datasets containing five different types of medical images are used, for each type of image, a methodology is developed that classifies according to a specific objective, which is unique for each type of image. Finally, all datasets are used to develop a system based on Meta-Learning, with the common aim of identifying the best classification techniques for each dataset using the features extracted from the images.

The chapter is organized by sections, in Section 5.1 is presented the datasets used, in Section 5.2 the proposed methodology is presented, Section 5.3 shows the results, finally in Section 5.4 presents final considerations. Figure 10 presents the proposed methodology, including datasets used.

5.1 Datasets

This section describes six datasets, as well as the specific classification objective for each type of medical image: For chest X-ray images the manifestation of Tuberculosis is identified. In retinal fundus images, those that present Glaucoma is recognized. For CT scan images the presence of Large cell carcinoma, Squamous cell carcinoma, Adenocarcinoma, Adenocarcinoma of mutation negative is identify. For Brain Magnetic Resonance Imaging (MRI), the presence the tumours is recognized. Finally, in images of mammography the classification is made based on the presence of tumours and whether they are benign or malignant. Initially, the images are used maintaining the original parameters with respect to size, color channels, format and histogram, only the modifications described in the proposed methodology are made.

Figure 10 – Proposed Methodology for Meta-Model Classifier.



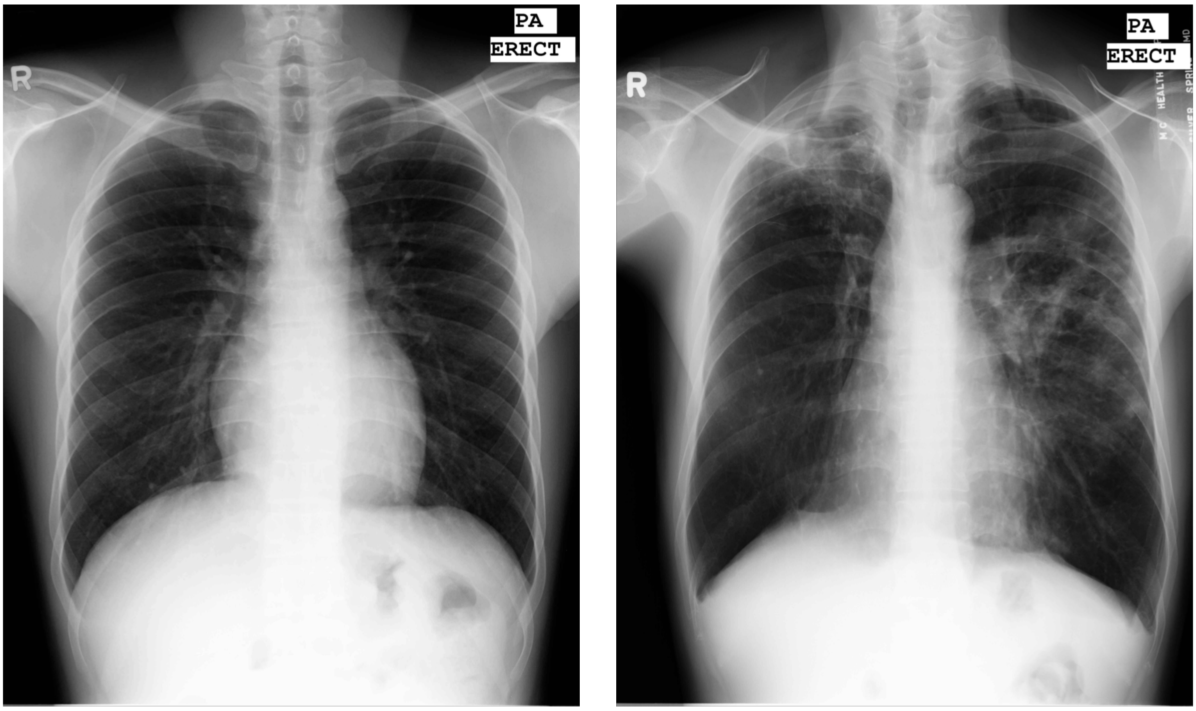
Source: Prepared by author

The Shenzhen set - Chest X-ray Database. This work uses the images available dataset from National Library of Medicine, National Institutes of Health, Bethesda, MD, USA and Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China (JAEGER et al., 2014),(CANDEMIR et al., 2014). The Chest X-rays are from outpatient clinics and were captured as part of the daily routine using Philips DR Digital Diagnose systems. There are 662 images, 336 cases present tuberculosis manifestation and 326 are normal cases, image size varies for each X-ray, it is approximately 3k x 3k.

Montgomery County - Chest X-ray Database. It is the second set of images used, available from National Library of Medicine, The National Institutes of Health, Bethesda, MD, USA (JAEGER et al., 2014),(CANDEMIR et al., 2014). The set contains data from X-rays collected under Montgomery County's Tuberculosis screening program. There are 138 images, 58 cases present tuberculosis manifestation and 80 are normal cases, matrix size is 4020 x 4892, or 4892 x 4020, the pixel spacing in vertical and horizontal directions is 0.0875 mm and number of gray levels is 12 bits. Figure 11 presents an example of Chest X-ray image datasets.

RIM-ONE-db-r2 is an open retinal fundus image database with accurate gold standards of the optic nerve head provided by different experts. It includes images from

Figure 11 – Chest X-ray Dataset Example.



(a) *Normal Case* (b) *Tuberculosis Manifestation*

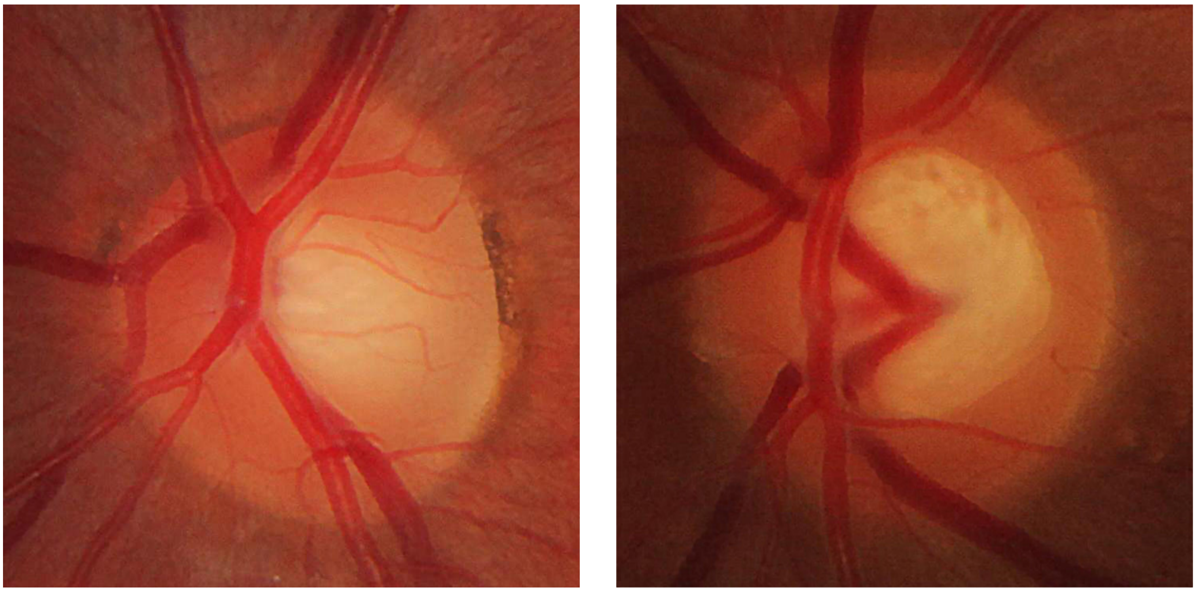
Source: Montgomery Database (JAEGER et al., 2014),(CANDEMIR et al., 2014)

healthy eyes as well as images from eyes with glaucoma at different stages. A variability measurement by zones of the optic disc is also proposed for the purpose of validation. Three hospitals have contributed to the development of this database: Hospital Universitario de Canarias, Hospital Clínico San Carlos and Hospital Universitario Miguel Servet (ALAYON et al., 2013),(PENA-BETANCOR et al., 2015). There are 455 images, these present 200 cases of glaucoma and 255 normal cases. Figure 12 presents an example of RIM-ONE datasets.

LIDC-IDRI The characteristics of dataset are described in Section 4.1.

The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), database used for the brain tumor segmentation challenge that provides multimodal scans: native (T1), post-contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. Dataset were acquired with different clinical protocols and various scanners from 19 institutions and their annotations were approved by experienced neuro-radiologists, annotations comprise the GD-enhancing tumor, the peritumoral edema, and the necrotic and non-enhancing tumor core (MENZE et al., 2015),(BAKAS et al., 2017). From this dataset, 4520 images were taken with 2445 cases

Figure 12 – Retinal Fundus Images Examples.

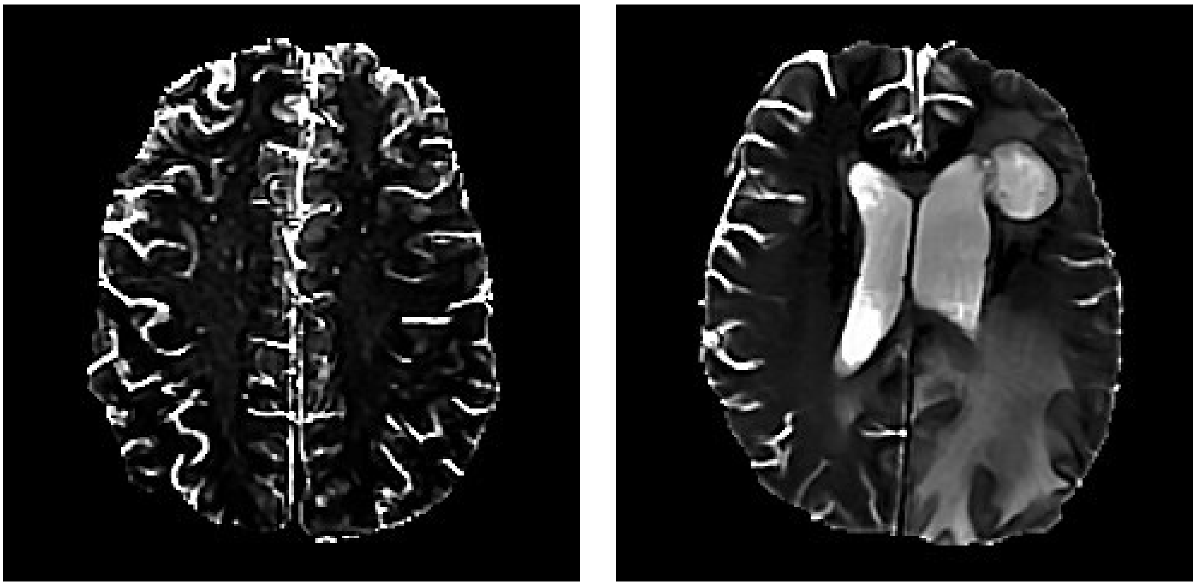


(a) *Normal Case* (b) *Glaucoma or Glaucoma Suspicious*

Source: RIM-ONE Dataset (ALAYON et al., 2013),(PENA-BETANCOR et al., 2015)

of brain tumor and 2072 normal cases. Figure 13 presents an example of BRATS datasets.

Figure 13 – Brain MRI Examples.



(a) *Normal Case* (b) *Tumor case*

Source: BRATS Dataset (MENZE et al., 2015),(BAKAS et al., 2017)

MIAS MiniMammographic Database, the original MIAS Database (digitised at 50 micron pixel edge) has been reduced to 200 micron pixel edge and clipped/padded so that every image is 1024 x 1024 pixels. Presents classification information that includes character of background tissue, class and severity of abnormality (SUCKLING; PARKER;

DANCE, 1994). There are 330 images, these present 67 cases with severity benign, 54 cases with severity malignant and 209 normal cases. An example of mammography can be seen in Figure 7 of Section 4.1.

5.2 Proposed Methodology

This chapter has five steps: 1) **Preprocessing**, this step is carried out in order to increase the number of available images, through the inclusion of noise and use of available segmentation information in **Training Image Datasets**. 2) **Feature Extraction**, information of each element belonging to datasets are extracted, building the data that is used in the processing. 3) **Classifier Evaluation for Each Dataset**, this step uses the extracted features to evaluate the performance of different **classification technique**, each dataset is evaluated independently, in order to identify the best methodology for each type of medical image, this information is part of the **Meta-Data**. 4) **Classifier Evaluation for Meta-Model**, the **classification techniques** that presents a better performance for **Meta-Model** is selected, using a new extraction of features and information of the best **classification techniques** for each type of medical image. 5) **Meta-Model Validation**, the **Meta-Model** is tested using a new set of test images. Each step of proposed methodology can be observed in Figure 10.

5.2.1 Preprocessing

In this step, new datasets are created from images of Shenzhen, Montgomery, RIM-ONE and BRATS datasets, the **preprocessing** is done in order to increase the amount of images available for **Meta-Model** training and use of additional information available, which allow segmentation of organs for Montgomery and separation of image according to scanning mode for BRATS. Increasing the number of available samples and adding noise to images, allows to increase the training set used by Machine Learning techniques and determine the level of susceptibility of Meta-Model classifier, establishing if it has the capacity to extract the features of small changes that imply a different classification methodology.

For Shenzhen, Montgomery and RIM-ONE dataset a first **preprocessing** is performed by adding Gaussian noise, using a 21x21 filter matrix, resulting in a new dataset that preserves the original dimensions. A second **preprocessing** applies salt and pepper noise. Additionally for Montgomery, the segmentation information provided in the original dataset is used to create two new datasets, with black and white backgrounds.

For BRATS dataset a **preprocessing** is performed to separate the images in each one of the scanning modes: T1, T1Gd, T2 and FLAIR, producing four new datasets. Table 6 summarizes the preprocessing performed and lists the resulting datasets.

Table 6 – List of Dataset After Preprocessing.

Preprocessing Type	**Dataset**	**Number of Dataset**
Gaussian noise	Shenzhen, Montgomery and RIM-ONE	3
Salt and peper noise	Shenzhen, Montgomery and RIM-ONE	3
Segmentation	Montgomery	2
Separate for Scan Mode	BRATS	4
	Originals Datasets	6
	Total Datasets	18

5.2.2 Feature Extraction

The features refer to the data that is extracted from images to represent them and are used to perform processing. Since the proposed methodology involves individually determining the best Machine Learning technique to classify each of five types of images under study, the features used may be different for each type of medical image. However, **Feature extraction** is also carried out for creation of **Meta-Data** that contains experience and training set for **Meta-Model**, since all types of images are combined into a single set, the features extracted for **Meta-Data** must be the same.

The features used are based on phylogenetic indexes described in Section 3.3.1, index q, RTD, EQI, EQE, AvTD, TTD, WMD and MNND, represented in Equation 3.11 to Equation 3.18 are used for all types of images, however, index SBP, SPN, PDNode, PDRoot and DFM, represented in Equation 3.19 to Equation 3.23 are only used in LIDC-IDRI. The reason for using different features is to demonstrate that for different types of images, it is not necessary to extract the same set of features.

5.2.3 Classifier Evaluation for Each Dataset

In this step, each **classification techniques** is applied to each dataset separately; the performance is evaluated, taking as performance metrics: accuracy, AUC and processing time. In the end, the technique that presents the best result with respect to performance metrics is selected, which implies that each dataset can have a different classifier; this information is considered experience and constitutes **Meta-Data**, .

An analysis of features is carried out in order to determine which ones to make the greatest contribution to classification and eliminate the least significant ones using the techniques PCA and CFS presented in Section 3.4. This information determines the data that is stored in **Meta-Data** as a training source.

Different techniques of classification are considered, discarding those that did not require the previous application of techniques of characterization and selecting those found more frequently in the state of the art, the parameters used are those defined as standard:

1. SVM presented in Section 3.2.14, the parameters used for processing are: loss:0.1, kernel type: radial function, termination criteria tolerance: 0.001 and number of decimals: 2.

2. NB presented in Section 3.2.9, the parameters used for processing are: one hundred instances, two decimals and without debugging.

3. MLP presented in Section 3.2.8, the parameters used for processing are: one hundred instances, learningRate 0.3, momentum 0.2, two decimals, five hundred times, number of hidden layers equal to half the sum of the attributes and classes and limit of validation twenty.

4. JDT presented in Section 3.2.5, the parameters used for processing are: confidence factor for pruning:0.25, new value of the distribution of the minority class: 50, number of samples: 99, minimum number of instances per leaf: 2, number of decimals: 2.

5. RT presented in Section 3.2.13, the parameters used for processing are: maximum depth: unlimited, minimum total weight of the instances in a leaf: 1, minimum proportion of the variance on all the data: 0.001, number of decimals: 2.

6. RC presented in Section 3.2.10, the parameters used for processing are: classifier: RT, number of iterations: 10, number of decimals: 2.

7. K-NN presented in Section 3.2.6, the parameters used for processing are: search algorithm: linear search, KNN: 11, number of decimals: 2.

8. RF presented in Section 3.2.11, the parameters used for processing are: maximum depth: unlimited, number of output decimals:2, number of iterations: 100.

9. RR presented in Section 3.2.12, the parameters used for processing are: alpha for partial reductions: 0.5, Confidence level to consider two neighboring intervals as different: 0.9, minimal frequency of decision class in each interval: 6, minimal number of intervals for each attribute: 3, number of intervals for each attribute: 5, number of decimals: 2.

5.2.4 Classifier Evaluation for Meta-Model

In this section, the **classification technique** used for **Meta-Model** is selected, to achieve this goal it is necessary to use a single **feature extraction** technique, which will be applied to all dataset, unlike the previous section in which each dataset can use a different **feature extraction** technique.

The **Meta-Data** is processed using the **classification techniques** presented in Section 5.2.3. Taking into account accuracy, AUC and processing time, the performance of each technique is evaluated and compared, selecting the one that presents the best performance as a **Meta-Model** classification technique.

The result of feature analysis determines the characterization that will be carried out on the new data using the proposed methodology based on Meta-Learning to determine the most appropriate **classification techniques** for each type of medical image. The accuracy, AUC and processing time of each **classification techniques** is analyzed and compared, as well as the amount of data eliminated by the features analysis techniques and then the **classification techniques** that performs best are selected.

5.2.5 Meta-Model Validation

In this section, the validation of developed methodology is carried out, for this purpose it has been randomly reserved 30% of images of each dataset. However, the class labelling information is retained to perform the verification of results.

Validation begins by extracting the most representative features, which are identified in the features analysis of Section 5.2.4, these data along with the class labelling information are the test dataset. Then, the classification technique selected for **Meta-Model** performs the classification of test set based on **Meta-Data** and assigns a class label, comparing with the original verification information. Finally, the counting of correctly classified data is carried out and presents the accuracy achieved by the methodology developed.

5.3 Result

In this section, the performance of **classification techniques** for each dataset and for **Meta-Model** are evaluated and compared, furthermore, the results of **Meta-Model Validation** step are presented. Data processing was carried out using a 64-bit operating system computer, with 2.8 Ghz Core i7 PC (16 GB RAM), and Nvidia GeForce GTX 1050 GPU (2 GB RAM).

Three software tools were developed, two for creation of new datasets as a result of **Preprocessing** step and one for **Feature Extraction** step; all tools are created using C++ language and OpenCV free library. The processing of data and applications of techniques based on Machine Learning are carried out using Waikato Environment for Knowledge Analysis (WEKA) (WITTEN et al., 2016). This tool was selected, since allows to use various Machine Learning and Deep Learning techniques, in the same way, it allows the analysis of the extracted features and the configuration of different test parameters, all these characteristics allow to perform the necessary tests, in addition, WEKA proved to be easy to install.

In **Feature Extraction** step the 70% of images of each of eighteen datasets shown in Table 6 are randomly selected to form training set, in total 17627 medical images. For each dataset is applied the process of **feature extraction** based in Section 5.2.2, for example, for the Shenzhen set - Chest X-ray dataset, eight features are extracted that produced 3696 registers, representing 462 images, 70% of dataset.

In **Classifier Evaluation for Each Dataset** step, the aim of classification process does not have to be the same, in the same way, that images can come from different capture systems and represent different organs. Each one of nine **classification techniques** described in Section 5.2.3 are applied independently to each dataset, Table 7 presents the average of five results of training for Shenzhen set - Chest X-ray dataset. Accuracy, AUC and processing time are used as parameters for performance evaluation using all extracted features.

For the Shenzhen set - Chest X-ray dataset, the feature analysis is applied on

Table 7 – Comparison of Average Performance Training Results of Classification Techniques Applied to Shenzhen set - Chest X-ray Dataset.

Dataset	**Classifiers**	**Accuracy %**	**AUC %**	**Processing time/sec**
The Shenzhen set - Chest X-ray Dataset	NB	0.565	0.642	0.008
	SVM	1	1	0.174
	MLP	0.701	0.725	0.422
	JDT	0.714	0.75	0.012
	RT	**1**	**1**	**0.005**
	RC	1	1	0.030
	K-NN	1	1	1.596
	RF	1	1	0.116
	RR	1	1	0.108

eight extracted features as described in Section 5.2.2. Original dataset contains 4620 registers, after applying the feature analysis techniques, the size of Shenzhen dataset is reduced to 1386 register for PCA and 924 for CFS. The new feature sets are tested individually with the **classification techniques** that present a better performance, for Table 7, the test is performed with the SVM, RT, RC, K-NN, RF and RR. Table 8 shows the average of five results obtained.

Table 8 – Comparison of Average Performance Training Results of Classification Techniques Applied to Most Relevant Features for Shenzhen set - Chest X-ray Dataset.

Feature Analysis	**Classifiers**	**Accuracy %**	**AUC %**	**Processing time/sec**
CFS	SVM	1	1	0.180
	RT	**1**	**1**	**0.008**
	RC	1	1	0.032
	K-NN	1	1	7.494
	RF	1	1	0.092
	RR	1	1	0.068
PCA	SVM	0.697	0.696	0.034
	RT	**1**	**1**	**0.001**
	RC	1	1	0.012
	K-NN	1	1	8.100
	RF	1	1	0.092
	RR	1	1	0.038

The identification of most relevant features for each dataset is an important part of **Meta-Data**, since it reduces the amount of data to be stored and in case of CFS technique, it is possible to identify which features are most representative of dataset for proposed classification objective. The results show that the RT classification technique is the most suitable for Shenzhen set - Chest X-ray dataset.

The comparison of results of features analysis techniques shows that both present a good performance for selected technique, being the PCA who presents a reduction in the processing time of 10% and 70% in the quantity of registers, however, CFS obtains a reduction in number of registers of 80% and identifies the MNND and EQI features as the most representative.

At the conclusion of **Classifier Evaluation for Each Dataset** step, extracted features and **classification techniques** selected for each dataset are stored, as well as the recommended features for each one, Table 9 presents classification and analysis features techniques chosen for each dataset, this information is considered experience, called **Meta-Data**.

Table 9 – Classification and Feature Analysis Techniques Recommended for Each Dataset.

	Recommended	
Dataset	**Features**	**Classifier**
Shenzhen		RT
Montgomery	CFS: MNND and EQI	K-NN
RIM-ONE		RC
LIDC-IDRI	CFS: DFM	SVM
MIAS	PCA	RR
BRATS	CFS: MNND	RT

Determining the appropriate classification methodology for each dataset, the necessary experience that Meta-Learning uses as a learning base has been developed, the experience is collected and stored in **Meta-Data**, this consists of two elements:

1. Information contained in Classification and Feature Analysis Techniques Recommended for Each Dataset (Table 9).
2. Configuration parameters for each classification technique selected in accordance with Section 5.2.3.

In **Classifier Evaluation for Meta-Model** step, **Meta-Model** is constructed, **Meta-Data** is used with the aim of identifying which of techniques described in Section 5.2.3 presents a better performance for identification of most suitable classification methodology based on features of images. In this step, all datasets are taken as a single dataset, which implies that all images form a single training set, this means that the training set contains all types of images used, however, each image contains a label that identifies the best methodology for processing.

Meta-Data is made up of all features extracted and experience obtained in the **Classifier Evaluation for Each Dataset** step. Accuracy and processing time are used as parameters for performance evaluation. Table 10 show average of five results.

Table 10 – Comparison of Average Performance Training Results of Classification Technique Applied to All Dataset.

Classifiers	Accuracy %	Processing time/sec
NB	0.123	0.958
SVM	0.999	1638.382
MLP	0.972	40.03
JDT	0.996	0.27
RT	**1**	**0.061**
RC	1	0.384
K-NN	1	1574.004
RF	1	3.266
RR	1	13.626

Datasets that have similar features must use the same classifier. When building **Meta-Data**, each dataset is processed independently to select its best classifier, however, if two different **classification techniques** are arbitrarily selected for two dataset with similar feature, the classification process performed by **Meta-Model** will be inefficient.

The PCA and CFS techniques are applied on all extracted features, that contain 264405 registers, the resulting datasets contain 158643 registers for PCA and 35254 for CFS, each new dataset is tested with the best performance classifiers, in case of Table 10, the test is performed with the RT, RC, K-NN, RF and RR. Table 11 shows the average of five results obtained.

Table 11 – Comparison of Average Performance Results of Classification Techniques Applied to Most Relevant Features for Meta-Model.

Feature Analysis	Classifiers	Accuracy %	Processing time/sec
CFS	**RT**	**1**	**0.060**
	RC	1	0.272
	K-NN	0.996	267.442
	RF	1	1.734
	RR	1	7.076
PCA	**RT**	**1**	**0.130**
	RC	1	1.092
	K-NN	1	511.84
	RF	1	6.188
	RR	1	19.418

When comparing the techniques PCA and CFS for analyze of features in Table 11, both allow to reduce the amount of data necessary to perform the classification of images correctly, however, the CFS has a higher rate of data reduction and allows to identify the most representative features of original set, which is an advantage over PCA technique. Similarly, based on results, RT is selected as classification technique for **Meta-Model**.

In the experiments carried out it is clear that K-NN and SVM techniques require a high processing time, with RT being the technique with the shortest processing time in each test, which implies that the approach to creation of class areas requires more time than creation of decision trees.

Until now, Meta-Data have been constructed, containing features of each type of image and its best classification methodology, in same way that Meta-Model has been constructed.

The next step is to validate the **Meta-Model**, which means to prove that classification system based on Meta-Learning to identify the best classification methodology for each type of medical image, obtains adequate performance with the set of tests, as described in **Meta-Model Validation** step (Section 5.2.5). Test set consists of 30% of images of each dataset, randomly selected, in total 7552 validation images. The MNND, EQI and DFM features and information in Table 9 are applied to validation images and used for construction of test set. Results of **Meta-Model Validation** step, present a 99.4% accuracy in prediction of best classification methodology.

5.4 Final Considerations

In order to create a methodology for selection of best methodology for classifying medical images, six public dataset are used, twelve new dataset are created as a result of original ones and nine **classification techniques** based on Machine Learning are used. Based on results of this chapter, it is concluded:

1. The results prove the effectiveness of use of techniques based on Meta-Learning for selection of the most appropriate classification methodology in medical images.

2. The methodology developed allows **Meta-Data** to be used to identify which feature are more significant, both for recommendation of best classification methodology and for its application.

3. The CFS technique allows a greater reduction of data necessary for correct classification of medical images and allows to identify which are the most significant features.

The use of Meta-Learning offers a high degree of freedom in relation to inclusion of new datasets, and updating of **Meta-Data**, which allows changing the current classification methodology for a type of image, for a new one that presents a better performance.

The main contributions of methodology proposed in this chapter are:

1. Characterization of medical images through phylogenetic indexes and their analysis to determine the most representative features to be stored in **Meta-Data**.

2. **Meta-Model** Construction to identify the most suitable classification methodology for images of chest X-Ray, thoracic CT scan, retinal fundus, brain MRI and mammography.

6 CONSTRUCTION OF META-DATA BASED ON EXPERIENCE REPORTED IN STATE OF THE ART, USING CNN AS A CLASSIFIER FOR META-MODEL.

In previous chapters, methodologies for classification of different types of medical images with specific classification aims have been developed, these methodologies and the features extracted from each type of image have been used as an experience, storing it in Meta-Data. In this chapter uses as experience the methodologies reported in the literature, transforming contributions made by scientific community into Meta-Data, instead of proposing a methodology as was done in previous research. In addition, this chapter uses CNN as a base technique, for construction of Meta-Model.

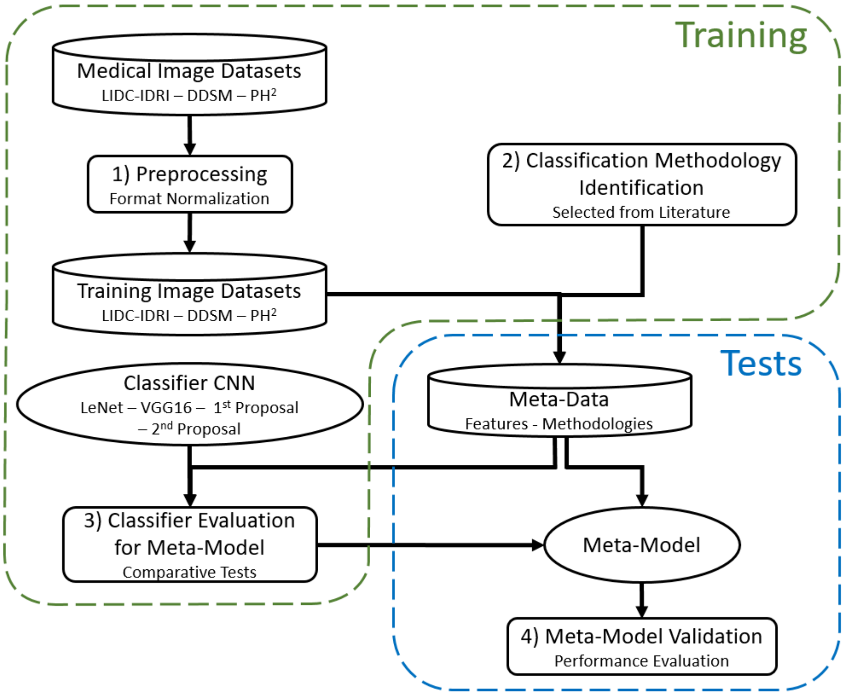
The aim of this chapter is to prove the effectiveness of Meta-Learning for selection of most appropriate classification methodology for different types of medical images, understanding as most appropriate, the methodology that presents a better classification performance for a specific type of medical image with respect to evaluation measures. Different types of medical images and the best classification methodology based on literature for each type of image are used as input data. Experience is created by labeling each image with its most appropriate classification methodology, which is stored in Meta-Data, then a Meta-Model is created that based on Meta-Data is able to extract features from new test images and recommend the most appropriate classification technique.

The chapter is organized by sections, in Section 6.1 is presented the datasets used, in Section 6.2 the proposed methodology is presented, Section 6.3 shows the results, finally in Section 6.4 final considerations are presented. Figure 14 presents the proposed methodology, including datasets used.

6.1 Datasets

Meta-Learning is applied to identify the most appropriate classification methodology for three different medical image datasets, each dataset has a specific classification objective, for thoracic CT scan images the aim is to perform a classification of lung nodules in malignant and benign; for mammography images, the objective is the classification of mass and calcification; finally, for the dermatoscopic images the objective is the identification of melanoma. Initially, the images are used maintaining the original parameters with respect to size, color channels, format and histogram, only the modifications described in

Figure 14 – Proposed Methodology for Meta-Learning Classifier Based on Experience from Literature.



Source: Prepared by author

the proposed methodology are made.

LIDC-IDRI The characteristics of dataset are described in Section 4.1.

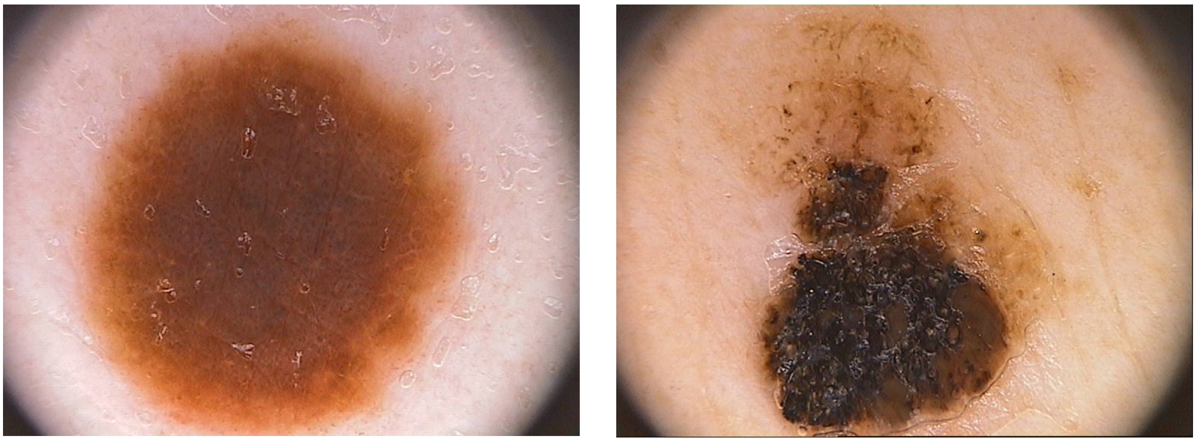
DDSM The characteristics of dataset are described in Section 4.1.

The $\mathbf{PH^2}$ dataset has been developed for research and benchmarking purposes, in order to facilitate comparative studies on both segmentation and classification algorithms of dermatoscopic images. The image dataset was acquired at the Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal. The PH^2 contains 200 dermatoscopic images of melanocytic lesions: 80 common nevi, 80 atypical nevi and 40 cases of melanoma, and includes medical annotation, segmentation of the lesion, clinical and histological diagnosis and evaluation of several dermatoscopic criteria: colors, pigment network, dots/globules, streaks, regression areas and blue-whitish veil (MENDONCA et al., 2013). Figure 15 presents an example of PH^2 datasets.

6.2 Proposed Methodology

The methodology consists of four steps: 1) **Preprocessing**, for each item of **Medical Image Datasets** a format change is made, the result of this step are the **Training Image Datasets**. 2) **Classification Methodology Identification**, based

Figure 15 – PH^2 Dataset Examples.



(a) *Healthy Skin* (b) *Melanoma case*

Source: PH^2 Dataset (MENDONCA et al., 2013)

on literature, the most appropriate methodology for each type of image are identified, the selected methodology and **Training Image Datasets** make up Meta-Knowledge, which is stored in **Meta-Data**. 3) **Classifier Evaluation for Meta-Model**, this step uses **Meta-Data** and different CNN from **Classifier CNN** to select the most suitable classifier for **Meta-Model**, the result of this step together with **Meta-Data** make up **Meta-Model**. 4) **Meta-Model Validation**, using a set of test images, the performs of proposed methodology is evaluated. Figure 14 shows the proposed methodology.

6.2.1 Preprocessing

For all datasets, a single preprocessing is carried out, which consists of normalization of image format to Portable Network Graphics, originally images of LIDC-IDRI and DDSM datasets are obtained in format Digital Imaging and Communications in Medicine (DICOM), the PH^2 dataset is obtained in bitmap format. Normalization is necessary because the DICOM format is not supported by the Waikato Environment for Knowledge Analysis (WEKA) tool (WITTEN et al., 2016), which is used to perform data processing and application of CNN. This tool was selected, since allows to use various Machine Learning and Deep Learning techniques, in the same way, it allows the analysis of the extracted features and the configuration of different test parameters, all these characteristics allow to perform the necessary tests, in addition, WEKA proved to be easy to install.

6.2.2 Classification Methodology Identification

In this step, for each type of medical image, the most appropriate classification methodology is identified from literature, to each selected methodology is associated the information of the most representative features that it uses to characterize the images. All this information is considered experience and is stored in **Meta-Data**. It is necessary to clarify that in this step, the identified methodology are not implemented and features of datasets are not extracted, both are only established for each datasets.

For **LIDC-IDRI** dataset, research in state of the art indicates that Neto et al. (2017) is best work for its classification in benign or malignant nodules. Using the linear logistic regression technique described in Section 3.2.7, together with use of descriptors based on analysis of form: spherical disproportion, compactness, statistical measures, Feret diameter, skeleton, sphericity, spherical density, measures of irregularity and circularity.

For **DDSM** dataset, research in state of the art indicates that Dhahbi, Barhoumi e Zagrouba (2015) is best work for its classification in mass or calcification. The K-NN technique described in Section 3.2.6 together with the use of feature extraction based on curvelet transform and moment theory has an adequate performance for classification process.

For **PH^2** dataset, research in state of the art indicates that Moura et al. (2017) is best work for identification of melanoma. Using MLP technique described in Section 3.2.8, together with selection of relevant attributes based on the gain radio information applied to descriptors: ABDC Rule, Gray level concurrency matrix, Gray Level Run Length Matrix, Histograms of Oriented Gradients, Local Binary Pattern, Tamura, Box-Counting and CNN.

6.2.3 Classifier Evaluation for Meta-Model

In this step, four different configurations of the CNN are tested to select the **Meta-Model** classification technique, processing **Meta-Data**, which contains **Training Image Datasets** and information of most appropriate classification methodology for each type of image and its characterization technique, according to results obtained in step of **Classification Method Identification** (Section 6.2.2).

The evaluation requires a training set that is initially made up of 70% of images of each dataset, which implies the processing of a large number of images with a high processing time, to face this challenge a test process is carried out in order to determine the minimum amount of images required, making changes in training set and validating the results. The test begins with one hundred images randomly selected from each dataset, increasing in each iteration of test and having as stopping criterion when the performance data does not present modifications, when selecting images, a method without restitution is used, in this way, it is guaranteed that there is no duplication of images.

CNN described in Section 3.2.2 are selected as a **Meta-Model** classification method due to ability to extract features of images and perform classification task (CHAIB et al., 2017). In this way, it is proved that features extracted for **Meta-Model** do not have to be the same used by the methods identified in Section 6.2.2.

Two of the configurations to be tested are selected based on their use in literature. In addition, two configurations are proposed in order to obtain performance information based on the number of layers and the size of the matrices:

1. LeNet Network presented in Section 4.2.3.

2. VGG16 Network is a convolutional neural network architecture named after the Visual Geometry Group from Oxford. VGG16 has 13 convolutional layers, five subsampling layers and one output layer. The first two convolution layers are made up of 64 nodes each one, the next two are made up of 128 nodes each one, the next three nodes each one have 256 nodes, and finally the last six layers are made up of 512 nodes each one. All layers use 3x3 convolution kernels and RELU activation function (SIMONYAN; ZISSERMAN, 2014).

3. First proposed configuration is composed of a single convolutional layer of 20 nodes with a 5x5 matrix and RELU activation, a subsampling layer that uses a 2x2 matrix with MAX pooling type, finally one classification output layer, the aim of this simple configuration is to determine if the use of more than one convolution layer is necessary.

4. Second proposed configuration is composed of a single convolutional layer of 20 nodes with a 3x3 matrix and RELU activation, a subsampling layer that uses a 2x2 matrix with MAX pooling type, finally one classification output layer, this configuration is designed to determine the effect of size of convolution kernel in performance of classification.

6.2.4 Meta-Model Validation

In this step, performance of proposed methodology is evaluated using a new test set consisting of 30% of images of each datasets, which has been reserved randomly, however, the class labelling information is preserved for validation.

The step begins with creation of test set, containing images and class labelling information obtained for each type of image in the step of **Classification Method Identification** (Section 6.2.2). Then, each of four configurations of CNN described in Section 6.2.3 are tested using **Meta-Data** and test set. Performance of each CNN configuration is stored and analyzed, finally, the performance comparison is carried out.

6.3 Result

In this section, the performance is evaluated and compared for the four configurations of CNN described in Section 6.2.3.

Training set: According to process described in Section 6.2.3, iterative tests are performed to determine the amount of images needed to form the training set, after seven iterations, the results verify that training dataset requires to be made up of 8% of the DDSM dataset, 70% of PH^2 dataset and 0.7% of LIDC-IDRI dataset, for a total of 540 images representing 1.7% of original training dataset.

Test set: As described in section 6.2.4, consists of a total of 8173 image tags in 3 classes, 645 images come from the DDSM dataset whose proper technique of classification is K-NN, 7468 images from the LIDC-IDRI dataset with Linear logistic regression as a classification technique, finally, 60 images of PH^2 dataset with MLP classification technique.

Determining the appropriate classification methodology for each dataset, the necessary experience that Meta-Learning uses as a learning base has been developed, in this case, the experience comes from literature as presented in **Classification Methodology Identification** step (Section 6.2.2), experience is collected and stored in **Meta-Data**, this consists of two elements:

1. Methodology for classification of each type of Medical Image presented in Section 6.2.2.

2. Given that methodology are based on CNN, the **Training Images Datasets** used are stored.

The next step is to evaluate different CNN configurations to determine **Meta-Model** classifier, this step uses all types of images. The aim is that **Meta-Model** can identify the best classification methodology for each type of image according to **Meta-Data**, as presented in **Classifier Evaluation for Meta-Model** step (Section 6.2.3). Table 12 shows the average of five results of **Classifier Evaluation for Meta-Model** step. Accuracy, AUC, F-measure and processing time are used as measures of the performance of each configuration of classification method.

Table 12 – Average Results of Test with Training Set in Step of Classifier Evaluation for Meta-Model.

Technique	**Accuracy %**	**AUC %**	**F-measure %**	**Processing Time/sec**
LeNet	**1**	**1**	**1**	359
VGG16	0.446	0.500	-	1301
Proposal 1	**1**	**1**	**1**	**176**
Proposal 2	**1**	**1**	**1**	178

The results of Table 12 show that LeNet, Proposal 1 and Proposal 2 configurations, are able to correctly classify each type of image identifying their most appropriate classification methodology, where Proposal 1 presents a better performance when considering its processing time, however, it is not possible to find relationships between configurations and their performance that allow selecting the best CNN for **Meta-Model**. Since the results of training test to determine the CNN configuration for **Meta-Model** classification technique are inconclusive, the test is performed using the set of test images.

Table 13 shows the average of five results of **Meta-Model Validation** step, using the same performance measures used for Table 12 and all CNN configuration.

Table 13 – Average Results with Test Set in Meta-Model Validation Step.

Technique	**Accuracy %**	**AUC %**	**F-measure %**	**Processing Time/sec**
LeNet	**0.995**	**0.999**	**0.996**	374
VGG16	0.079	0.005	-	1204
Proposal 1	0.993	**0.999**	0.993	**317.82**
Proposal 2	**0.995**	**0.999**	0.995	331.13

When analyzing results of Table 13, it is possible to find:

First, when analyzing the processing carried out by configuration of parameters of VGG16 it is possible to deduce that the loss of relevant features for the classification is caused by the high number of convolution layers, which leads to VGG16 being inadequate, the tests show that the loss of information begins from third layer, being critical in the eleventh layer.

Second, when comparing configurations of Proposal 1 and Proposal 2, it can be found that change in convolution kernel produces a loss of important information for classification objective, being more efficient the use of a small kernel. Similarly, it implies a change in processing time, reducing by using a larger kernel.

Third, configuration that presents a better performance taking as the most important parameter the correct classification of images is LeNet, this deduction is possible thanks to the F-measure and the unbalanced nature of dataset.

Fourth, changes in kernel size and the number of convolution layers have an effect on processing time, by using smaller kernels, time increases, as well as increasing convolution layers.

Finally, when comparing proposed configurations with LeNet, it is evident that it is necessary to use more than one convolution layer to increase the performance, in same way a greater number of neurons in each layer is necessary.

In **Classification Method Identification** step, a state of the art investigation are carried out to select the best classification methods for each type of medical image under study, including features that are used. This information constitutes experience that Meta-Learning uses to predict which classification methodology is the most appropriate for each image of test set. Experiments carried out showed that **Meta-Model** has the advantage of easily adapting to changes that occur in state of the art, which implies that, if a new classification methodology is presented with a better performance for a type of image, it is only necessary, update the class in **Meta-Data**.

The results show that use of Meta-Learning for identification of the proper methodology of classification of medical images is very efficient, likewise, use of CNN as a **Meta-Model** classification technique obtains high results, however, the use of CNN for extraction of features has advantages and limitations: as an advantage it is possible to extract the features without an additional technique, however at the same time, it generates an increase in size of **Meta-Data**, since the training images are stored and not the characterization of images and also causes an increase in computational cost.

6.4 Final Considerations

A new methodology based on Meta-Learning has been presented for identification of the most suitable methodology of classification for different types of medical images based on experience presented in literature, for this purpose three public access datasets are used and four different CNN parameter settings are tested as a **Meta-Model** classification technique. Based on the results it is possible to conclude:

1. Use of methodology based on Meta-Learning is effective for selection of most appropriate methodology of classification and its most representative features for medical images classification.

2. CNN is effective as a characterization and classification technique for **Meta-Model**.

3. Given the properties of datasets under study, it is not necessary to have a large training set.

When analyzing the information contained in **Meta-Data**, identification of the most representative features to be used by classification methodology is possible, this implies that the use of Meta-Learning can identify an entire structure of data processing, based on the particular features of input data.

The main contributions of methodology proposed in this chapter are:

1. Collect the experience presented in literature, transforming it into Meta-Data to be used in a classification system based on Meta-Learning.

2. using CNN as a method of characterization and classification of Meta-Model for images of thoracic computed tomography (CT) scan, mammography and dermostopic.

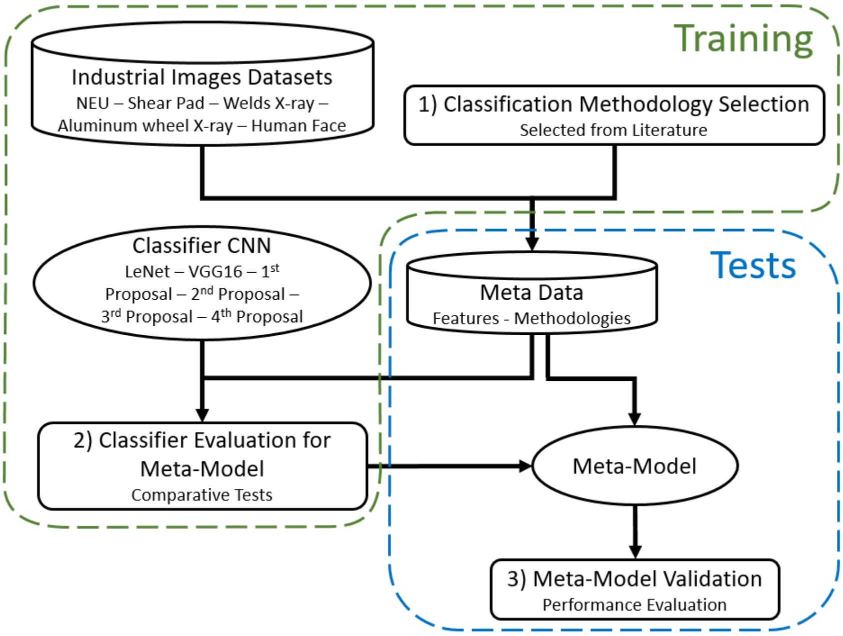
7 META-LEARNING APPLIED TO SELECTION OF CLASSIFICATION METHODOLOGY IN INDUSTRIAL IMAGES

In previous chapters, datasets with medical images have been used as an object of study, in order to test the effectiveness of methodology based on Meta-Learning in a different type of image, this chapter presents as an object of study, images of industrial field.

The aim of this chapter is to demonstrate the efficiency of use of Meta-Learning for identification of most appropriate classification methodology for different types of industrial images, understanding as more appropriate the method that presents a better performance for a type of image with respect to the evaluation measures.

The chapter is organized by sections, in Section 7.1 is presented the datasets used, in Section 7.2 the proposed methodology is presented, Section 7.3 shows the results, finally in Section 7.4 Final Considerations are presented. Figure 16 presents the proposed methodology, including datasets used.

Figure 16 – Proposed Methodology for Meta-Learning Classifier for Industrial Images.



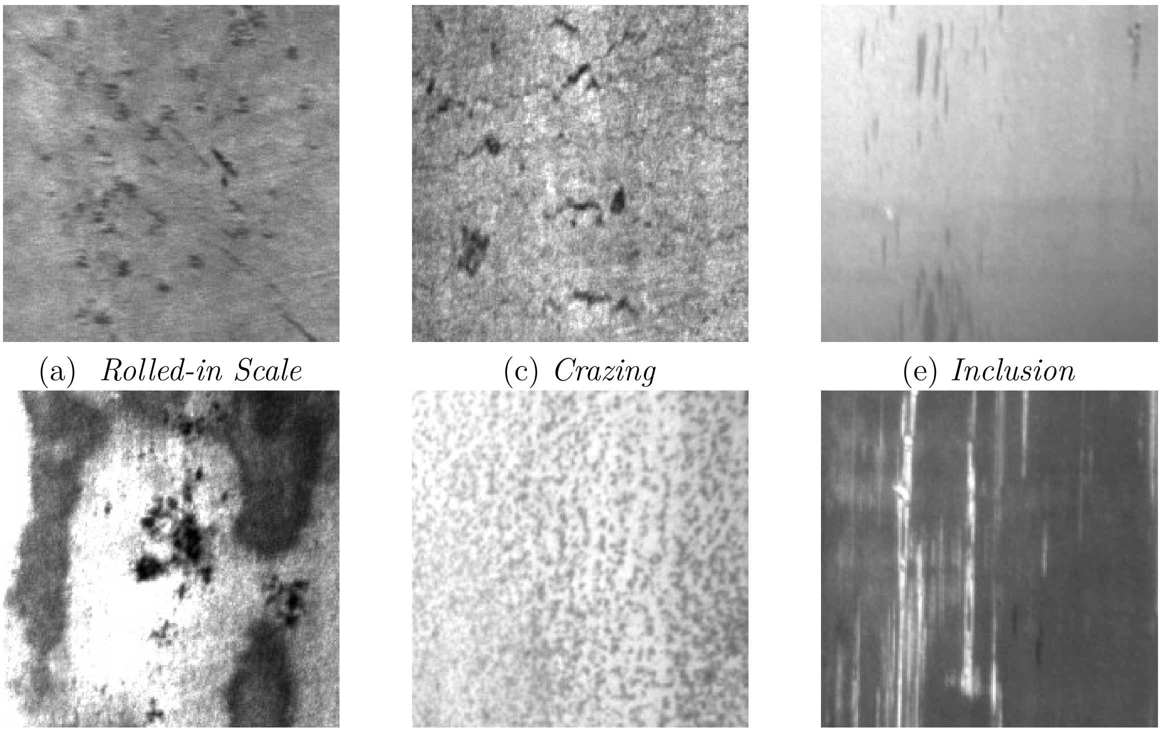
Source: Prepared by author

7.1 Datasets

Meta-Learning is used for selection of the most appropriate classification methodology for five datasets with different types of industrial images, each dataset has a specific objective: for hot-rolled steel strip the aim is classification of defects; for shear pad of wagon train the objective is classification of state of the pad; for welds x-rays the target is identification of geometric faults; for aluminum wheel x-rays the objective is detection of defects; finally for human faces the objective is classification of emotion based on expression. Images are used maintaining the original parameters with respect to size, color channels, format and histogram.

NEU surface defect dataset was created by Northeastern University, contains six types of surface defects of Hot-rolled Steel Strip: rolled-in scale, patches, crazing, pitted surface, inclusion and scratches. The dataset consists of 1800 grayscale images with 300 samples of each type of defect, each image has a resolution of 200x200 pixels. The NEU Surface defect dataset includes different difficult challenges that include image defects given the influence of lighting and the change in materials (SONG; YAN, 2013). Figure 17 presents an example of NEU datasets.

Figure 17 – NEU Dataset Examples.



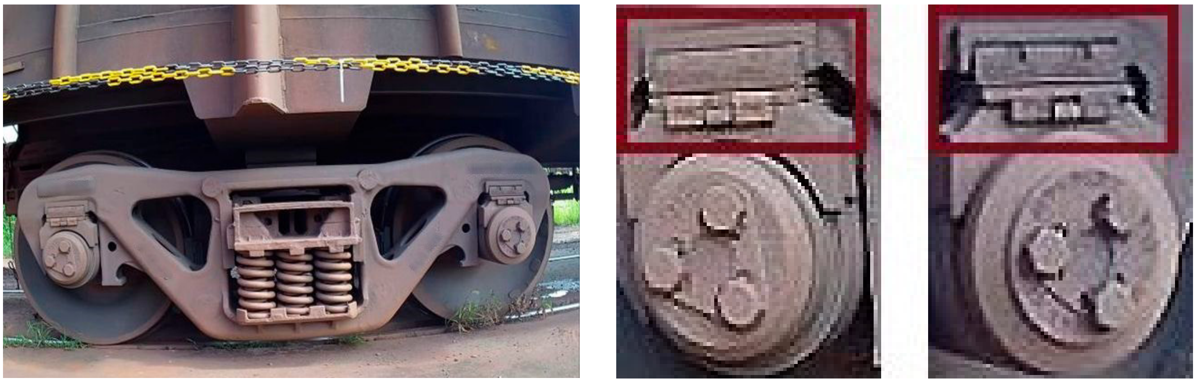
(a) *Rolled-in Scale* (c) *Crazing* (e) *Inclusion*

(b) *Patches* (d) *Pitted Surface* (f) *Scratches*

Source: NEU Dataset (SONG; YAN, 2013)

The **Shear pad of wagon train** dataset based on Rocha et al. (2017), is constituted by 1081 colorful images, with a resolution of 640x480 pixels and include the whole wagon truck, each image contains the elements: Pad, suspension springs and roller bearing screw set. Red rectangles of Figure 18 presents an example of shear pad.

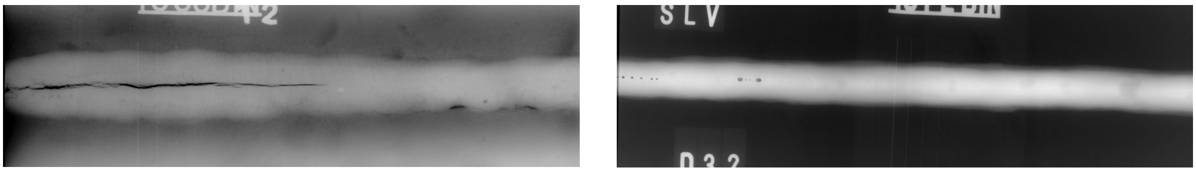
Figure 18 – Shear Pad of Wagon Train Dataset Examples.



Source: Shear Pad of Wagon Train Dataset (ROCHA et al., 2017)

The **Welds X-ray** dataset was taken by the BAM Federal Institute for Materials Research and Testing, Berlin, Germany and is part of X-ray images for X-ray testing and Computer Vision (GDXray). The dataset contains 88 images arranged in 3 series, and have 641 defects. the images are in 8-bit gray scale with a pixel size of 40.3 microns (630 dpi) (MERY et al., 2015). Figure 19 presents an example of dataset.

Figure 19 – Welds X-ray Dataset Examples.

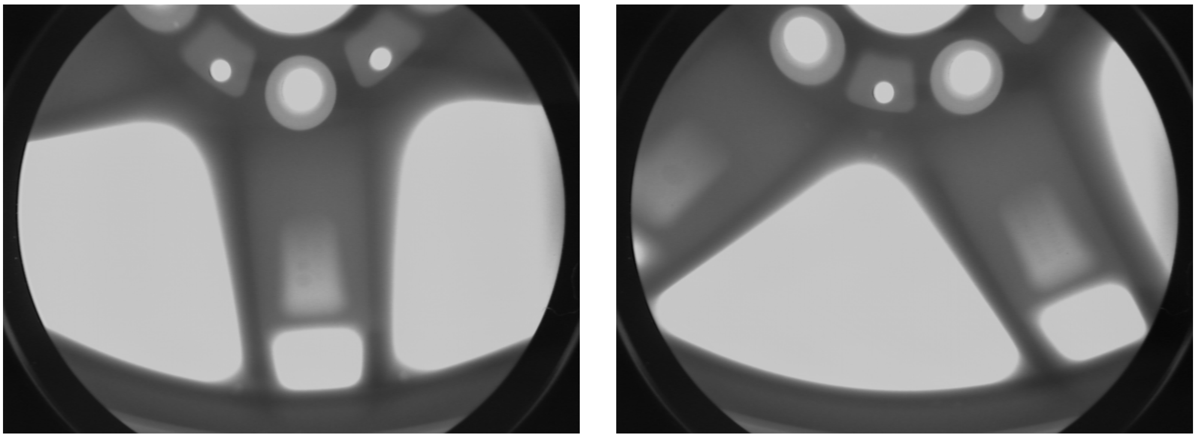


Source: Welds X-ray Dataset (MERY et al., 2015)

The **Aluminum wheel X-ray** dataset consists of 72 images taken with 5-degree rotation and using an image intensifier, additionally includes defect annotations and calibration. The selected dataset is a subset of the Castings group that contains 2727 images and belongs to GDXray (MERY et al., 2015). Figure 20 presents an example of dataset.

The **Human Face** dataset is made up of 300 images of human faces, selected from the MIRFLICKR-25000 collection. The dataset is made up of images with different

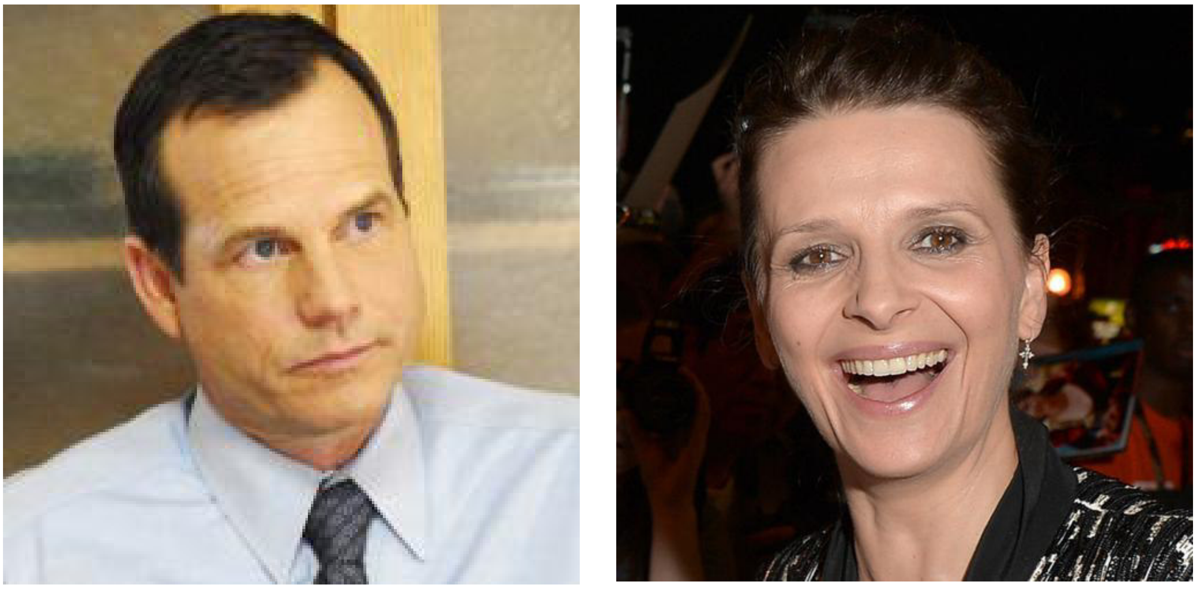
Figure 20 – Aluminum Wheel X-ray Dataset Examples.



Source: Aluminum Wheel X-ray Dataset (MERY et al., 2015)

size, lighting and sharpness characteristics, and includes colourful and gray scale images. The faces in the images belong to different people that vary in gender and age (HUISKES; LEW, 2008),(HUISKES; LEW, 2010). Figure 21 presents an example of dataset.

Figure 21 – Human Face Dataset Examples.



Source: Human Face Dataset (HUISKES; LEW, 2008),(HUISKES; LEW, 2010)

7.2 Proposed Methodology

Proposed methodology consists of three steps: 1) **Classification Methodology Selection**, based on literature is identified the most appropriate classification methodology for each type of dataset; result of this step together with **Industrial Images Datasets**, make up the Meta-knowledge that is contained in **Meta-Data**. 2) **Classifier Evaluation for Meta-Model**, this step uses different CNN configurations from **Classifier CNN** to process **Meta-Data** and identify the configuration that presents a better performance; the result of this step together with **Meta-Data** make up **Meta-Model**. 3) **Meta-Model Validation**, the proposed methodology is validated using a new test set. The proposed methodology can be seen in Figure 16.

7.2.1 Classification Methodology Selection

At this step a classification methodology based on literature is selected for each dataset, the selected methodology do not have to be similar, this quality implies that each methodology can use a different characterization form to represent the dataset. Information of each selected classification methodology, as well as type of characterization used, is considered experience and is part of **Meta-data**. It is necessary to clarify that in this step the classification methodology found are not implemented or recommended characterization techniques are applied, both are only identified and established for each dataset.

For **NEU surface defect** dataset according to Song e Yan (2013), the SVM technique presented in Section 3.2.14, is identified as the most suitable for performing defect classification, together with use of a proposed new characterization technique called adjacent evaluation completed local binary patterns presented in Section 3.3.2.

For **Shear pad of wagon train** dataset according to Rocha et al. (2017) the use of CNN presented in Section 3.2.2 in conjunction with SVM presented in Section 3.2.14 to replace the softmax layer is efficient for pad classification, when used in grayscale images. In the proposed CNN the first convolution layer has 32 neurons with 3x3 kernel, step 1 and ReLU activation. the next convolution layer has 64 neurons, with 3x3 kernel, step 3 and ReLU activation. The fully connected layer has 128 neurons with ReLU activation and 50% abandonment, final layer consists of SVM.

For **Welds X-ray** dataset according to Soares et al. (2017) the HMM technique presented in Section 3.2.4 contributes to the identification of geometric faults, when using as input data the information obtained after applying the method of Gaussian mixtures on the features resulting from applying the PCA technique presented in Section 3.4.1 on images.

For **Aluminum wheel X-ray** dataset the methodology proposed by Mery e Filbert (2002) presents efficient results for automatic detection of defects in sequential images taken with rotation of piece in small intervals of time. The methodology basically consists of two steps: identification and monitoring of potential failures.

Identification of potential faults uses two features of defects to perform their individualization: 1) A defect can be considered as a connected subset of the image. 2) There is a significant difference in intensity of gray level between the fault and its neighbours, however, it is necessary to perform a preprocessing to identify the signal-to-noise ratio and establish a classification threshold for intensity of the gray level. To carry out identification of potential faults, two processes are applied: 1) Edge detection, where a Laplacian of Gaussian kernel and a zero-crossing algorithm are used to detect edges present in images. 2) Segmentation and classification of potential faults, regions with closed limits whose average gray level is 2.5% greater than average gray level of their neighbours and whose area is greater than 15 pixels are labelled as potential faults. The tracking of potential faults allows separating real faults from the false positives, so that a failure is considered true, must comply with the follow-up in three steps: 1) Matching in two views, where the faults detected in two images of the same piece with rotation, they coincide in relative position of the object and extracted features. 2) Tracking in more views, where the path of the fault is established in three images of the same piece with rotation. 3) Verification, where minimum squares are used to establish the centers of gravity of the faults, projecting a point in each trajectory (MERY; FILBERT, 2002).

For **Human Face** dataset according to Rivera et al. (2017) the use of features based on Units of Action (AU) and classifier K-NN presented in Section 3.2.6 allows the detection of six basic facial expressions.

AU allow the description of perceptible facial muscle movements based on AU functions, in Rivera et al. (2017) twenty action units are used: two InnerBrowRaiser, two OuterBrowRaiser, two BrowLowerer, UpperLipRaiser, two LipCornerPuller, two CheekPuffer, two LipStretcher, two LipCornerDepressor, jawLowerer , two EyesClosed, two JawLeftRight. The AU can take values between 1 and -1, and are treated as a vector of twenty dimensions. The employed AUs are written in facial action coding system proposed by (EKMAN; FRIESEN, 1976).

7.2.2 Classifier Evaluation for Meta-model

In this step, different CNN configurations are tested to identify the one that presents a better performance with respect to evaluation measures. Selected configuration is used as a classification technique for **Meta-Model**. Each configuration is tested using the same **Meta-Data**, which contains a set of training images, information of the most appropriate classification methodology for each type of image and its characterization technique, according to information obtained in step of **Classification Methodology Selection** (Section 7.2.1).

CNN is selected as a classification technique for **Meta-Model** because of its ability to extract features directly from image and perform the classification task without intervention of alternative characterization techniques (CHAIB et al., 2017). Given the variety of configurations that can be used for classification task, identifying the most appropriate implies tests, two configurations known in literature and fourth proposed configurations are tested and compared:

1. LeNet Network presented in Section 4.2.3.

2. VGG16 Network presented in Section 6.2.3.

3. First proposed configuration is designed to be simple and determine if more than one convolution layer is necessary and if a low number of neurons is adequate to perform the characterization; the first convolution layer is made up of 20 neurons with a 5x5 matrix and RELU activation, the second subsampling layer uses a 2x2 matrix and a maximum value selection function, the last layer is full connected and is responsible for classification.

4. Second proposed configuration is composed of a single convolutional layer of 20 nodes with a 3x3 matrix and RELU activation, a subsampling layer that uses a 2x2 matrix with MAX pooling type, finally one classification output layer, this configuration is designed to determine the effect of size of convolution kernel in performance of classification.

5. Third proposed configuration is based on first proposed, changing the size of the subsampling matrix to determine its effect on performance of classification, the subsampling layer uses a 3x3 matrix and maximum value as selection function.

6. Fourth proposed configuration performs a combination of two previous ones, in order to determine the effects on performance; first convolution layer is made up of 20 neurons with a RELU activation, and a 3x3 convolution matrix; second subsampling layer uses 3x3 matrix and maximum value choice function; the final layer is full connected and performs classification.

The step begins with creation of training set, containing images and class labelling information. Then, each of six configurations of CNN are tested using **Meta-Data**. Performance of each CNN configuration is stored and analysed. Finally, performance comparison is carried out.

7.2.3 Meta-Model Validation

In this step, **Meta-Model** validation is performed, using **Meta-Data** as training data and a new set of test images. The test set is made up of 30% of the images of each dataset, selected randomly, the label that identifies which is the best classification and characterization technique for each type of image is the same one recognized for dataset in step of **Classification Method Selection** (Section 7.2.1), however, this information is used only to verify the result of classification given by **Meta-Model**.

The step begins with creation of test set, containing images and class labelling information. Then performance of **Meta-Model** is stored and analysed.

7.3 Result

In this section, performance of different configurations for CNN is evaluated and compared for classification task, and performance of **Meta-Model** with a set of test images is also evaluated. Processing of data and applications of CNN are carried out using Waikato Environment for Knowledge Analysis (WEKA) (WITTEN et al., 2016). This tool was selected, since allows to use various Machine Learning and Deep Learning techniques, in the same way, it allows the analysis of the extracted features and the configuration of different test parameters, all these characteristics allow to perform the necessary tests, in addition, WEKA proved to be easy to install.

For construction of **Meta-Data**, 70% of images of each dataset are randomly selected: 1260 from NEU surface dataset, 756 from Shear Pad of Wagon Train dataset, 62 from Welds X-ray dataset, 50 from Aluminum Wheel X-ray dataset, finally 210 from Human Face dataset.

Determining the appropriate classification methodology for each dataset, the necessary experience that Meta-Learning uses as a learning base has been developed, in this case, the experience comes from literature as presented in **Classification Methodology Selection** step (Section 7.2.1), experience is collected and stored in **Meta-Data**, this consists of two elements:

1. Methodology for classification of each type of industrial images presented in Section 7.2.1.

2. Given that methodology are based on CNN, the **Industrial Images Datasets** used are stored.

The next step is to evaluate different CNN configurations to determine **Meta-Model** classifier, this step uses all types of images. The aim is that **Meta-Model** can identify the best classification methodology for each type of image according to **Meta-Data**, as presented in **Classifier Evaluation for Meta-Model** step (Section 7.2.2). Table 14 shows the average of five results, Accuracy, AUC and F-measure are used as measures of performance of each configuration of CNN.

Table 14 – Average Test Result for Different CNN Configurations Applied to Industrial Images.

Technique	Accuracy %	AUC %	F-measure %	Processing Time/sec
LeNet	**0.994**	**0.999**	**0.994**	306.22
VGG16	0.543	0.5	-	1932.11
Proposal 1	0.967	0.998	0.967	201.51
Proposal 2	0.977	**0.999**	0.976	189.25
Proposal 3	0.982	0.997	0.982	**177.88**
Proposal 4	0.958	**0.999**	0.955	183.49

When analyzing results of Table 14 it is possible to find: 1) VGG16 presents a loss of relevant information, caused by the high number of convolutions carried out, which leads to VGG16 being inadequate, the tests show that the loss of information begins from fourth layer and becomes critical in ninth layer. 2) when comparing proposals 1, 2, 3 and 4, it is found that there is a loss of relevant information when using a larger convolution matrix, conversely it is possible to increase the size of subsampling matrix without losing relevant information, however, when using both resources, as in case of proposal 4, it is found that loss of information increases, indicating that loss of information due to change in size of subsampling matrix increases. 3) Changes in kernel size and the number of convolution layers have an effect on processing time, by using smaller kernels, time increases, as well as increasing convolution layers. Finally, LeNet configuration is the most suitable to be used as a classifier of **Meta-Model**, given its superior value of F-measure, which indicates a better classification given the unbalanced nature of classes.

When determining CNN configuration used as a **Meta-Model** classification technique, training shown in Figure 16 is concluded.

The next step is to validate the **Meta-Model**, which means to prove that classification system based on Meta-Learning to identify the best classification methodology for each type of industrial image, obtains adequate performance with the set of tests, as described in **Meta-Model Validation** step (Section 7.2.3).

A test set consisting of 1004 images taken at random from datasets is constructed: 540 from NEU surface dataset, 325 from Shear Pad of Wagon Train dataset, 26 from Welds X-ray dataset, 22 from Aluminum Wheel X-ray dataset and 90 from Human Face dataset. 96% accuracy, 99.7% AUC and 96.5% F- measure are obtained.

Results show that use of Meta-Learning is efficient to identify the most appropriate methodology of image classification to perform tasks such as quality control in products and welding, identify defective parts or expressions on faces of staff.

7.4 Final Considerations

A new methodology based on Meta-learning has been presented for identification of the most appropriate industrial image classification methodology, for this purpose five dataset and six different configurations of parameters for CNN are used. Based on the results it is possible to conclude:

1. Proposed methodology is effective for identification of most appropriate methodology of classification and characterization in industrial images.

2. CNN is effective as a characterization and classification technique for **Meta-model**.

3. Use of Meta-Learning allows to change or update the information of classification and characterization techniques quickly and easily, since characterization is done only on images under study.

When comparing the different configurations for CNN, it is concluded that for the classification of images under study it is necessary to use two convolution layers.

The main contributions of methodology proposed in this chapter are:

1. A new application of Meta-Learning in industrial images for selection of classification methodology.

2. Application of CNN as a method for classifying and characterizing the Meta-Model for hot-rolled steel strip, the shear pad of wagon train, welds x-rays, aluminum wheel x-rays and human faces images.

Information contained in this chapter is presented in article "Meta-Learning Applied to the Selection of the Classification Methods in Industrial Images", accepted for 14º Simpósio Brasileiro de Automação Inteligente (SBAI); 27 a 30 de octubro de 2019; Ouro Preto - MG.

8 META-LEARNING FOR SELECTION OF CNN PARAMETERS APPLIED TO MEDICAL IMAGES

In previous chapters different applications of Meta-Learning have been explored, changing the techniques, images under study and sources of experience, but always with the aim of recommending the best classification methodology. This chapter uses Meta-Learning to create a CNN configuration parameter recommendation methodology for image classification, this means that Meta-Learning is used to identify the features of dataset and CNN parameter configuration that presents a better performance for classification of each type of image under study, which implies a change in the aim of recommendation, now being the configuration of parameters more suitable for a single classification technique, instead of identifying a technique within a set of different techniques.

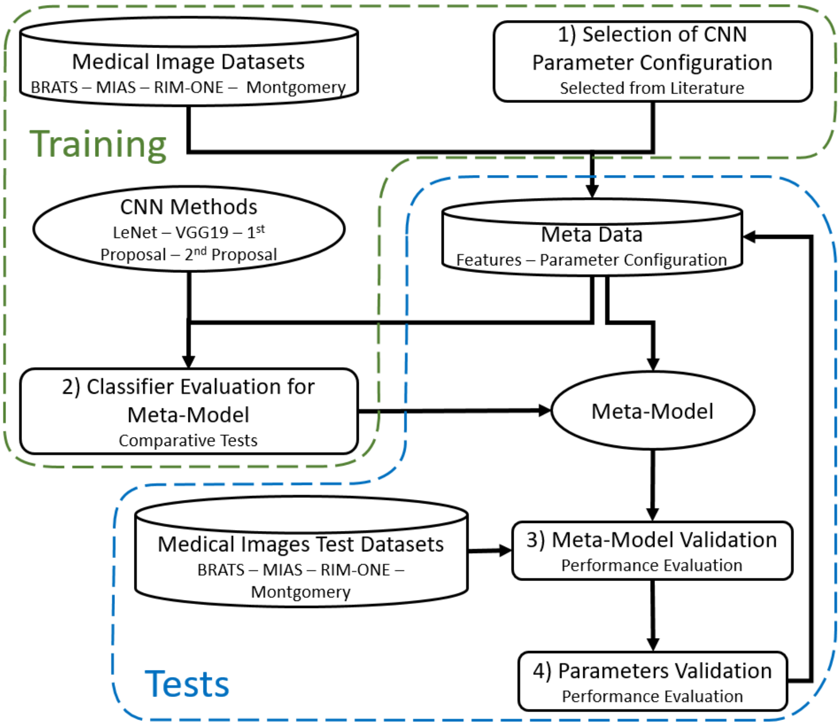
The aim of this chapter is to develop a methodology based on Meta-Learning that allows the identification of most suitable configuration parameters for a CNN-based methodology that allows the classification of different types of images with different classification objectives, understanding as configuration of most suitable parameters for those that present a better performance with respect to evaluation metrics. Used as input data, different types of medical images and different configurations of CNN reported in literature as most appropriate to classify the medical images, then relationship between images and configuration of CNN parameters is considered as experience and it is used to build Meta-Data. Finally, Meta-Model is trained using Meta-Data and has the ability to select the most appropriate CNN parameter configuration for new images.

The chapter is organized by sections, in Section 8.1 is presented the datasets used, in Section 8.2 the proposed methodology is presented, Section 8.3 shows the results, finally in Section 8.4 final considerations are presented. Figure 22 presents the proposed methodology, including datasets used.

8.1 Datasets

Four datasets of medical imaging are used, each set has different types of images with different types of classification objectives: for brain images tumors are identified, for mammography images the aim is classification of tissue abnormalities, diabetic retinopathy is identified for retinal images; finally, for X-ray images of chest, the presence of tuberculosis cases is detected. Images are used maintaining the original parameters with respect to

Figure 22 – Proposed Methodology for Meta-Learning Classifier for Parameters Configuration.



Source: Prepared by author

size, color channels, format and histogram.

BRATS The characteristics of dataset are described in Section 5.1.

MIAS The characteristics of dataset are described in Section 5.1.

RIM-ONE The characteristics of dataset are described in Section 5.1.

Montgomery The characteristics of dataset are described in Section 5.1.

8.2 Proposed Methodology

Methodology proposed consists of four steps: 1) **Selection of CNN Parameters Configuration**, according to literature, works that present configurations of CNN parameters that allow classification of each type of medical image are selected. At the end of this step, relationship between each element of **Medical Image Datasets** and selected configurations are stored in **Meta-Data**. 2) **Classifier Evaluation for Meta-Model**, two CNN selected from literature and two proposed are tested to determine the most suitable for **Meta-Model**, result of this step together with **Meta-Data** make up **Meta-Model**. 3) **Meta-Model validation**, using **Medical Images Test Datasets**, runs performance tests to validate the effectiveness of proposed methodology. 4) **Parameters Validation**, using results of Meta-Model classification, CNN configuration parameters are tested with images, validating the performance. Figure 22 shows the proposed methodology.

8.2.1 Selection of CNN Parameters Configuration

In this step, CNN parameters configurations suitable for classification of each type of medical image are identified, based on works found in literature. Some of these works consist of methodologies that include preprocessing, this information is also considered experience and is part of **Meta-Data**. It is necessary to clarify that in this step the implementation of found works is not carried out, only relation with each dataset is established. The works are selected taking into account their CNN application and performance for classification of images under study.

For **BRATS** research in state of the art indicates that Sedlar (2018) presents a work that allows the identification of brain tumor. Based on detection of local and contextual information pieces of multiple scale patches centered around the voxel. Large patches provide information about the general context, while patches in local region contain information about the details of nearby neighborhood. The model exploits information about cerebral asymmetry that often produces a tumor.

The methodology has a preprocessing in which all the non-zero voxels of a scan are normalized and cut to predefined limits. The configuration of CNN parameters for large patches uses three convolutional layers with a 5x5 matrix with 32, 64 and 125 neurons (nodes) respectively, each convolutional layer is followed by a Rectified Linear Unit (ReLU) activation and a Max Pool operation with a 2x2 matrix. For small patches, three convolutional layers of matrix 5x5 are used, and with 16, 32 and 64 neurons respectively, each convolution layer is followed by a ReLU activation. After the convolution layers in both paths, there are two layers fully connected with ReLU activation. At the end, local and contextual features are joined by two fully connected layers, the first layer is followed by a ReLU activation, and the second is responsible for carrying out the classification process, it uses a Softmax function that calculates the probability of belonging to each class.

For **MIAS** research in state of the art indicates that Bakkouri e Afdel (2018) presents a work that allows the recognition of mammographic patterns of different dimensions. The methodology used extracts representative regions of mammography using the determinant of Hessian matrix, then three different scales of images are obtained using the decomposition of Gaussian pyramid, for each scale a different CNN is used.

Configuration of first CNN scale is designed for processing of images of size 64x64, and is formed by two convolutional layers, with matrices of size 7x7 and 4x4 respectively, after each convolution layer is followed by a Max Pool layer of size 29x29x29 and 50x13x13 respectively, then two fully connected layers of size 1924 and 5 respectively. Second CNN scale is designed for processing of images of size 32x32, and is formed by two convolutional layers, with matrices of size 5x5 and 3x3 respectively, after each convolution layer is followed by a Max Pool layer of size 20x14x14 and 50x6x6 respectively, then two fully connected layers of size 247 and 5 respectively. The last scale of CNN pyramid is designed for processing of images of size 16x16, and is formed by two convolutional layers, with matrices of size 3x3 and 2x2 respectively, after each convolution layer is followed by a Max Pool layer of size 20x7x7 and 50x3x3 respectively, then two fully connected layers of size 78 and 5 respectively. All the features are processed by a softmax classifier.

For **RIM-ONE** research in state of the art indicates that Diaz-Pinto et al. (2019) presents a work that validates the configuration of VGG19 CNN is the most suitable for evaluation of glaucoma in fundus images.

VGG19 Network consists of 25 layers, all convolution layers use a 3x3 size matrix, first two convolution layers use 64 neurons each, followed by a Max Pool layer, next two convolution layers are made up of 128 neurons each, followed by a Max Pool layer, next four convolution layers use 256 neurons each, followed by a Max Pool layer, next four convolution layers are made up of 512 neurons each, followed by a max layer pool, last four convolution layers are made up of 512 neurons each, followed by a Max Pool layer, next two layers are fully connected with 4096 neurons each, followed by a fully connected layer with 1000 neurons, the final layer is softmax (SIMONYAN; ZISSERMAN, 2015).

For **Montgomery** research in state of the art indicates that Liu et al. (2018) presents a work that shows that CNN AlexNet is efficient for detection of Tuberculosis in images of chest X-ray.

AlexNet is made up of 5 convolutional layers and 3 fully connected layers, the first convolutional layer is formed by 96 neurons, 11x11 matrix with step 4, followed by a Max Pool overlapping with 3x3 matrix and step 2, the second convolutional layer is made up of 256 neurons of matrix 5x5 and step 2, followed by ovelarpping Max Pool of matrix 3x3 and step 2, the third convolutional layer is of 384 neurons, matrix 3x3 and step 1, the fourth convolutional layer is of 384 neurons, matrix 3x3 and step 1, the Fifth convolutional layer is 256 neuronal, 3x3 matrix and step 1, followed by Max Pool overlapping of 3x3 matrix and step 2, the next two fully connected layers of size 4096, followed by the last softmax layer.

Construction of **Meta-Data** that is used as a training source for **Meta-Model** is carried out by linking each image of **Medical Image Datasets** with one of four CNN configurations described in this step.

8.2.2 Classifier Evaluation for Meta-Model

Although in section 8.2.1, four different CNN configurations have been found that present a good performance when classifying each of the types of images under study, according to their respective analysis objective, it is necessary for Meta-Learning work correctly, find a classification technique that works well by classifying all types of images, since **Meta-Model** must classify all images into four classes that identify which CNN parameter configuration works best for each image.

In this step, four different CNNs are analyzed and compared to identify the most suitable for **Meta-Model**. Tests are performed using **Medical Image Datasets**, which consist of 70% of randomly selected images of each dataset presented in Section 8.1, and information that identifies the best configuration for each type of image that is found in **Selection of CNN Parameter Configuration** step (Section 8.2.1).

CNN is selected as a classifier for **Meta-Model** given its ability to perform the characterization of images without use of auxiliary methods (CHAIB et al., 2017).

Two proven methods of CNN are selected based on their use in literature. In addition, two configurations are proposed in order to obtain performance information:

1. LeNet Network presented in Section 4.2.3.
2. VGG19 Network presented in Section 8.2.1.
3. First proposed configuration consists of only a convolutional layer of 5x5 matrix with 5 neurons and a step 1, followed by a subsampling layer of size 2x2, finally a softmax layer. This configuration is designed to be simple and determine if more complex configurations are required to perform the classification.

4. Second proposed configuration consists only of a convolutional layer of 3x3 matrix with 5 neurons and a step 1, followed by a subsampling layer of size 2x2, and finally a layer of softmax. This configuration is designed to determine the effect of the size of the convolution kernel on the performance of the classification.

8.2.3 Meta-Model Validation

In this step, validation of proposed methodology is performed, using a new set of images in **Meta-Model**.

The step begins by linking **Medical Images Test Datasets**, which is composed of 30% of images of each dataset presented in Section 8.1, with most appropriate CNN configuration for each type of image according to results of **Selection of Parameter CNN Configuration** step (Section 8.2.1) Then, each of four CNNs presented in Section 8.2.2 is tested and compared, using **Meta-Data** as a training source for **Meta-Model**, finally analysis of results is performed.

8.2.4 Parameters Validation

In this step, two configurations found in state of the art are tested, applying them to images according to classification of **Meta-Model**. The purpose of validation is not to verify the results presented by authors of works selected in Section 8.2.1, the aim is to verify if Transfer Learning application is possible, which applies a model designed for a dataset in another (LI et al., 2020). Since **Meta-Model** is designed to select the best configuration of CNN parameters for different types of medical images, it is feasible to test configuration designed for a dataset and validate its performance in another, in this way the tests are performed in configurations that have already been tested before.

Validation is done using Montgomery and RIM-ONE datasets as object of study, these are selected because the configurations of CNN parameters of works found for their classification do not require preprocessing.

The step begins by applying parameter configuration recommended by **Meta-Model** for Montgomery dataset, to RIM-ONE images, results are compared with those reported by Diaz-Pinto et al. (2019), if the parameter configuration recommended for Montgomery is more effective than one recommended in Section8.2.1, means that Transfer Learning exists and **Meta-Data** must be updated. Test is repeated using CNN parameter configuration recommended by **Meta-Model** for RIM-ONE in Montgomery dataset. For test, a training set consisting of 70% of randomly selected images and 30% for set of tests are used.

8.3 Result

In this section, performance is evaluated and compared for four configurations of CNN presented in Section 8.2.2. Processing are executed using a 64-bit operating system, 2.8 Ghz Core i7 PC (16GB RAM), and Nvidia GeForce GTX 1050 GPU (2GB RAM). Processing of data and applications of CNN are carried out using Waikato Environment for Knowledge Analysis (WEKA) (WITTEN et al., 2016). This tool was selected, since allows to use various Machine Learning and Deep Learning techniques, in the same way, it allows the analysis of the extracted features and the configuration of different test parameters, all these characteristics allow to perform the necessary tests, in addition, WEKA proved to be easy to install.

Since the Meta-Model's classification technique is CNN, the resulting training **Meta-Data** is made up of images of each of datasets presented in Section 8.1, each image is linked to a methodology that describes the configuration of parameters of CNN with which a good classification performance is achieved, as explained in Section 8.2.1.

Determining the appropriate parameters configuration of CNN for each dataset, the necessary experience that Meta-Learning uses as a learning base has been developed, in this case, the experience comes from literature as presented in **Selection of CNN Parameter Configuration** step (Section 8.2.1), experience is collected and stored in **Meta-Data**, this consists of two elements:

1. Parameter configuration for CNN of each type of images presented in Section 8.2.1.
2. Given that methodology are based on CNN, the **Medical Image Datasets** used are stored.

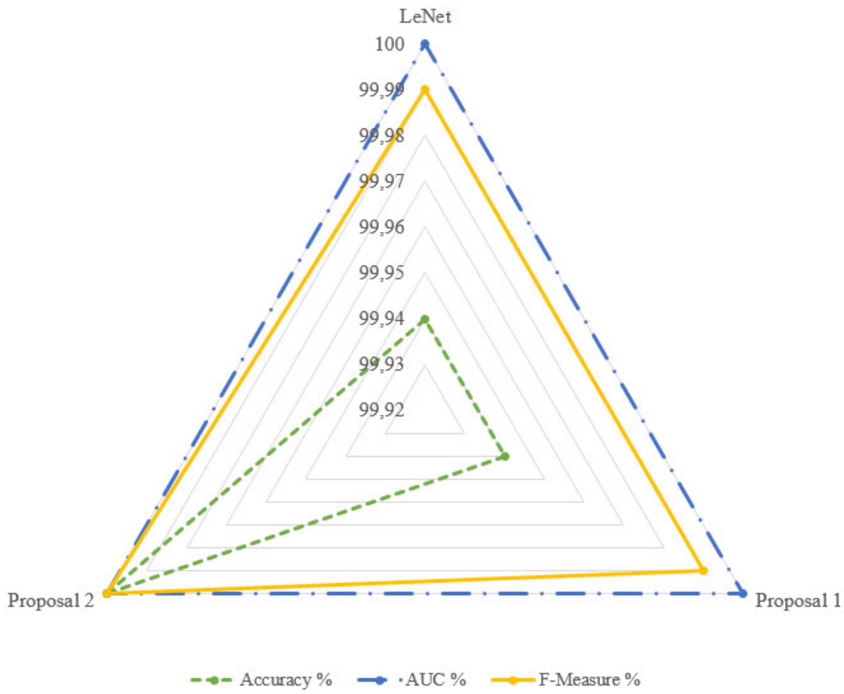
Results obtained for **Classifier Evaluation for Meta-Model** in training section shown in Figure 22 they are inconclusive, for this reason, the set of test images is used to identify the classifier for **Meta-Model**, although it involves the use of information from the Test section as it is shown in Figure 22.

Table 15 shows the average of five results of **Classifier Evaluation for Meta-Model** and **Meta-Model Validation** steps (Section 8.2.2 and Section 8.2.3). Accuracy, AUC, F-measure and processing time are used as measures of performance. Figure 23 shows the radar graph with CNN configurations that present a better performance.

Table 15 – Average Result with Set of Test Images and Meta-Data for Meta-Model.

Technique	**Accuracy %**	**AUC %**	**F-measure %**	**Processing Time/sec**
LeNet	0.999	**1**	0.999	471
VGG19	0.854	0.500	-	3740
Proposal 1	0.999	**1**	0.999	**268**
Proposal 2	**1**	**1**	**1**	287

Figure 23 – Radar Chart with Best CNN Configurations.



Source: Prepared by author

When analyzing results presented in Table 15 and Figure 23, it is found:

First, VGG19 has a significant loss of relevant information, which starts from second convolution layer and becomes critical in third layer of Max Pool, in same way it manifests that given unbalanced nature of dataset contained in **Meta-Data**, all test images are classified into a single class, which does not allow calculation of F-measure and made the VGG19 highly inappropriate as a classifier of **Meta-Model**.

Secondly, when comparing LeNet with Proposal 1, it is understood that it is only necessary to use a single convolution layer for classification of test images, in same way when execution time is considered, Proposal 1 presents a better performance.

Third, when comparing proposals 1 and 2, it is found that there is a loss of relevant information when using a larger convolution kernel size, although its use implies a shorter processing time.

Fourth, changes in kernel size and the number of convolution layers have an effect on processing time, by using smaller kernels, time increases, as well as increasing convolution layers.

Finally, when considering the types of images used as an object of study, Proposal 2 is most appropriate, since it uses a single layer of convolution with a low number of neurons, which implies that a low number of features are sufficient to carry the classification, in same way, results show that it is necessary to use a small size Kernel to preserve most important features of images. Proposal 2 presents a better performance according to evaluation measures and is the most appropriate CNN configuration for **Meta-Model**.

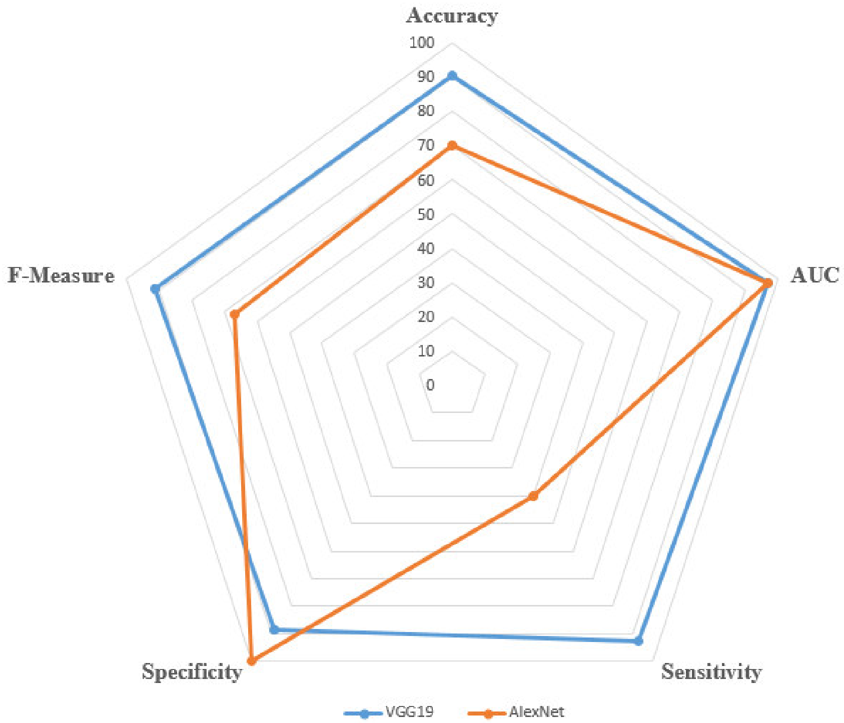
When determining CNN configuration used as a **Meta-Model** classification technique, training shown in Figure 22 is concluded.

Results validate that proposed **Meta-Model** has the ability to identify for each type of medical image being studied, the most appropriate CNN parameter configuration according to classification objective associated with each dataset.

The next step is validation of CNN configuration parameters, in which, the parameters identified by literature as the most suitable for one type of image, are applied for classification of a different type of medical image, Figure 24 shows the results obtained for RIM-ONE dataset, comparing results presented by Diaz-Pinto et al. (2019) and those obtained by using CNN parameter configuration selected by **Meta-Model** for Montgomery dataset according to **Parameters Validation** step (Section 8.2.4). Accuracy, AUC, sensitivity, specificity and F-measure are used as performance evaluation measures.

When analyzing the Figure 24 it is evident that configuration VGG19 proposed in state of the art for classification of images of retinal fundus images is the most indicated, which means that in this case the configuration designed for chest X-Ray image, does not obtain a better performance, in this way result of **Meta-Model**, correctly identifies the best parameter configuration for RIM-ONE dataset and it is not necessary to alter

Figure 24 – Result of Classification for RIM-ONE.



Source: Prepared by author

Meta-Data.

When using configuration of CNN parameter identified as most appropriate for retinal fundus images in Montgomery dataset, an accuracy of 42.26% is obtained, which is well below the result reported by Liu et al. (2018) of 85.68%. In addition, a sensitivity of 100% is obtained with a specificity of 0% that indicates that all samples are classified into a single class, in this way result of **Meta-Model**, correctly identifies the best parameter configuration for chest X-Ray images and it is not necessary to alter **Meta-Data**.

Although the results of **Validation of Parameters step** (Section 8.2.4) showed that when comparing only RIM-ONE and Montgomery datasets, **Meta-Model** correctly identifies the CNN configuration parameters, realization of verification makes possible analysis of Transfer Learning, given that proposed methodology uses Meta-Learning in different types of medical images.

8.4 Final Considerations

A methodology based on Meta-Learning for selection of CNN configuration parameters for classification of medical images is proposed, for this aim four public access datasets with different types of images, different CNN configurations informed in literature and two proposed CNN configurations are used. Based on results it is possible to affirm:

1. Proposed methodology is appropriate for selection of CNN configuration parameters applied to classification of medical images.

2. Use of CNN as a technique of characterization and classification for **Meta-Model** is efficient.

3. Use of Meta-Learning to determine configuration parameters of CNN for classification of different types of images, allows to verify the existence of Transfer Learning.

Use of Meta-Learning offers a great facility to add new types of images or change parameters of CNN configurations identified as the most suitable by new ones that present a better performance or that have been found in process of parameter verification and are result of Transfer Learning.

CNN has been used as an object of study given the large number of applications found in state of the art, however, the methodology proposed in this chapter can be applied to recommend the configuration parameters of different techniques.

The main contributions of methodology proposed in this chapter are:

1. Construction of Meta-Data for images of Brain MRI Scans, Mammograms, Retinal Fundus Image and Chest X-ray.

2. A new application of Meta-Learning in medical images to identify the best CNN parameters configuration.

3. Development of cross-test configuration model to verify the existence of Transfer Learning.

9 DISCUSSION

In previous chapters several investigations have been carried out to demonstrate the usefulness of Meta-Learning in field of image processing, each chapter presents considerations related to its particular aims, this chapter takes the considerations to perform a general analysis and present the lessons learned in research development:

- Considering that Meta-Learning is used as a recommendation system for methodologies and configuration parameters applied to different types of images, data characterization is a fundamental part of Meta-Data, if different types of images have similarities, it is necessary to improve the extraction of features, this is explained as presented in Section 5.3, where it is evident that images with similar features should have only one recommended methodology, which implies that if two different types of images have some features in common, e.g., coming from the same capture system, it is necessary that characterization discards similar features and highlights differences.

- By using different types of images in the methodologies proposed of each chapter, it is evident that Meta-Data that belong to a specific type of image, allows its update or change quickly without affecting the Meta-Data of other types of images, this means that it is possible to quickly add Meta-Data belonging to new sets of images or update a methodology whose performance is surpassed by a new one.

- Meta-Learning uses the experience collected to predict the performance of different methodologies when used in new datasets and to make a recommendation, this task is carried out based on Meta-Data. Since the experience is created for a specific type of data, Meta-Data have a static nature, however, as presented in Chapter 8, it is possible to use them to carry out a Transfer Learning process, which allows verify, if the methodology developed for a specific type of image has a better performance for another type of image than the methodology initially identified in Meta-Data.

- Based on results of Chapters 4 and 5, when comparing CNN with classification techniques that require the use of characterization, a diverse learning is obtained. 1) CNN increases computational cost, size of Meta-Data and processing time. 2) The use of characterization techniques and feature analysis allows to reduce the size of Meta-Data, however, it implies a greater analysis of the data that results in an increase in processing time. 3) Both classification techniques have high performance when used in Meta-Model.

10 CONCLUSION

In this dissertation, different applications of Meta-Learning in the field of digital image processing were explored, to achieve this aim, five investigations were carried out, using twelve different types of images belonging to industrial and medical field: breast tissue biopsy slides image, mammography, lung CT, chest X-ray, retinal fundus images, brain magnetic resonance imaging, dermatoscopic images, hot-rolled steel strip, shear pad of wagon train, welds X-ray, aluminum wheel X-ray and human faces. For each type of image, the most appropriate methodology for classification was developed or identified in the literature, creating Meta-Data and Meta-Models.

Each investigation carried out aims to demonstrate the applicability of Meta-Learning, as well as creation of its components, building Meta-Data and Meta-Models, varying the source of experience and purpose of recommendation.

Based on results it is possible to affirm:

1. Proposed methodology based on Meta-Learning makes the recommendation of the most appropriate classification methodologies for each type of image used with a high level of precision according to performance metrics.

2. Proposed methodology based on Meta-Learning is efficient for recommendation of configuration parameters in techniques based on Machine Learning, according to performance metrics.

3. The experience for construction of Meta-Data can be obtained from different sources, such as development of proposed methodologies or research reported by scientific community.

4. Proposed methodology can be used in several fields that require image processing, which allows identifying the most appropriate techniques and methodologies to perform a specific task based on nature of images.

5. Meta-Learning, when used in a process of cross-validation of Meta-Data, allows to verify the existence of Transfer Learning.

Determining the most appropriate methodologies for the execution of specific tasks, taking into account the features of the data, is an activity that implies an important consumption of resources, both personally and computationally. constantly this whole process is repeated in different investigations with images. Meta-Learning based on results obtained in this dissertation, has proven to be a methodology that allows gathering all the experience generated in various investigations to avoid reprocessing, thus creating a recommendation system that constitutes an initial starting point for future research to avoid repetition of studies already carried out.

10.1 Contributions

1. Development of a methodology that presents a better performance than those found in state of the art for identification of DC in biopsy images of breast tissue.

2. Construction of Meta-Data for Meta-Learning that allows to determine the best methodology for identification of DC in biopsy images of breast tissue.

3. Characterization of medical images through phylogenetic indexes and their analysis to determine the most representative features to be stored in the Meta-Data.

4. Meta-Model construction to identify the most suitable classification methodology for images of chest X-Ray, thoracic CT scan, retinal fundus, brain MRI and mammography.

5. Collect the experience presented in literature, transforming it into Meta-Data to be used in a classification system based on Meta-Learning.

6. using CNN as a method of characterization and classification of Meta-Model for images of thoracic CT scan, mammography and dermostopic.

7. A new application of Meta-Learning in industrial images for selection of classification methodology.

8. Application of CNN as a method for classifying and characterizing the Meta-Model for hot-rolled steel strip, the shear pad of wagon train, welds x-rays, aluminum wheel x-rays and human faces images.

9. Construction of Meta-Data for images of Brain MRI Scans, Mammograms, Retinal Fundus Image and Chest X-ray.

10. A new application of Meta-Learning in medical images to identify the best CNN parameters configuration.

11. Development of cross-test configuration model to verify the existence of Transfer Learning.

10.2 Future Works

Validation of existence of Transfer Learning is a process that allows Meta-Learning to create a new experience to feed back Meta-Data and improve recommendation, however, performing validation is an arduous process that offers several future works:

1. Creation of a system based on proposed methodology that allows the execution of recommended methodologies for each type of images.

2. Investigate and test different characterization techniques to determine the relationship between data features and existence of Transfer Learning.

3. Perform tests with new pattern recognition techniques.

4. Investigate and test feature analysis techniques applied to CNN.

10.3 Scientific Productions

Table 16 lists the scientific articles based on proposed methodology.

Table 16 – Published Articles Based on Proposed Methodology

Type	**Paper**	**Status**
Congress	Luis Fernando Marin Sepulveda, Aristófanes Correâ Silva, João Otávio Bandeira Diniz (2019). Meta-Data Construction for Selection of Breast Tissue Biopsy Slides Image Classifier to Identify Ductal Carcinoma. In Brazilian Conference on Intelligent Systems (BRACIS 2019).	Published
Congress	Luis Fernando Marin Sepulveda, Aristófanes Correâ Silva, João Otávio Bandeira Diniz (2019). Meta-Learning Applied to the Selection of the Classification Methods in Industrial Images. In 14º Simpósio Brasileiro de Automação Inteligente (SBAI).	Published

REFERENCES

ALAYON, S.; GONZALEZ, D. L. R. M.; FUMERO, F. J.; SAAVEDRA, J. F. S.; SANCHEZ, J. L. Variability between experts in defining the edge and area of the optic nerve head. **Archivos de la Sociedad Española de Oftalmología (English Edition)**, 2013. ISSN 2173-5794.

AMIRI, S. M.; POURAZAD, M. T.; NASIOPOULOS, P.; LEUNG, V. C. Human action recognition using meta learning for RGB and depth information. In: **2014 International Conference on Computing, Networking and Communications, ICNC 2014**. [S.l.: s.n.], 2014.

ARAUJO, T.; ARESTA, G.; CASTRO, E.; ROUCO, J.; AGUIAR, P.; ELOY, C.; POLONIA, A.; CAMPILHO, A. Classification of breast cancer histology images using convolutional neural networks. **PLoS ONE**, v. 12, n. 6, 2017.

ARMATO, S. G.; MCLENNAN, G.; BIDAUT, L.; MCNITT-GRAY, M. F.; MEYER, C. R.; REEVES, A. P.; ZHAO, B.; ABERLE, D. R.; HENSCHKE, C. I.; HOFFMAN, E. A.; KAZEROONI, E. A.; MACMAHON, H.; VAN BEEK, E. J.; YANKELEVITZ, D.; BIANCARDI, A. M.; BLAND, P. H.; BROWN, M. S.; ENGELMANN, R. M.; LADERACH, G. E.; MAX, D.; PAIS, R. C.; QING, D. P.; ROBERTS, R. Y.; SMITH, A. R.; STARKEY, A.; BATRA, P.; CALIGIURI, P.; FAROOQI, A.; GLADISH, G. W.; JUDE, C. M.; MUNDEN, R. F.; PETKOVSKA, I.; QUINT, L. E.; SCHWARTZ, L. H.; SUNDARAM, B.; DODD, L. E.; FENIMORE, C.; GUR, D.; PETRICK, N.; FREYMANN, J.; KIRBY, J.; HUGHES, B.; VANDE CASTEELE, A.; GUPTE, S.; SALLAM, M.; HEATH, M. D.; KUHN, M. H.; DHARAIYA, E.; BURNS, R.; FRYD, D. S.; SALGANICOFF, M.; ANAND, V.; SHRETER, U.; VASTAGH, S.; CROFT, B. Y.; CLARKE, L. P. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. **Medical Physics**, 2011.

BAKAS, S.; AKBARI, H.; SOTIRAS, A.; BILELLO, M.; ROZYCKI, M.; KIRBY, J. S.; FREYMANN, J. B.; FARAHANI, K.; DAVATZIKOS, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. **Scientific Data**, 2017. ISSN 20524463.

BAKKOURI, I.; AFDEL, K. Multi-scale cnn based on region proposals for efficient breast abnormality recognition. **Multimedia Tools and Applications**, p. 1–22, 2018.

BAXEVANIS, A. D.; OUELLETTE, B. F. F. **BIOINFORMATICS A Practical Guide to the Analysis of Genes and Proteins SECOND EDITION**. Wiley. [S.l.: s.n.], 2004. 504 p. ISBN 0471383902.

BELGIU, M.; DRăGUţ, L. Random forest in remote sensing: A review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, apr 2016. ISSN 09242716. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0924271616000265>.

BOOTS, B. N.; GETIS, A. Point pattern analysis. **Sage Publications**, 1988.

BRAZDIL, P.; CARRIER, C. G.; SOARES, C.; VILALTA, R. **Metalearning: Applications to Data Mining**. [S.l.: s.n.], 2008. ISSN 1611-2482. ISBN 978-3-540-73262-4.

BRAZDIL, P.; GIRAUD-CARRIER, C. Metalearning and Algorithm Selection: progress, state of the art and introduction to the 2018 Special Issue. **Machine Learning**, Springer US, v. 107, n. 1, p. 1–14, 2018.

CAMPOS, G. F.; BARBON, S.; MANTOVANI, R. G. A meta-learning approach for recommendation of image segmentation algorithms. **Proceedings - 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI 2016**, p. 370–377, 2017.

CANDEMIR, S.; JAEGER, S.; PALANIAPPAN, K.; MUSCO, J. P.; SINGH, R. K.; XUE, Z.; KARARGYRIS, A.; ANTANI, S.; THOMA, G.; MCDONALD, C. J. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. **IEEE Transactions on Medical Imaging**, 2014. ISSN 02780062.

CHAIB, S.; YAO, H.; GU, Y.; AMRANI, M. Deep feature extraction and combination for remote sensing image classification based on pre-trained CNN models. **Proceedings of SPIE - The International Society for Optical Engineering**, v. 10420, 2017.

CHEPLYGINA, V.; MOESKOPS, P.; VETA, M.; DASHTBOZORG, B.; PLUIM, J. P. Exploring the Similarity of Medical Imaging Classification Problems. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2017.

CLARKE, K. R.; WARWICK, R. M.; ; R.M.WARWICK, K. **Change in marine communities: An Approach to Statistical Analysis and Interpretation**. [S.l.: s.n.], 2007.

CLARKE, K. R.; WARWICK, R. M. A taxonomic distinctness index and its statistical properties. **Journal of Applied Ecology**, 1998.

CUNHA, T.; SOARES, C.; CARVALHO, A. C. de. Metalearning and Recommender Systems: A literature review and empirical study on the algorithm selection problem for Collaborative Filtering. **Information Sciences**, Elsevier Inc., v. 423, p. 128–144, 2018.

DE MELO, C. E. C.; PRUDÊNCIO, R. B. C. Cost-sensitive measures of algorithm similarity for meta-learning. **2014 Brazilian Conference on Intelligent Systems, BRACIS 2014**, p. 7–12, 2014.

DE MORAIS, R. F.; MIRANDA, P. B.; SILVA, R. M. A Meta-Learning Method to Select Under-Sampling Algorithms for Imbalanced Data Sets. **2016 5th Brazilian Conference on Intelligent Systems, BRACIS 2016**, p. 385–390, 2017.

DENG, Z.; ZHU, X.; CHENG, D.; ZONG, M.; ZHANG, S. Efficient kNN classification algorithm for big data. **Neurocomputing**, 2016.

DHAHBI, S.; BARHOUMI, W.; ZAGROUBA, E. Breast cancer diagnosis in digitized mammograms using curvelet moments. **Computers in Biology and Medicine**, Elsevier, v. 64, p. 79–90, 2015.

DIAZ-PINTO, A.; MORALES, S.; NARANJO, V.; KöHLER, T.; MOSSI, J.; NAVEA, A. Cnns for automatic glaucoma assessment using fundus images: An extensive validation. **BioMedical Engineering Online**, v. 18, n. 1, 2019.

DRISS, S. B.; SOUA, M.; KACHOURI, R.; AKIL, M. A comparison study between mlp and convolutional neural network models for character recognition. In: N., C. M. K. (Ed.). [S.l.]: SPIE, 2017. v. 10223.

EKMAN, P.; FRIESEN, W. V. Measuring facial movement. **Environmental Psychology and Nonverbal Behavior**, 1976. ISSN 03613496.

FAITH, D. P. Conservation evaluation and phylogenetic diversity. **Biological Conservation**, 1992. ISSN 00063207.

_____. Phylogenetic Pattern and the Quantification of Organismal Biodiversity. **Philosophical Transactions of the Royal Society B: Biological Sciences**, 1994.

GARDNER, M. W.; DORLING, S. R. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. **Atmospheric Environment**, v. 32, n. 14-15, p. 2627–2636, 1998.

GELASCA, E. D.; BYUN, J.; OBARA, B.; MANJUNATH, B. **Evaluation and Benchmark for Biological Image Segmentation**. [S.l.: s.n.], 2008.

GIRAUD-CARRIER, C. Metalearning - A Tutorial. **Tutorial at the 7th international conference on machine learning and applications (ICMLA)**, San Diego, California, USA, n. December, 2008.

GONZALEZ, R. C.; WOODS, R. E. **Digital image processing**. Upper Saddle River, N.J.: Prentice Hall, 2008. ISBN 9780131687288 013168728X 9780135052679 013505267X. Available: <http://www.amazon.com/Digital-Image-Processing-3rd-Edition/dp/013168728X>.

GRABCZEWSKI, K. Meta-learning in decision tree induction. **Studies in Computational Intelligence**, Springer International Publishing, Cham, v. 498, p. 1–358, 2014. ISSN 1860949X. Available: <http://link.springer.com/10.1007/978-3-319-00960-5>.

HALL, M.; SMITH, L. A. Feature Selection for Machine Learning : Comparing a Correlation-based Filter Approach to the Wrapper CFS : Correlation-based Feature. **International FLAIRS Conference**, 1999.

HEATH, M.; BOWYER, K.; KOPANS, D.; KEGELMEYER, W.; MOORE, R.; CHANG, K.; MUNISHKUMARAN, S. Current Status of the Digital Database for Screening Mammography. In: **Digital Mammography. Computational Imaging and Vision**. [S.l.]: Springer, Dordrecht, 1998. v. 13, p. 457–460.

HUISKES, B. T. M. J.; LEW, M. S. New trends and ideas in visual concept detection: The mir flickr retrieval evaluation initiative. In: **MIR '10: Proceedings of the 2010 ACM International Conference on Multimedia Information Retrieval**. New York, NY, USA: ACM, 2010. p. 527–536.

HUISKES, M. J.; LEW, M. S. The mir flickr retrieval evaluation. In: **MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval**. New York, NY, USA: ACM, 2008.

IZSÁK, J.; PAPP, L. A link between ecological diversity indices and measures of biodiversity. **Ecological Modelling**, 2000.

JAEGER, S.; KARARGYRIS, A.; CANDEMIR, S.; FOLIO, L.; SIEGELMAN, J.; CALLAGHAN, F.; XUE, Z.; PALANIAPPAN, K.; SINGH, R. K.; ANTANI, S.; THOMA, G.; WANG, Y. X.; LU, P. X.; MCDONALD, C. J. Automatic tuberculosis screening using chest radiographs. **IEEE Transactions on Medical Imaging**, 2014. ISSN 02780062.

JIANG, L.; ZHANG, L.; YU, L.; WANG, D. Class-specific attribute weighted naive Bayes. **Pattern Recognition**, Elsevier Ltd, v. 88, p. 321–330, 2019. ISSN 00313203. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0031320318304205>.

JOLLIFFE, I. T. **Principal Component Analysis, Second Edition**. [S.l.: s.n.], 2002. ISSN 00401706. ISBN 0387954422.

KARGER, D.; LEHMAN, E.; LEIGHTON, T.; PANIGRAHY, R.; LEVINE, M.; LEWIN, D. Consistent hashing and random trees. In: **Proceedings of the twenty-ninth annual ACM symposium on Theory of computing - STOC '97**. New York, New York, USA: ACM Press, 1997. p. 654–663. ISBN 0897918886. ISSN 0012821X. Available: <http://portal.acm.org/citation.cfm?doid=258533.258660>.

KEITH, M.; CHIMIMBA, C. T.; REYERS, B.; JAARSVELD, A. S. V. Taxonomic and phylogenetic distinctiveness in regional conservation assessments: A case study based on extant South African Chiroptera and Carnivora. **Animal Conservation**, 2005. ISSN 13679430.

KHAKZAD, N. Application of dynamic bayesian network to risk analysis of domino effects in chemical infrastructures. **Reliability Engineering and System Safety**, v. 138, p. 263–272, 2015.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. 2015.

LEE, R. S.; GIMENEZ, F.; HOOGI, A.; MIYAKE, K. K.; GOROVOY, M.; RUBIN, D. L. Data Descriptor: A curated mammography data set for use in computer-aided detection and diagnosis research. **Scientific Data**, 2017.

LEMKE, C.; BUDKA, M.; GABRYS, B. Metalearning: a survey of trends and technologies. **Artificial Intelligence Review**, 2015. ISSN 15737462.

LI, J.; HUANG, S.; ZHANG, X.; FU, X.; CHANG, C.-C.; TANG, Z.; LUO, Z. Facial expression recognition by transfer learning for small datasets. **Advances in Intelligent Systems and Computing**, v. 895, p. 756–770, 2020.

LIRA, M. M.; DE AQUINO, R. R.; FERREIRA, A. A.; CARVALHO, M. A.; NÓBREGA NETO, O.; SANTOS, G. S. Combining multiple artificial neural networks using random committee to decide upon electrical disturbance classification. In: **IEEE International Conference on Neural Networks - Conference Proceedings**. [S.l.: s.n.], 2007. ISBN 142441380X. ISSN 10987576.

LIU, C.; CAO, Y.; ALCANTARA, M.; LIU, B.; BRUNETTE, M.; PEINADO, J.; CURIOSO, W. Tx-cnn: Detecting tuberculosis in chest x-ray images using convolutional neural network. In: . [S.l.: s.n.], 2018. v. 2017-September, p. 2314–2318.

MAGURRAN, A. E. **Measuring biological diversity: Rejoinder**. [S.l.: s.n.], 2004. 285–286 p. ISSN 1608-5914. ISBN 0632056339.

MAICAS, G.; BRADLEY, A. P.; NASCIMENTO, J. C.; REID, I.; CARNEIRO, G. Training medical image analysis systems like radiologists. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2018.

MANTOVANI, R. G.; ROSSI, A. L.; VANSCHOREN, J.; CARVALHO, A. C. Meta-learning recommendation of default hyper-parameter values for SVMs in classifications tasks. In: **CEUR Workshop Proceedings**. [S.l.: s.n.], 2015.

MENDONCA, T.; FERREIRA, P. M.; MARQUES, J. S.; MARCAL, A. R.; ROZEIRA, J. PH2- A dermoscopic image database for research and benchmarking. In: **Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS**. [S.l.: s.n.], 2013.

MENZE, B. H.; JAKAB, A.; BAUER, S.; KALPATHY-CRAMER, J.; FARAHANI, K.; KIRBY, J.; BURREN, Y.; PORZ, N.; SLOTBOOM, J.; WIEST, R.; LANCZI, L.; GERSTNER, E.; WEBER, M. A.; ARBEL, T.; AVANTS, B. B.; AYACHE, N.; BUENDIA, P.; COLLINS, D. L.; CORDIER, N.; CORSO, J. J.; CRIMINISI, A.; DAS, T.; DELINGETTE, H.; DEMIRALP, Ç.; DURST, C. R.; DOJAT, M.; DOYLE, S.; FESTA, J.; FORBES, F.; GEREMIA, E.; GLOCKER, B.; GOLLAND, P.; GUO, X.; HAMAMCI, A.; IFTEKHARUDDIN, K. M.; JENA, R.; JOHN, N. M.; KONUKOGLU, E.; LASHKARI, D.; MARIZ, J. A.; MEIER, R.; PEREIRA, S.; PRECUP, D.; PRICE, S. J.; RAVIV, T. R.; REZA, S. M.; RYAN, M.; SARIKAYA, D.; SCHWARTZ, L.; SHIN, H. C.; SHOTTON, J.; SILVA, C. A.; SOUSA, N.; SUBBANNA, N. K.; SZEKELY, G.; TAYLOR, T. J.; THOMAS, O. M.; TUSTISON, N. J.; UNAL, G.; VASSEUR, F.; WINTERMARK, M.; YE, D. H.; ZHAO, L.; ZHAO, B.; ZIKIC, D.; PRASTAWA, M.; REYES, M.; VAN LEEMPUT, K. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). **IEEE Transactions on Medical Imaging**, 2015. ISSN 1558254X.

MERCAN, E.; MEHTA, S.; BARTLETT, J.; WEAVER, D.; ELMORE, J.; SHAPIRO, L. Automated diagnosis of breast cancer and pre-invasive lesions on digital whole slide images. In: . [S.l.: s.n.], 2018. v. 2018-January, p. 60–68.

MERY, D.; FILBERT, D. Automated flaw detection in aluminum castings based on the tracking of potential defects in a radioscopic image sequence. **IEEE Transactions on Robotics and Automation**, 2002. ISSN 1042296X.

MERY, D.; RIFFO, V.; ZSCHERPEL, U.; MONDRAGÓN, G.; LILLO, I.; ZUCCAR, I.; LOBEL, H.; CARRASCO, M. GDXray: The Database of X-ray Images for Nondestructive Testing. **Journal of Nondestructive Evaluation**, 2015. ISSN 15734862.

MOURA, N.; VERAS, R.; AIRES, K.; SANTOS, L.; MACHADO, V. Proposta de um descritor híbrido para aprimoramento da identificação automática de melanoma. In: **17º Workshop de Informática Médica (WIM 2017)**. Porto Alegre, RS, Brasil: SBC, 2017. v. 17. Available: <http://portaldeconteudo.sbc.org.br/index.php/sbcas/article/view/3721>.

NETO, A. C. D. S.; RAMOS, A. R. C.; FILHO, A. O. D. C.; SOUSA, A. D. D.; DRUMOND, P. M. L. D. L. Desenvolvimento de descritores baseado em análise de forma para diagnóstico de lesões pulmonares. In: **17º Workshop de Informática Médica (WIM 2017)**. [S.l.]: SBC, 2017. v. 17.

NETO, A. C. S.; DINIZ, P. H.; DINIZ, J. O.; CAVALCANTE, A. B.; SILVA, A. C.; PAIVA, A. C. D.; ALMEIDA, J. D. D. Diagnosis of Non-Small Cell Lung Cancer Using Phylogenetic Diversity in Radiomics Context. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2018.

PENA-BETANCOR, C.; GONZALEZ-HERNANDEZ, M.; FUMERO-BATISTA, F.; SIGUT, J.; MEDINA-MESA, E.; ALAYON, S.; DE LA ROSA, M. G. Estimation of the relative amount of hemoglobin in the cup and neuroretinal rim using stereoscopic color fundus images. **Investigative Ophthalmology and Visual Science**, 2015. ISSN 15525783.

PIMENTEL, B. A.; CARVALHO, A. C. D. A new data characterization for selecting clustering algorithms using meta-learning. **Information Sciences**, Elsevier Inc., v. 477, p. 203–219, 2019. ISSN 00200255. Available: <https://doi.org/10.1016/j.ins.2018.10.043>.

POSADAS, P.; ESQUIVEL, D. R. M.; CRISCI, J. V. Using phylogenetic diversity measures to set priorities in conservation: An example from southern South America. **Conservation Biology**, 2001. ISSN 08888892.

POWERS, D. M. W. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. **Journal of Machine Learning Technologies**, v. 2, n. 1, p. 37–63, 2011.

QUINLAN, J. R. Improved Use of Continuous Attributes in C4.5. **Journal of Artificial Intelligence Research**, v. 4, p. 77–90, mar 1996. ISSN 1076-9757. Available: <https://jair.org/index.php/jair/article/view/10157>.

RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. **Proceedings of the IEEE**, v. 77, n. 2, p. 257–286, 1989. ISSN 00189219. Available: <http://ieeexplore.ieee.org/document/18626/>.

RAHMAN, M. M.; BHATTACHARYA, P. Biomedical image classification with multi response linear regression (MLR) as meta-learner combiner and its effectiveness on small to large data sets. In: **Proceedings - 2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016**. [S.l.: s.n.], 2017.

REYNOLDS, D. Gaussian mixture models. In: _____. **Encyclopedia of Biometrics**. Boston, MA: Springer US, 2015. p. 827–832. ISBN 978-1-4899-7488-4. Available: <https://doi.org/10.1007/978-1-4899-7488-4_196>.

RIVERA, H.; GOULART, C.; FAVARATO, A.; VALADÃO, C.; CALDEIRA, E.; BASTOS, T. Development of an Automatic Expression Recognition System Based on. **Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS_CSBC)**, p. 615–620, 2017.

ROCHA, R.; SIRAVENHA, A.; C. S. GOMES, A.; L. SEREJO, G.; SILVA, A.; MOUSINHO RODRIGUES, L.; BRAGA, J.; DIAS, G.; CARVALHO, S. Avaliação de técnicas de Deep Learning aplicadas à identificação de peças defeituosas em vagões de trem. In: CLUA, E.; PÁDUA, F. L. C. (Ed.). **Workshop of Industry Applications (WIA) in the 30th Conference on Graphics, Patterns and Images (SIBGRAPI'17)**. Niterói, RJ, Brazil: [s.n.], 2017. Available: <http://sibgrapi2017.ic.uff.br/>.

RODRIGUES, A. S. L.; GASTON, K. J. Maximising phylogenetic diversity in the selection of networks of conservation areas. **Biological Conservation**, 2002. ISSN 00063207.

SAMMUT, C.; WEBB, G. I. **Encyclopedia of Machine Learning and Data Mining**. 2nd. ed. [S.l.]: Springer Publishing Company, Incorporated, 2017. ISBN 148997685X, 9781489976857.

SANTOS, A. B.; DE ARAUJO, A. A.; DOS SANTOS, J. A.; SCHWARTZ, W. R.; MENOTTI, D. Combination techniques for hyperspectral image interpretation. In: **International Geoscience and Remote Sensing Symposium (IGARSS)**. IEEE, 2017. v. 2017-July, p. 3648–3651. ISBN 9781509049516. Available: <http://ieeexplore.ieee.org/document/8127789/>.

SEDLAR, S. Brain tumor segmentation using a multi-path cnn based method. **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**, v. 10670 LNCS, p. 403–422, 2018.

SIMONYAN, K.; ZISSERMAN, A. Very deep ConvNets for large-scale image recognition. **CoRR**, 2014.

______. Very deep convolutional networks for large-scale image recognition. In: **International Conference on Learning Representations**. [S.l.: s.n.], 2015.

SMITH, M. R.; MITCHELL, L.; GIRAUD-CARRIER, C.; MARTINEZ, T. Recommending learning algorithms and their associated hyperparameters. In: **CEUR Workshop Proceedings**. [S.l.: s.n.], 2014.

SOARES, L. B.; ATILA, W.; GUTERRES, B.; NAGEL RODRIGUES, R.; BOTELHO, S.; DREWS-JR, L. J, P.; MOR, L, J.; FONSECA, T. Sistema De Visão Computacional Para Análise Geométrica De Cordões De Solda. In: **XIII Simpósio Brasileiro de Automação Inteligente**. [S.l.: s.n.], 2017.

SONG, K.; YAN, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. **Applied Surface Science**, v. 285, p. 858–864, nov 2013. ISSN 01694332. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0169433213016437>.

SPANHOL, F.; OLIVEIRA, L.; PETITJEAN, C.; HEUTTE, L. A dataset for breast cancer histopathological image classification. In: . [S.l.]: IEEE Computer Society, 2016. v. 63, n. 7, p. 1455–1462.

SUCKLING, J.; PARKER, J.; DANCE, D. The mammographic image analysis society digital mammogram database. In: **Exerpta Medica. International Congress Series.** [S.l.: s.n.], 1994. ISBN 0531-5131.

SUGUMARAN, V.; MURALIDHARAN, V.; RAMACHANDRAN, K. Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing. **Mechanical Systems and Signal Processing**, v. 21, n. 2, p. 930–942, feb 2007. ISSN 08883270. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0888327006001142>.

SUMNER, M.; FRANK, E.; HALL, M. Speeding up Logistic Model Tree Induction. In: **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)**. [S.l.: s.n.], 2005.

TANFILEV, I.; FILCHENKOV, A.; SMETANNIKOV, I. Feature selection algorithm ensembling based on meta-learning. In: **Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017**. IEEE, 2018. v. 2018-Janua, p. 1–6. ISBN 9781538619377. Available: <http://ieeexplore.ieee.org/document/8302301/>.

VANE-WRIGHT, R. I.; HUMPHRIES, C. J.; WILLIAMS, P. H. What to protect?-Systematics and the agony of choice. **Biological Conservation**, 1991. ISSN 00063207.

VAPNIK, V. N. **The Nature of Statistical Learning Theory. Springer-Verlag**. [S.l.: s.n.], 1995. ISSN 10762787. ISBN 0-471-03003-1.

WEI, W.; LIANG, J. Information fusion in rough set theory : An overview. **Information Fusion**, v. 48, p. 107–118, aug 2019. ISSN 15662535. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1566253518300071>.

WEISS, K.; KHOSHGOFTAAR, T.; WANG, D. A survey of transfer learning. SpringerOpen, v. 3, n. 1, 2016.

WEITZMAN, M. L. On diversity. **The Quarterly Journal of Economics**, v. 107, p. 363–405, 1992.

WITTEN, I.; FRANK, E.; HALL, M.; PAL, C. Data mining: Practical machine learning tools and techniques. **Elsevier Inc**, p. 1–621, 2016.

XIA, Z.; WANG, X.; ZHANG, L.; QIN, Z.; SUN, X.; REN, K. A Privacy-Preserving and Copy-Deterrence Content-Based Image Retrieval Scheme in Cloud Computing. **IEEE Transactions on Information Forensics and Security**, v. 11, n. 11, p. 2594–2608, nov 2016. ISSN 1556-6013. Available: <http://ieeexplore.ieee.org/document/7511677/>.

YADAV, A. K.; CHANDEL, S. S. Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model. **Renewable Energy**, 2015. ISSN 18790682.