Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas

Guilherme Freire Roberto

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2021

Guilherme Freire Roberto

Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas

Tese de doutorado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Doutor em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Marcelo Zanchetta do Nascimento
Coorientador: Leandro Alves Neves

Uberlândia
2021

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

R642
2021

Roberto, Guilherme Freire, 1993-
Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas [recurso eletrônico] / Guilherme Freire Roberto. - 2021.

Orientador: Marcelo Zanchetta do Nascimento.
Coorientador: Leandro Alves Neves.
Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.te.2021.656
Inclui bibliografia.
Inclui ilustrações.

1. Computação. I. Nascimento, Marcelo Zanchetta do, 1976-, (Orient.). II. Neves, Leandro Alves,1975-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. IV. Título.

CDU: 681.3

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Ciência da Computação
Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpgfacom@ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Tese de Doutorado, 25/2021, PPGCO				
Data:	04 de novembro de 2021	Hora de início:	08:40	Hora de encerramento:	11:45
Matrícula do Discente:	11723CCP004				
Nome do Discente:	Guilherme Freire Roberto				
Título do Trabalho:	Associação entre atributos manuais e aprendizado profundo baseada em geometria fractal para classificação de imagens histológicas				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciências de Dados				
Projeto de Pesquisa de vinculação:	-				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Rita Maria da Silva Julia - FACOM/UFU, Maria Camila Nardini Barioni - FACOM/UFU, Rodrigo Pereira Ramos - CPGCSB/ UNIVASF, Ana Carolina Lorena - PPGPO/ITA, Leandro Alves Neves - DCCE/UNESP (Coorientador) e Marcelo Zanchetta do Nascimento - FACOM/UFU orientador do candidato.

Os examinadores participaram desde as seguintes localidades: Rodrigo Pereira Ramo - Juazeiro/BA; Ana Carolina Lorena - São José dos Campos/SP; Leandro Alves Neves - São José do Rio Preto/SP; Rita Maria da Silva Julia, Maria Camila Nardini Barioni e Marcelo Zanchetta do Nascimento - Uberlândia/MG. O discente participou da cidade de Uberlândia/MG.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. Marcelo Zanchetta do Nascimento, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Maria Camila Nardini Barioni**, **Professor(a) do Magistério Superior**, em 04/11/2021, às 14:57, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Ana Carolina Lorena**, **Usuário Externo**, em 05/11/2021, às 08:51, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Rodrigo Pereira Ramos**, **Usuário Externo**, em 05/11/2021, às 09:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Leandro Alves Neves**, **Usuário Externo**, em 05/11/2021, às 11:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Rita Maria da Silva Julia**, **Professor(a) do Magistério Superior**, em 06/11/2021, às 16:25, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Marcelo Zanchetta do Nascimento**, **Professor(a) do Magistério Superior**, em 08/11/2021, às 08:21, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3143384** e o código CRC **310C7B61**.

Referência: Processo nº 23117.072950/2021-67 SEI nº 3143384

Criado por ludimila.andrade, versão 10 por ludimila.andrade em 04/11/2021 13:30:41.

À minha mãe Maria de Fátima

Ao meu pai Walter

À minha irmã Thaís

Às minhas tias

Agradecimentos

Agradeço primeiramente à minha família pelo apoio incondicional que sempre me deram desde quando decidi seguir na carreira acadêmica.

Agradeço ao meu orientador Prof. Dr. Marcelo Zanchetta do Nascimento e meu co-orientador Prof. Dr. Leandro Alves Neves que sempre se dispuseram a me auxiliar de todas as formas possíveis em minha trajetória desde antes do início deste doutorado.

Agradeço aos amigos Álvaro Magri, Bruna Tibúrcio, Carolina da Costa, Cléber Oliveira, David Candelero, Gabriela Camilo, Guilherme Botazzo, Jaqueline Tenguam, Leandro Valentim, Leonardo Longo, Matheus Carreira, Matheus Gonçalves, Matheus Popp e Rafaela Polizel, cuja amizade tornaram ainda mais agradável o ambiente de pesquisa dentro do laboratório.

Agradeço aos amigos Beatriz Maniero, Bianca Oliveira, Bruno Caldeira, Carol Signorini, Juliana Garcia, Mariele Carvalho, Renan Cebola e Vinícius Cucolo pela amizade e por sempre estarem ao meu lado. E um agradecimento mais que especial aos senhores Daniel Belzer e Igor Blanco por me permitirem usufruir do espaço da Mansão Piruleta.

Agradeço à todos do GEU-UFU e do LIPAI, por fazerem com que eu me sentisse em casa desde o primeiro momento em que cheguei em Uberlândia.

Agradeço à Profa. Dra. Alessandra Lumini por todo o apoio dedicado à este trabalho e por ter me recebido tão bem como membro de seu grupo de pesquisa na *Università di Bologna*. Agradeço também aos colegas do BioLab e do 3022 pela recepção e pelos momentos de descontração durante os horários de almoço. E claro, não posso deixar de esquecer Francisco Carranza, Juliane Machado e Tayná Trigo pela amizade em Cesena. E mais um agradecimento especial a Nicola Corrado, por todo o apoio durante o período de *lockdown*.

Agradeço ao Prof. Dr. Alessandro Santana Martins, ao Dr. Paulo R. Faria e à Profa. Dra. Thaína Aparecida Azevedo Tosta pelo apoio no desenvolvimento deste trabalho.

Finalmente, agradeço o apoio financeiro fornecido pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) referente aos projetos #88882.461704/2019-01 e #88882.429128/2019-01, Fundação de Amparo à Pesquisa do Estado de Minas Gerais

referente aos projetos APQ-00578-18 e APQ-01129-21, ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) referente aos projetos #304848/2018-2, #430965/2018-4 e #313365/2018-0 e ao IFTM referente ao Projeto Edital Nº 01/2020.

"A PhD is a doctorate. It's literally describing a doctor."
Captain Raymond Holt

Resumo

A classificação de imagens histológicas é uma tarefa que tem sido amplamente explorada nas recentes pesquisas de visão computacional. A abordagem mais estudada para esta tarefa tem sido a aplicação de aprendizado profundo por meio de modelos de CNN. Entretanto, o uso de CNN no contexto da classificação de imagens histológicas tem ainda algumas limitações, como a necessidade de grandes conjuntos de dados e a dificuldade de implementar um modelo generalizado capaz de classificar diferentes tipos de tecido histológico. Neste trabalho, propõe-se um modelo de comitê de classificadores baseado em atributos fractais e aprendizado profundo que consiste em combinar a classificação de duas CNN e a classificação de atributos manuais locais e globais aplicando a regra da soma. A extração das características é aplicada para obter 300 atributos fractais de diferentes conjuntos de dados histológicos. Estes atributos são reorganizados em uma matriz a fim de compor uma imagem RGB. São avaliados quatro procedimentos diferentes para efetuar esta reorganização, que geram modelos de representação dos atributos fractais que são dados como entrada para uma primeira CNN. Outra CNN recebe como entrada a imagem original correspondente. Depois de combinar os resultados de ambas as CNN com a classificação dos atributos manuais utilizando abordagens clássicas de aprendizado de máquina, foram obtidas acurácias que variam de 88,45% a 99,77% em cinco conjuntos de dados diferentes. Além disso, o modelo foi capaz de classificar imagens de conjuntos de dados com classes desbalanceadas, sem a necessidade de imagens possuírem a mesma resolução, e com um treinamento de 10 épocas. Também foi verificado que os resultados obtidos são complementares aos estudos mais relevantes publicados recentemente no contexto da classificação de imagens histológicas.

Palavras-chave: aprendizado profundo; atributos fractais; comitê de classificadores; imagens histológicas.

Ensemble of handcrafted and deep learning features based on fractal geometry for the classification of histology images

Guilherme Freire Roberto

Universidade Federal de Uberlândia
Faculdade de Computação
Programa de Pós-Graduação em Ciência da Computação

Uberlândia
2021

Abstract

Classification of histology images is a task that has been widely explored on recent computer vision researches. The most studied approach for this task has been the application of deep learning through CNN models. However, the use of CNN in the context of histological images classification has yet some limitations such as the need of large datasets and the difficulty to implement a generalized model able to classify different types of histology tissue. In this project, an ensemble model based on handcrafted fractal features and deep learning that consists on combining the classification of two CNN and the classification of local and global handcrafted features by applying the sum rule is proposed. Feature extraction is applied to obtain 300 fractal features from different histological datasets. These features are reshaped into a matrix in order to compose an RGB feature image. Four different reshaping procedures are evaluated, wherein each generates a representation model of fractal features which is given as input to a CNN. Another CNN receives as input the correspondent original image. After combining the results of both CNN with the classification of the handcrafted features using classical machine learing approaches, accuracies that range from 88.45% up to 99.77% on five different datasets were obtained. Moreover, the model was able to classify images from datasets with imbalanced classes, without the need of images having the same resolution, and using 10 epochs for training. It was also verified that the obtained results are complementary to the most relevant studies recently published in the context of histology image classification.

Keywords: deep learning; fractal features; classification ensemble; histology images.

List of Figures

List of Tables

Acronyms list

ADASYN Adaptive synthetic sampling approach

ANN Artificial neural network

CAD Computer aided diagnosis

CLL Chronic lymphocyte leukemia

CNN Convolutional neural networks

DT Decision tree

FD Fractal dimension

FL Follicular lymphoma

GA Genetic algorithm

GADF Gramian difference angular field

GAF Gramian angular fields

GASF Gramian summation angular field

GAN Generative adversarial networks

GLCM Grey level co-occurrence matrix

Grad-CAM Gradient weighting class activaton mapping

H&E Hematoxylin and eosin

INCA Instituto Nacional de Câncer

LAC Lacunarity

LBP Local binary pattern

LIME Locally-interpretable model-agnostic explanation

LOG Logistic

MCL Mantle cell lymphoma

MLP Multilayer perceptron

NHL Non-*Hodgkin* lymphoma

PERC Percolation

RaF Random forest

RP Recurrence plot

ReLU Rectified linear unit

SVM Support vector machine

SURF Speeded up robust features

XAI Explainable artificial intelligence

Contents

Chapter 1

Introduction

Histopathology comprises the study of diseases affecting biological tissues, through the analysis of cell samples. The analysis is performed by a specialist, via a microscope, where tissues stained with specific materials, for example Hematoxylin and eosin (H&E), are observed. The analysis process can be summarised as a detailed visual assessment aimed at counting and checking the shape of cells (BENTAIEB; HAMARNEH, 2018). Furthermore, to ensure greater accuracy in diagnosis, techniques such as double reading, in which two specialists analyse tissue samples independently, increase the cost of the procedure (BENVENISTE; FERREIRA; AGUILLAR, 2006). This task is complex, can demand a high time consumption and the diagnosis can be highly dependent on the level of experience of the specialist (DOBBS et al., 2015; MUELLER et al., 2016). In this context, the use of a Computer aided diagnosis (CAD) system can support and automate some of these tasks, assisting the professionals involved in the decision-making process.

A CAD system can be divided into steps, from image acquisition, through preprocessing, segmentation, feature extraction and classification (JOTHI; RAJAM, 2017). According to the authors in (FUKUMA et al., 2016), a CAD system for tumour detection requires an effective feature extraction method and, consequently, a classifier capable of distinguishing such attributes. Furthermore, computer vision methods should be able to represent an object-level analysis of the features present in histological images, based on the criteria used by pathologists such as shape or size of cell nuclei (GURCAN et al., 2009). In general, the spatial distribution of cellular structures is distorted according to the degree of malignancy of a tumour and the appearance of cell clusters in the tissues. To quantify such information present in the tissues, computational algorithms must be able to extract features from histological images that explore their morphology, texture, topology and colouration (JOTHI; RAJAM, 2017). Due to the importance of these steps, feature extraction and, consequently, the classification of such features are the focus of the model proposed in this work.

Several techniques can be applied to extract attributes from histological images, commonly referred to as handcrafted features. Among the most common techniques in the

literature, Local binary pattern (LBP), Grey level co-occurrence matrix (GLCM), Speeded up robust features (SURF) and fractal geometry can be highlighted, which have been applied for renal tissue analysis (SIMON et al., 2018), breast cancer classification (YU et al., 2019), colon cell nucleus detection (AMALINA; RAMADHANI; STHEVANIE, 2019) and lymphoma classification (RIBEIRO et al., 2018), respectively. Among the available feature extraction methods, fractal geometry stands out in the context of histological images (LI et al., 2012; ROBERTO et al., 2019). Fractal geometry consists of the study of shapes that do not appear to have continuity or well-defined patterns and therefore cannot be described by Euclidean geometry (DHANSAY; BRANDL; WIT, 2016). The main proposals related to the calculation of properties in complex structures, which can be associated with fractals, were first presented by Mandelbrot (MANDELBROT, 1983). The Fractal dimension (FD) and the Lacunarity (LAC) are the best known examples in the context of fractal geometry (JOTHI; RAJAM, 2017). The FD is a metric that quantifies the irregularity and complexity of a fractal, indicating how much the space is filled. LAC, on the other hand, measures how such space is filled. Approaches based on these measures have indicated interesting results in quantifying psoriatic lesions (IVANOVICI; RICHARD; DECEAN, 2009), prostate cancer (NEVES et al., 2014), brain tumours (AL-KADI, 2015) and colorectal cancer (RIBEIRO et al., 2019). More recently, the concept of Percolation (PERC) has also started to be explored to define fractal features of images, even showing that vascular and histological structures can present similar behaviours to those observed in models explored in percolation theory (BAISH; JAIN, 2000). The information obtained with percolation has the ability to describe properties related to the presence, the dimension and the quantity of clusters in the images. These properties are related to features such as shape, size and quantity of objects, which are relevant information for experts in histopathology (HE et al., 2012).

Handcrafted features were the main way of obtaining information from an image until approximately the beginning of the 2010s. It was during this period that, due to the increased availability of large image sets and advances in hardware technology such as GPU's, Convolutional neural networks (CNN) became a powerful tool in the field of computer vision (KHAN et al., 2020). In the last decade, CNN-based models have been proposed for image processing in various types of applications, including CAD systems.

When CNNs receive an image as input, the information analysis can be treated in two ways: a) feature extraction in the initial layers followed by classification through an external classifier; and/or b) classification of the features in the final layers of the CNN. In the extraction stage, a series of convolutional filters and normalization techniques are applied to each image in order to build a feature vector. This vector is used for the pattern classification stage, but convolutional filters present some limitations in the extraction of information from images of specific contexts, such as in histopathology, causing CNNs to have variable performances in this environment, which may result in classification rates

lower than 70%, as observed in some of the studies (XU et al., 2016; KHOSHDELI; CONG; PARVIN, 2017; YI et al., 2018; WINKENS et al., 2018).

The main difficulties regarding the use of CNNs include the need for a training set with a significant number of samples (YAMASHITA et al., 2018), the initialisation of suitable values for the parameters (SEVERYN; MOSCHITTI, 2015) and a standardised input set (HE et al., 2015). In the context of histological images, these limitations are even more evident due to the fact that most public image datasets have a limited set of samples, and these, on several occasions, do not have standardised resolutions. To deal with this problem, data augmentation can be applied to the images in order to generate more samples through rotation, mirroring or region clipping. Another option is to use Adaptive synthetic sampling approach (ADASYN), which is a technique applied to generate synthetic data from minority classes in order to deal with imbalance (HE et al., 2008). Nevertheless, this increase in data further raises the computational cost of operations with a CNN (LIU et al., 2018).

One of the possible solutions to reduce the processing time is to simplify the CNN architecture by reducing the number of layers. However, the removal of deeper layers may impair the image analysis from a global perspective (ARAÚJO et al., 2017), which may compromise the performance of the whole network. So, some alternative approaches such as hybrid networks have been explored in the literature (WANG et al., 2014; YU et al., 2019; BAI et al., 2019; NANNI et al., 2019; TRIPATHI; SINGH, 2020). These approaches associate feature extraction techniques such as Gabor filters or LBP operators with CNN convolution operations, replacing some of the network layers (JUEFEI-XU; BODDETI; SAVVIDES, 2017; JIANG; SU, 2018). Recent research has shown that a combination between feature extraction techniques with deep learning models can improve the results obtained by applying these approaches separately (NANNI; GHIDONI; BRAHNAM, 2017; BAI et al., 2019; TRIPATHI; SINGH, 2020).

Other approaches generate feature images from the obtained attributes and provide them as input to a CNN. In (LUMINI; NANNI, 2018), a technique is proposed to generate feature images from a feature vector in order to classify them via CNN. In (KAUSAR et al., 2019), the authors applied *Haar-wavelet* decomposition to breast histological images and used the decomposed images as inputs to a CNN.

Fractal geometry-based descriptors can also be associated with hybrid CNNs. In (XU et al., 2017), CNNs were applied to extract values from a FD invariant filter for curve detection in grayscale image objects. The authors in (MOHAMMED et al., 2018) applied multifractal analysis to quantify and detect breast cancer by classifying the generated feature vectors using a deep learning technique. However, an approach similar to that proposed in (LUMINI; NANNI, 2018) or (KAUSAR et al., 2019) in which CNN receives as input feature images generated by a specific technique is still an open challenge in the context of fractal geometry. Methods that directly associate fractal features with CNN

through a classifier ensemble in the context of histological images were not found in the literature.

1.1 Motivation

Cancer consists of a disease that is caused by the disordered growth of malignant tumours. There are several types of cancer, however, breast cancer, colorectal cancer and lymphomas are among the most common and have resulted in higher mortality rates (NOONE et al., 2018).

Breast cancer is a disease that starts with a tumour in the breast region, which after some time can grow and spread to adjacent tissue. This is the most common type of cancer among women, although it can also affect men. According to the Instituto Nacional de Câncer (INCA), 66,280 new cases are expected in 2020 in Brazil (SAÚDE, 2020). Colorectal cancer consists of the growth of malignant polyps in the area of the colon or rectum. This is one of the most common types of cancer, with 40,990 new cases expected in 2020 in Brazil (SAÚDE, 2020). Lymphomas are a type of cancer that affects the cells of the immune system, in which the most common occurrence is of the Non-*Hodgkin* lymphoma (NHL) type. According to statistics, 12,030 new cases are expected in 2020 in Brazil (SAÚDE, 2020). NHL are divided into categories, each requiring specific treatments.

Regardless of the classes of cancer described herein, histopathological analysis is inevitable for the detection of the disease (JOTHI; RAJAM, 2017). Early diagnosis is essential for efficient treatment and an effective detection system is critical for advances in this area (DABEER; KHAN; ISLAM, 2019). Furthermore, algorithms in CAD systems extend a pathologist's vision beyond the microscope, according to (NIAZI; PARWANI; GURCAN, 2019), enabling the integration of human knowledge with artificial intelligence. Therefore, the accuracy of diagnosis provided by pathologists can be improved with the integration of computational techniques applied to images.

Feature extraction techniques based on fractal geometry can provide relevant performances in the quantification of histological images, as demonstrated in (LI et al., 2012; NEVES et al., 2014; ROBERTO et al., 2017; RIBEIRO et al., 2019; ROBERTO et al., 2019). On the other hand, the small number of samples available in public histological image datasets is a factor that hinders the application of CNN for the classification of this type of image (YAMASHITA et al., 2018). As this factor has not been shown as a limitation for fractal feature classification, the association of these techniques to a CNN-based network architecture may improve the results obtained in histological image classifications.

The implementation of a CNN that receives the three main types of fractal features as input can automate the task of selecting parameters for the various metrics calculated

from the different existing fractal features. Furthermore, recent research has shown that the association of handcrafted features with CNN was able to provide good results in the context of histological images (WANG et al., 2014; YU et al., 2019; BAI et al., 2019; NANNI et al., 2019; TRIPATHI; SINGH, 2020). The combination of CNN models are one of the main directions of the researches in this area, according to (KHAN et al., 2020). However, fractal geometry has not yet been explored in this type of application, especially in approaches similar to (LUMINI; NANNI, 2018; KAUSAR et al., 2019), where the extracted attributes form a feature image that is given as input to a CNN.

The generation of a feature image is usually done from reshaping a handcrafted feature vector into a matrix. Methods such as the proposed in (AFONSO et al., 2019; FARIA et al., 2016; WANG; OATES, 2015) applied different procedures to convert a 1-D feature vector into an image, which were later used for tasks such as Parkinson's disease detection and plant recognition. These methods could be evaluated using fractal descriptors as inputs to support the classification of histological images.

1.2 Objectives

The general objective of this work is to develop a model based on the ensemble of deep learning and handcrafted features, using fractal descriptors for the classification of breast and colorectal tumours, non-Hodgkin lymphomas, and liver tissue. The specific objectives are listed below:

1. construction of a feature image model to represent fractal properties;
2. development of an approach based on CNN and fractal geometry to identify deficiencies and improve performance in unbalanced histological image datasets with different resolutions;
3. development of a hybrid CNN model based on fractal and convolutional features;
4. investigating shortcomings and enhancements to improve the performance of a classifier ensemble with fractal descriptor approaches for different histological image samples, using learned and handcrafted features.

1.3 Contributions

The implementation of the proposed model may bring the following contributions to the literature:

1. New representations based on fractal descriptors for histological lesion classification;

2. The availability of an adaptive method, capable of classifying different sets of histology images, including datasets with unbalanced classes and variations in image resolution;

3. A support tool for pathologists able to provide a diagnosis from a computational perspective and reduce the costs of an eventual second reading;

4. A deep learning based approach where there is no requirement for a long training of the network every time a new set of images has to be classified;

5. Multiclass classifier ensemble methods to improve predictive performance for lesions in histological images.

1.4 Text Organisation

Chapter 2 of this text presents the theoretical foundations necessary for the development of the method. The bibliographical survey is discussed in Chapter 3, with current and relevant works in the area. Chapter 4 presents the methodology applied for the development of the work. The tests performed and the results obtained are presented in Chapter 5. Chapter 6 presents the conclusions obtained from the results achieved, as well as the overview of the work developed during this doctorate.

Chapter 2

Technical Background

The objective of this chapter is to present theoretical concepts that are important for the understanding of the work. The section was divided into concepts of histology, computer aided diagnosis system, fractal geometry applied to digital images, basic concepts about CNN and metrics to evaluate the performance of the techniques described.

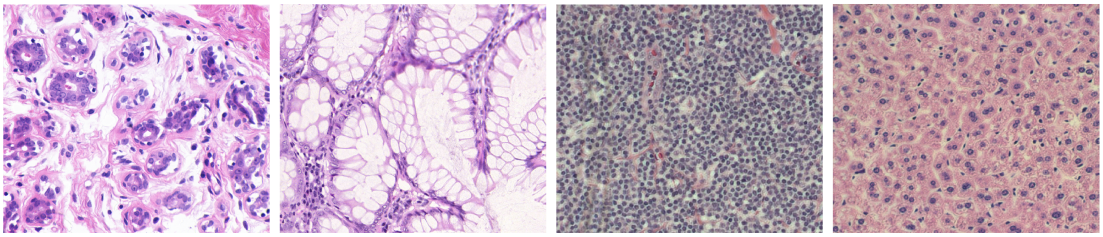
2.1 Histology images

Histology is a subfield of medicine that consists of the microscopic analysis of cells and tissues. In the procedure for diagnosing cancer, for example, histopathologists analyse histological samples under a microscope. The specialists visually examine the regularity, shape and distribution of structures in order to determine whether the tissue being analysed has some kind of lesion or can be considered healthy.

The process of acquiring a histological image begins with the collection of a tissue sample *in vivo*. As described in (HE et al., 2012), after obtaining the sample, it must be fixed to preserve decomposition. Depending on the type of tissue, different fixation methods can be applied, such as fixation by heat or by immersion in substances such as ethanol, methanol, chloroform or acetone. Besides preserving the sample, immersion in these compounds also causes tissue dehydration by removing lipids and reducing protein solubility. This procedure is necessary to slow down the degeneration of the specimen (HAM; CORMACK, 1983). Next, a series of treatments are applied to the tissue with the aim of fixing it in a solid mould and thus facilitating the cutting of the sample into thin sections for analysis under the microscope. This procedure is known as sectioning, in which $5\mu m$ or 80 to $100nm$ samples are sectioned for analysis under a light or electron microscope. The sections are placed separately on glass slides, where they will receive dyes to enhance contrast and facilitate identification of intra- and extra-cellular structures. In light microscopy, the most commonly used stains are H&E. Hematoxylin marks the cell nuclei with dark blue shades, while eosin acts on the stroma and cytoplasm organelles by marking them with pink, red or orange shades (KIERNAN, 2008). Finally, the slides

containing the samples are covered with a thin layer of plastic or glass to facilitate visual analysis via a microscope. To enable various experts to analyse these samples at different locations or instants, they are digitized and stored in computer files. Figure 1 shows examples of breast, colorectal, lymphoma and liver tissue samples after the acquisition and digitization processes described.

Figure 1 – Examples from breast (a), colorectal (b), lymphoma (c) and liver (d) histology images.



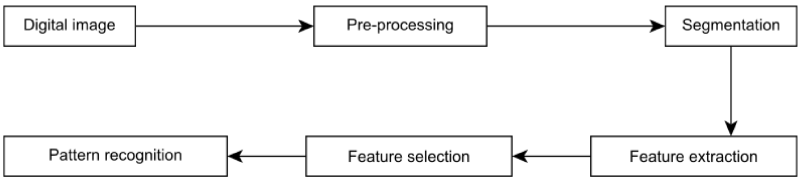
(a) (b) (c) (d)

(Source: (GELASCA et al., 2008; SIRINUKUNWATTANA et al., 2017; SHAMIR et al., 2008; AGEMAP, 2020))

2.2 Computer Aided Diagnosis

CADs are built with computer algorithms to analyse medical images with the aim of providing the specialist with a diagnostic aid or even a teragnosis, to predict the likely outcome of applying a specific treatment (CHAN; HADJIISKI; SAMALA, 2020). This type of support system is commonly divided into a few steps, as shown in Figure 2.

Figure 2 – Steps of a CAD system.



(Source: Adapted from (JOTHI; RAJAM, 2017))

The quality of the image given as input to a CAD system is determinant to identify the need to apply or not techniques for image enhancement or segmentation. Pre-processing techniques, such as colour normalization, are commonly applied in the context of histological images, since corrections need to be made due to the way each sample was prepared, or even due to eventual alterations in the environment illumination at the moment of the image acquisition (ROBERTO et al., 2019).

The segmentation consists of the selection or detection of image regions that are the focus for lesion identification. The most common segmentation techniques are based

on thresholding or edge detection. Thresholding approaches consist of selecting one or more values, usually based on the intensities of pixels, to separate objects from the background (SEZGIN; SANKUR, 2004). In histological images, these methods are efficient for separating nuclei from the stromal region (VETA et al., 2013). Edge detection-based approaches apply algorithms to analyse the difference between the intensity levels of pixels in order to define the boundary between the objects of interest and the background, such as the Sobel filter (BEN-COHEN et al., 2017). Recently, bio-inspired algorithms have also obtained relevant results in the segmentation of histological images. The segmentation of the nuclei of histological images of lymphomas was performed by applying a genetic algorithm that adopted Rényi's entropy as a fitness function (TOSTA et al., 2018).

The feature extraction step, or information quantification, consists in obtaining numerical attributes that are able to describe, in a colour model, properties such as shape, size or texture present in the regions of interest of the images. In the context of histological images, the most commonly evaluated features are: morphological attributes, based on object shape and size (DIMITROPOULOS et al., 2017); topological attributes (BEJNORDI et al., 2016); colour and brightness intensity attributes (NGUYEN et al., 2017; NASCIMENTO et al., 2018); attributes extracted from the frequency domain (LIN et al., 2016; RIBEIRO et al., 2019); and texture attributes, which highlight those based on Haralick features (HARDER et al., 2016) and fractals (ROBERTO et al., 2017; ROBERTO et al., 2019). In histological images, the use of fractal features can indicate the growth pattern of tumours and vascular structures (LOPES; BETROUNI, 2009; LI et al., 2012; RIBEIRO et al., 2019), (BAISH; JAIN, 2000) *apud* (JOTHI; RAJAM, 2017).

The last stage of a CAD system consists in the pattern recognition from the obtained features. Generally, this recognition is done based on the use of a classifier algorithm, which consists in a supervised learning method, whose application occurs in two main stages: training and test. In the training stage, a classifier model is built based on attributes of previously labeled images. The test stage consists in applying on unlabelled images the model built in the training step, with the purpose of assigning them a class label. Several types of classifiers can be used in the context of histological images, among which stand out Bayesian classifiers (KURMI; CHAURASIA; GANESH, 2019; MURTAZA et al., 2019), function-based (SIMON et al., 2018; RIBEIRO et al., 2019), instance-based (ROBERTO et al., 2017), decision trees (LIU; YAN; WANG, 2018; ROBERTO et al., 2019) and neural networks, which include CNN (TAVOLARA et al., 2019; TRIPATHI; SINGH, 2020).

2.3 Feature extraction

In image processing, the feature extraction stage consists in obtaining relevant information present in the image (KUMAR; BHATIA, 2014). This procedure is applied

in CAD systems to quantify elements present in cellular tissues. According to (JOTHI; RAJAM, 2017), histological images representing malignant tissue have a more irregular spatial distribution of cellular components compared to normal tissue. The main feature extraction techniques applied in the context of histopathology consist in obtaining:

- Morphological features: these describe simple characteristics such as the shape and size of cell structures. Examples of this type of information are area, perimeter, concavity and Fourier shape descriptors;
- Topological features: they describe the spatial distribution of cellular components. The main ones include Delaunay triangulation, Voronoi diagram and minimum spanning tree;
- Texture features: they describe the intensity variation in different regions of the image. According to (JOTHI; RAJAM, 2017), the main methods for obtaining texture features are GLCM, Gabor filter, *wavelet* transform and fractal geometry.

Although there are several ways to quantify a histological image, fractal geometry is one of the approaches that has recently stood out most to the classification of histological images through handcrafted features (RIBEIRO et al., 2019), due to the existing similarity between the growth pattern of tumors and the architecture of a fractal (BAISH; JAIN, 2000). Therefore, this approach was selected for application in the groups of images investigated in this work.

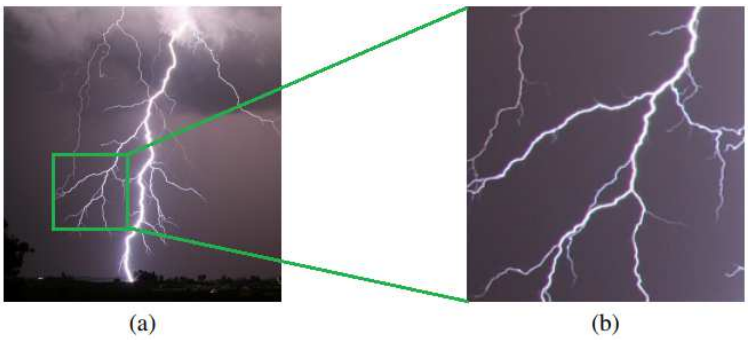
2.4 Fractal features

Fractal geometry is a concept presented for the study of shapes that cannot be defined by Euclidean geometry, as described in (MANDELBROT, 1975). The shapes present in nature, such as the coastline of an island, clouds, trees and lightning, are examples of structures that do not present patterns to be described by representations with geometric structures such as circle, polygon or polyhedron, which are well defined by classical mathematics.

In this context, for shapes with irregular representation, it is possible to apply rules that can define them, in which these rules are related to the observation scale of the object. In Figure 3 the visualization of a natural form of a lightning that cannot be described by the Euclidean geometry is presented, in which in Figure 3(a) there is the complete representation of a lightning and in Figure 3(b) this representation has its scale reduced, presenting only a part of the structure of the lightning. The images are similar, but have different scales. This property is one of the foundations of fractal geometry, known as self-similarity (CASTRO, 2006). Although there is still no formal definition for self-similarity, the authors at (BRACHMANN; REDIES, 2017) describe that this property

can be defined as features present in the image that can be observed either at smaller scales, in sub-regions, or in the complete image.

Figure 3 – Picture of a lightining under two different observation scales: full picture (a) and picture under a scale four times smaller (b).



(Source: Adapted from (POHJOLA; GRANDELL, 2016))

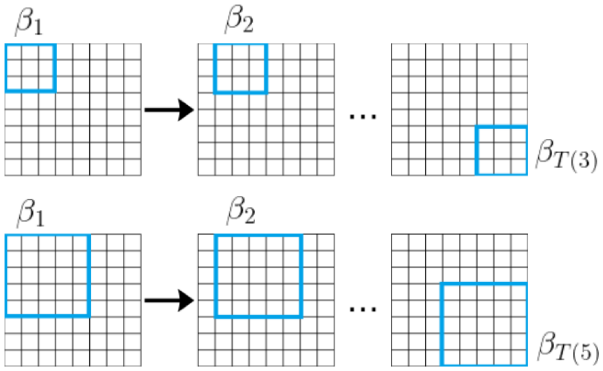
In general, a fractal should have the following characteristics: detail at small scales; inability to be described by traditional geometry; approximate or statistical self-similarity; possibly recursive definition; and fractal dimension larger than the topological one. In computer vision, there are some approaches for the observation of self-similarity properties in images, known as multi-scale. Among the most common ones, the algorithms *box-counting* (NIKOLAIDIS; NIKOLAIDIS; TSOUROS, 2011) and *gliding-box* (IVANOVICI; RICHARD; DECEAN, 2009) stand out. These algorithms aim to divide images into different scales and then extract features from each sub-image, generating local features, for each scale, and global features to represent the complete image. FD, LAC and percolation (PERC) are among the main features for representing properties in images with fractal data (BAISH; JAIN, 2000; LOPES; BETROUNI, 2009; ROBERTO et al., 2017; JOTHI; RAJAM, 2017; RIBEIRO et al., 2019).

In the context of medical area, there are some techniques that use colour images for abnormalities analysis. As the model proposed by Mandelbrot presents limitations, new adaptations and improvements motivated researchers to investigate approaches for the context of colour images (IVANOVICI; RICHARD; DECEAN, 2009; IVANOVICI; RICHARD, 2009; ROBERTO et al., 2017).

According to these adaptations, fractal features can be obtained from coloured images initially by applying the *gliding-box* algorithm. This algorithm consists in positioning a box of size $L \times L$ in the upper left corner of the image, where L represents the side dimension of the box in pixels. This box moves from left to right to the bottom region of the image, passing through all pixels. After gliding through the entire image, the box is repositioned to the starting point and the value of L is incremented by 2 units, as illustrated in Figure 4. The value of L must always be an odd number, since the existence of a central pixel is required for next steps of the method. In an image of size $H \times W$, the total number T of boxes β_i as a function of L is given by Equation 1:

$$T(L) = (H - L + 1) \times (W - L + 1) \quad | \quad L \leq min(H, W). \tag{1}$$

Figure 4 – Application of the gliding-box algorithm in an image with dimensions 8×8.



(Source: author)

Each time the box β_i is shifted, a colour similarity analysis is performed for each pixel of β_i. This analysis is performed by fixing the central pixel and assigning it to a vector $f_c = \{r_c, g_c, b_c\}$, where r_c, g_c and b_c correspond to the luminous intensity of each of the colour channels, considering RGB images. The other pixels of the box are assigned to another vector $f_i = \{r_i, g_i, b_i\}$ and compared to the central pixel by calculating a colour distance Δ, which allows verifying which pixels belong to the RGB hyperspace formed by the central pixel of the box β_i (IVANOVICI; RICHARD; DECEAN, 2009). In this work, the distances *Chessboard* (Δ_h), *Euclidean* (Δ_e) and *Manhattan* (Δ_m) were used, wherein calculations are performed according to the Equations 2-4.

$$\Delta_h = max(|f_i(k_i) - f_c(k_c)|), k \in \{r, g, b\}. \tag{2}$$

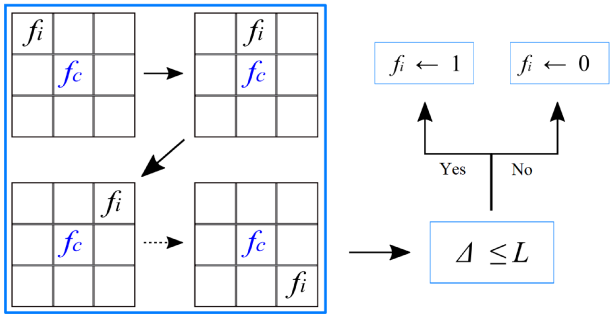
$$\Delta_e = \sqrt{\sum_{k=r,g,b} \left(f_i(k_i) - f_c(k_c)\right)^2}. \tag{3}$$

$$\Delta_m = \sum_{k=r,g,b} |f_i(k_i) - f_c(k_c)|. \tag{4}$$

If the value of Δ corresponding to the distance between f_i and f_c is less than or equal to the scale L, then f_i is labelled 1, otherwise f_i is labelled 0. In Figure 5 this pixel labelling procedure is exemplified. In this case, a distance Δ is calculated for pixels in a box at scale $L = 3$.

After calculating the number of pixels that satisfy Δ for different values of L, the information obtained is used to construct a structure known as a probability matrix. Each element of the matrix corresponds to the probability P that m pixels on a scale L are labelled as 1. A visual representation of this matrix is given in Table 1. The matrix is normalized so that the sum of all elements in a column equals 1, as shown in Equation 5.

Figure 5 – Pixel labelling in a 3×3 box after the calculation of a distance Δ.



(Source: author)

Table 1 – Representation of a probability matrix.

	3	5	$\cdots$	L_{max}
1	$P(1,3)$	$P(1,5)$	$\cdots$	$P(1,L_{max})$
2	$P(2,3)$	$P(2,5)$	$\cdots$	$P(2,L_{max})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
L^2	$P(L^2,3)$	$P(L^2,5)$	$\cdots$	$P(L^2,L_{max})$

$$\sum_{m=1}^{L^2} P(m,L) = 1, \forall L. \tag{5}$$

It is important to point out that the probability matrix does not have the common format of a rectangular matrix, due to the fact that the number of lines grows exponentially for each L value. After the matrix is completely filled, local values of the fractal measures can be obtained.

2.4.1 Fractal dimension

FD is the most common metric for assessing fractal properties in an image. One of the first research approaches occurred in 1961, when Richardson was trying to measure the length of the coastline of Great Britain. He observed that the smaller the measurement scale, the greater the length of the coastline (RICHARDSON, 1961). This was because, at smaller scales, more details such as small inlets, streams or peninsulas, contributed to the sum of the final length of the analysed structure. Mandelbrot, in 1983, observed a similar behaviour in fractals (MANDELBROT, 1983). It was also observed that a straight line is obtained if the logarithm of the measurement scale was projected onto a graph in relation to the logarithm of the total length. The value of the angular coefficient of this straight line allows to define the FD measurement of the object, and the individual values for each measurement scale correspond to the local fractal dimensions. In its initial model, FD was calculated only for objects with up to two topological dimensions or binary images.

With the development of new approaches, it became possible to calculate the FD for gray-scale or colour images. According to (IVANOVICI; RICHARD, 2011), local values

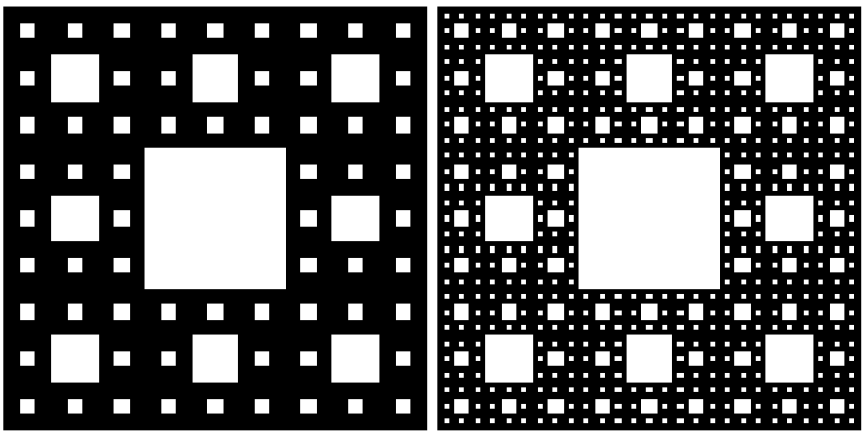
of FD defined as $D(L)$ for a colour image are given from the generation of the respective probability matrix according to Equation 6:

$$D(L) = \sum_{m=1}^{L^2} \frac{P(m, L)}{m}. \tag{6}$$

Besides the (IVANOVICI; RICHARD, 2011) approach, which is based on the algorithm *gliding-box* for the generation of the probability matrix, another method available in the literature to determine the FD of colour images is the (NIKOLAIDIS; NIKOLAIDIS; TSOUROS, 2011) based on the algorithm *box-counting*. This had been implemented for N-dimensional images of size $L_x \times L_y \times ... \times L_n$ and uses a scale factor to divide the image into boxes. The FD is calculated from a partition table, in which each column corresponds to one dimension of the set. This method also generates a graph that relates the evolution of the local FD values with respect to the box size. In both approaches, for cases where the graph does not generate a straight line but a nonlinear set of points, the authors in (IVANOVICI; RICHARD, 2011) tested different linear regression methods, all based on the least squares technique. The method of (NIKOLAIDIS; NIKOLAIDIS; TSOUROS, 2011) addresses this problem in a less specific way, by performing a cut-off on values that are outside the trend line of the generated straight line, which solves the problem in a general way. However, this solution may make the model more susceptible to errors.

The main limitation of the measures based on FD occurs when different images present similar or even identical FD values, as exemplified in Figure 6 by the *Sierpinski* fractals. To solve this restriction, other measures based on fractal geometry were elaborated, such as the lacunarity, to complement the information obtained by the similarity analysis.

Figure 6 – Sierpinski's carpet. The images (a) and (b) have similar FD values, however they also have different LAC values.



(a) (b)

(Source: (MANDELBROT, 1983))

2.4.2 Lacunarity

LAC is a complementary measure to FD and allows to evaluate how the space of a fractal is filled. As in the FD measure, (IVANOVICI; RICHARD, 2009) also proposed a method to obtain the lacunarity in colour images. This method presents steps similar to the FD method for the construction of the probability matrix. From this, the first and second order moments are calculated using Equations 7 and 8. The values m and $P(m, L)$ are obtained in the same way as described for FD in Equation 5.

$$\mu(L) = \sum_{m=1}^{L^2} mP(m, L). \tag{7}$$

$$\mu^2(L) = \sum_{m=1}^{L^2} m^2P(m, L). \tag{8}$$

Then, the LAC measure on a L scale is given by $\Lambda(L)$, obtained with Equation 9.

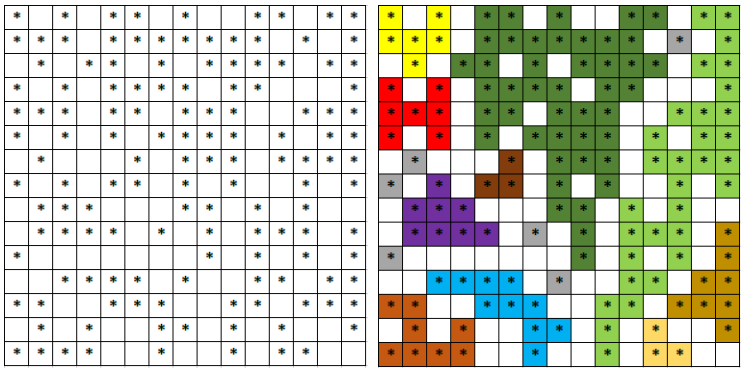
$$\Lambda(L) = \frac{\mu^2(L) - (\mu(L))^2}{(\mu(L))^2}. \tag{9}$$

2.4.3 Percolation

Percolation is a physical phenomenon, defined by (BROADBENT; HAMMERSLEY, 1957), referring to the transport of fluids in porous media. Several natural systems exhibit this type of behaviour, such as water flowing through coffee powder or gases through a rocky medium. If one of these fluids, given by water or gas, is able to cross the whole system from one end to the other, a phenomenon known as percolation occurs. These concepts can also be adapted to image analysis. In Figure 7 an example of a grid of size 15×15 pixels is presented which shows how percolation occurs. In the example, some pixels of the grid have been labelled with the symbol (∗). In an analogy to percolation theory, these labelled pixels correspond to pores, which are spaces through which a supposed fluid is able to pass over a medium. It can be observed that some pixels form clusters, when a neighbourhood of 4 pixels is applied, as shown in Figure 7(b). When a cluster occurs where it extends from one end of the image to the other, the occurrence of the percolation phenomenon is defined.

In natural systems, the distribution of pores happens randomly. Based on this principle, studies have found that each system has a value p, which corresponds to the probability that a space is or is not a pore (MALARZ, 2015). In addition to indicating what the ratio between the existence of pores and non-pores in a system is, p has an important property known as the percolation threshold. This property denotes that, after a certain value of p, there is the guarantee of percolation occurrence. These concepts can also be adapted for images and avoid efforts in searching for clusters that indicate percolation. Each type of system has a different percolation threshold value, as shown in Table 2. The

Figure 7 – Matrix of size 15 × 15 (a) wherein percolation occurs due to the existence of a cluster that extends between two extremities of the matrix, represented in light-green (b).



(a) (b)

(Source: author)

threshold for square matrices, which also fits for images, is 0.59275 based on (DEAN, 1963; ZIFF, 1986; BIRD; PERRIER, 2010).

Table 2 – Percolation thresholds for different structures.

Grid	**Percolation threshold**
Square	0.593
Triangular	0.486
Hexagonal	0.679
Four-eight	0.739
Kagome	0.655

In 2017, an approach was proposed to extract percolation features from an image using concepts from fractal geometry (ROBERTO et al., 2017). The method is based on Ivanovici and Richard's models for obtaining FD and LAC from colour images (IVANOVICI; RICHARD; DECEAN, 2009; CĂLIMAN; IVANOVICI, 2012). This approach follows the steps of the *gliding-box* algorithm and calculation of the *Minkowski* distance between a central pixel and the other pixels of a box for similarity analysis of the coloured pixels. Then, the pixels that satisfied the *Minkowski* distance, which can be defined as pores, are identified through the Hoshen-Kopelman labelling algorithm (HOSHEN; BERRY; MINSER, 1997). For each scale L, three different features are extracted: average number of clusters per bin, percolation occurrence rate, and average area of the largest cluster. The average number of clusters $C(L)$ is given by Equation 10, which consists of the sum of the total number of clusters c in each bin i divided by the total number of boxes T.

$$C(L) = \frac{\sum_{i=1}^{T(L)} c_i}{T(L)}. \tag{10}$$

In Equation 12, it is shown how the percolation occurrence rate $Q(L)$ is calculated.

This corresponds to the sum of boxes β_i whose number of pixels labelled as pores (Ω_i) satisfy the percolation threshold.

$$q_i = \begin{cases} 1, & \frac{\Omega_i}{L^2} \geqslant 0.59275. \\ \\ 0, & \frac{\Omega_i}{L^2} < 0.59275. \end{cases} \tag{11}$$

$$Q(L) = \frac{\sum_{i=1}^{T(L)} q_i}{T(L)}. \tag{12}$$

Equation 13 corresponds to the calculation of the average coverage area of the largest cluster $M(L)$, given by the sum of the occupancy rates of the largest clusters in each box $\frac{M_i}{L^2}$ divided by the total number of boxes T.

$$M(L) = \frac{\sum_{i=1}^{T(L)} \frac{M_i}{L^2}}{T(L)}. \tag{13}$$

The number of local features obtained depends on the total observation scale L. Considering that L varies between 3 and L_{max} with an increment of 2 units, the amount of local features corresponds to $5 \times (\frac{L_{max}-3}{2} + 1)$ for each Δ. A summary of these features is shown in Table 3.

Table 3 – Summary of the obtained local features for each Δ.

FD	**LAC**	**PERC**		
$D(3)$	$\Lambda(3)$	$C(3)$	$Q(3)$	$M(3)$
$D(5)$	$\Lambda(5)$	$C(5)$	$Q(5)$	$M(5)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$D(L_{max})$	$\Lambda(L_{max})$	$C(L_{max})$	$Q(L_{max})$	$M(L_{max})$

2.4.4 Global Features

Fractal Dimension

A global FD value can also be obtained by generating a plot from the values $logL \times logD(L)$. With the use of linear regression to estimate the slope of the $log - log$ curve, the resultant slope value can be considered as a global FD indicator.

Lacunarity

Similar to what is done for the FD, global LAC features can be obtained from a curve containing the local values of Λ. Using the method proposed in (CĂLIMAN; IVANOVICI, 2012), global values can be obtained from such curve: area under curve ($A(\Lambda)$), skewness

($S(\Lambda)$), area ratio ($A_r(\Lambda)$), maximum point ($Max(\Lambda)$) and scale of the maximum point ($\sigma(\Lambda)$).

The area under the function curve can be obtained by applying the trapezoidal method, as given by Equation 14, where a and b are the minimum and maximum values of L respectively, Λ is the lacunarity function and N is the number of local LAC values:

$$A(\Lambda) = \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(\Lambda_n) + f(\Lambda_{n+1})). \tag{14}$$

The second global LAC feature obtained is the skewness, which corresponds to a measure of asymmetry. For a LAC curve wherein N local values were calculated, the skewness is obtained according to Equation 15, wherein Λ_i is the i_{th} value of the sample and $\overline{\Lambda}$ is its average value.

$$S(\Lambda) = \frac{\frac{1}{N} \sum_{i=a}^{b} (\Lambda_i - \overline{\Lambda})^3}{\sqrt[2]{[\frac{1}{N-1} \sum_{i=a}^{b} (\Lambda_i - \overline{\Lambda})^2]^3}}. \tag{15}$$

Another feature that can be calculated is the ratio between the right and left halves of the curve as shown in Equation 16. According to (CĂLIMAN; IVANOVICI, 2012), a greater value on this feature indicates a more heterogeneous texture.

$$A_R(\Lambda) = \frac{A_{(\frac{b}{2}+1,b)}}{A_{(a,\frac{b}{2})}}. \tag{16}$$

The values of $Max(\Lambda)$ and $\sigma(\Lambda)$ can also be obtained, which gives a total of five LAC global features.

Percolation

To obtain percolation global features, the same metrics applied to the LAC curves can be applied to the three PERC functions (C, Q and M). Therefore, a number of 21 global features are obtained for each Δ. The summary of these features is shown in Table 4.

Table 4 – Summary of the obtained global features for each Δ.

FD	**LAC**	**PERC**		
	$A(\Lambda)$	$A(C)$	$A(Q)$	$A(M)$
	$S(\Lambda)$	$S(C)$	$S(Q)$	$S(M)$
$Fd(D)$	$A_R(\Lambda)$	$A_R(C)$	$A_R(Q)$	$A_R(M)$
	$Max(\Lambda)$	$Max(C)$	$Max(Q)$	$Max(M)$
	$\sigma(\Lambda)$	$\sigma(C)$	$\sigma(Q)$	$\sigma(M)$

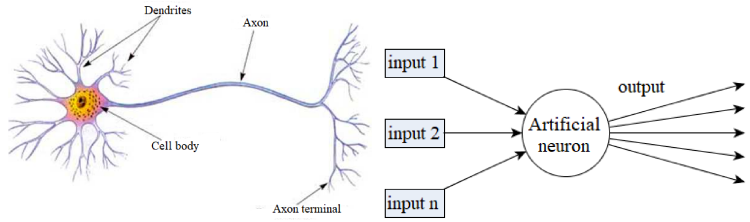
2.5 Artificial neural networks

Artificial neural network (ANN) are a subtype of machine learning algorithms that consists of the application of techniques for pattern recognition and sample classification. Learning algorithms are divided into three categories:

- **Supervised Learning:** applied when the expected responses in a series of samples are known;
- **Non-Supervised Learning:** applied when the desired outputs are not known, but the intention is to find patterns of behaviour in the input samples;
- **Reinforcement Learning:** applied when an algorithm is desired to interact with an environment to make decisions. There is practice by trial and error, receiving rewards or punishments as *feedback* for decisions made.

ANN consist in an approach that simulates the human brain, in the sense of learning to perform a task. These methods are built using variations of a basic concept, the artificial neuron, which is a mechanism capable of processing a set of signals it receives as input to produce an output signal. In Figure 8, structures for representing neurons are presented.

Figure 8 – Neuron representation: biological neuron (a) artificial neuron (b).



(Source: (a) adapted from (PINTO, 2010), (b) author)

The input of the artificial neuron is received in the form of a vector $X = \{x_1, x_2, ..., x_N\}$, where x_i represents the value of a feature i in a set of N features. In (ROSENBLATT, 1958), a basic artificial neuron model called perceptron was proposed. This model consists of associating a weight to each input connection of the artificial neuron to weight its influence on a given task. For this, an activation function defined by $f(z)$ is responsible for determining the output value of a perceptron, which is obtained by the product of the weight vector, given by W, and the feature vector X, added to a bias value b, as presented in Equations 17 and 18. The vector X consists of the set of features, given as input, and the vector W is composed of weights that must be associated to each feature of X.

$$z = W \times X + b = \sum_{i=1}^{N} w_i x_i + b. \tag{17}$$

$$f(z) = \begin{cases} 1, & W \times X + b > 0. \\ \\ 0, & W \times X + b \leq 0. \end{cases} \tag{18}$$

Unlike the programming of traditional algorithms specific to solving a problem, neural networks have a learning process. This process starts with a set of input samples where the desired outputs are known. This is called the training set, and each sample is a pair $(X; Y)$, where X is an n-dimensional vector describing an input signal, and Y is the corresponding desired response. The parameters W and b configure the decision frontier modeled by the neuron, and thus are the parameters to be learned during training. To start training a perceptron, the n weights of the set W are given random values in the range $[-1, 1]$ and 0 for b. Some works propose that these initial values be different (IJJINA; CHALAVADI, 2016; AN, 2019).

The training step consists of two phases to be alternated repeatedly: (a) processing an input signal X, in order to produce an output value $o(t)$; and, (b) adjusting the neuron parameters, according to the comparison between the obtained output $o(t)$ and the desired result Y. When performing the first step, the parameter values W and b are kept fixed and used to weight the elements of a training sample. These values are summed to the bias b and applied to the activation function, producing an output $o(t) = f(z(X))$. After an output, the second step must be performed based on the learning process. Then, the perceptron adjusts the weights and the bias based on the difference between these parameters given by the term $o(t)$. The error obtained in iteration t is calculated by $e(t) = Y - o(t)$, whose value is employed to update the parameters given by Equation 19:

$$W(t+1) = W(t) + \alpha \times (Y - o(t)) \times X, \tag{19}$$

where α represents the learning rate, t represents the iteration step, X represents the feature vector of the input sample plus $x_0 = 1$ and W represents the vector of weights plus $w_0 = b$.

2.5.1 Feedforward neural network

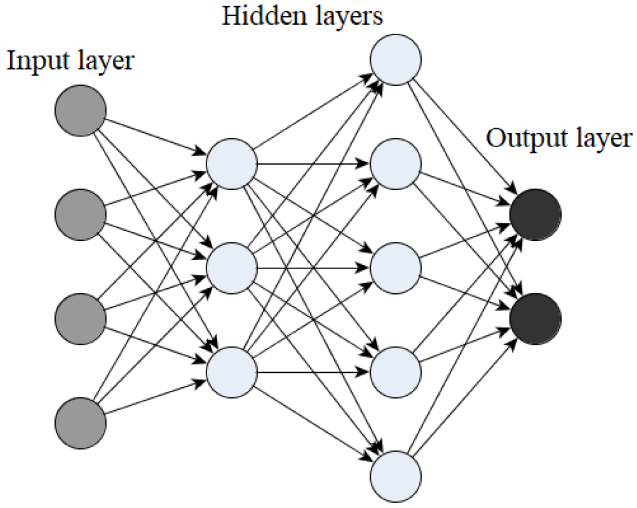
In this type of neural network, which is also known as a Multilayer perceptron (MLP), neurons are organized in sequential layers so that the signal provided as input is transmitted in only one direction to the next layer. Each neuron receives input signals coming from neurons in previous layers and, in turn, transmits the signal it produces to neurons in subsequent layers. This layer structure is illustrated in Figure 9, which is composed of:

- ❑ An initial layer defined as the input layer, which is responsible for reading the data to be processed;

- One or more intermediate layers, called "hidden" layers, which are responsible for the processing itself;

- A last layer responsible for issuing the result of the processing, which is the output of the network.

Figure 9 – Example of a neural network architecture in the MLP model.



(Source: author)

In MLP, learning occurs by applying a training algorithm called backpropagation. In this algorithm, an extension is added after its output layer, responsible for calculating an error function called E. The main goal of this phase is to minimize the E function during training. At each iteration, after the propagation of the signal from the input layer to the output layer, there is a backpropagation of the error in the opposite direction to the established by the connections, until it reaches the neurons of the input layer. Once is estimated how much each connection influenced the error, the weights are updated. The update of the weight of a feature, defined by i in a layer j, therefore given by w_{ji}, is calculated by the tangent of the error function $E(t)$ relative to $w_{ji}(t)$. This represents the gradient of the function $E(t)$, as presented in Equation 20. During the training stage in MLPs, the weights are updated according to the ratio of the gradient vector of the error function by the learning rate α. This calculation is shown in Equation 21:

$$\nabla E(t) = \left\langle \frac{\partial E(t)}{\partial w_{1,1}(t)}, \frac{\partial E(t)}{\partial w_{1,2}(t)}, ..., \frac{\partial E(t)}{\partial w_{N,N}(t)} \right\rangle, \tag{20}$$

$$w_{ji}(t+1) = w_{ji}(t) - \alpha \frac{\partial E(t)}{\partial w_{ji}(t)}. \tag{21}$$

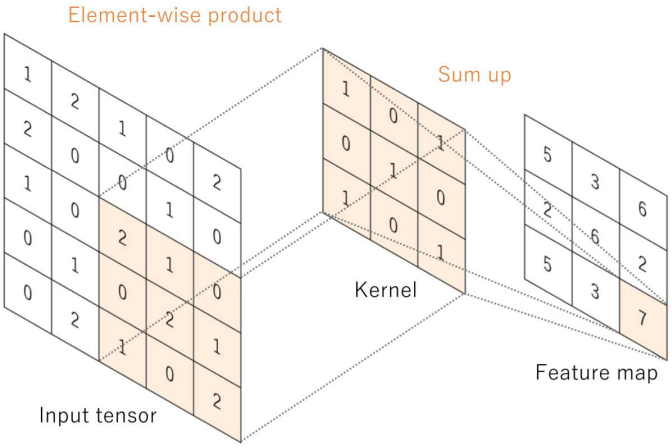
2.6 Convolutional neural networks

CNNs are special deep learning models that obtain features from low- and high-level patterns in grid format data (YAMASHITA et al., 2018). The core of CNNs is usually built from layers: convolution; *pooling*; and fully connected (FC). While the first two are responsible for feature extraction, the FC layers perform feature classification and generate as output a label to be assigned to the input data.

The convolution layers are the structures related to the base of the network, which generally consist of two operations. The first is the convolution itself, a linear procedure that performs the multiplication of the input data elements and small numerical matrices called *kernels*, which are the trainable parameters in this type of layer. In Figure 10, an example of the operation performed by the convolution layer in a CNN is shown. The second operation consists of applying a non-linear activation function to the feature map resulting from the convolution. Different functions, such as logistic or sigmoid, are applied in CNN models, but the Rectified linear unit (ReLU) has become the most popular due to its tendency to reduce training time by eliminating negative values (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). This function is calculated according to Equation 22:

$$f(x) = max(0, x). \tag{22}$$

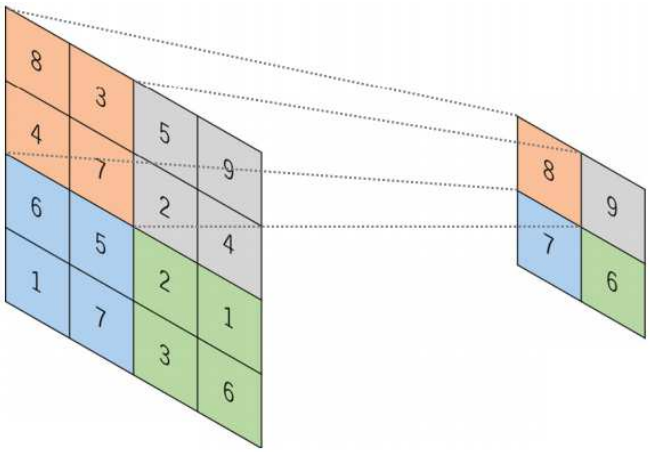
Figure 10 – Application of a kernel sized 3×3 pixels for obtaining a feature map.



(Source: (YAMASHITA et al., 2018))

The *pooling* layers perform operations to reduce the dimensionality of the data. This usually occurs by selecting the maximum (*Max Pooling*) or average (*Global Average Pooling*) value among the elements of a sub-region of the feature map. The aim of this operation is not only to reduce the number of features but also to introduce a certain invariance to translations and distortions of structures in the input data. In Figure 11, a *Max Pooling* operation is illustrated with a 2×2 pixels filter on a feature map of size 4×4 pixels. In this operation, it is possible to observe that in the regions marked in red, grey, blue and green the largest values were selected.

Figure 11 – *Max Pooling* operation performed on the feature map of a CNN.



(Source: adapted from (YAMASHITA et al., 2018))

FC layers consist of a series of one or more layers in which all outputs are connected to all inputs of the next layer by using a weight adapted through training. The adjustment of the weights is usually done by using the backpropagation algorithm, since the sequence of one or more FC layers in a CNN has operation close to that of an MLP. Generally, the number of perceptrons in the last FC layer is equal to the number of classes in the dataset under training. In classification problems, its outputs correspond to class probabilities, which are obtained by means of an activation function. For multi-class problems, the most suitable function for this task is *softmax* (YAMASHITA et al., 2018). This function is applied in order to normalize the results transmitted to the output layer. For each perceptron, the probability calculation is given by Equation 23:

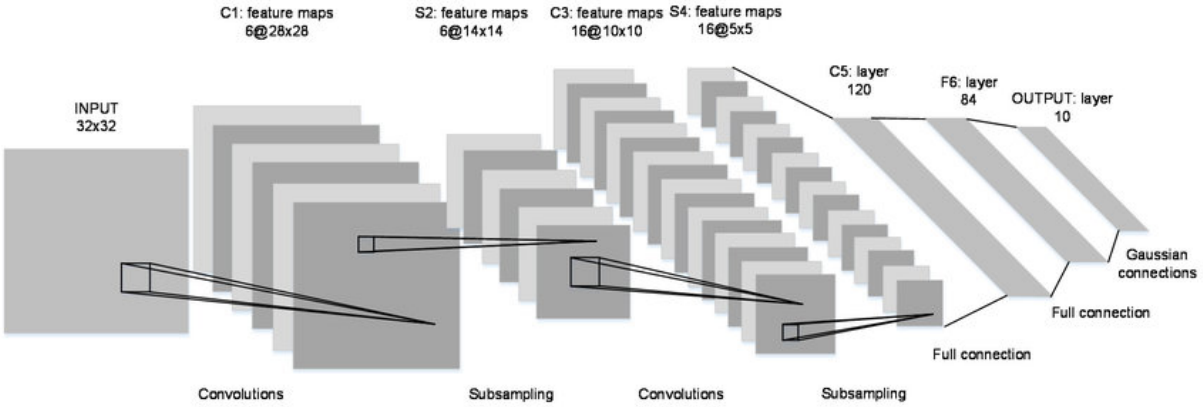
$$f(u_j) = \frac{e^{u_j}}{\sum_{k=1}^{N} e^{u_k}}, \tag{23}$$

wherein u_j is the output perceptron representing class j and N is the number of classes in the problem under analysis. The final prediction of the network model is given by the class that obtained the highest probability value.

2.6.1 CNN Models

A CNN consists of an alternating sequence of convolutional and *pooling* layers, together with activation functions, with at least one FC layer at the end. Several ways of combining such layers have been published in the literature, making the most varied CNN models available to researchers in the field. The first of these, the LeNet model, was proposed in (LECUN et al., 1998) for the classification of handwritten characters. LeNet was a simple network consisting of only 7 layers, as illustrated in Figure 12. Due to its simplicity, the model had very limited applications and was not suited to more complex problems (ALOYSIUS; GEETHA, 2017). It took almost 15 years for computer capability to evolve enough to support more complex models, which will be described in this section.

Figure 12 – LeNet architecture.

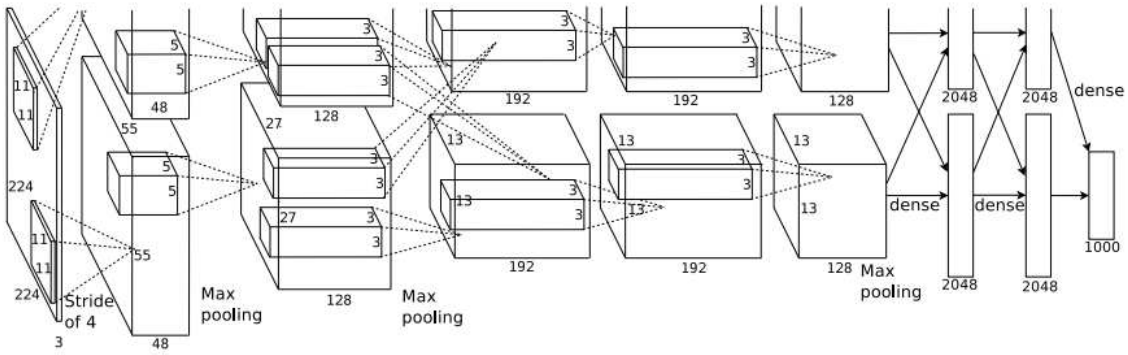


(Source: (TRA et al., 2017))

2.6.1.1 AlexNet

The publication of the AlexNet CNN model (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) was a milestone in the evolution of image recognition algorithms by significantly reducing the image classification error rate of the project *ImageNet* (ALOYSIUS; GEETHA, 2017). The AlexNet model consists of a 25-layer sequential architecture, illustrated in Figure 13, totaling 61 million parameters. In the following years, the performance of AlexNet was outperformed by other more complex CNN models.

Figure 13 – AlexNet architecture.



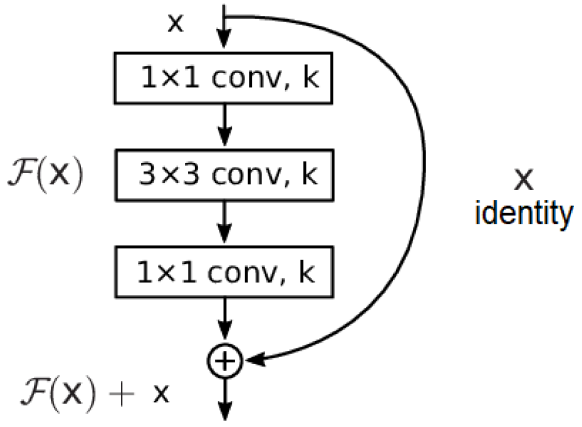
(Source: (KRIZHEVSKY; SUTSKEVER; HINTON, 2012))

2.6.1.2 ResNet-50 and ResNet-101

CNN models proposed after the success of AlexNet consisted only of incrementing the number of layers. However, there is a degradation point in which if the network becomes too deep, the accuracy is saturated (HE; SUN, 2015). Thus, in (HE et al., 2016) the concept of residual blocks has been applied, consisting of the use of "skip connections", as can be observed in Figure 14. The residual blocks are always composed of three layers that have k convolutional filters 1×1, 3×3 and 1×1, respectively. Given an input x,

the output corresponds to the sum of the residual mapping $\mathcal{F}(x)$ with the input itself (x identity).

Figure 14 – Residual block with skip connections.



(Source: adapted from (BEZERRA et al., 2019))

The CNNs ResNet-50 and ResNet-101, containing 50 and 101 layers, respectively, were proposed based on the concept of residual blocks. Moreover, these are among the first models to use batch normalization, which consists in normalizing the values received by the input layer in order to speed up the training process.

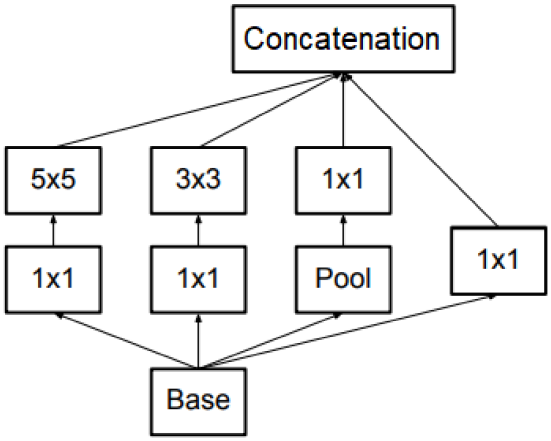
2.6.1.3 InceptionV3

Besides residual blocks, another concept that was proposed to solve the problem of accuracy convergence in CNN was the *Network in Network* (LIN; CHEN; YAN, 2013), which consists of inserting a micro MLP between two convolution layers. Based on this type of network, the concept of inception modules was introduced in (SZEGEDY et al., 2016). These modules consist of inserting a series of convolutional layers with *kernels* of different sizes between two layers of the network, as illustrated in Figure 15. The *inception* modules receive in the input layer, referred to as the base, a feature map that will pass through the module's internal convolution layers to be concatenated in the final layer into a single map. CNN InceptionV3, a model formed by 48 layers, is composed of this type of module.

2.6.1.4 Xception

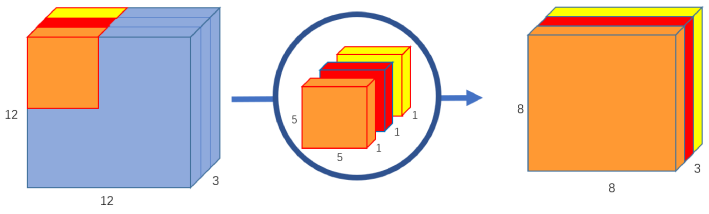
The Xception model, proposed in (CHOLLET, 2017), represents an abbreviation of the term "extreme inception", since its architecture is inspired by the InceptionV3 networks. Xception networks apply the concept of deep separable convolutions, which are operations consisting in the application of convolutional filters in two steps, as illustrated in Figure 16.

Figure 15 – Inception module.

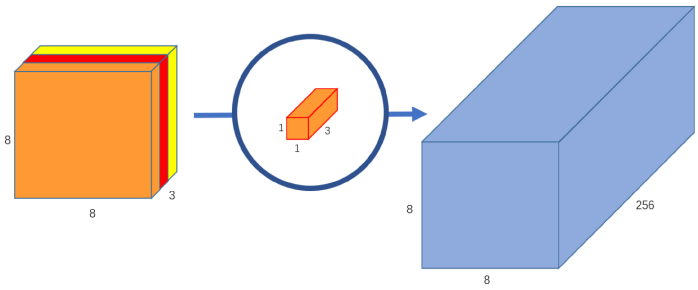


(Source: (SZEGEDY et al., 2016))

Figure 16 – Deep separable convolution operation appled in a $12 \times 12 \times 3$ image. The operation is split into two steps: depthwise convolution (a) and pointwise convolution (b).



(a)



(b)

(Source: (WANG, 2018))

Although this operation provides the same result as a classical convolution if the product of the kernels of each step was applied on the original image, a smaller number of multiplications are performed in this type of convolution. The Xception model is composed of 71 layers.

In Table 5, a summary of the CNN models detailed in this subsection is presented with number of layers and parameters, resolution of the input image and performance in the *ImageNet* challenge dataset.

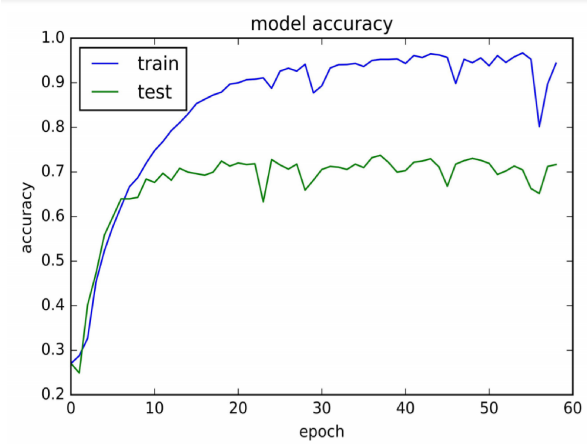
Table 5 – CNN models and performance in the *ImageNet* dataset.

Model	Layers	Parameters	Input resolution	Accuracy
LeNet	7	6.0×10^4	32×32	-
AlexNet	25	6.1×10^7	$227 \times 227 \times 3$	63.3%
ResNet-50	50	2.6×10^7	$224 \times 224 \times 3$	77.2%
ResNet-101	101	4.5×10^7	$224 \times 224 \times 3$	78.3%
InceptionV3	48	2.4×10^7	$299 \times 299 \times 3$	78.8%
Xception	71	2.3×10^7	$299 \times 299 \times 3$	79.0%

2.6.2 Transfer learning

One of the main limitations in the CNN context occurs when a set of images different from the one in which the model training occurred must be classified (RIBANI; MARENGONI, 2019). This occurs due to the fact that the test and training data must belong to a very similar feature space (PAN; YANG, 2009). The first solution would be to re-train the model from the images in the new dataset. However, this highlights two of the main problems in CNN. The first one is that training a CNN without established parameterization requires a set of tens of thousands of images, which is an even bigger problem in histopathology where histological image datasets with more than 1,000 samples are rare (YAMASHITA et al., 2018). Secondly, training a CNN is a high computational cost procedure. Furthermore, if the CNN is trained on a set with an insufficient number of samples, the phenomenon known as *overfitting* is likely to occur. In this case, the CNN is modelled to fit the training set very well, yet being unable to classify other data, such as a test set. A striking feature of CNN models with *overfitting* is a high training accuracy accompanied by a low accuracy on the test or validation set, as illustrated in Figure 17.

Figure 17 – Training of a CNN with overfitting.

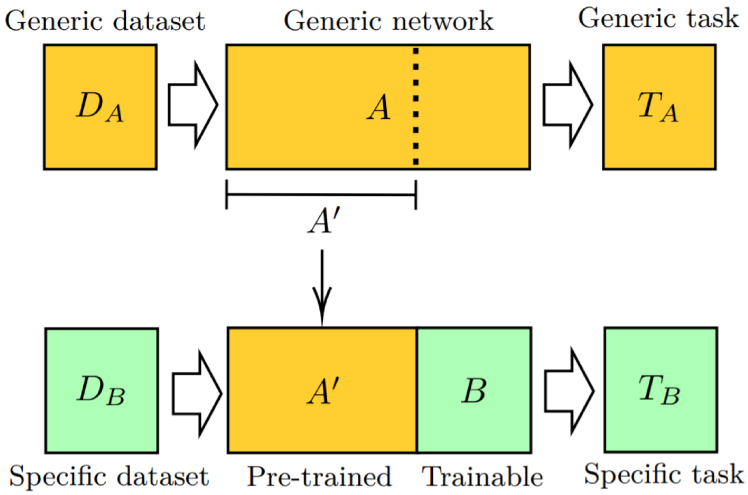


(Source: (LU, 2016))

One of the possible solutions is the use of data augmentation techniques, whose objective consists of enlarging the training set. There are several ways to apply such techniques, but the most commonly used in images involve geometric transformations such

as mirroring, cropping, rotation and translation (SHORTEN; KHOSHGOFTAAR, 2019). However, a lower computational cost solution consists in the use of transfer learning, an approach that uses pre-trained CNN models on an image dataset with a large number of samples as part of a new model that should be applied on a new set of images (RIBANI; MARENGONI, 2019). The basic transfer learning procedure is illustrated in Figure 18, where the CNN A was trained on a set of images D_A for a classification task T_A. For the classification task T_B, part of the model A was reused in training the set D_B. Thus, only a part B of the CNN needs to be trained on the new set.

Figure 18 – Transfer learning procedure.



(Source: (MARI et al., 2019))

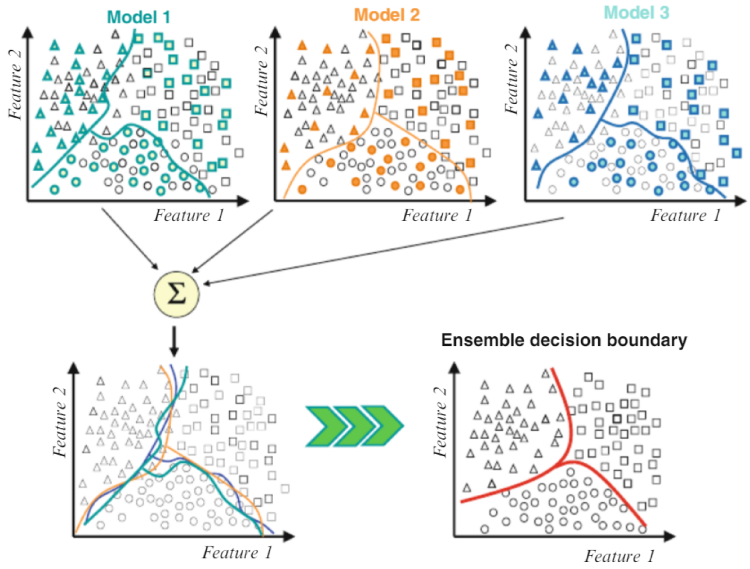
There are several transfer learning approaches, such as instance-based (LIU et al., 2017), mapping-based (LONG et al., 2017), network-based (CHANG et al., 2017) or adversarial (TZENG et al., 2017). In the most common classification problems where the use of a CNN model such as ResNet, AlexNet or Inception is enough, the network-based transfer learning is recommended (TAN et al., 2018). In this type of transfer, the initial layers of the network are kept, as the extraction of low-level features such as basic geometric shapes and contours occurs in these layers, which can be utilized to classify various image types (YOSINSKI et al., 2014). Generally, only the FC layers are changed and trained to suit the new dataset (GHAZI; YANIKOGLU; APTOULA, 2017).

Transfer learning applied for histological image classification is one of the main solutions to the problem of image datasets with few samples, since training a CNN whose weights are initialized based on a pre-trained model does not require a considerable number of (BEEVI; NAIR; BINDU, 2019) labeled samples.

2.7 Classifier ensemble

A classifier ensemble is among the most successful approaches in machine learning applications in recent years, due to its ability to train different models and combine their predictions (SAGI; ROKACH, 2018), with relevant results obtained in histopathology applications (KASSANI et al., 2019). This approach consists in applying a predetermined procedure to join the results generated by two or more classifiers and, from this union, provide a decision made based on the behavior of all classifiers employed, as illustrated in Figure 19.

Figure 19 – Illustration of the behaviour of an ensemble composed by three classifiers in a dataset described by two features.



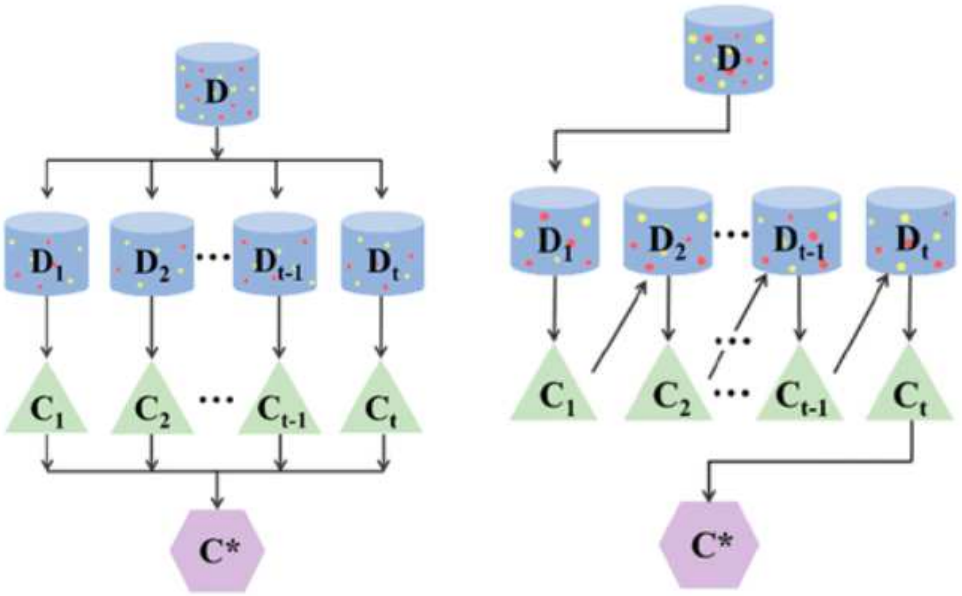
(Source: (POLIKAR, 2012))

There are two main approaches in the context of classifier ensembles: *bagging* and *boosting*. The former, illustrated in Figure 20 (a), consists in the parallel classification of a dataset D using t independent C classifiers. The *boosting* approach, illustrated in Figure 20 (b), consists in the sequential classification of the dataset D, where every time a C classifier performs a classification, the D features are enhanced by using weights, aiming to fix incorrectly classified instances (POLIKAR, 2012).

Although techniques based on *boosting* obtained better performance in binary classifications, it was observed that this approach is more susceptible to *overfitting* and noise, besides that the *bagging* approach generally tends to produce better models relative to separately applied classifiers (FERREIRA; FIGUEIREDO, 2012).

Among the main factors that motivate the use of classifier ensembles are the ability to avoid *overfitting*, greater representativeness of the problem and adaptability to unbalanced classes (SAGI; ROKACH, 2018). There are several approaches to perform the combination of the results obtained by the different classifiers of the committee, among which stand out:

Figure 20 – Classifier ensembles based on the approaches *bagging* (a) and *boosting* (b).



(a) (b)

(Source: Adapted from (YANG et al., 2019))

1. Sum rule: consists in combining the result of one or more classifiers by summing the prediction probabilities obtained by each classifier in each class (NANNI; LUMINI; GHIDONI, 2018);

2. Weighted sum: similar to the sum rule, with the inclusion of weight values that prioritise specific models or classes. This technique requires a pre-training step to determine the weights (ONAN; KORUKOĞLU; BULUT, 2016);

3. Majority voting: technique that simulates a voting system among the classifiers, where the class selected by the majority of the classifiers consists in the final prediction. According to (LEON; FLORIA; BĂDICĂ, 2017), the probability of the set of classifiers predicting the correct class is greater than that of which individual classifier, if the hit probability of each classifier is greater than 50%.

In general, a classifier ensemble is an approach that aims to make decisions based on opinions from different sources (SAGI; ROKACH, 2018). These approaches have generated relevant results in computer vision applications from the field of histopathology in classifications by means of CNN (NANNI; LUMINI; GHIDONI, 2018) or fractal features (ROBERTO et al., 2019).

2.8 Considerations

In this chapter, several computer vision methods applied in histological images were presented. Regarding the feature extraction step, techniques for obtaining fractal features

capable of describing different properties present in histological tissue were presented (IVANOVICI; RICHARD; DECEAN, 2009; CĂLIMAN; IVANOVICI, 2012; ROBERTO et al., 2017). These features aim to represent characteristics such as shape, irregularity or quantity of cell nuclei, in order to simulate the microscopic analysis techniques performed manually by pathologists (GURCAN et al., 2009).

Furthermore, CNN models, which serve both as feature extraction and classification, were presented and were able to provide good results in the context of histopathology (RAKHLIN et al., 2018; MAHBOD et al., 2018; JIANG et al., 2019). Using transfer learning and classifier ensemble approaches, the fractal features and image classification strategies via CNN can be investigated, aiming to highlight the strengths of each approach.

Chapter 3

Related Work

This chapter aims to present recent studies related to image processing in histology, the application of techniques based on fractal geometry for feature extraction and the use of CNN for quantification and classification of histological images. The described works allow to ground the model proposed in this thesis. On each section, the studies are grouped according to the type of histology image wherein they were applied, and then presented in chronological order.

3.1 Methods based on the classification of fractal features

Before the advent of CNNs as the most common tool in computer vision, feature extraction methods were the main way to obtain information from images (TRIPATHI; SINGH, 2020). However, even with the increasing use of deep learning techniques, the extraction of "handcrafted features" is still an important way to obtain descriptors of the structures present in histological images, such as shape, size, regularity and connectivity of cellular structures (KURMI; CHAURASIA; GANESH, 2019). Moreover, these techniques present a certain advantage in relation to deep learning models when applied to image sets with few samples, since CNNs depend on a large number of examples for a good generalization in the weight training stage (TRIPATHI; SINGH, 2020).

Techniques based on fractal geometry are part of this set of methods that obtain texture information. In (CHAN; TUSZYNSKI, 2016), FD was applied to differentiate eight classes of tumours from breast histological images. The method was explored on a public image dataset (SPANHOL et al., 2016), which consists of 7,909 images at 4 magnification scales. Initially, each image was binarized and the edge detected using the algorithm EdgeDetect, available in the software Mathematica (WOLFRAM, 2013). The approach used to obtain the FD was box-counting, with two box sizes ($L = 3$ and $L = 4$). Although the obtained *F-score* ($F1$) of 0.979 is relevant in a binary investigation, when

applied to multiclass classification with 8 groups the accuracy (A_c) metric was only 55.6%. All tests were performed using the SVM classifier. According to the authors, there is a need for new descriptors that can complement the FD metric in a multiclass investigation.

Fractal features can also be applied in conjunction with other texture features. Based on this concept, the authors at (FONDÓN et al., 2018) developed a method that associates FD with Haralick's texture descriptors, LBP, among other colour descriptors. These descriptors were applied to a dataset of 156 histological images of breast tumours, labelled in four distinct grades. Then, a feature vector containing 260 features was generated and given as input to nine different classifiers. The best classification result obtained among the four tumour grades was an accuracy of 75.8%, via the SVM classifier. Besides the association of several descriptors for the composition of the feature vector, this method presented as innovation the application of a histogram equalization algorithm in a pre-processing step. The applied algorithm was the contrast level adaptive histogram equalization (CLAHE). This algorithm consisted in adapting the histogram of all images from the dataset so that they are as similar as possible to a reference image. According to the authors, the application of this technique standardized the tested images in terms of dyes coloration and the main contribution of this work was the ability to classify a problem involving four classes, in which tumors were not only labeled as malignant or benign, but in different degrees of evolution.

Fractal features have also shown a good performance in the context of colorectal tumors classification. The method proposed in (RIBEIRO et al., 2019) applies the *curvelet* transform on colorectal histological images to obtain 41 *curvelet* sub-images from each original image. Then, fractal features given by the techniques proposed in (IVANOVICI; RICHARD; DECEAN, 2009; NIKOLAIDIS; NIKOLAIDIS; TSOUROS, 2011; CĂLIMAN; IVANOVICI, 2012; ROBERTO et al., 2017) and Haralick descriptors were extracted from the original images and the *curvelet* sub-images, which generated a vector containing 1,512 features. The ReliefF feature selection algorithm was applied and the set was reduced to 43 features, which were given as input to a polynomial classifier. The accuracy obtained by this method was 97.68% for this lesion classification model. It is worth mentioning that among the 43 selected features, 40 corresponded to fractal attributes, 26 from PERC and 12 from LAC, which indicates the relevance of fractal geometry in the context of histology. The authors suggested that different approaches to select the most relevant attributes could improve the results.

Thus, a method based on Genetic algorithm (GA) was proposed in (TAINO et al., 2019) aiming to perform a more complete evaluation of the 1,512 features obtained in the work (RIBEIRO et al., 2019) to classify histological images of colorectal tumors. In this proposal, the GA is applied to choose the classifier, feature selection method and number of features that are closest to an AUC of value 1.0. After 50 iterations with a population of 50 individuals, an accuracy of 90.82% was obtained with the selection of 29

features using the ReliefF algorithm and the *Random Forest* classifier. Although the same initial set of features was used, this approach provided lower results than (RIBEIRO et al., 2019). As the main difference between the two methods is in the classifier used, it can be concluded that the choice of a polynomial classifier was a decisive factor for obtaining the results reported in (RIBEIRO et al., 2019).

Another fractal-based approach was also applied in (ROBERTO et al., 2019), where fractal percolation features were extracted from histological images of colorectal and breast tumors. However, the application of the *curvelet* transform was replaced by a colour normalisation technique, where no sub-images are generated and features are extracted directly from the normalised images. The final feature set was composed of 63 percolation features, which were given as input to different classifiers. For breast tumour classification, the best accuracy (86.2%) was obtained with the *Logistic* classifier. The best accuracy value for colorectal tumors classification was 90.9% and was obtained by the *Rotation Forest* classifier. The authors also found that while global PERC features obtained better performance on colorectal images, the local PERC features at scales corresponding to low L observation values generated better classification results for breast tumors.

Although it is an approach with a lower computational cost, the performance obtained was also lower than that proposed in (RIBEIRO et al., 2019), which, besides percolation, also used LAC and FD features, as well as Haralick features, which may have been one of the reasons that made it possible to obtain a higher accuracy. As future works, the authors suggested tests in different image contexts and also new ways of exploring the attributes, such as organizing them as a feature matrix and providing this as input for a CNN.

Fractal geometry was also able to provide promising results in the classification of lymphomas. In (ROBERTO et al., 2017), PERC was used for the first time as a descriptor for colour histological images. The method was applied to segmented and non-segmented NHL images and global features were extracted from the curves of each of the three PERC functions generated, which was based on metrics proposed in (CĂLIMAN; IVANOVICI, 2012) and the calculation of the *Minkowski* distance. For non-segmented images, the best AUC value obtained was 0.940 with the K* classifier. However, there was no significant improvement in the AUC value obtained from images segmented by an expert (TOSTA et al., 2017), which was 0.944 using the DECORATE classifier. These results were obtained from a set of only 15 features, which indicates the feasibility of the PERC metrics for the classification of NHL images. The analysis of the method in different sets of histological images and the evaluation of different distance calculation approaches to obtain the features may improve the results obtained with this technique.

In (MARTINS et al., 2019), other fractal features (FD and LAC) were obtained from NHL images coloured in RGB and LAB models. The features extracted from the original images, without data augmentation, composed feature vectors containing 18 attributes

each, which were provided as input to a polynomial classifier. For the binary classification, an accuracy of up to 97% was obtained. The use of fractal features together with different colour models proved to be efficient in this context of images. Moreover, as occurred in (RIBEIRO et al., 2019), which consisted in the analysis of colorectal tumors, the polynomial classifier was also able to obtain good results in the context of NHL images, which is a probable indicator of its good performance when associated with fractal features. However, according to the authors, the main limitation of the classifier is the processing time, which increases exponentially with the increase in the number of features.

Although fractal geometry has shown good performance as a metric for quantification of histological lesions, we have not found any applications of these approaches in other histopathology problems outside the context of disease identification. In (SHAMIR et al., 2008) a series of biological image classification challenges were proposed. One of these consisted in the differentiation of gender and age based on images of liver tissue obtained from mice. Some works stood out in this context, such as in (WATANABE; KOBAYASHI; WADA, 2016), where a semi-supervised feature transformation method was proposed through fusion between PCA and LDA techniques. This method was applied to features extracted from liver tissue images through the GIST global descriptor, one of the most popular feature extraction techniques (DOUZE et al., 2009). The features were given as input to a nearest mean classifier, which provided accuracies of 88.4% and 93.7% for age and gender classification respectively. In (RUBERTO et al., 2016), three statistical approaches were applied for texture analysis in different colour spaces. After using SVM to classify the generated features, an accuracy of 100% was obtained for gender and age classes. Despite the relevant classification performance, employing fractal descriptors may be relevant to discover the behaviour characteristics of the method on using other texture descriptors.

In Table 6, a summary is presented with information on methods that have applied handcrafted features in the context of breast, colorectal, NHL and liver histological images.

3.2 Methods based on deep learning

In the context of medical image classification with CNN, breast tumour analysis was one of the first applications to be investigated. In (CIREŞAN et al., 2013), the authors implemented a CNN for mitosis detection in breast cancer. In this approach, each *pixel* of the image was labelled as mitosis or non-mitosis. Then, square windows were generated representing different regions of the image, where a window was labelled as mitosis if its *pixel* was at most 10 *pixels* away from a *pixel* initially labelled as mitosis. The windows were given as input to two CNN models, initially proposed in (CIREŞAN; MEIER; SCHMIDHUBER, 2012), with 12 and 10 layers, respectively. As the mitosis detection problem is rotation invariant, eight variations of each image were constructed by applying

Table 6 – Overview of methods based on handcrafted features applied in histology images.

Method	Image	Approach	Results
(CHAN; TUSZYNSKI, 2016)	Breast tumours	FD	0,979 (F1)
(FONDÓN et al., 2018)	Breast tumours	FD, Haralick, LBP and other descriptors	75,80% (Ac)
(PAPASTERGIOU; ZACHARAKI; MEGA-LOOIKONOMOU, 2018)	Breast tumours	Spatial decomposition, tensors	84,67% (Ac)
(JØRGENSEN et al., 2018)	Colorectal tumours	Different colour models, histogram and GLCM	0,950 (AUC)
(SANTOS et al., 2018)	Colorectal tumours	Sample entropy, fuzzy function	0,983 (AUC)
(RIBEIRO et al., 2019)	Colorectal tumours	Fractal features, Haralick, curvelet transform	97,68% (Ac)
(ROBERTO et al., 2019)	Breast tumours	PERC, colour normalizatiopn	86,21% (Ac)
	Colorectal tumours		90,91% (Ac)
(ROBERTO et al., 2017)	NHL	PERC	92,00% (Ac)
(MARTINS et al., 2019)	NHL	FD and LAC	97,00% (Ac)
(JIANG et al., 2018)	NHL	Morphology, entropy, GLCM and other descriptors	97,96% (Ac)
WATANABE et al., 2016	Liver (gender)	GIST descriptors, PCA and LDA	93,70% (Ac)
	Liver (age)		88,40% (Ac)
(RUBERTO et al., 2016)	Liver (gender)	Statistical analysis of texture features	100,00% (Ac)
	Liver (age)		100,00% (Ac)

four 90° rotations with and without mirroring for evaluation of the two CNNs. After combining the classification results of the two CNNs, the obtained *F-score* was 0.758. According to the authors, the main limitation of this approach lies in the need for a bulky image set, a problem that occurs in much of CNN applications (YAMASHITA et al., 2018).

In (ARAÚJO et al., 2017), the authors developed a custom 13-layer CNN for cancer classification in breast histological images. A dataset of 70,000 images of resolution 512×512 were given as input for training the CNN. The images belonged to four different classes: normal tissue, benign abnormality, carcinoma *in situ* and invasive carcinoma. Importantly, these 70,000 images were generated from 250 available breast histological samples, by cutting and rotating them. Initially, the colour channel normalization algorithm available at (MACENKO et al., 2009) was applied to all images in the dataset. This

algorithm aimed to standardise the colours of the H&E dyes in an initial process of the analysis. According to the authors, for a good classification of breast cancer images, CNN should extract features at different scales, especially regarding the analysis of nuclei and histological tissue as a whole. Thus, the CNN architecture was implemented to identify these features in different layers: a) the identification of the edges occurs in the first two layers; b) the analysis of the nuclei was performed in layers 3, 4, 5 and 6; c) layers 7, 8, 9 and 10 were responsible for exploring the general structure of the tissue represented in the image; and d) the classification occurred in the last three layers. The obtained accuracies were 77.8% and 80.6% for the classification of 4 (normal, benign, carcinoma *in situ* and invasive carcinoma) and 2 (benign and malignant) classes, respectively. The authors also tested the performance of the convolution filters when classified in an SVM algorithm. The values obtained after the 12th layer of the network were employed to build a feature vector given as input to the SVM. The obtained accuracy values exceeded the classifications performed in the CNN in relation to the 2 classes (83.3%) and maintained the same values in the comparison of 4 classes (77.8%). The results obtained were relevant, with highlight to the good accuracy indexes with experiments of more than 2 classes and the association of an external SVM classifier to complement the CNN. However, according to the authors, accuracy values can be improved with the use of other feature extraction methods besides convolution-based features.

Recently, with greater availability of computational resources, the implementation of more complex CNN models, as well as the combination of different CNNs has become a recurrent practice in research in the area. In (LI et al., 2019), the authors used data augmentation to increase the number of samples from the breast cancer dataset, generating sub-images of size 112×112 *pixels*. To improve classification, the authors applied a six-layer CNN called *RefineNet* to remove incorrectly labelled sub-images. Then, the sub-images were given as input to a new CNN model named *Atrous DenseNet*. This model consists of the use of the atrous convolution operation, which consists of applying a dilation to the *kernels* based on a λ parameter, in order to increase the information capture area for the construction of the feature maps. With the combined use of the atrous convolution with the traditional convolution, this model was able to obtain an accuracy of 100%. According to the authors, this operation is capable of highlighting multiscale attributes in an image, besides providing a balance between the network depth and the learning capacity, which is effective for sets with few samples.

In the context of colorectal tumor classification, the use of CNN has also provided excellent results. In (DABASS; VIG; VASHISTH, 2019), the authors proposed a custom 31-layer CNN to perform classification of colorectal histological images, achieving an accuracy of 93.24% and 96.97% for the classification of 5 and 2 classes, respectively. The proposed model resembles the AlexNet network architecture (KRIZHEVSKY; SUTSKEVER; HINTON, 2012), however the dimensionality of the convolution *kernels* were changed so that

it was not necessary to change the original resolution of the images, as usually occurs in good parts of CNN applications where each model requires a different input resolution. Similar results were obtained in (SENA et al., 2019) when classifying 4 categories of colorectal tumors with a custom 12-layer network, in which an accuracy of 93.08% was obtained after 400 training epochs. Both in the work of (DABASS; VIG; VASHISTH, 2019) and (SENA et al., 2019), the optimization of hyper-parameters and the increase in the number of CNN layers are indicated as necessary factors for improving the performance obtained by CNN models for the classification of colorectal images.

Some researchers have recently experimented with approaches consisting of ensembles of different CNN models. In the method proposed in (AWAN et al., 2019), for example, a normalisation of the colour channels was initially applied to the colorectal images. Then, the normalised images were provided as input to a U-Net model CNN to perform segmentation, in order to remove non-glandular areas. The CNN models GoogLeNet, DenseNet and ResNet-50 were tested to classify the images after the segmentation step. However, due to the depth of these networks, the classification would be susceptible to *overfitting* if a large enough sample set was not trained. To avoid this problem, the data were augmented by mirroring and rotation operations. During classification on CNN, each image was divided into six sub-images of resolution 256×256, where the probability vectors obtained for each sub-image were summed and the final prediction was given by the class that obtained the highest probability after applying the sum rule. The best accuracy was obtained by CNN GoogLeNet, by classifying four classes of segmented colorectal images, with a value of 85%. According to the authors, segmentation was a decisive factor for the improvement in the classification of this type of images, however, superior results were obtained in non-segmented colorectal images (DABASS; VIG; VASHISTH, 2019; SENA et al., 2019).

In regard to the study of colorectal tumour detection methods, a set of different CNNs was also proposed in (TAVOLARA et al., 2019). The authors developed an approach based on Generative adversarial networks (GAN), where the generator was implemented as a U-Net and the discriminator is a standard CNN. With 3,000 colorectal tumour image samples available, the method provided a *F-score* of 0.940. According to the authors, this approach allows exploring data in unbalanced classes due to the fact that, instead of training a generic CNN model for the whole set of images, a different model must be trained for each class. This strategy allows new classes to be added without the need to re-train the entire set, but it is computationally expensive since several models must be trained initially instead of just one. However, this study does not provide details on the description of the model implemented, such as, for instance, which parameters or which number of layers of the CNN architecture was used as discriminator.

In (JANOWCZYK; MADABHUSHI, 2016), NHL images were also investigated for CNN model-based approaches. These images were split into sub-images of resolution $32 \times$

32 using the Caffe architecture to generate 825,000 training samples. These samples were given as input to an AlexNet CNN model that applied a classification scheme similar to that used in (AWAN et al., 2019), however instead of combining the subimage classification by the sum rule, a majority vote scheme was used, where the class with the highest number of votes in the subimage classification was considered as the class of the complete image. For the classification, cross-validation of 5 *folds* was used, which were composed of 300 images for training and 75 images for testing each, which provided an accuracy of 96.58%. The authors observed that in cases where the vote was won by a small margin, the chance of the image being incorrectly classified was higher. In such cases, it was suggested that the classification should be done visually by an expert. The authors suggested the use of handcrafted features to complement and improve the CNN classification.

Using a denser CNN, the authors at (TAMBE et al., 2019) were able to outperform the results of (JANOWCZYK; MADABHUSHI, 2016) by applying a model similar to InceptionV3. Data augmentation was applied to the image set through rotation and displacement operations, resulting in a set of 225 images for training and 75 images for testing. After running tests with different values for the hyperparameters, the method was able to obtain an accuracy of 97.33% for differentiating the three classes of NHL, considering a fixed learning rate of 10^{-3} and 35 training epochs. The method was able to obtain relevant results even without applying segmentation techniques and with a relatively low number of training epochs. The use of handcrafted features and segmented images was indicated by the authors as a possible improvement of the work.

Gender and age identification through histological tissue analysis has also been investigated with the application of CNN models. In (ANDREARCZYK; WHELAN, 2017), a new deep learning approach called *Texture*-CNN was presented, which consists of a 12-layer architecture developed specifically for texture classification. Each image was divided into 24 non-overlapping sub-images, which were then rescaled to the input resolution of the model which was defined as a *patch* of size 227×227 pixels. Each of the 24 sub-images was classified by CNN and then the probability vectors obtained by the classification are summed and the class with the highest value corresponds to the classification of the original image. After applying this approach, together with a voting classification scheme, accuracies of 99.1% and 98.2% were obtained for the classification of 2 gender and 4 age classes, respectively. The approach adopted resembles the sum rule used in the classifiers ensembles. The division of the image into 24 sub-images also served as a way to increase the dataset, since the images analysed were of regular textures and the same can also be used for training without loss of information, since the textures of the sub-images generated were similar to the texture of the original image. The authors suggested that a deeper network could generate better results, but highlighted the main limitation of the application of CNN models in histological images as being the difficulty in obtaining good results with a small number of samples.

A classifier ensemble can also be applied in approaches based only on deep learning. After selecting the ResNet-50 model, the authors in (NANNI et al., 2020) generated 9 variations of this model by replacing the activation function ReLU by the functions *LeakyReLU*, *SeLu*, *SReLu*, *APLu*, *GaLu*, *preLu*, *MeLu* and *wMeLu*. The set of liver tissue images was augmented by mirroring and scaling operations. The resulting images were given as input to the 10 available CNN models and trained for 30 epochs with a fixed learning rate of 10^{-4}. When analysed separately, the standard ResNet-50 model, which applies the function *ReLu* as the activation function, provided the best results with accuracies of 93% and 85.1% for gender and age classification, respectively. However, when the results of the 10 models are combined using the sum rule, accuracy values of 98.3% and 91.16% were obtained for gender and age classification, respectively. According to the authors, this approach introduces diversity to CNN models, which would be ideal in a classifier ensemble context. Moreover, due to the transfer learning applied, a relatively low number of training epochs was sufficient to obtain relevant results. An evaluation of this approach on other CNN models could provide more information about the behaviour of different activation functions in CNNs and how they can complement the classification.

In Table 7, a summary is presented with information about methods that have applied deep learning in the context of breast, colorectal, NHL and liver histological images.

Table 7 – Overview of methods based on deep learning applied in histology images.

Method	**Image**	**Approach**	**Results**
(CIREŞAN et al., 2013)	Breast tumours	12- and 10-layer CNNs	0,782 (F1)
(ARAÚJO et al., 2017)	Breast tumours	Colour normalization, 13-layer CNN and SVM	83,30% (Ac)
(LI et al., 2019)	Breast tumours	RefineNet and Atrous-DenseNet	100% (Ac)
(DABASS; VIG; VASHISTH, 2019)	Colorectal tumours	31-layer CNN, similar to AlexNet	96,97% (Ac)
(SENA et al., 2019)	Colorectal tumours	12-layer CNN	93,28% (Ac)
(AWAN et al., 2019)	Colorectal tumours	Colour normalization, U-Net and GoogLeNet	85,00% (Ac)
(TAVOLARA et al., 2019)	Colorectal tumours	GAN and U-Net	0,940 (F1)
(JANOWCZYK; MADABHUSHI, 2016)	NHL	Caffe and AlexNet	96,58% (Ac)
(TAMBE et al., 2019)	NHL	InceptionV3	97,33% (Ac)
(ANDREARCZYK; WHELAN, 2017)	Liver (gender)	Texture CNN	99,10% (Ac)
	Liver (age)		98,20% (Ac)
(NANNI et al., 2020)	Liver (gender)	ResNet-50 and different activation functions	98,30% (Ac)
	Liver (age)		91,16% (Ac)

3.3 Hybrid methods

As occurred with deep learning-based methods, hybrid methods also had in breast tumour images one of the first contexts of application in histological images. In (WANG et al., 2014), the authors proposed a combination of attributes generated by a CNN and handcrafted features obtained via morphological, intensity and texture descriptors for the identification of mitosis in breast cancer histological images. A total of 253 handcrafted features such as area, perimeter, Hausdorff dimension, mean, variance, skewness and Haralick descriptors were extracted. Due to the high computational cost of a CNN, the authors opted for a custom CNN implementation with only 3 layers. Thus, three classifications were performed in this approach: first, the handcrafted features composed a feature vector for a *Random Forest* classifier. Then, another classification was performed using the same three-layer CNN. Finally, in case of conflict in the result of the first two classifications, a third classification is performed combining the two sets of features, handcrafted and convolutional, which were again classified by the *Random Forest* classifier. The *F-score* obtained was 0.735, and when comparing only the CNN without complementary attributes the *F-score* was 0.573. Despite being a relatively fast method, when compared to other CNN-based approaches, the accuracy obtained was not expressive, according to the authors. As a solution, it was suggested for future works the addition of more layers to improve the performance of the classifications and the use of GPU.

In (YU et al., 2019), an automatic breast cancer detection method was developed by applying handcrafted feature extraction and segmentation via CNN. The segmentation of breast tumour nuclei was performed by applying a custom 9-layer CNN whose output consists of three binary images highlighting the nuclei, the contour of the nuclei, and the image background, respectively. Then, texture, morphological, and spatial features such as LBP, SURF, GLCM, Delaunay diagram and number of nuclei, were obtained from the segmented images and provided as input to an SVM classifier. After applying the *Relief* feature selection method, the approach provided an accuracy of 96.7%. This result was higher than studies that employed only handcrafted features or CNN for breast cancer image classification. Moreover, the authors showed that the proposed nine-layer CNN performs a more accurate segmentation than that of other classical techniques such as Otsu's method or *watershed* (VETA et al., 2011). However, no comparison was made regarding the performance of the same set of features in non-segmented images, which hinders an analysis about the importance of using segmentation in this type of image.

Besides the combination of different types of handcrafted features, different CNN models can be used for the composition of a set of classifiers, as proposed in (TRIPATHI; SINGH, 2020) for colorectal image classification. In this method, the CNN models named AlexNet, VGG16, VGG19, ResNet50, DenseNet121 and InceptionV3 were applied on a colorectal histological tissue base. The features generated by the FC layers of each CNN are extracted and stored in feature vectors. Classical feature extraction techniques such as

SURF, LBP and histogram oriented gradients were also applied on the colorectal image base. The authors propose two models to perform the fusion of different types of attributes. The first, named as "cascade model", applied the PCA technique on each set of features to reduce the dimensionality of the data and then perform classifications through an MLP network. The fusion of the probability values obtained for each class by each set of features is done using the sum rule, where the set of classifications performed by the six CNNs has the same weight as the results of the classifications obtained by using the classical features. The cascade model provided a *F-score* of 0.981. The second proposed model, named as "concatenated model", consists of the concatenation of all the obtained features, which includes the six CNNs and all the classical feature extraction techniques. The *F-score* value obtained with this model was 0.985. The results obtained with both models indicate that the combination of classifications obtained by handcrafted features and classifications via CNNs can provide important contributions to the literature. However, the use of six CNN models makes this method highly costly, and difficult to apply in larger sets of images.

In (BAI et al., 2019), the authors proposed a method that applies transfer learning and handcrafted feature extraction for NHL classification. In this approach, colour and texture features such as LBP and histogram statistics, were extracted from cropped samples of the original images and provided as input to a *Random Forest* classifier and a pre-trained CNN from the GoogLeNet model. Both the *Random Forest* and CNN provided class predictions that were combined using three different techniques: majority vote; sum rule; and weighted sum. The best accuracy obtained was 99.10% with the use of the weighted sum technique. According to the authors, this work was able to show the efficiency of the transfer learning in the classification of lymphomas, besides the relevance of the assignment of weights to perform the combination of different classifiers. The application of this method in different image bases and the test with other pre-trained CNN models were suggested for future works.

More recently, classifier ensemble techniques for handcrafted features have been applied to obtain better results from this image dataset. The authors in (NANNI et al., 2019) proposed a combination of 12 texture features and deep learning approaches. In addition, new data augmentation techniques based on principal component analysis and discrete cosine transform were also presented. Using a set of 6 CNN models trained with different data augmentation approaches and a set of features, the method was also able to obtain an accuracy of 100% for gender and age classification from liver histological images. According to the authors, one of the main contributions of the work consisted in the indication that different data augmentation methods can be used to compose a set of CNNs in order to combine their respective classification results. The authors also suggest that the work could be enhanced with methods to train CNNs on sets with a small number of images.

In Table 8, a summary is presented with information on hybrid methods applied in the context of breast, colorectal, NHL and liver histological imaging.

Table 8 – Overview of hybrid methods applied in histology images.

Method	Image	Approach	Results
(WANG et al., 2014)	Breast tumours	3-layer CNN, intensity, texture and morphological features	0,735 (F1)
(YU et al., 2019)	Breast tumours	CNN, LBP, SURF, GLCM and other descriptors	96,70% (Ac)
(TRIPATHI; SINGH, 2020)	Colorectal tumours	6 CNN models, LBP, SURF, gradient histograms	0,985 (F1)
(BAI et al., 2019)	NHL	GoogLeNet, intensity, texture and morphology features	99,10% (Ac)
(NANNI et al., 2019)	Liver (gender)	6 CNN models and other descriptors	100,00% (Ac)
	Liver (age)		100,00% (Ac)

3.4 Considerations

This chapter presented several computer vision methods for the classification of histological images of breast tumours, colorectal tumours, NHL and liver tumours.

Although handcrafted features have performed well in various histological imaging contexts, the increasing evolution of deep learning techniques cannot be ignored. The improvement of approaches such as transfer learning, data augmentation and *fine-tuning* have served as options to circumvent the limitations imposed by sparse medical image sets (TRIPATHI; SINGH, 2020). Thus, researchers have increasingly delved into the use of deep learning in recent years, especially with the exploration of different CNN models.

Regarding deep learning-based methods, although these have provided excellent results in several application contexts, becoming the most popular approach in recent research in computer vision, CNNs still have certain limitations. Especially when used for the classification of histological images, where datasets with significant numbers of samples are rare, making it difficult to train the networks (YAMASHITA et al., 2018). Thus, other types of features have been used to complement the classifications performed by deep learning models (NANNI et al., 2019).

On the other hand, hybrid models attempt to exploit the strengths of each approach, combining the great discriminative power of CNNs with the adaptability to specific features of different histological lesion types of the handcrafted features. This association have shown promising results in recent years (YU et al., 2019; BAI et al., 2019; NANNI

et al., 2019; TRIPATHI; SINGH, 2020). However, these methods were implemented for specific classification tasks, so only the method from (NANNI et al., 2019) was able to provide relevant results in different histological imaging contexts.

In general, there are few computer vision approaches that have been able to perform well on different categories of histological images (SHAMIR et al., 2008; ILSE; TOMCZAK; WELLING, 2018; ROBERTO et al., 2019; NANNI et al., 2019). Fractal features (RIBEIRO et al., 2019; ROBERTO et al., 2019; ROBERTO et al., 2017) and CNN models (LI et al., 2019; ANDREARCZYK; WHELAN, 2017; BAI et al., 2019) were able to provide high accuracy rates in several CAD systems for histopathology tasks, but usually in separate applications. These features also performed well on image sets with few samples, and their application in conjunction with a CNN would be a possible solution to circunvent the problem of insufficient training samples. Therefore, a hybrid method that addresses fractal geometry and deep learning can be investigated to improve these results when applied to different histological datasets. Nevertheless, models usually consist of classifying the handcrafted features with a common classifier such as SVM or decision tree and then combining the result with the classification of one or more CNNs using the sum rule or majority vote.

A novel approach of converting the handcrafted features into a synthetic model can be investigated as input to a CNN. Studies based on this approach are being addressed in the literature as highlighted in the works (LUMINI; NANNI, 2018; KAUSAR et al., 2019). Some works have applied well known techniques such as Recurrence plot (RP) and Gramian angular fields (GAF) in order to convert handcrafted features into a visual representation and classify it with a CNN. These approaches have been applied recently for tasks such as Parkinson's disease identification (AFONSO et al., 2019), time series classification (YANG; CHEN; YANG, 2020), epileptic seizure detection (SHANKAR et al., 2020) and cardiac arrhythmia classification (ZHANG et al., 2021). However, there is still no research exploring this kind of approach with methods based on fractal geometry, while we also noticed a lack of studies regarding the use of RP or GAF as a tool for features representation in histology images classification. We believe that new investigations can contribute in the process of histology lesions classification, addressing this gap in the literature.

Chapter 4

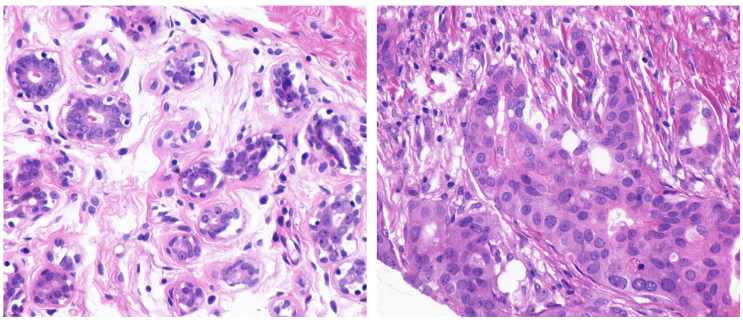
Methodology

In this chapter, the proposed method is presented. Firstly, the datasets wherein the evaluation was performed are introduced. Then, a detailed step-by-step description of the method is described. At last, the performance evaluation section presents the testing procedures, tools, and metrics that were used to verify the method's performance.

4.1 Image datasets

In order to investigate the performance of the proposed approach, five histological image datasets were selected. The first consists of the breast cancer set of images provided by the *Center of Bio-Image Informatics* of the *University of California, Santa Barbara* (UCSB) (GELASCA et al., 2008). This dataset consists of 58 images of breast tissue divided into two groups: benign (32) and malignant (26). Examples of this dataset are illustrated in Figure 21.

Figure 21 – Examples of a benign (a) and a malignant (b) case from the UCSB dataset.



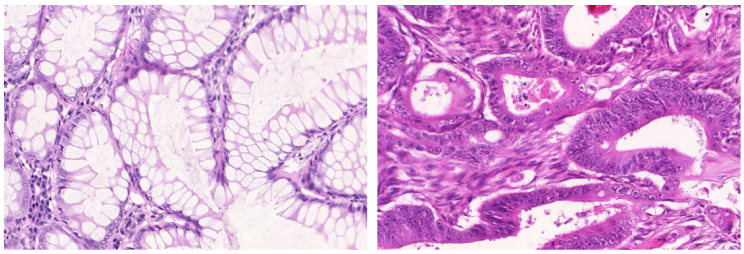
(a) (b)

(Source: (GELASCA et al., 2008))

The second base consists of 165 colorectal (CR) tissue images, also divided into benign (74) and malignant (91) tumour groups (SIRINUKUNWATTANA et al., 2017). To obtain these images, histological sections were digitally photographed using a *Zeiss MIRAX*

MIDI Slide Scanner at a pixel resolution of 0.620μm, which corresponds to a magnification of 20x. In Figure 22, an example of each class is shown.

Figure 22 – Examples of a benign (a) and a malignant (b) case from the CR dataset.

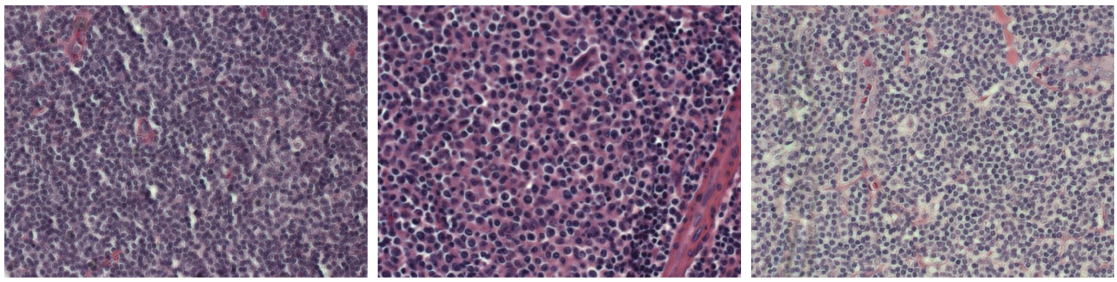


(a) (b)

(Source: (SIRINUKUNWATTANA et al., 2017))

The third dataset is composed of 173 non-Hodgkin's lymphoma (NHL) images divided into three classes: Mantle cell lymphoma (MCL) (99); Follicular lymphoma (FL) (62); and Chronic lymphocyte leukemia (CLL) (12). For image acquisition, a *Zeiss Axioscope* light microscope with 20x objective and a digital colour camera (*AXio Cam MR5*) were used. The obtained images were stored uncompressed, with a quantization rate of 24 bits in the RGB colour model. Then, regions of interest were selected by an expert. This image set was publicly available from the *National Cancer Institute* and the *National Institute on Aging* (SHAMIR et al., 2008). An example of each case is shown in Figure 23.

Figure 23 – Examples of a CLL (a), FL (b) and MCL (c) case from the NHL dataset.



(a) (b) (c)

(Source: (SHAMIR et al., 2008))

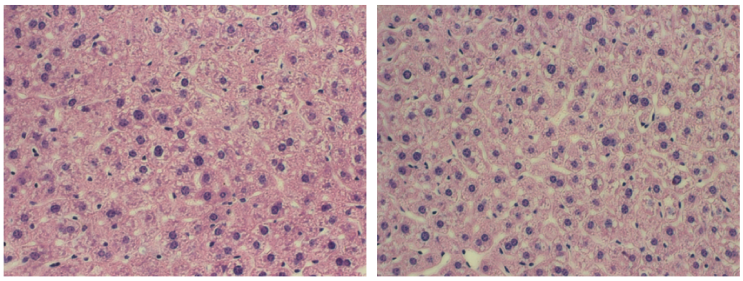
The following two image datasets were both provided by the Atlas of Gene Expression in Mouse Aging Project (AGEMAP) and are composed of images of liver tissue obtained from mice (AGEMAP, 2020). Images were obtained using a *Carl Zeiss Axiovert 200* microscope and a 40x objective. All images have the same resolution of 417×312 *pixels*. The dataset *Liver Gender* (LG) consists of 265 images of liver tissue obtained from male (150) and female (115) mice subjected to calorie restricted diets. Examples of each class are illustrated in Figure 24. The *Liver Age* (LA) dataset is composed of 529 images

Table 9 – Overview of the tested image datasets.

Datasets	Images	Classes	Samples	Resolution
UCSB	Breast tumours	2	58 (32/26)	896×768
CR	Colorectal tumours	2	165 (74/91)	from 567×430 to 775×522
NHL	Non-*Hodgkin* lymphomas	3	374 (113/139/122)	1388×1040
LG	Liver tissue	2	265 (150/115)	417×312
LA	Liver tissue	4	529 (100/115/162/152)	417×312

divided into four classes, which each class represents a different age group of female mice on *ad-libitum* diets: one month (100), six months (115), 16 months (162) and 24 months (152) of age. In Figure 25, an example of each age group is presented. An overview of the five image bases employed in this work is presented in Table 9.

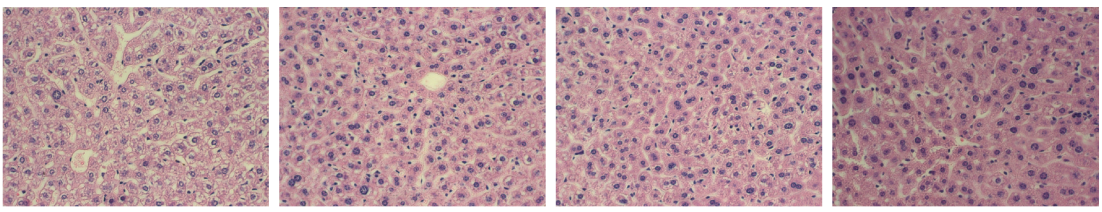
Figure 24 – Samples of liver tissue from male (a) and female (b) mice from the LG dataset.



(a) (b)

(Source: (AGEMAP, 2020))

Figure 25 – Samples of liver tissue from female mice aged one (a), six (b), 16 (c) and 24 (d) months, from the LA dataset.



(a) (b) (c) (d)

(Source: (AGEMAP, 2020))

4.2 Method overview

The approach proposed in this work was divided into two main phases. The first phase performs the extraction of local and global features by applying the fractal techniques described in Section 2.4. The output of this phase consists of a set of 300 local and 63 global features, which were obtained by calculating the values of FD, LAC and PERC for each of the Δ distances investigated. The second phase is where the classifications of the

features and images obtained in the first phase are carried out, whose aim is to combine these classifications using an ensemble to obtain class probability vectors.

There are four classifications that occur in the second phase. The first one is performed by a CNN, hereafter called O-CNN (Original image CNN), which receives as input the original image. The second classification is also performed by a CNN, hereafter named F-CNN (Feature image CNN), and it consists in the classification of an artificial image generated from the 300 local features obtained by the application of the FD, LAC and PERC techniques. This set of features is transformed into a square RGB image through a reshaping procedure (LUMINI; NANNI, 2018; KAUSAR et al., 2019). Three different reshaping procedures were evaluated: sequential reshaping (ROBERTO et al., 2021); recurrence plot (AFONSO et al., 2019); and gramian angular fields (GAF) using the summation and difference approaches (WANG; OATES, 2015). The third and fourth classifications were performed by traditional machine learning algorithms and consist in the classifications of the local and global feature vectors, respectively. Then, the values of class probabilities obtained by each of these classifications were combined in an ensemble using the sum rule, similarly to (NANNI; LUMINI; GHIDONI, 2018). After this step, the highest probability value indicated the class prediction. An overview of this approach is presented in Figure 26, with each step described in details in the following sections.

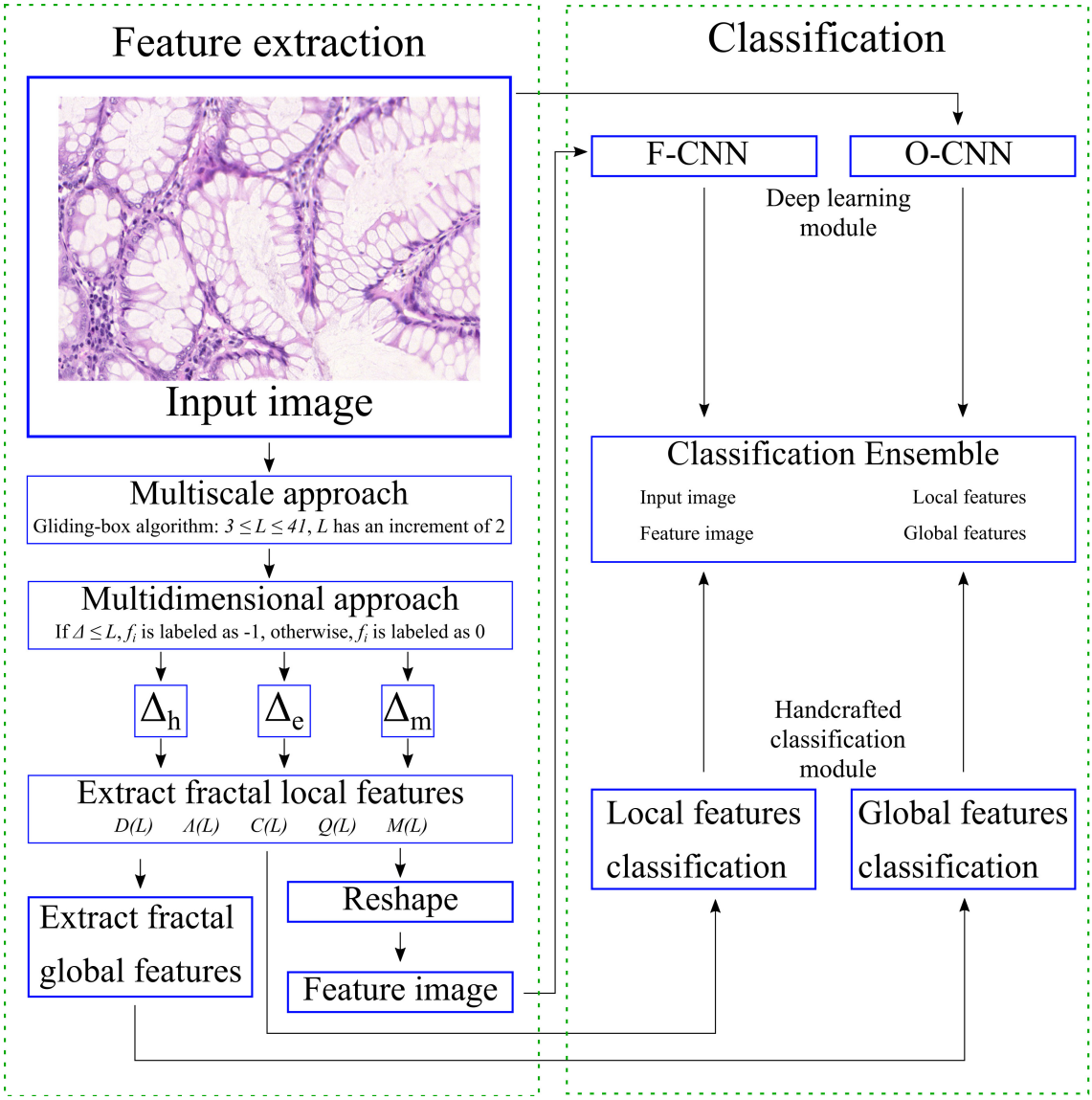
4.3 Feature extraction

The main stage of the proposed method consists in applying techniques based on fractal geometry, described in Section 2.4, on the images under investigation. Local and global FD, LAC and PERC features are extracted by applying multi-scale and multidimensional approaches.

4.3.1 Local features

Each input image is divided into different scales according to the *gliding-box* algorithm, which generates a set of matrices for each region of the image at each assigned value of L, ranging from 3 to 41 with an increment of 2 (RIBEIRO et al., 2019; ROBERTO et al., 2019). This provides a total of 20 features per function ($C(L)$, $Q(L)$, $M(L)$, $\Lambda(L)$ and $D(L)$), which results in 100 features for each Δ. In order not to generate distortion when resizing the image when it is given as input to CNN, the value 41 was chosen for L_{max} since it generates an amount of features that can be organized as a square matrix after the rearrangement procedures, which will be detailed in Section 4.4. Furthermore, this value for the parameter L_{max} has demonstrated relevant results in classification as presented in (RIBEIRO et al., 2019). Then, the multidimensional approach described in Section 2.4.1 is applied to the generated matrices, which consists in the computation of a different Δ distance between the *pixels* of the region, as illustrated in Figure 5 and Equations 2-4.

Figure 26 – Overview of the proposed method.



(Source: author)

The application of the multidimensional approach outputs a set of binary matrices, whose values labelled 1 correspond to the *pixels* that have satisfied the Δ distance criterion. These matrices are given as input to the third stage of the proposed method, where the techniques described in Section 2.4 for obtaining the local values of FD, LAC and PERC are finally applied according to Equations 6-13. The resulting output consists of a set of 300 local features. These features are classified on two different occasions. First, they are given as input to a classic machine learning algorithm as a regular 300-dimensional handcrafted feature vector. On a second occasion, the 300 features are reshaped into a feature image that consists of a square RGB matrix, which is given as input to a CNN. This reshaping procedure is detailed in Section 4.4.

4.3.2 Global features

After obtaining the 300 local features, the metrics described in Subsection 2.4.4 are applied in order to generate FD, LAC and PERC global features. As the number of global features is not dependent on the value of L_{max}, the same 21 features presented in Table 4 are obtained for each Δ, which results in a total of 63 global features.

These features are given as inputs to be classified by a machine learning algorithm, wherein the classification results will be part of the final ensemble. The reshaping procedure is not applied on the global features.

4.4 Reshape

In order to generate a feature image to be given as input to a CNN, a reshaping procedure is applied to the set of 300 local features. We have tested four different reshaping approaches, which are detailed in the following subsections.

4.4.1 Sequential reshape

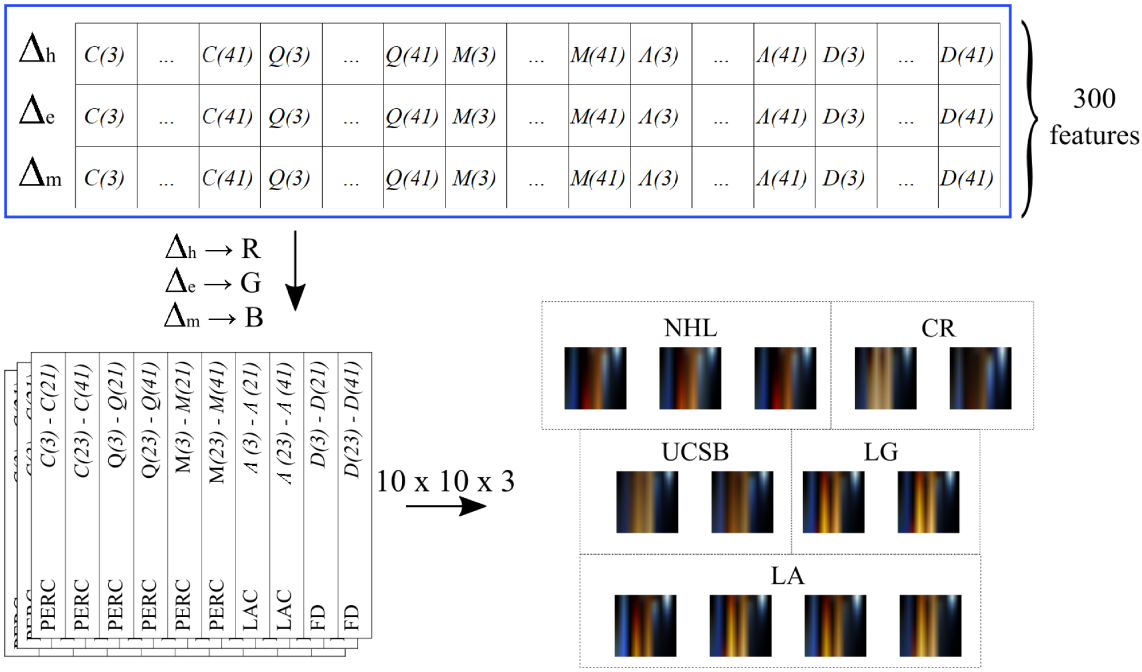
In order to be classified by a CNN, the local feature vectors generated in the previous stages must be converted into feature matrices, in a similar way to the method available in (LUMINI; NANNI, 2018). In this way, the 300 local features obtained from the calculation of each of the three Δ distances are reshaped to compose different dimensions of the matrix, in order to represent colour model channels. In this case, for this stage of the experiments, the conversion to RGB colour channels was employed. These features were sequentially arranged in a matrix $10 \times 10 \times 3$. The matrices generated by Δ_h, Δ_e and Δ_m correspond to the red, green and blue colour channels, respectively. In Figure 27, images are presented to exemplify the process of building the feature image generated with the RGB colour channels. Since each of the functions $C(L)$, $Q(L)$, $M(L)$, $\Lambda(L)$ and $D(L)$, obtained from a specific Δ, generate 20 features, each function fits exactly 2 columns of the feature image.

4.4.2 Recurrence plot

Another way of representing feature vectors as an image is through the use of recurrence plot (RP), which is a technique that has been proposed in (ECKMANN et al., 1995) for the projection of repeated events into two or three dimensional spaces. It has been used in applications such as the detection of Parkinson's disease and plant recognition (AFONSO et al., 2019; FARIA et al., 2016).

The application of this technique on a feature vector containing N features outputs a squared matrix $N \times N$ wherein each element $R_{i,j}$ of the matrix is obtained through Equation 24:

Figure 27 – Sequential reshaping applied to the feature vectors in order to generate feature images.



(Source: author)

$$R_{i,j} = \|x_i - x_j\| \quad \forall \quad i,j = 1...N. \tag{24}$$

in which x_i and x_j are the i^{th} and j^{th} features in the vector, and $\|x_i - x_j\|$ is the norm of the Euclidean distance.

In our proposed approach, the 100 features obtained from each of the distances Δ represents a feature vector wherein the RP is calculated. As the generated matrices for each Δ are two-dimensional, the combination of the three distances outputs an RGB image, as illustrated in Figure 28.

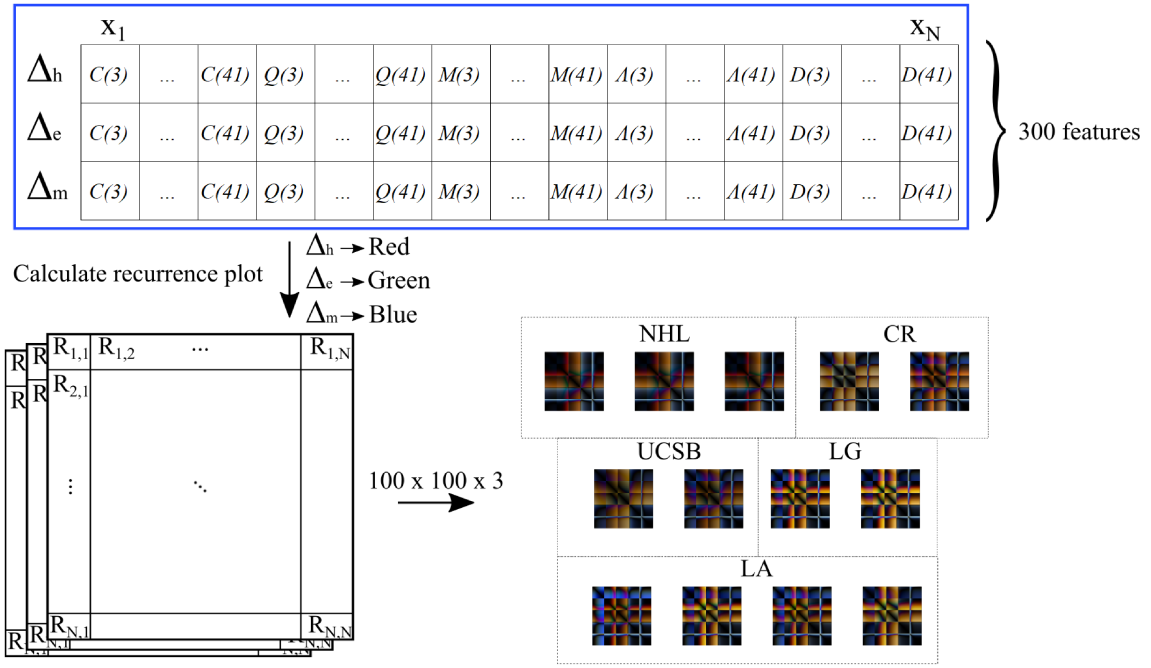
4.4.3 Gramian angular field

Feature vectors can also be converted into RGB images through the use of the GAF technique. This method was initially proposed in (WANG; OATES, 2015) and consists in representing a time series in polar coordinates and then generating a Gramian matrix through the application of trigonometric operations.

In order to represent a feature vector $X = \{x_1, x_2, ..., x_n\}$ composed by N elements in polar coordinates, it must first be re-scaled to fit in the $[-1, 1]$ interval, which can be done through Equation 25:

$$\tilde{x}_i = \frac{(x_i - max(X) + (x_i - min(X))}{max(X) - min(X)}. \tag{25}$$

Figure 28 – RP applied to the feature vectors in order to generate feature images.



(Source: author)

Then, the re-scaled feature vector $\tilde{X}$ can be represented as polar coordinates by encoding the value $\tilde{x}_i$ as the angular cosine according to Equation 26:

$$\theta_i = arccos(\tilde{x}_i), \quad \forall \quad \tilde{x}_i \in \tilde{X}. \tag{26}$$

This θ_i value is used to obtain the GAF matrices, which can be of two different types: Gramian summation angular field (GASF) and Gramian difference angular field (GADF). These matrices are of dimension $N \times N$ and each of its elements are obtained as shown in Equations 27 and 28:

$$\Phi_{i,j} = [\cos(\theta_i + \theta_j)] \quad \forall \quad i, j \in \{1, 2, ..., N\}, \tag{27}$$

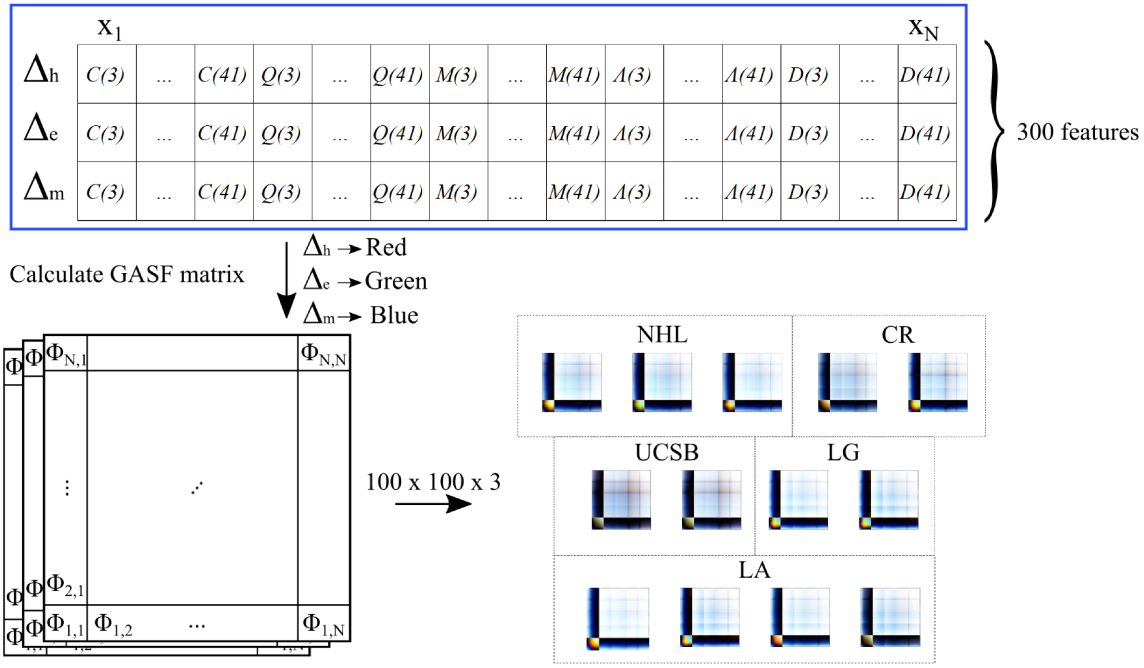
$$\Psi_{i,j} = [\sin(\theta_i - \theta_j)] \quad \forall \quad i, j \in \{1, 2, ..., N\}, \tag{28}$$

wherein $\Phi_{i,j}$ and $\Psi_{i,j}$ corresponds to the elements of the GASF and GADF matrices, respectively.

Similarly to what was done in the two previous representations, the feature vectors for each distance Δ represents a vector X, which will have its GAF matrices calculated, being later converted to an RGB image wherein each colour channel is obtained from one of the distances Δ. The summary of this procedure is illustrated in Figures 29 and 30 for the GASF and GADF images, respectively.

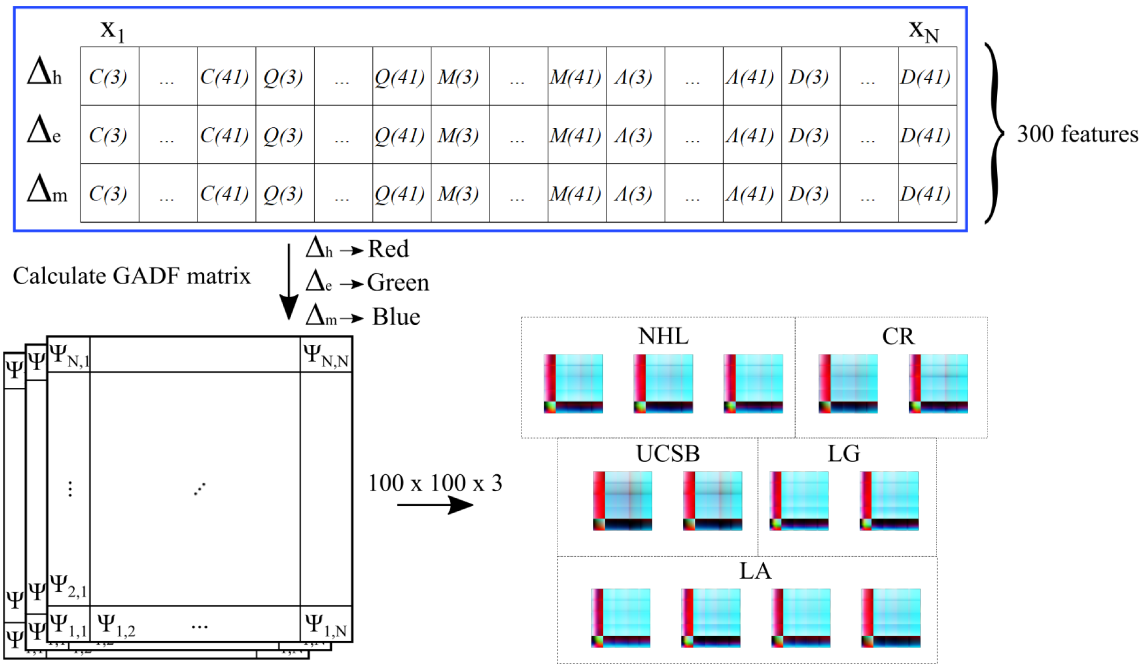
The obtained feature images were given as inputs to the F-CNN, which outputs a probability vector for the classes of the tested dataset. The pseudo-code corresponding to the feature extraction module of the proposed method is presented in Algorithm 1.

Figure 29 – GASF applied to the vectors in order to generate feature images.



(Source: author)

Figure 30 – GADF applied to the feature vectors in order to generate feature images.



(Source: author)

4.5 Classification

In the proposed model, the classification is performed in two modules: deep learning and; handcrafted classification. The deep learning module is composed by two CNN, wherein the first performs the classification of the original image (O-CNN) and the second performs the classification of the feature image (F-CNN), which is obtained after reshaping

Algorithm 1 Feature extraction module

```
input ← originalImage
for L = 3 to L = 41, increment L in 2 do
    boxes(L) ← glidingBox(input,L)
    for dist = i ∀ i ∈ {Chessboard, Euclidian, Manhattan} do
        binaryBoxes(dist) ← Δ_dist(boxes(L))
        PERCfeatures ← {PERCfeatures, obtainPERC(binaryBoxes(dist))}
        LACfeatures ← {LACfeatures, obtainLAC(binaryBoxes(dist))}
        FDfeatures ← {FDfeatures, obtainFD(binaryBoxes(dist))}
    end for
end for
```

the local features. A transfer learning approach from pre-trained models in the *ImageNet* dataset (RUSSAKOVSKY et al., 2015) was used, with the aim of improving accuracy as well as reducing training time. In the second module, the feature vectors containing the local and global handcrafted fractal features are classified by classical machine learning algorithms.

4.5.1 Deep learning module

Original images CNN

In the O-CNN classification, the original images were given as inputs and the class probabilities obtained in the *softmax* layer were considered as classification results. These results were later combined with the class probabilities obtained from the other three classifications.

Feature images CNN

The classification of the feature images obtained after reshaping the local features is performed on the F-CNN. We firstly evaluated how the use of the four reshaping approaches (sequential, RP, GASF and GADF) affects the classification outcome. The approach that provided the best classification outcome was selected to be part of the classifier ensemble.

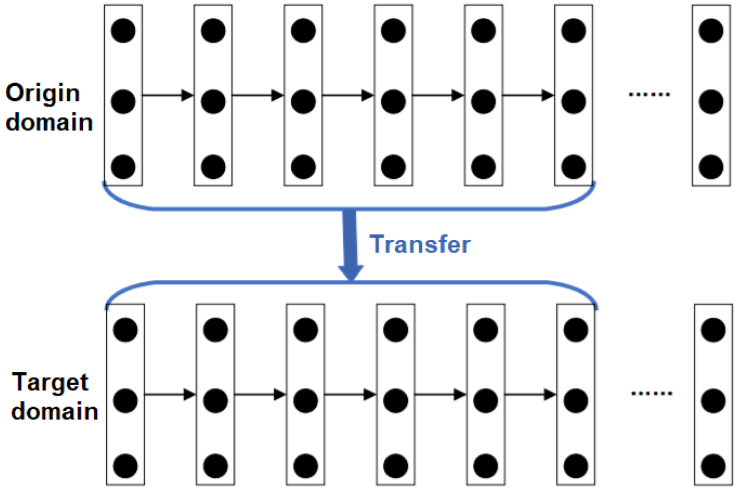
It is important to emphasize that O-CNN and F-CNN are not different networks, but a same CNN model that is employed twice in the proposed model: once to analyze the original image; and another time to analyze the feature image.

4.5.1.1 Transfer learning

In order to reduce the training time of CNN and obtain promising results with a smaller number of epochs, the transfer learning strategy was adopted to decrease the evaluation time of a model without presetting the weights. Therefore, four candidate CNN

that recently provided relevant results in histological image classification were selected: ResNet-50, ResNet-101, InceptionV3 and Xception, based on the works of (RAKHLIN et al., 2018; MAHBOD et al., 2018; JIANG et al., 2019). The used transfer learning model was the network-based transfer, in which the parameters of the first CNN layers are set from a model trained on a source domain, as illustrated in Figure 31. The source model on which the CNN have been pre-trained consists of the *ImageNet* dataset.

Figure 31 – Network-based transfer learning.



(Source: Adapted from (TAN et al., 2018))

4.5.1.2 Explainable artificial intelligence

Explainable artificial intelligence (XAI) is the concept of using certain tools to make deep learning predictions easier to understand for humans. In the context of image classification through CNN, the tool is usually applied in a specific layer or in the network as a whole and outputs a heat map indicating the image regions that had a greater impact in the model prediction.

Three of the main XAI tools are Occlusion sensitivity, Gradient weighting class activaton mapping (Grad-CAM) and Locally-interpretable model-agnostic explanation (LIME). We have chosen Grad-CAM since it was designed specifically to interpret image classification and it has performed well when applied to histology images (GULUM; TROMBLEY; KANTARDZIC, 2021). We applied Grad-CAM on the testing set in order to better visualize the classifications performed by O-CNN and F-CNN models. The default settings were used, which means that the feature maps were extracted from the output of the last ReLU layer of the model, as it is the last layer with nonsingleton spatial dimensions.

4.5.2 Handcrafted classification module

Apart from the CNN classifications of the original and feature images, the obtained feature vector containing local and global fractal features were also classified, separately, by

classical machine learning algorithms. The evaluated classifiers were: Decision tree (DT); Logistic (LOG); Random forest (RaF); Support vector machine (SVM); and Multilayer perceptron (MLP). We selected these classifiers as they are the most common representatives of two of the main supervised learning approaches: tree- and function-based (HSIEH et al., 2011; ROBERTO et al., 2019; CHEN et al., 2020). Then, the class probabilities obtained from the classifications of both feature sets are included in the classification ensemble, as illustrated in Figure 26. After applying the sum rule to all probability vectors included in the ensemble, the class which obtained the highest value is considered the predicted class.

4.6 Ensemble model

In order to investigate a relevant ensemble model that properly associates the modules previously described, some approaches were evaluated. In this stage, we firstly investigated the association composed by O-CNN and F-CNN, considering all possible feature image representations. Then, a model combining the proposed ensemble composed by two CNN classifiers (O-CNN and F-CNN) and handcrafted feature classifiers (local and global) was investigated. This association employed the LOG and MLP classifiers to fully evaluate the ensemble based in the obtained results (which will be presented at Section 5.3). Moreover, as preliminary tests have shown that O-CNN significantly outperforms the other classification approaches present in the ensemble, it has been assigned a weight of 3, which represents 50% of the ensemble sum, as described in (ONAN; KORUKOĞLU; BULUT, 2016; HSU; CHIEN, 2020). This weight value was obtained through empirical testing. For simplification, the part of the ensemble composed by the CNN classification module was defined as Φ, according to Equation 29:

$$\Phi = 3 \times p(O\text{-}CNN) + p(F\text{-}CNN). \tag{29}$$

, wherein the probability vectors obtained after each classification is denoted by $p()$. The final proposed ensemble is composed by the following elements:

- ❏ CNN for the classification of the original images, with weight 3, representing 50% of the ensemble composition ($O\text{-}CNN$);
- ❏ CNN for the classification of the feature images, generated by one of the tested reshaping procedures ($F\text{-}CNN$);
- ❏ Classification of handcrafted local fractal features (HC_{local});
- ❏ Classification of handcrafted global fractal features (HC_{global});

Therefore, using the weighted sum rule, the proposed ensemble Υ can be summarized according to Equation 30:

$$\Upsilon = 3 \times p(O\text{-}CNN) + p(F\text{-}CNN) + p(HC_{local}) + p(HC_{local}). \quad (30)$$

The pseudo-code corresponding to the classification module, as well as the ensemble composition of the proposed method is presented in Algorithm 2.

Algorithm 2 Classification module and ensemble

```
O-CNN ← classifyCNN(input)
localFeatureVector ← {PERCfeatures, LACfeatures, FDfeatures}
featureImage ← reshape(localFeatureVector)
F-CNN ← classifyCNN(featureImage)
localClassification ← classify(localFeatureVector)
globalClassification ← classify(globalFeatureVector)
result ← argmax(3×O-CNN + F-CNN + localClassification + globalClassification)
```

4.7 Performance evaluation

In order to investigate the performance of the approaches proposed in this research, details of the experiments conducted are presented.

In the first stage, an investigation of the most relevant network model for histological image classification was performed. Each of the CNN models presented in Section 2.6.1, apart from the LeNet, was applied to the five tested datasets. In the second stage, in order to ensure that there is not a negative transfer problem, which occurs when the use of transfer learning impairs the performance of the model (TORREY; SHAVLIK, 2010), tests were performed without the use of transfer learning and the results obtained were compared. In the third testing stage, the performance of F-CNN and handcrafted classifications were evaluated separately. Then, in the fourth stage, different ensemble compositions were evaluated, containing the results previously obtained by O-CNN, F-CNN and handcrafted classifications. Then, the proposed method was compared with other techniques such as geometric data augmentation, ADASYN and other handcrafted feature-based approaches. To obtain features from these other techniques, the respective methods were implemented in *software* Matlab R2019b and applied on the same image bases. For classification of the feature vectors, the classifiers available in the *Weka 3.6.13* platform were used with their default parameters.

All tests were run on a *Intel Xeon Silver 4116 CPU* of $2.10GHz$, 128GB RAM and a *NVIDIA GeForce RTX 2080Ti* card using *software MATLAB R2019b*. As some of the tested bases have a small number of samples, the cross-validation technique with 10-*folds* was adopted in all test steps to avoid problems such as *overfitting*. The training/test split for each fold is 90%/10%. Moreover, all the tests were performed 10 times, where

each time the set of 10 folds was generated differently. For training the CNN models, the default values of hyper-parameters available from the *Deep Learning Toolbox* of the *software MATLAB R2019b* were used, which are listed below:

1. Solver: sgdm;
2. Initial learning rate: 0.01;
3. Learning rate drop period: 2 epochs;
4. Learning rate drop factor: 0.75;
5. Mini-batch size: 32;
6. L2 regularization: 0.0001;
7. Epochs: 10.

4.7.1 Evaluation metrics

The evaluation of results obtained from the application of a new proposal is an important step in the process of analysing its feasibility. For that purpose, it is necessary to define the metrics to be evaluated and the statistical analysis procedure of the obtained values. In this section, two of the most commonly used performance metrics in image classification problems will be described, as well as a non-parametric statistical test to evaluate the results.

4.7.1.1 Accuracy

Accuracy is among the most common metrics for the analysis of results in the context of image classification because the calculation is simple and the interpretation of the results is easy. After data classification, a structure called confusion matrix can be generated. This matrix is able to provide the number of True Positives (TP), positive values that the system correctly classified as positive; True Negatives (TN), negative values that the system correctly classified as negative; False Positives (FP), negative values that the system incorrectly classified as positive; and False Negatives (FN), positive values that the system incorrectly classified as negative. The accuracy A_c is given based on these values, as shown in Equation 31.

$$A_c = \frac{TP + TN}{TP + TN + FP + FN} \tag{31}$$

Accuracy corresponds to a value between 0 and 1. If the classifier has correctly labelled all samples, the accuracy will have a value equal to 1 (or 100%). However, in problems where classes are unbalanced, accuracy may not be a reliable measure. An accuracy of

90% for a given class A, where this class contains 95% of the total number of samples in the base, is not a good result. Therefore, the use of complementary performance measures is necessary to evaluate methods.

4.7.1.2 F-score

The F-score is a metric that indicates the general quality of a method, independent of the number of classes in the input sets or its unbalance. This metric consists of the harmonic mean between two other metrics: precision ($Prec$) and recall (Rec). The $Prec$ indicates the ratio between samples classified as a given class and the total samples of the same class, given by Equation 32.

$$Prec = \frac{TP}{TP + FP} \tag{32}$$

The calculation of Rec is done in a similar way, but number of samples classified as A is used instead, as indicated in Equation 33.

$$Rec = \frac{TP}{TP + FN} \tag{33}$$

Thus, the F-score, on several occasions also called $F1$, is given by Equation 34.

$$F1 = \frac{2 \times Prec \times Rec}{Prec + Rec} \tag{34}$$

4.7.1.3 Statistical significance test

The performance measures A_c and $F1$ are adequate measures to evaluate the performance of a given computational approach. However, for a comparison across different approaches, it is necessary to apply a statistical test in order to verify whether the differences in the results obtained by each approach are relevant or not (JAPKOWICZ; SHAH, 2011).

There are several tests of statistical significance, and the choice of the optimal test depends on the type of comparison that is intended. In a context where different approaches are evaluated on different input sets independent of each other, Friedman's non-parametric test is the most appropriate (JAPKOWICZ; SHAH, 2011). The aim of this test is to check whether there is a statistically relevant difference between the compared algorithms given an α value. If this difference is verified, a *Post-hoc* analysis is performed with the objective of finding in which pairs of compared methods such difference occurs. For the statistical analysis, the software *StatsDirect 3* was used.

4.8 Considerations

On this chapter, the proposed method was presented in details. The method consists in an ensemble among the CNN classification of the original image, the CNN classification of a feature image, which is generated from the set of local fractal features, and the classification of local and global handcrafted fractal features. The chosen datasets for evaluating the proposed method were CR, LA, LG, NHL and UCSB, which are representatives of four different types of histology images: colorectal, breast, lymphomas and liver tissue. The method was evaluated using the A_c and $F1$ metrics and the results are presented in the following chapter.

Chapter 5

Results and Discussion

In this section, the results obtained from the experiments described in Section 4.7 are presented. The chapter is divided in three sections. First, the deep learning classification module is evaluated. Then, the tests regarding the classification of local and global fractal features are performed for evaluating the handcrafted classification module. At last, the two modules are combined and the proposed ensemble is evaluated. In some cases, the standard deviation values were omitted for better formatting. These values can be consulted in Appendix A.

5.1 Deep learning module

5.1.1 O-CNN

The first set of experiments aimed to identify the most suitable CNN model for classifying the original input images. The five datasets were given as input to the CNN models described in Section 2.6.1, except for the LeNet model which only supports grayscale images. These models were chosen not only due to their good performance in the classification of histology datasets (RAKHLIN et al., 2018; MAHBOD et al., 2018; JIANG et al., 2019), but also as they have a significantly lower computational cost than the recent deeper models. The parameters described in Section 4.7 were used in the experiments presented here. The classification results are shown in Table 10. The evaluated metrics were CNN loss for the testing set and training time in seconds.

To analyse the statistical relevance between the investigated models, the Friedman non-parametric test was employed in this step to verify the difference between the loss values (JAPKOWICZ; SHAH, 2011). With $\alpha = 0.05$, the value $P_k = 0.4144$ was obtained, which indicates the absence of a significant difference in the values compared between the four tested CNN models once $P_k > \alpha$. Then, the training time was analysed and the Friedman test indicated a significant difference ($P_k < 0.0001$) in all pairwise comparisons (Conover) involving the ResNet-50 model. Therefore, this model was chosen to be utilized

Table 10 – Loss and training time (s) for the classification of the five datasets using the original images in different CNN models (O-CNN).

Dataset		ResNet-50	ResNet-101	InceptionV3	Xception
CR	Loss	**0.018**	0.022	0.045	0.043
	Time	**50.21**	91.26	93.82	83.37
LA	Loss	0.048	0.128	**0.031**	0.051
	Time	**122.47**	247.67	259.65	247.22
LG	Loss	**0.005**	0.017	**0.005**	0.041
	Time	**67.25**	133.84	139.99	128.66
NHL	Loss	0.767	0.585	0.497	**0.409**
	Time	**42.69**	85.79	87.58	77.01
UCSB	Loss	0.318	**0.271**	0.305	0.606
	Time	**21.99**	39.65	39.01	29.76

in the deep learning module of the proposed method. Moreover, this model was adopted due to its reduced training time and the relevant results recently obtained in the classification of histological images (RAKHLIN et al., 2018; MAHBOD et al., 2018; WANG et al., 2019; WANG et al., 2020).

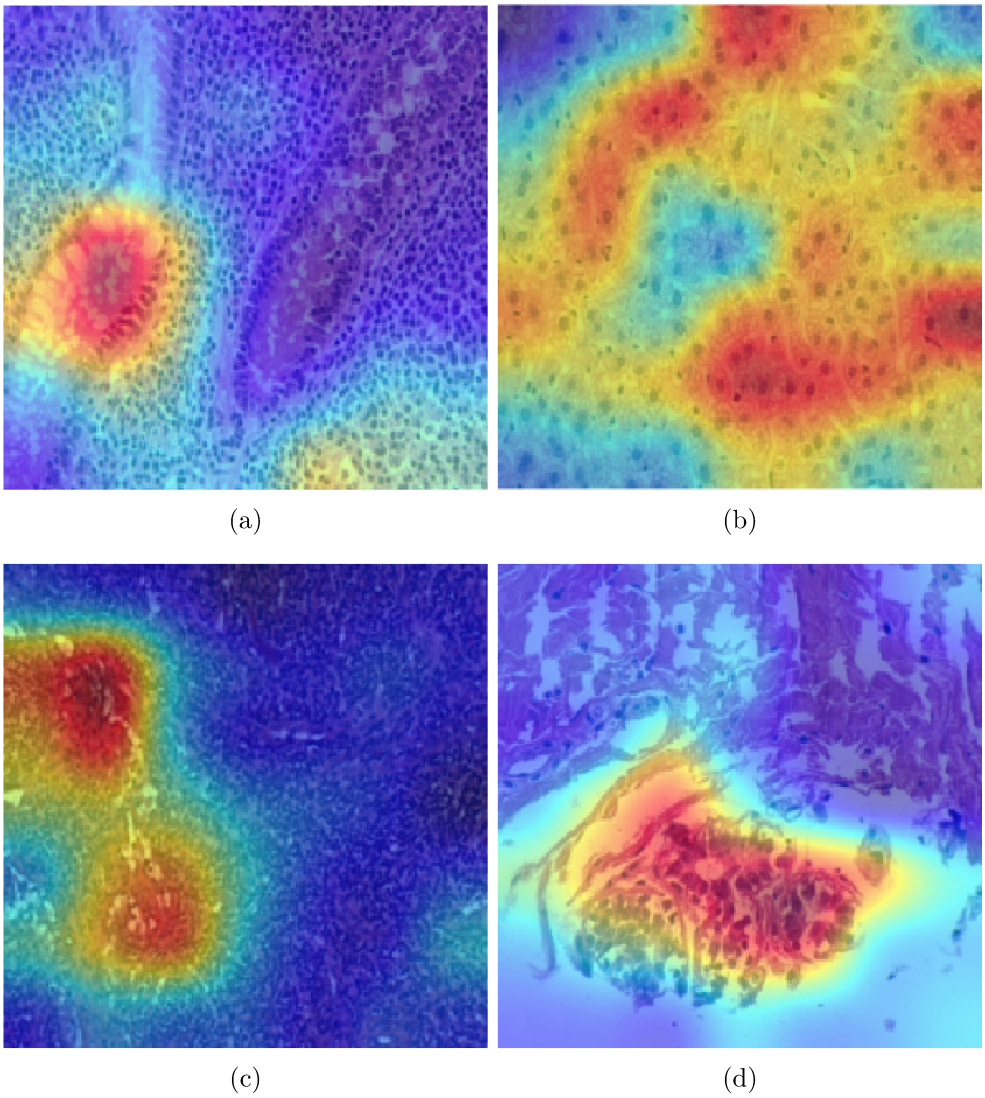
In order to better visualize how the ResNet-50 model is interpreting the images, we can use the Grad-CAM tool to obtain heat maps that illustrate which regions of the images were more relevant for the network's prediction, as described in Section 4.5.1.2. An example of each type of tissue of the obtained class activation maps are shown in Figure 32.

For the colorectal tissue example, shown in Figure 32 (a), the network seems to focus on regions wherein structures called goblet cells are present (LEEUWEN et al., 2007). These regions are often less homogeneous than the stroma and, therefore, can be more relevant for the model to identify them. The opposite can be stated about the breast tissue example shown in Figure 32 (d), wherein the focus is on the nuclei cells. On the other hand, the liver and NHL images have a more regular texture, which makes it less clear to visualize which patterns are being detected by the CNN. For the liver example shown in Figure 32 (b), the model's focus is spread throughout different areas, while in the NHL example in Figure 32 (c), the area with less nuclei cells appear to be more relevant. It is noteworthy that these generated heat maps are highly dependent on the dataset and the trained model.

Finally, the occurrence of negative transfer was evaluated. In order to do so, the five datasets were classified using a pre-trained version of the ResNet-50 model, as described in Section 4.5.1.1, and an untrained version, without pre-loaded weights. The results are shown in Table 11.

These results show that the pre-trained model has a better performance in all datasets by a significant margin, which indicates the absence of negative transfer. We also considered increasing the number of training epochs from 10 to 100 and using geometric data

Figure 32 – Class activation maps of colorectal (a), liver (b), non-Hodgkin lymphoma (c) and breast (d) tissue, after being classified on a trained ResNet-50 model.



(a) (b)

(c) (d)

(Source: author)

Table 11 – Negative transfer evaluation in the O-CNN based on ResNet-50.

Dataset	Pre-trained (transfer learning)		Training from scratch	
	A_c	$F1$	A_c	$F1$
CR	**98.61%**	0.986 ± 0.009	64.85%	0.643 ± 0.037
LA	**99.03%**	0.991 ± 0.005	76.80%	0.760 ± 0.021
LG	**99.36%**	0.993 ± 0.004	95.58%	0.955 ± 0.014
NHL	**91.55%**	0.914 ± 0.010	66.90%	0.649 ± 0.020
UCSB	**80.34%**	0.797 ± 0.045	58.28%	0.564 ± 0.046
Average	**93.78%**	0.936 ± 0.015	72.48%	0.714 ± 0.028

augmentation on the scenario without transfer learning. The obtained accuracy rates were 75.27%, 84.15%, 97.40%, 70.56% and 69.83% for CR, LA, LG, NHL and UCSB datasets, respectively. These values are also significantly lower than the ones obtained using trans-

fer learning without data augmentation and only 10 training epochs. Therefore, transfer learning was applied to all of the procedures involving CNN classification.

5.1.2 F-CNN

The following experiment was aimed to investigate the most appropriate reshaping procedure that generates a feature image to be given as input to a CNN (F-CNN). As detailed in Section 4.4, the four evaluated procedures are: sequential, RP, GASF, and GADF. The generate feature images are given as inputs to the ResNet-50 CNN model. The accuracy and F1 values obtained from the five datasets are shown in Table 12.

Table 12 – Results obtained from the classification of feature images using the ResNet-50 CNN model.

Dataset	Sequential		Rec. Plot		GASF		GADF	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	82.73%	0.825	84.79%	0.845	**85.21%**	**0.849**	84.97%	0.847
LA	**95.40%**	**0.954**	94.81%	0.948	93.52%	0.936	93.66%	0.937
LG	94.91%	0.948	**95.58%**	**0.955**	94.19%	0.941	94.34%	0.943
NHL	69.71%	0.693	72.43%	0.719	72.25%	0.717	**73.45%**	**0.729**
UCSB	72.76%	0.724	72.76%	0.724	71.03%	0.708	**72.93%**	**0.726**
Average	83.10%	0.829	84.08%	0.838	83.24%	0.830	83.87%	0.836

The sequential reshape provided the best results for the LA dataset, with an accuracy of 95.40%. For the LG dataset, the best result was obtained using the RP reshape, which provided an accuracy of 95.58% and an F-score of 0.955. The CR dataset had the best performance using the GASF reshape, with an accuracy of 85.21%. The GADF reshape was the only one that provided the best results for two datasets: NHL and UCSB, with F1 values of 0.729 and 0.726, respectively. From an overall perspective, there is not a specific reshaping procedure that clearly outperforms the other three, in regard to the classification of the feature images. Hence, the Friedman test was applied for a more detailed analysis and the results are shown in Table 13.

Table 13 – p-values obtained for all pairwise comparisons obtained from the classifications of feature images using different reshaping procedures.

p-values	GADF	Rec. Plot	Sequential	GASF	Avg. Ranking
GADF	-	0.8142	0.4850	0.1751	2.00
Rec. Plot	0.8142	-	0.6396	0.2529	2.20
Sequential	0.4850	0.6396	-	0.4850	2.60
GASF	0.1751	0.2529	0.4850	-	3.20

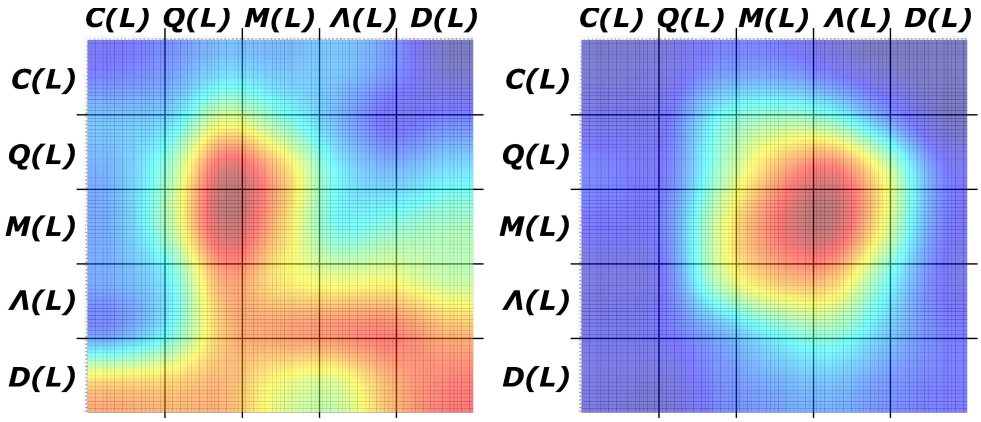
These p-values indicate that none of the reshaping procedures are clearly more suitable than the others, since all obtained p-values were above the significance threshold ($\alpha = 0.05$). According to the average ranking, the GADF reshape performed slightly better,

which can be observed by the results presented in Table 12 specially in the NHL and UCSB datasets. These datasets were also the ones whose classification has shown to be more challenging, considering both original and feature images.

Since the feature images are generated based on well-defined rules, it is possible to know what features each pixel is composed of, according to the type of reshaping procedure applied. Therefore, Grad-CAM can be applied in order to identify which regions of the image contributed more to the CNN classification and, consequently, verify which features these regions are composed of. For each class, the output maps generated for each image were combined and a map containing the average values was generated. However, since misclassifications could affect the analysis, we only considered activation maps from images that obtained a score greater than 0.95 in the softmax layer. This provides an overview of what regions, in general, are more relevant for a correct classification.

For the CR dataset, whose Grad-CAM images are shown in Figure 33, the network does not focus on the regions related to the $C(L)$ function, which represents the number of percolating clusters per box, in neither classes, except when combined with the FD features, represented by $D(L)$, in the benign class. For the malignant class, the focus is primarily on the central area of the feature image, which corresponds to the $Q(L)$ $M(L)$ and $\Lambda(L)$ functions, which represent occurrence of percolation, size of the largest cluster and LAC, respectively.

Figure 33 – Class activation maps of the benign (a) and malignant (b) classes from the CR dataset classified in the F-CNN using the GADF reshape.



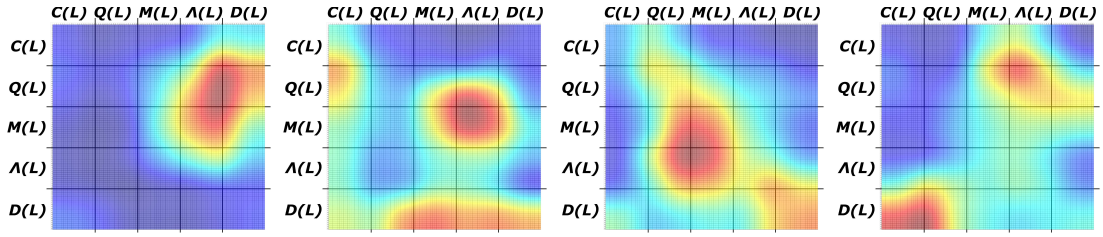
(a) (b)

(Source: author)

The LA dataset is split into four classes, whose Grad-CAM images are shown in Figure 34. For the first three (one, six and 16 months), the focus is mainly on the combinations of the $Q(L)$, $M(L)$ and $\Lambda(L)$ functions. The FD features are more relevant for the 24 months class, when combined with the $C(L)$ and $Q(L)$ function.

The class activation maps obtained from the LG dataset are similar to the ones from the LA dataset, as these are composed by the same type of images. The average class

Figure 34 – Class activation maps of liver tissue from female mice aged one (a), six (b), 16 (c) and 24 (d) months, from the LA dataset classified in the F-CNN using the GADF reshape.

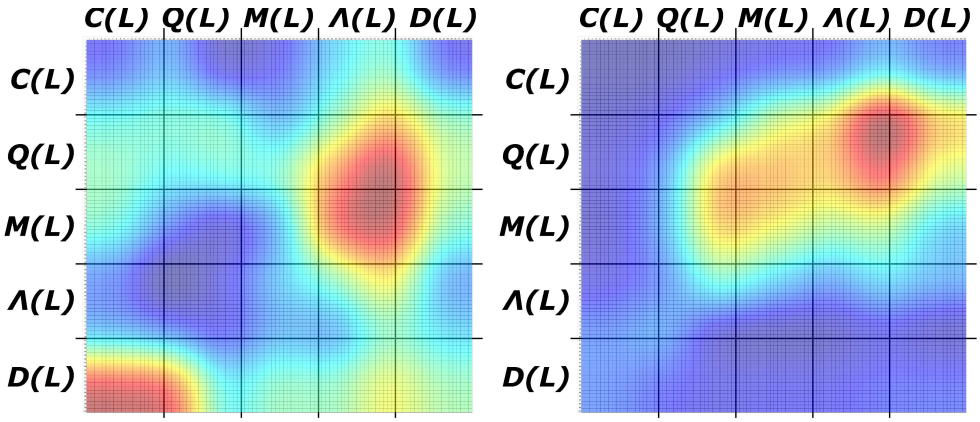


(a) (b) (c) (d)

(Source: author)

activation maps are shown in Figure 35. Therefore, the areas represented by the combinations of the $Q(L)$, $M(L)$ and $\Lambda(L)$ functions were once again the network's focus for classifying both male and female classes. The FD features were also important for the classification of the male class, specially when combined with the $C(L)$ function.

Figure 35 – Class activation maps of liver tissue from male (a) and female (b) mice from the LG dataset classified in the F-CNN using the GADF reshape.



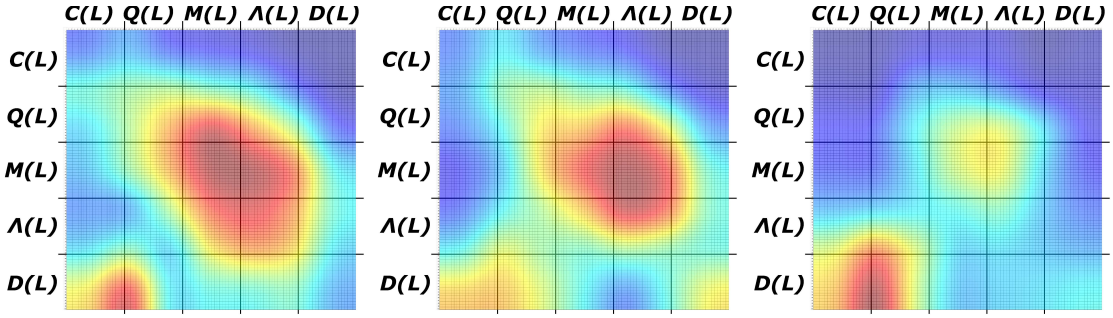
(a) (b)

(Source: author)

For the NHL dataset, the combination pairs $M(L)$, $\Lambda(L)$ and $C(L)$, $D(L)$ were the focus of the network for all three classes, as seen in Figure 36. For the MCL class though, the region represented by the combination of the FD features with the $C(L)$ and $Q(L)$ functions seems to be more relevant.

For the UCSB dataset shown in Figure 37, the activation areas are more spread throughout all different regions of the feature images representing the benign class, with focus on $D(L)$ function combined with $Q(L)$ and $M(L)$ functions. We believe that this is related to the lower accuracy obtained for this dataset in relation to the others. On the other hand, there is only one region in the average image representing the malignant class that is composed mainly by the $M(L)$ and $\Lambda(L)$ functions.

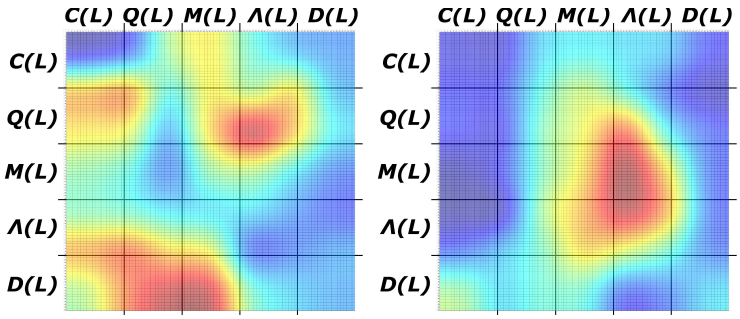
Figure 36 – Class activation maps of a CLL (a), FL (b) and MCL (c) case from the NHL dataset classified in the F-CNN using the GADF reshape.



(a) (b) (c)

(Source: author)

Figure 37 – Class activation maps of a benign (a) and a malignant (b) case from the UCSB dataset classified in the F-CNN using the GADF reshape.



(a) (b)

(Source: author)

Overall, these class activation maps show that all five functions were relevant for the F-CNN classification in at least one of the tested datasets. PERC features were mainly represented by the $Q(L)$ and $M(L)$ functions, although the areas represented by $C(L)$ function were also focused in some cases, specially when combined with the $D(L)$ function, which is composed by the FD features. Finally, LAC features, represented by the $\Lambda(L)$ function, seem to have worked better when combined with the $Q(L)$ and $M(L)$ functions.

5.2 Handcrafted classification module

In this section, the results obtained from the classification of the feature vectors composed by the handcrafted local and global fractal features are presented. Firstly, the results obtained from classifying the local fractal features are shown in Table 14. The best results for each dataset are highlighted in bold.

The LOG classifier provided the best results for the CR, NHL and UCSB datasets with accuracy values of 0.948, 0.818 and 0.802 respectively. The best results for the LA

Table 14 – Results obtained from the classification of handcrafted local fractal features.

Dataset	DT		LOG		RaF		SVM		MLP	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	70.67%	0.703	**94.85%**	**0.948**	81.39%	0.809	82.06%	0.811	92.67%	0.926
LA	91.50%	0.917	96.86%	0.969	94.07%	0.942	94.68%	0.947	**97.78%**	**0.979**
LG	91.81%	0.916	98.23%	0.982	95.43%	0.954	96.08%	0.960	**99.85%**	**0.998**
NHL	61.95%	0.613	**81.82%**	**0.816**	68.45%	0.677	65.21%	0.642	71.55%	0.712
UCSB	73.10%	0.730	**80.17%**	**0.800**	68.10%	0.678	65.17%	0.635	77.41%	0.771
Average	77.81%	0.776	90.38%	0.903	81.48%	0.812	80.64%	0.799	87.85%	0.877

and LG datasets were provided by the MLP classifier. In order to analyse the overall performance of these classifiers on the five datasets, the Friedman test was also applied. The average ranking considering the F1 metric and the p-values are shown in Table 15. Comparisons with a significant statistical difference ($\alpha = 0.05$) are highlighted in bold.

Table 15 – p-values obtained for all pairwise comparisons obtained from the classifications of handcrafted local fractal features.

p-values	LOG	MLP	SVM	RaF	DT	Avg. Ranking
LOG	-	0.4313	**<0.0001**	**<0.0001**	**<0.0001**	1.33
MLP	0.4313	-	**0.0003**	**<0.0001**	**<0.0001**	1.67
SVM	**<0.0001**	**0.0003**	-	0.4313	**0.0108**	3.50
RaF	**<0.0001**	**<0.0001**	0.4313	-	0.0583	3.83
DT	**<0.0001**	**<0.0001**	**0.0108**	0.0583	-	4.67

These p-values indicate that the MLP and LOG classifiers outperform the other three by a significant margin, with p-values smaller than 0.001 in all comparisons. When the classifiers MLP and LOG are compared, the latter seems to have a slightly better performance, however it is not statistically significant. The LOG classifier obtained the best average ranking (1.33). These same evaluations were performed using global fractal features. The classification results for each of the five datasets are shown in Table 16 and the statistical evaluation using the Friedman test is shown in Table 17.

Table 16 – Results obtained from the classification of handcrafted global fractal features.

Dataset	DT		LOG		RaF		SVM		MLP	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	72.67%	0.723	**90.42%**	**0.902**	85.15%	0.849	87.45%	0.871	90.30%	**0.902**
LA	86.08%	0.860	94.49%	0.944	93.60%	0.937	93.30%	0.932	**97.25%**	**0.973**
LG	89.32%	0.892	97.58%	0.975	93.96%	0.939	94.26%	0.942	**98.75%**	**0.987**
NHL	62.17%	0.616	71.44%	0.710	72.54%	0.718	67.73%	0.669	**72.67%**	**0.721**
UCSB	72.76%	0.726	**77.93%**	**0.777**	72.76%	0.722	70.52%	0.700	75.86%	0.754
Average	76.60%	0.763	86.37%	0.862	83.60%	0.833	82.65%	0.823	86.97%	0.867

Similarly to what was observed in the local features classification, LOG and MLP classifiers have provided the best accuracy and F-score values when classifying global fractal features. The LOG classifier provided accuracy values of 90.42% and 77.93% for the CR and UCSB datasets, respectively. The LA, LG and NHL datasets had their best results provided by the MLP classifier.

Table 17 – p-values obtained for all pairwise comparisons obtained from the classifications of handcrafted global fractal features.

p-values	MLP	LOG	RaF	SVM	DT	Avg. Ranking
MLP	-	0.3185	**0.0003**	**0.0001**	**<0.0001**	1.33
LOG	0.3185	-	**0.0028**	**0.0013**	**<0.0001**	1.83
RaF	**0.0003**	**0.0028**	-	0.7367	**0.0270**	3.50
SVM	**0.0001**	**0.0013**	0.7367	-	0.0541	3.67
DT	**<0.0001**	**<0.0001**	**0.0270**	0.0541	-	4.67

The Friedman test has also indicated that the statistical differences of the classifiers MLP and LOG is significant when compared to RaF, SVM and DT. However, there is not a significant difference between LOG and MLP classifiers regarding the classification of global fractal features, although MLP has achieved a slightly better average ranking.

It is noteworthy that the local features provided better classification results overall when compared to the global features, which indicates a higher discriminating power. Despite not being the main reason they were selected for generating the feature images, this shows that the local features are more suitable for this task.

5.3 Ensemble model evaluation

In this section, the evaluation of different combinations of reshaping procedures and classifiers for the handcrafted features are presented. The first tests consisted in evaluating the *O-CNN + F-CNN* ensemble. Since the statistical analysis presented in Table 13 has shown that all representations of a feature image provided similar results, the four reshaping methods were reevaluated in an ensemble. The classification results are presented in Table 18 and the statistical analysis using the Friedman test is shown in Table 19.

Table 18 – Evaluation of different reshaping procedures for the O-CNN + F-CNN ensemble model.

Dataset	Sequential		Rec. Plot		GASF		GADF	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	98.48%	**0.985**	**98.55%**	**0.985**	98.18%	0.982	98.42%	0.984
LA	99.07%	0.991	99.09%	0.991	99.15%	**0.992**	**99.17%**	**0.992**
LG	**99.47%**	**0.995**	99.17%	0.992	98.91%	0.989	99.13%	0.991
NHL	89.41%	0.892	**89.57%**	**0.894**	88.85%	0.887	89.06%	0.889
UCSB	83.62%	0.832	82.59%	0.822	83.97%	0.835	**85.34%**	**0.850**
Average	94.01%	0.939	93.79%	0.937	93.81%	0.937	94.22%	0.941

Once again, none of the four reshaping procedures evaluated clearly outperformed the other three. In the classification results shown in Table 18, the sequential reshape provided the best accuracy for the LG dataset. CR and NHL datasets had their best results

Table 19 – p-values obtained for all pairwise comparisons obtained from the classifications of the O-CNN + F-CNN ensemble using different reshaping procedures.

p-values	GADF	Rec. Plot	Sequential	GASF	Avg. Ranking
GADF	-	>0.9999	0.8176	0.2614	2.20
Rec. Plot	>0.9999	-	0.8176	0.2614	2.20
Sequential	0.8176	0.8176	-	0.3644	2.40
GASF	0.2614	0.2614	0.3644	-	3.20

with the RP reshape while the GADF reshape provided the best results for the LA and UCSB datasets. The Friedman test presented in Table 19 has also indicated the absence of a significant statistical difference among these four procedures, with GADF and RP obtaining the best average rankings (2.20). Since the GADF reshaping has also obtained the highest average ranking in the individual evaluation (see Table 12), this procedure has been chosen for the following evaluations. However, it must be noted that the four reshaping procedures were able to provide good results in the overall performance in the O-CNN + F-CNN ensemble, considering the five datasets. This ensemble model significantly outperforms the results obtained with the use of the the handcrafted descriptors only. However, it provided only a marginal improvement when compared to the results obtained with a single ResNet-50: the average accuracy increased from 93.78% to 94.22% using the GADF representation. Therefore, the classifications with both local and global handcrafted descriptors were included in the model.

Finally, the proposed ensemble composed by two CNN classifiers (O-CNN and F-CNN) and handcrafted feature classifiers (local and global) was evaluated. For the CNN classifiers, the ResNet-50 model and the GADF reshaping procedure were used, as indicated on the previous tests. Since the handcrafted classification tests presented in Section 5.2 have shown that LOG and MLP classifiers outperformed the other three in both local and global features classification, these two algorithms were evaluated in the full ensemble. The classification results for all possible variations between the LOG and MLP classifiers for local and global feature classification are shown in Table 20. For the definition of Φ see Equation 29.

Table 20 – Evaluation of the proposed ensemble with different classifier combinations for the handcrafted features.

Dataset	$\Phi + LOG(local) + ... + LOG(global)$		$\Phi + LOG(local) + ... + MLP(global)$		$\Phi + MLP(local) + ... + LOG(global)$		$\Phi + MLP(local) + ... + MLP(global)$	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	99.27%	0.993	99.27%	0.993	**99.45%**	**0.995**	99.27%	0.993
LA	**99.60%**	**0.996**	99.55%	**0.996**	99.45%	0.995	99.45%	0.995
LG	99.58%	0.996	99.58%	0.996	**99.77%**	**0.998**	99.70%	0.997
NHL	92.81%	0.927	**93.16%**	**0.931**	92.41%	0.923	92.33%	0.922
UCSB	**88.45%**	**0.882**	87.41%	0.871	**88.45%**	**0.882**	86.21%	0.856
Average	95.94%	0.959	95.79%	0.957	95.91%	0.959	95.39%	0.953

Apart from the ensemble wherein the MLP classifier was used for the classification of

both local and global features, the other three combinations seemed to have presented similar results overall. The Friedman test was applied for a more complete analysis. The results are shown in Table 21.

Table 21 – p-values for all pairwise comparisons obtained from the classifications of the proposed ensemble including Φ and using different classifiers for the handcrafted features.

p-values	MLP+LOG	LOG+LOG	LOG+MLP	MLP+MLP	Avg. Ranking
MLP+LOG	-	0.6027	0.4383	0.0859	1.90
LOG+LOG	0.6027	-	0.7938	0.2062	2.30
LOG+MLP	0.4383	0.7938	-	0.3061	2.50
MLP+MLP	0.0859	0.2062	0.3061	-	3.30

Besides the lack of a statistically significant difference among the evaluated combinations, the ensemble composed by $\Phi + MLP(local) + LOG(global)$ has obtained the highest average ranking. Moreover, the classification results obtained for all datasets are relevant, with the smallest accuracy being of 88.45% for the UCSB dataset. Hence, despite most of the evaluated combinations presenting results similarly relevant, we chose the combinations the provided the highest average ranking.

The results obtained with the proposed ensemble were compared with some other approaches commonly used in the literature for the classification of histological images. For a fair comparison, the approaches were applied in the five datasets using exactly the same folds configuration that were used for all the previously presented experiments. The compared approaches were: ResNet-50; ResNet-50 with ADASYN, a technique that generates synthetic data to fix imbalanced datasets (HE et al., 2008); and ResNet-50 with data augmentation through geometric transformations. The comparison is shown in Table 22.

Table 22 – Comparison of the proposed ensemble with other approaches commonly used in the literature.

Dataset	Proposed ensemble		ResNet-50		ADASYN		Data augmentation	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	**99.45%**	**0.995**	98.61%	0,986	99.41%	0,994	99.03%	0,990
LA	**99.45%**	**0.995**	99.03%	0,991	99.28%	0,993	98.05%	0,981
LG	**99.77%**	**0.998**	99.36%	0,993	99.17%	0,992	99.25%	0,992
NHL	92.41%	0.923	91.55%	0,914	92.11%	0,920	**93.58%**	**0,935**
UCSB	**88.45%**	**0.882**	80.34%	0,797	80.34%	0,798	86.55%	0,862
Average	95.91%	0.959	93.78%	0.936	94.06%	0.939	95.29%	0.952

The proposed ensemble was able to provide the best results for the CR, LA, LG and UCSB datasets when compared to the ResNet-50, ADASYN and data augmentation approaches. However, the approach using data augmentation provided the best results for the NHL dataset. These results show that the proposed method is able to achieve

promising performance in histological image classification. With the exception of the UCSB dataset, the accuracy values were higher than 90%, which shows an approach with good adaptive characteristics to different histological tissue categories. Despite the evaluations in unbalanced samples in all the tested bases, the proposed method was also able to provide *F-score* values above 0.880 in all cases. It is worth highlighting the relevant results obtained in the CR, LG and LA datasets with performance values close to 100% and 1.0 with these metrics. However, using only ResNet-50 also provided accuracy rates close to 99% in these datasets. Therefore, it can be said that the contribution provided by the proposed method was higher in the NHL and UCSB datasets. A statistical analysis using the Friedman test was also performed. According to the results shown in Table 23, the proposed ensemble not only has the best average ranking, but it also provides results statistically more significant than the compared approaches, since the p-values were all smaller than α.

Table 23 – p-values for all pairwise comparisons obtained from the classifications of the five datasets using different approaches.

p-values	Proposed	Data Aug.	ADASYN	ResNet-50	Avg. Ranking
Proposed	-	**0.0470**	**0.0264**	**0.0046**	1.20
Data Aug.	**0.0470**	-	0.7573	0.2299	2.60
ADASYN	**0.0264**	0.7573	-	0.3615	2.80
ResNet-50	**0.0046**	0.2299	0.3615	-	3.40

Despite the proposed ensemble having obtained relevant results even in small datasets, as is the case for UCSB, data augmentation through geometric transformations could be applied for a slight enhancement in the accuracy. Table 24 shows the accuracy and F-score values when data augmentation is applied to images trained in the O-CNN. The proposed ensemble described in Equation 30 was used. It can be noted that, despite there being a slight improvement in the accuracy obtained for the CR, NHL and UCSB datasets, this improvement falls within the error margin. Therefore, it is unlikely that the use of data augmentation through geometric transformations significantly improves the performance of the proposed ensemble.

To show the contribution of the feature images classification (F-CNN) to the ensemble, we have performed tests without including F-CNN. The evaluated ensemble consists of the O-CNN and the local and global handcrafted features classification with either LOG or MLP classifiers. We also evaluated the proposed ensemble without assigning any weights. The results are shown in Table 25, which indicate that the proposed ensemble in the configuration presented in Equation 30 outperforms the other scenarios in all of the five datasets. This highlights the contribution of the feature images in the F-CNN and how assigning a greater weight to the O-CNN improved the classification accuracy.

Finally, an overview of the results obtained with the proposed ensemble, in relation to other approaches presented in the literature for the context of histology image classifica-

Table 24 – Classification results from the proposed ensemble with data augmentation applied to the O-CNN.

Dataset	Proposed ensemble with data augmentation		Proposed ensemble without data augmentation	
	A_c	$F1$		
CR	$99.70\% \pm 0.43$	0.997 ± 0.004	$99.45\% \pm 0.67$	0.995 ± 0.007
LA	$99.15\% \pm 0.29$	0.992 ± 0.003	$99.45\% \pm 0.36$	0.995 ± 0.004
LG	$99.62\% \pm 0.36$	0.996 ± 0.004	$99.77\% \pm 0.36$	0.998 ± 0.004
NHL	$93.58\% \pm 0.56$	0.935 ± 0.006	$92.41\% \pm 1.05$	0.923 ± 0.011
UCSB	$90.69\% \pm 2.72$	0.905 ± 0.028	$88.45\% \pm 2.16$	0.882 ± 0.022
Average	$96.55\% \pm 0.87$	0.965 ± 0.009	$95.91\% \pm 0.92$	0.959 ± 0.010

Table 25 – Evaluation of the F-CNN contribution to the ensemble and the use of weights on the O-CNN.

Dataset	Proposed ensemble		No F-CNN (LOG)		No F-CNN (MLP)		No weights	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	**99.45%**	**0.995**	98.48%	0,985	99.27%	0,993	96.85%	0,968
LA	**99.45%**	**0.995**	99.38%	0,994	99.36%	0,994	99.09%	0,991
LG	**99.77%**	**0.998**	99.58%	0,996	**99.77%**	**0,998**	99.70%	0,997
NHL	**92.41%**	**0.923**	89.68%	0,895	91.44%	0,913	86.10%	0,859
UCSB	**88.45%**	**0.882**	83.28%	0,830	83.97%	0,835	85.69%	0,855
Average	95.91%	0.959	94.08%	0.940	94.76%	0.947	93.49%	0.934

tion is shown in Tables 26, 27, 28 and 29. The methods are ordered chronologically and the best results on each dataset are highlighted in bold. It can be noted that the methods that provided the best results on each classification task applied both deep learning (DL) and handcrafted (HC) features, except for breast image classification.

According to Table 26, we noted a lack of methods that used both HC features and DL for NHL image classification, which indicates a demand for new research in the area, although the best accuracy for this dataset was reported by (BAI et al., 2019), wherein both DL and HC approaches were applied. Breast cancer classification remains a challenging task in computer vision since few methods were able to obtain accuracies above 95% when classifying these types of images, as observed in Table 27. It is likely that the main reason for this is related to the short number of samples available on public breast tumour datasets. The methods that obtained higher accuracies on this dataset applied data augmentation techniques through geometric transformations, wherein they were able to generate and train 200,000 (YU et al., 2019) and 896,000 (LI et al., 2019) image patches, respectively. When geometric data augmentation was applied to the proposed ensemble, an accuracy of 90.69% was obtained for the classification of breast cancer. For colorectal images classification, the proposed method obtained the best performance in relation to other approaches, as seen in Table 28. The second, third and fourth methods ranked in the comparison are all based on CNN approaches, without the use of HC features (DABASS; VIG; VASHISTH, 2019; TAVOLARA et al., 2019; SENA et al., 2019). The results obtained with the proposed ensemble may encourage new research on the com-

bination of HC and DL features for colorectal images classification since few papers on literature have applied similar approaches. Lastly, different papers that applied computer vision approaches for the classification of gender and age from liver tissue images were compared, as seen in Table 29. This is a context that has not been as explored as cancer detection challenges. However, it has already provided relevant results with high accuracy rates.

Table 26 – Overview of the accuracy values (%) obtained by different approaches for NHL image classification.

Method	Approach	Accuracy
(JANOWCZYK; MADABHUSHI, 2016)	Caffe and AlexNet (DL)	96.58
(ROBERTO et al., 2017)	PERC (HC)	92.00
(JIANG et al., 2018)	Morphology, entropy, GLCM and other handcrafted features (HC)	97.96
(RIBEIRO et al., 2018)	Colour normalization, PERC, LAC, GLCM and curvelet (HC)	86.14
(MARTINS et al., 2019)	FD and LAC (HC)	97.00
(BAI et al., 2019)	**GoogLeNet, intensity, texture and morphological features (DL+HC)**	**99.10**
(ZHANG et al., 2020)	Static and Dynamic training method for VGG-16 and PCA (DL)	98.93
(MARTINS et al., 2021)	FD, LAC and Hermite polynomial classifier (HC)	91.61
Proposed	ResNet-50, GADF, FD, LAC, PERC, LOG and MLP (DL+HC)	92.41

Table 27 – Overview of the accuracy values (%) obtained by different approaches for breast histology image classification (UCSB).

Method	Approach	Accuracy
(ARAÚJO et al., 2017)	Colour normalization, 13-layer CNN and SVM (DL)	83.30
(PAPASTERGIOU; ZACHARAKI; MEGALOOIKONOMOU, 2018)	Spatial decomposition, tensors (DL)	84.67
(FENG; ZHANG; YI, 2018)	Stacked denoising autoencoder (DL)	94.41
(LI et al., 2019)	**RefineNet and Atrous DenseNet (DL)**	**97.63**
(YU et al., 2019)	CNN, LBP, SURF, GLCM and other handcrafted features (DL+HC)	96.67
(ROBERTO et al., 2019)	PERC (HC)	86.20
(KAUSAR et al., 2019)	Colour normalization, Haar wavelet decomposition and 16-layer CNN (DL)	91.00
Proposed	ResNet-50, GADF, FD, LAC, PERC, LOG and MLP (DL+HC)	88.45

Overall, despite providing relevant results, most of these methods were implemented for specific classification tasks. Few computer vision approaches were able to perform well on different histological image categories (ROBERTO et al., 2019; NANNI et al., 2019). Moreover, recent papers have presented new approaches to deal with handcrafted fractal features. (ROBERTO et al., 2017) proposed an approach to obtain global PERC features from RGB images. Later, in (ROBERTO et al., 2019), local PERC features were also used for image classification. In (RIBEIRO et al., 2019), the authors used different

Table 28 – Overview of the accuracy values (%) obtained by different approaches for colorectal histology image classification (CR).

Method	Approach	Accuracy
(SANTOS et al., 2018)	Sample entropy and fuzzy (HC)	91.39
(BENTAIEB; HAMARNEH, 2018)	U-Net and AlexNet (DL)	87.50
(DABASS; VIG; VASHISTH, 2019)	31-layer CNN (DL)	96.97
(TAVOLARA et al., 2019)	GAN and U-Net (DL)	94.02
(SENA et al., 2019)	12-layer CNN (DL)	93.28
(ROBERTO et al., 2019)	PERC (HC)	90.90
(AWAN et al., 2019)	Colour normalization, U-Net and GoogLeNet (DL)	85.00
(CANDELERO et al., 2020)	LeNet-5, FD, LAC, PERC, Haralick descriptors, K* and GA (DL+HC)	91.06
Proposed	**ResNet-50, GADF, FD, LAC, PERC, LOG and MLP (DL+HC)**	**99.45**

Table 29 – Overview of the accuracy values (%) obtained by different approaches for gender and age classification from liver histology images.

Method	Approach	Acc. (gender)	Acc. (age)
(WATANABE; KOBAYASHI; WADA, 2016)	GIST descriptors, PCA and LDA (HC)	93.70	88.40
(RUBERTO et al., 2016)	**Statistical analysis and texture features (HC)**	**100.00**	**100.00**
(ANDREARCZYK; WHELAN, 2017)	Texture CNN (DL)	99.10	98.20
(NANNI et al., 2019)	**6 CNN models and handcrafted descriptors (DL+HC)**	**100.00**	**100.00**
Proposed	ResNet-50, GADF, FD, LAC, PERC, LOG and MLP (DL+HC)	99.77	99.45

types of handcrafted features, including FD, LAC and PERC, to classify colorectal tumours. However, these methods applied only the Δ_h metric to extract the fractal features. Moreover, none of these approaches explored the contribution of these features along with CNN models, which were able to provide high accuracy rates in several CAD systems for histopathology tasks (LI et al., 2019; BAI et al., 2019; ANDREARCZYK; WHELAN, 2017). Besides, each of these authors have chosen a different validation approach for their methods, including k-fold cross validation with different values for k or hold-out validation. We have used 10-fold cross validation performed 10 times with different fold variations, which increases the reliability of the results. Nevertheless, these values are rather complementary than comparable since the experiments were performed under different situations. Hence, an ensemble method that addresses both fractal geometry and deep learning, which is the core of the proposed method, could improve these results when applied to different histology datasets.

5.4 Considerations

On this chapter, we have presented a detailed evaluation of the proposed method. It was shown that the ResNet-50 is the most adequate model for the deep learning module, as well as the LOG and MLP classifiers are for the handcrafted classification module. We have also noted that the GADF reshape provides better classification results on average when compared to the other three reshaping procedures evaluated, although there is not a statistically significant difference. We have also shown that, in general, our method is able to obtain better results than other common computer vision approaches such as ADASYN and data augmentation. At last, we have shown that our results are on par with recent published research in this area, and we believe that our method can serve as a complementary approach for the development of new histology image classification studies.

Chapter 6

Conclusion

In this PhD work, a new approach was presented whose objective was to investigate representations based on fractal geometry for the classification of lesions in histology images. The evaluated representations were GASF and GADF (WANG; OATES, 2015), sequential (ROBERTO et al., 2021) and RP (AFONSO et al., 2019). The obtained results have shown that the use of an image representation of fractal features through the GADF method in association with the classification of local and global handcrafted fractal features contributed significantly to the improvement of the classification accuracy when compared to traditional techniques such as ADASYN or geometric data augmentation. The proposed approach was able to provide relevant results, with accuracy rates above 88%, for the five sets of histological images analysed. Moreover, accuracy values higher than 99% were obtained for three of the five datasets evaluated (CR, LG and LA). The evaluations performed with the class activation mapping using Grad-CAM have shown that the three fractal approaches (FD, LAC and PERC) have played a significant role in the classification of the feature images (F-CNN) using the GADF reshape in all tested datasets.

A classifier ensemble was proposed, composed by four classification elements: O-CNN, F-CNN, local and global handcrafted fractal features classification. Despite the fact that the highest accuracy rates were provided by the O-CNN, we have shown that using an ensemble, each of these elements play a significant role in the classification of the five datasets, since the removal of the F-CNN causes a reduction of up to 1.83% in the average accuracy and 1.90% if only the CNN classifiers are used. We have also shown that assigning a greater weight to the O-CNN increased the overall accuracy by 2.42%. This is a relevant improvement since any misdiagnosis in the medical area could lead to inadequate treatment, hindering the patient's life quality. However, a more detailed study on weight assignment is required for better understanding the contribution of each of the ensemble elements. This could be explored in future works, as well as other open points that are described Section 6.1. Nevertheless, despite requiring two CNN classifications, we have shown that with the use of transfer learning, both O-CNN and F-CNN require only 10

training epochs for obtaining the expected classification accuracy rates. Therefore, we believe that the objectives listed at the beginning of the project have been met:

1. the feature image models based on RP, GASF, GADF and sequential reshape were able to represent fractal properties;
2. the proposed model has performed well on five different histology image datasets;
3. the proposed model is based on both convolutional and fractal features;
4. handcrafted local and global fractal features were able to enhanced the model's performance.

Overall, we believe that the proposed method could provide to the literature the expected contributions. The new feature image representation of handcrafted fractal features based on GADF has shown to increase the classification performance in an ensemble model. Our method was also able to obtain relevant results on five datasets from four different types of histology tissue: colorectal, breast, lymphoma and liver. These results are compatible with state-of-the-art approaches and we believe that our method can provide good information as a supporting tool for pathologists. Moreover, we have shown that the association of fractal features and deep learning is able to classify different types of histology lesions even in unbalanced and multiclass contexts.

6.1 Future works

The proposed approach shows that the values obtained are promising for the research area, allowing adaptability to different domains of histological tissues and with relatively low computational cost. Moreover, this approach brings the reshaping of fractal feature vectors for applications in CNN. This method, which consists of building a feature image from handcrafted fractal features, has still open points that must be explored for a more robust analysis for the classification of lesions in histological images. Nevertheless, some improvements in the proposed model can be explored so that results can be enhanced such as the UCSB dataset and other image sets. The adjustment of hyperparameters, such as the learning rate and the way it varies, can play an important role to improve the model's accuracy. It is also suggested to evaluate a pre-processing step based on H&E colour channel normalisation for dye uniformisation. The colour normalisation techniques proposed in (REINHARD et al., 2001; KHAN et al., 2014; VAHADANE et al., 2016) have provided promising results in the classification of breast, colorectal and NHL tumour images when applied in association with fractal features (RIBEIRO et al., 2018; ROBERTO et al., 2019). However, this approach has not yet been explored in a CNN context, as proposed in this work. At last, we encourage the application of this method on other types of image, including classification tasks that are not related to the medical area.

6.2 Research topics

The following papers were published in local and international conferences, as well as in relevant journals, as an outcome of the results obtained during the research developed in this doctorate course.

- ❏ Publication of the paper *An investigation of Jaya optmization for non-Hodgkin lymphoma classification* at the Computer Vision Workshop, held in november 2018 at Ilhéus-BA, which was awarded as 3rd best poster in the event (ROBERTO et al., 2018);

- ❏ Presentation of the poster *Classificação de tumores de mama e colorretais baseada em percolação de imagens normalizadas* at the XIII Workshop of thesis and dissertations in computer science – FACOM TechWeek, which was awarded as the best doctorate project of the event;

- ❏ Publication of the paper *Classification of breast and colorectal tumors based on percolation of color normalized images* in the journal *Computer and Graphics*, which was also presented at the event SIBGRAPI 2019, held in Rio de Janeiro-RJ in october 2019 (ROBERTO et al., 2019);

- ❏ Publication of the paper *Fractal Neural Network: a new ensemble of fractal geometry and convolutional neural networks for the classification of histology images* in the journal *Expert Systems with Applications*, volume 166, 2021, in collaboration with Prof. Dr. Alessandra Lumini, from *Università di Bologna*.

- ❏ Submission of the paper *Classification of histology images using an association of fractal geometry and deep learning* in the journal *Biomedical Signal Processing and Control*.

During the development of this doctorate, some collaborative research, in the field of histology images classification, have also been published through the knowledge acquired while implementing the proposed approach. These papers are listed below:

- ❏ Publication of the paper *Unsupervised method for normalization of hematoxylin-eosin stain in histological images* in the journal *Computerized Medical Imaging and Graphics* (TOSTA et al., 2019);

- ❏ Publication of the paper *A Model Based on Genetic Algorithm for Colorectal Cancer Diagnosis* in the *Iberoamerican Congress on Pattern Recognition* (TAINO et al., 2019);

- Publication of the paper *Multidimensional and multiscale Higuchi dimension for the analysis of colorectal histological images* in the *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (TENGUAM et al., 2020);
- Publication of the paper*Selection of CNN, Haralick and Fractal Features Based on Evolutionary Algorithms for Classification of Histological Images* in the *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (CANDELERO et al., 2020);
- Publication of the paper *Analysis of cancer in histological images: employing an approach based on genetic algorithm* in the journal *Pattern Analysis and Applications* (TAINO et al., 2021);
- Publication of the paper *A Hermite polynomial algorithm for detection of lesions in lymphoma images* in the journal *Pattern Analysis and Applications* (MARTINS et al., 2021).

Bibliography

AFONSO, L. C. et al. A recurrence plot-based approach for parkinson's disease identification. **Future Generation Computer Systems**, Elsevier, v. 94, p. 282–292, 2019. Disponível em: <https://doi.org/10.1016/j.future.2018.11.054>.

AGEMAP, N. I. o. A. **The Atlas of Gene Expression in Mouse Aging Project (AGEMAP)**. 2020. <https://ome.grc.nia.nih.gov/iicbu2008/agemap/index.html>. Acesso em: 04/05/2020.

AL-KADI, O. S. A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours. **Computerized Medical Imaging and Graphics**, Elsevier, v. 41, p. 67–79, 2015. Disponível em: <https://doi.org/10.1016/j.compmedimag.2014.05.013>.

ALOYSIUS, N.; GEETHA, M. A review on deep convolutional neural networks. In: IEEE. **2017 International Conference on Communication and Signal Processing (ICCSP)**. 2017. p. 0588–0592. Disponível em: <https://doi.org/10.1109/ICCSP.2017.8286426>.

AMALINA, N. N.; RAMADHANI, K. N.; STHEVANIE, F. Nuclei detection and classification system based on speeded up robust feature (surf). **EMITTER International Journal of Engineering Technology**, v. 7, n. 1, p. 1–13, 2019. Disponível em: <https://doi.org/10.24003/emitter.v7i1.288>.

AN, F.-P. Medical image classification algorithm based on weight initialization-sliding window fusion convolutional neural network. **Complexity**, Hindawi, v. 2019, 2019. Disponível em: <https://doi.org/10.1155/2019/9151670>.

ANDREARCZYK, V.; WHELAN, P. F. Deep learning for biomedical texture image analysis. In: IRISH PATTERN RECOGNITION & CLASSIFICATION SOCIETY (IPRCS). **Proceedings of the Irish Machine Vision & Image Processing Conference**. [S.l.], 2017.

ARAÚJO, T. et al. Classification of breast cancer histology images using convolutional neural networks. **PloS one**, Public Library of Science, v. 12, n. 6, p. e0177544, 2017. Disponível em: <https://doi.org/10.1371/journal.pone.0177544>.

AWAN, R. et al. Glandular structure-guided classification of microscopic colorectal images using deep learning. **Computers & Electrical Engineering**, Elsevier, p. 106450, 2019. Disponível em: <https://doi.org/10.1016/j.compeleceng.2019.106450>.

BAI, J. et al. Nhl pathological image classification based on hierarchical local information and googlenet-based representations. **BioMed research international**, Hindawi, v. 2019, 2019. Disponível em: <https://doi.org/10.1155/2019/1065652>.

BAISH, J. W.; JAIN, R. K. Fractals and cancer. **Cancer research**, AACR, v. 60, n. 14, p. 3683–3688, 2000.

BEEVI, K. S.; NAIR, M. S.; BINDU, G. Automatic mitosis detection in breast histopathology images using convolutional neural network based deep transfer learning. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 39, n. 1, p. 214–223, 2019. Disponível em: <https://doi.org/10.1016/j.bbe.2018.10.007>.

BEJNORDI, B. E. et al. Automated detection of dcis in whole-slide h&e stained breast histopathology images. **IEEE transactions on medical imaging**, IEEE, v. 35, n. 9, p. 2141–2150, 2016. Disponível em: <https://doi.org/10.1109/TMI.2016.2550620>.

BEN-COHEN, A. et al. Retinal layers segmentation using fully convolutional network in oct images. **RSIP Vision**, 2017.

BENTAIEB, A.; HAMARNEH, G. Adversarial stain transfer for histopathology image analysis. **IEEE transactions on medical imaging**, IEEE, v. 37, n. 3, p. 792–802, 2018. Disponível em: <https://doi.org/10.1109/TMI.2017.2781228>.

BENVENISTE, A. P. A.; FERREIRA, A. H. P. G.; AGUILLAR, V. L. N. Dupla leitura no rastreamento mamográfico. **Radiologia Brasileira**, SciELO Brasil, v. 39, n. 2, p. 85–89, 2006. Disponível em: <https://doi.org/10.1590/S0100-39842006000200003>.

BEZERRA, C. S. et al. Uma abordagem de segmentação de íris para fins biométricos usando aprendizagem profunda. 2019.

BIRD, N.; PERRIER, E. Multiscale percolation properties of a fractal pore network. **Geoderma**, Elsevier, v. 160, n. 1, p. 105–110, 2010. Disponível em: <https://doi.org/10.1016/j.geoderma.2009.10.009>.

BRACHMANN, A.; REDIES, C. Defining self-similarity of images using features learned by convolutional neural networks. **Electronic Imaging**, Society for Imaging Science and Technology, v. 2017, n. 14, p. 188–194, 2017. Disponível em: <https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-142>.

BROADBENT, S. R.; HAMMERSLEY, J. M. Percolation processes: I. crystals and mazes. In: CAMBRIDGE UNIVERSITY PRESS. **Mathematical Proceedings of the Cambridge Philosophical Society**. 1957. v. 53, n. 3, p. 629–641. Disponível em: <https://doi.org/10.1017/S0305004100032680>.

CĂLIMAN, A.; IVANOVICI, M. Psoriasis image analysis using color lacunarity. In: IEEE. **Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on**. 2012. p. 1401–1406. Disponível em: <https://doi.org/10.1109/OPTIM.2012.6231850>.

CANDELERO, D. et al. Selection of cnn, haralick and fractal features based on evolutionary algorithms for classification of histological images. In: IEEE. **2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. 2020. p. 2709–2716. Disponível em: <https://doi.org/10.1109/BIBM49941.2020.9313328>.

CASTRO, L. N. D. **Fundamentals of natural computing: basic concepts, algorithms, and applications**. CRC Press, 2006. Disponível em: <https://doi.org/10.1201/9781420011449>.

CHAN, A.; TUSZYNSKI, J. A. Automatic prediction of tumour malignancy in breast cancer with fractal dimension. **Royal Society open science**, The Royal Society, v. 3, n. 12, p. 160558, 2016. Disponível em: <https://doi.org/10.1098/rsos.160558>.

CHAN, H.-P.; HADJIISKI, L. M.; SAMALA, R. K. Computer-aided diagnosis in the era of deep learning. **Medical physics**, Wiley Online Library, v. 47, n. 5, p. e218–e227, 2020. Disponível em: <https://doi.org/10.1002/mp.13764>.

CHANG, H. et al. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 40, n. 5, p. 1182–1194, 2017. Disponível em: <https://doi.org/10.1109/TPAMI.2017.2656884>.

CHEN, W. et al. Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 10, n. 2, p. 425, 2020. Disponível em: <https://doi.org/10.3390/app10020425>.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2017. p. 1251–1258. Disponível em: <https://doi.org/10.1109/CVPR.2017.195>.

CIREŞAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. In: IEEE. **2012 IEEE conference on computer vision and pattern recognition**. 2012. p. 3642–3649. Disponível em: <https://doi.org/10.1109/CVPR.2012.6248110>.

CIREŞAN, D. C. et al. Mitosis detection in breast cancer histology images with deep neural networks. In: SPRINGER. **International Conference on Medical Image Computing and Computer-assisted Intervention**. 2013. p. 411–418. Disponível em: <https://doi.org/10.1007/978-3-642-40763-5_51>.

DABASS, M.; VIG, R.; VASHISTH, S. Five-grade cancer classification of colon histology images via deep learning. In: CRC PRESS. **Communication and Computing Systems: Proceedings of the 2nd International Conference on Communication and Computing Systems (ICCCS 2018), December 1-2, 2018, Gurgaon, India**. 2019. p. 18. Disponível em: <https://doi.org/10.1201/9780429444272-3>.

DABEER, S.; KHAN, M. M.; ISLAM, S. Cancer diagnosis in histopathological image: Cnn based approach. **Informatics in Medicine Unlocked**, Elsevier, v. 16, p. 100231, 2019. Disponível em: <https://doi.org/10.1016/j.imu.2019.100231>.

DEAN, P. A new monte carlo method for percolation problems on a lattice. In: CAMBRIDGE UNIVERSITY PRESS. **Mathematical Proceedings of the Cambridge Philosophical Society**. 1963. v. 59, n. 2, p. 397–410. Disponível em: <https://doi.org/10.1017/S0305004100037026>.

DHANSAY, T.; BRANDL, G.; WIT, M. D. Fractal geometry of the fault network across the soutpansberg mountains, limpopo, south africa. **South African Journal of Geology 2016**, Geological Society of South Africa, v. 119, n. 1, p. 235–242, 2016. Disponível em: <https://doi.org/10.2113/gssajg.119.1.235>.

DIMITROPOULOS, K. et al. Automated detection and classification of nuclei in pax5 and h&e-stained tissue sections of follicular lymphoma. **Signal, Image and Video Processing**, Springer, v. 11, n. 1, p. 145–153, 2017. Disponível em: <https://doi.org/10.1007/s11760-016-0913-6>.

DOBBS, J. L. et al. Micro-anatomical quantitative optical imaging: toward automated assessment of breast tissues. **Breast Cancer Research**, Springer, v. 17, n. 1, p. 105, 2015. Disponível em: <https://doi.org/10.1186/s13058-015-0617-9>.

DOUZE, M. et al. Evaluation of gist descriptors for web-scale image search. In: **Proceedings of the ACM International Conference on Image and Video Retrieval**. [s.n.], 2009. p. 1–8. Disponível em: <https://doi.org/10.1145/1646396.1646421>.

ECKMANN, J.-P. et al. Recurrence plots of dynamical systems. **World Scientific Series on Nonlinear Science Series A**, WORLD SCIENTIFIC PUBLISHING, v. 16, p. 441–446, 1995. Disponível em: <https://doi.org/10.1142/9789812833709_0030>.

FARIA, F. A. et al. Fusion of time series representations for plant recognition in phenology studies. **Pattern Recognition Letters**, Elsevier, v. 83, p. 205–214, 2016. Disponível em: <https://doi.org/10.1016/j.patrec.2016.03.005>.

FENG, Y.; ZHANG, L.; YI, Z. Breast cancer cell nuclei classification in histopathology images using deep neural networks. **International journal of computer assisted radiology and surgery**, Springer, v. 13, n. 2, p. 179–191, 2018. Disponível em: <https://doi.org/10.1007/s11548-017-1663-9>.

FERREIRA, A. J.; FIGUEIREDO, M. A. Boosting algorithms: A review of methods, theory, and applications. In: **Ensemble machine learning**. Springer, 2012. p. 35–85. Disponível em: <https://doi.org/10.1007/978-1-4419-9326-7_2>.

FONDÓN, I. et al. Automatic classification of tissue malignancy for breast carcinoma diagnosis. **Computers in biology and medicine**, Elsevier, v. 96, p. 41–51, 2018. Disponível em: <https://doi.org/10.1016/j.compbiomed.2018.03.003>.

FUKUMA, K. et al. A study on feature extraction and disease stage classification for glioma pathology images. In: IEEE. **2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**. 2016. p. 2150–2156. Disponível em: <https://doi.org/10.1109/FUZZ-IEEE.2016.7737958>.

GELASCA, E. D. et al. Evaluation and benchmark for biological image segmentation. In: **IEEE International Conference on Image Processing**. [S.l.: s.n.], 2008.

GHAZI, M. M.; YANIKOGLU, B.; APTOULA, E. Plant identification using deep neural networks via optimization of transfer learning parameters. **Neurocomputing**, Elsevier, v. 235, p. 228–235, 2017. Disponível em: <https://doi.org/10.1016/j.neucom.2017.01.018>.

GULUM, M. A.; TROMBLEY, C. M.; KANTARDZIC, M. A review of explainable deep learning cancer detection models in medical imaging. **Applied Sciences**, Multidisciplinary Digital Publishing Institute, v. 11, n. 10, p. 4573, 2021. Disponível em: <https://doi.org/10.3390/app11104573>.

GURCAN, M. N. et al. Histopathological image analysis: A review. **IEEE reviews in biomedical engineering**, IEEE, v. 2, p. 147–171, 2009. Disponível em: <https://doi.org/10.1109/RBME.2009.2034865>.

HAM, A.; CORMACK, D. **Histologia. Ed**. [S.l.]: Guanabara Koogan, Rio de Janeiro, 1983.

HARDER, N. et al. Co-occurrence features characterizing gland distribution patterns as new prognostic markers in prostate cancer whole-slide images. In: IEEE. **Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on**. 2016. p. 807–810. Disponível em: <https://doi.org/10.1109/ISBI.2016.7493389>.

HE, H. et al. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE. **2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)**. [S.l.], 2008. p. 1322–1328.

HE, K.; SUN, J. Convolutional neural networks at constrained time cost. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2015. p. 5353–5360. Disponível em: <https://doi.org/10.1109/CVPR.2015.7299173>.

HE, K. et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 37, n. 9, p. 1904–1916, 2015. Disponível em: <https://doi.org/10.1109/TPAMI.2015.2389824>.

HE, K. et al. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2016. p. 770–778. Disponível em: <https://doi.org/10.1109/CVPR.2016.90>.

HE, L. et al. Histology image analysis for carcinoma detection and grading. **Computer methods and programs in biomedicine**, Elsevier, v. 107, n. 3, p. 538–556, 2012. Disponível em: <https://doi.org/10.1016/j.cmpb.2011.12.007>.

HOSHEN, J.; BERRY, M.; MINSER, K. Percolation and cluster structure parameters: The enhanced hoshen-kopelman algorithm. **Physical Review E**, APS, v. 56, n. 2, p. 1455, 1997. Disponível em: <https://doi.org/10.1103/PhysRevE.56.1455>.

HSIEH, C.-H. et al. Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. **Surgery**, Elsevier, v. 149, n. 1, p. 87–93, 2011. Disponível em: <https://doi.org/10.1016/j.surg.2010.03.023>.

HSU, C.-Y.; CHIEN, J.-C. Ensemble convolutional neural networks with weighted majority for wafer bin map pattern classification. **Journal of Intelligent Manufacturing**, Springer, p. 1–14, 2020. Disponível em: <https://doi.org/10.1007/s10845-020-01687-7>.

IJJINA, E. P.; CHALAVADI, K. M. Human action recognition using genetic algorithms and convolutional neural networks. **Pattern recognition**, Elsevier, v. 59, p. 199–212, 2016. Disponível em: <https://doi.org/10.1016/j.patcog.2016.01.012>.

ILSE, M.; TOMCZAK, J. M.; WELLING, M. Attention-based deep multiple instance learning. In: INTERNATIONAL MACHINE LEARNING SOCIETY (IMLS). **35th International Conference on Machine Learning, ICML 2018**. [S.l.], 2018. p. 3376–3391.

IVANOVICI, M.; RICHARD, N. The lacunarity of colour fractal images. In: IEEE. **Image Processing (ICIP), 2009 16th IEEE International Conference on**. 2009. p. 453–456. Disponível em: <https://doi.org/10.1109/ICIP.2009.5414394>.

IVANOVICI, M.; RICHARD, N. Fractal dimension of color fractal images. **IEEE Transactions on Image Processing**, IEEE, v. 20, n. 1, p. 227–235, 2011. Disponível em: <https://doi.org/10.1109/TIP.2010.2059032>.

IVANOVICI, M.; RICHARD, N.; DECEAN, H. Fractal dimension and lacunarity of psoriatic lesions-a colour approach. **medicine**, v. 6, n. 4, p. 7, 2009.

JANOWCZYK, A.; MADABHUSHI, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. **Journal of pathology informatics**, Wolters Kluwer–Medknow Publications, v. 7, 2016. Disponível em: <https://doi.org/10.4103/2153-3539.186902>.

JAPKOWICZ, N.; SHAH, M. **Evaluating learning algorithms: a classification perspective**. Cambridge University Press, 2011. Disponível em: <https://doi.org/10.1017/CBO9780511921803>.

JIANG, C.; SU, J. Gabor binary layer in convolutional neural networks. In: IEEE. **2018 25th IEEE International Conference on Image Processing (ICIP)**. 2018. p. 3408–3412. Disponível em: <https://doi.org/10.1109/ICIP.2018.8451298>.

JIANG, H. et al. An effective multi-classification method for nhl pathological images. In: IEEE. **2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**. 2018. p. 763–768. Disponível em: <https://doi.org/10.1109/SMC.2018.00138>.

JIANG, Y. et al. Classification of h&e stained breast cancer histopathology images based on convolutional neural network. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. 2019. v. 1302, n. 3, p. 032018. Disponível em: <https://doi.org/10.1088/1742-6596/1302/3/032018>.

JØRGENSEN, A. S. et al. Exploiting multiple color representations to improve colon cancer detection in whole slide h&e stains. In: **Computational Pathology and Ophthalmic Medical Image Analysis**. Springer, 2018. p. 61–68. Disponível em: <https://doi.org/10.1007/978-3-030-00949-6_8>.

JOTHI, J. A. A.; RAJAM, V. M. A. A survey on automated cancer diagnosis from histopathology images. **Artificial Intelligence Review**, Springer, v. 48, n. 1, p. 31–81, 2017. Disponível em: <https://doi.org/10.1007/s10462-016-9494-6>.

JUEFEI-XU, F.; BODDETI, V. N.; SAVVIDES, M. Local binary convolutional neural networks. In: IEEE. **Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on**. 2017. v. 1. Disponível em: <https://doi.org/10.1109/CVPR.2017.456>.

KASSANI, S. H. et al. Classification of histopathological biopsy images using ensemble of deep learning networks. In: **Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering**. [S.l.: s.n.], 2019. p. 92–99.

KAUSAR, T. et al. Hwdcnn: Multi-class recognition in breast histopathology with haar wavelet decomposed image based convolution neural network. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 39, n. 4, p. 967–982, 2019. Disponível em: <https://doi.org/10.1016/j.bbe.2019.09.003>.

KHAN, A. et al. A survey of the recent architectures of deep convolutional neural networks. **Artificial Intelligence Review**, Springer, p. 1–62, 2020.

KHAN, A. M. et al. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 61, n. 6, p. 1729–1738, 2014. Disponível em: <https://doi.org/10.1109/TBME.2014.2303294>.

KHOSHDELI, M.; CONG, R.; PARVIN, B. Detection of nuclei in h&e stained sections using convolutional neural networks. In: IEEE. **Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on**. 2017. p. 105–108. Disponível em: <https://doi.org/10.1109/BHI.2017.7897216>.

KIERNAN, J. A. Nuclear stains. **Cold Spring Harbor Protocols**, Cold Spring Harbor Laboratory Press, v. 2008, n. 7, p. pdb–top50, 2008. Disponível em: <https://doi.org/10.1101/pdb.top50>.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2012. p. 1097–1105.

KUMAR, G.; BHATIA, P. K. A detailed review of feature extraction in image processing systems. In: IEEE. **2014 Fourth international conference on advanced computing & communication technologies**. 2014. p. 5–12. Disponível em: <https://doi.org/10.1109/ACCT.2014.74>.

KURMI, Y.; CHAURASIA, V.; GANESH, N. Tumor malignancy detection using histopathology imaging. **Journal of medical imaging and radiation sciences**, Elsevier, v. 50, n. 4, p. 514–528, 2019. Disponível em: <https://doi.org/10.1016/j.jmir.2019.07.004>.

LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 1998. Disponível em: <https://doi.org/10.1109/5.726791>.

LEEUWEN, I. M. V. et al. Towards a multiscale model of colorectal cancer. **World journal of gastroenterology: WJG**, Baishideng Publishing Group Inc, v. 13, n. 9, p. 1399, 2007. Disponível em: <https://doi.org/10.3748/wjg.v13.i9.1399>.

LEON, F.; FLORIA, S.-A.; BĂDICĂ, C. Evaluating the effect of voting methods on ensemble-based classification. In: IEEE. **2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)**. 2017. p. 1–6. Disponível em: <https://doi.org/10.1109/INISTA.2017.8001122>.

LI, L. et al. Multifractal analysis and lacunarity analysis: A promising method for the automated assessment of muskmelon (cucumis melo l.) epidermis netting. **Computers and electronics in agriculture**, Elsevier, v. 88, p. 72–84, 2012. Disponível em: <https://doi.org/10.1016/j.compag.2012.06.006>.

LI, Y. et al. Reverse active learning based atrous densenet for pathological image classification. **BMC bioinformatics**, Springer, v. 20, n. 1, p. 445, 2019. Disponível em: <https://doi.org/10.1186/s12859-019-2979-y>.

LIN, M.; CHEN, Q.; YAN, S. Network in network. **arXiv preprint arXiv:1312.4400**, 2013.

LIN, W.-C. et al. Curvelet-based texture classification of critical gleason patterns of prostate histological images. In: IEEE. **Computational Advances in Bio and Medical Sciences (ICCABS), 2016 IEEE 6th International Conference on**. 2016. p. 1–6. Disponível em: <https://doi.org/10.1109/ICCABS.2016.7802768>.

LIU, L. et al. A survey of recent advances in texture representation. **arXiv preprint arXiv:1801.10324**, 2018.

LIU, M.; YAN, X.; WANG, L. Identification of cell pathology by using stacked auto-encoders combination with rotation forest. In: IEEE. **2018 5th International Conference on Information Science and Control Engineering (ICISCE)**. 2018. p. 261–265. Disponível em: <https://doi.org/10.1109/ICISCE.2018.00062>.

LIU, X. et al. Ensemble transfer learning algorithm. **IEEE Access**, IEEE, v. 6, p. 2389–2396, 2017. Disponível em: <https://doi.org/10.1109/ACCESS.2017.2782884>.

LONG, M. et al. Deep transfer learning with joint adaptation networks. In: JMLR. ORG. **Proceedings of the 34th International Conference on Machine Learning-Volume 70**. [S.l.], 2017. p. 2208–2217.

LOPES, R.; BETROUNI, N. Fractal and multifractal analysis: a review. **Medical image analysis**, Elsevier, v. 13, n. 4, p. 634–649, 2009. Disponível em: <https://doi.org/10.1016/j.media.2009.05.003>.

LU, Y. Food image recognition by using convolutional neural networks (cnns). **arXiv preprint arXiv:1612.00983**, 2016.

LUMINI, A.; NANNI, L. Convolutional neural networks for atc classification. **Current pharmaceutical design**, Bentham Science Publishers, v. 24, n. 34, p. 4007–4012, 2018. Disponível em: <https://doi.org/10.2174/1381612824666181112113438>.

MACENKO, M. et al. A method for normalizing histology slides for quantitative analysis. In: IEEE. **Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on**. 2009. p. 1107–1110. Disponível em: <https://doi.org/10.1109/ISBI.2009.5193250>.

MAHBOD, A. et al. Breast cancer histological image classification using fine-tuned deep network fusion. In: SPRINGER. **International Conference Image Analysis and Recognition**. 2018. p. 754–762. Disponível em: <https://doi.org/10.1007/978-3-319-93000-8_85>.

MALARZ, K. Simple cubic random-site percolation thresholds for neighborhoods containing fourth-nearest neighbors. **Physical Review E**, APS, v. 91, n. 4, p. 043301, 2015. Disponível em: <https://doi.org/10.1103/PhysRevE.91.043301>.

MANDELBROT, B. B. **Les objets fractals: forme, hasard et dimension**. [S.l.]: Flammarion, 1975.

MANDELBROT, B. B. **The fractal geometry of nature**. [S.l.]: WH freeman New York, 1983. v. 173.

MARI, A. et al. Transfer learning in hybrid classical-quantum neural networks. **arXiv preprint arXiv:1912.08278**, 2019. Disponível em: <https://doi.org/10.22331/q-2020-10-09-340>.

MARTINS, A. S. et al. Colour feature extraction and polynomial algorithm for classification of lymphoma images. In: SPRINGER. **Iberoamerican Congress on Pattern Recognition**. 2019. p. 262–271. Disponível em: <https://doi.org/10.1007/978-3-030-33904-3_24>.

MARTINS, A. S. et al. A hermite polynomial algorithm for detection of lesions in lymphoma images. **Pattern Analysis and Applications**, Springer, v. 24, n. 2, p. 523–535, 2021. Disponível em: <https://doi.org/10.1007/s10044-020-00927-z>.

MOHAMMED, M. A. et al. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. **Computers & Electrical Engineering**, Elsevier, 2018. Disponível em: <https://doi.org/10.1016/j.compeleceng.2018.01.033>.

MUELLER, J. L. et al. Rapid staining and imaging of subnuclear features to differentiate between malignant and benign breast tissues at a point-of-care setting. **Journal of cancer research and clinical oncology**, Springer, v. 142, n. 7, p. 1475–1486, 2016. Disponível em: <https://doi.org/10.1007/s00432-016-2165-9>.

MURTAZA, G. et al. Breast cancer classification using digital biopsy histopathology images through transfer learning. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. 2019. v. 1339, n. 1, p. 012035. Disponível em: <https://doi.org/10.1088/1742-6596/1339/1/012035>.

NANNI, L. et al. General purpose (genp) bioimage ensemble of handcrafted and learned features with data augmentation. **arXiv preprint arXiv:1904.08084**, 2019.

NANNI, L.; GHIDONI, S.; BRAHNAM, S. Handcrafted vs. non-handcrafted features for computer vision classification. **Pattern Recognition**, Elsevier, v. 71, p. 158–172, 2017. Disponível em: <https://doi.org/10.1016/j.patcog.2017.05.025>.

NANNI, L.; LUMINI, A.; GHIDONI, S. Ensemble of deep learned features for melanoma classification. **arXiv preprint arXiv:1807.08008**, 2018.

NANNI, L. et al. Stochastic activation function layers for convolutional neural networks. 2020. Disponível em: <https://doi.org/10.20944/preprints202002.0231.v1>.

NASCIMENTO, M. Z. do et al. Lymphoma images analysis using morphological and non-morphological descriptors for classification. **Computer methods and programs in biomedicine**, Elsevier, v. 163, p. 65–77, 2018. Disponível em: <https://doi.org/10.1016/j.cmpb.2018.05.035>.

NEVES, L. A. et al. Multi-scale lacunarity as an alternative to quantify and diagnose the behavior of prostate cancer. **Expert Systems with Applications**, Elsevier, v. 41, n. 11, p. 5017–5029, 2014. Disponível em: <https://doi.org/10.1016/j.eswa.2014.02.048>.

NGUYEN, L. et al. Spatial statistics for segmenting histological structures in h&e stained tissue images. **IEEE transactions on medical imaging**, IEEE, v. 36, n. 7, p. 1522–1532, 2017. Disponível em: <https://doi.org/10.1109/TMI.2017.2681519>.

NIAZI, M. K. K.; PARWANI, A. V.; GURCAN, M. N. Digital pathology and artificial intelligence. **The lancet oncology**, Elsevier, v. 20, n. 5, p. e253–e261, 2019. Disponível em: <https://doi.org/10.1016/S1470-2045(19)30154-8>.

NIKOLAIDIS, N.; NIKOLAIDIS, I.; TSOUROS, C. A variation of the box-counting algorithm applied to colour images. **arXiv preprint arXiv:1107.2336**, 2011.

NOONE, A. et al. **Cronin (eds) KA. SEER Cancer Statistics Review. 1975–2015, National Cancer Institute**. 2018.

ONAN, A.; KORUKOĞLU, S.; BULUT, H. A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. **Expert Systems with Applications**, Elsevier, v. 62, p. 1–16, 2016. Disponível em: <https://doi.org/10.1016/j.eswa.2016.06.005>.

PAN, S. J.; YANG, Q. A survey on transfer learning. **IEEE Transactions on knowledge and data engineering**, IEEE, v. 22, n. 10, p. 1345–1359, 2009. Disponível em: <https://doi.org/10.1109/TKDE.2009.191>.

PAPASTERGIOU, T.; ZACHARAKI, E. I.; MEGALOOIKONOMOU, V. Tensor decomposition for multiple-instance classification of high-order medical data. **Complexity**, Hindawi, v. 2018, 2018. Disponível em: <https://doi.org/10.1155/2018/8651930>.

PINTO, T. **Modelagem do Potencial Elétrico através da Membrana do Neurônio Ganglionar e Células de Neuroblastoma: Efeitos das Cargas Superficiais**. Tese (Doutorado) — Universidade do Estado do Rio de Janeiro, 2010.

POHJOLA, H.; GRANDELL, J. **Meteosat Third Generation Lightning Imager (MTG-LI): Flash and Accumulated products and test data for user readiness activities**. 2016. Disponível em: <http://www.eumetrain.org/data/4/433/Session_6.pdf>.

POLIKAR, R. Ensemble learning. In: **Ensemble machine learning**. Springer, 2012. p. 1–34. Disponível em: <https://doi.org/10.1007/978-1-4419-9326-7_1>.

RAKHLIN, A. et al. Deep convolutional neural networks for breast cancer histology image analysis. In: SPRINGER. **International Conference Image Analysis and Recognition**. 2018. p. 737–744. Disponível em: <https://doi.org/10.1007/978-3-319-93000-8_83>.

REINHARD, E. et al. Color transfer between images. **IEEE Computer graphics and applications**, IEEE, v. 21, n. 5, p. 34–41, 2001. Disponível em: <https://doi.org/10.1109/38.946629>.

RIBANI, R.; MARENGONI, M. A survey of transfer learning for convolutional neural networks. In: IEEE. **2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)**. 2019. p. 47–57. Disponível em: <https://doi.org/10.1109/SIBGRAPI-T.2019.00010>.

RIBEIRO, M. G. et al. Classification of colorectal cancer based on the association of multidimensional and multiresolution features. **Expert Systems with Applications**, Elsevier, v. 120, p. 262–278, 2019. Disponível em: <https://doi.org/10.1016/j.eswa.2018.11.034>.

RIBEIRO, M. G. et al. Analysis of the influence of color normalization in the classification of non-hodgkin lymphoma images. In: IEEE. **2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)**. 2018. p. 369–376. Disponível em: <https://doi.org/10.1109/SIBGRAPI.2018.00054>.

RICHARDSON, L. Fractal growth phenomena. **Ann. Arbor, Mich.: The Society**, v. 6, p. 139, 1961.

ROBERTO, G. F. et al. Fractal neural network: A new ensemble of fractal geometry and convolutional neural networks for the classification of histology images. **Expert Systems with Applications**, Elsevier, v. 166, p. 114103, 2021. Disponível em: <https://doi.org/10.1016/j.eswa.2020.114103>.

ROBERTO, G. F. et al. Classification of breast and colorectal tumors based on percolation of color normalized images. **Computers & Graphics**, Elsevier, v. 84, p. 134–143, 2019. Disponível em: <https://doi.org/10.1016/j.cag.2019.08.008>.

ROBERTO, G. F. et al. An investigation of jaya optimization for non-hodgkin lymphoma classification. In: **2018 XIV Workshop de Visão Computacional, Ilhéus-BA**. [S.l.: s.n.], 2018. p. 39–44.

ROBERTO, G. F. et al. Features based on the percolation theory for quantification of non-hodgkin lymphomas. **Computers in biology and medicine**, Elsevier, v. 91, p. 135–147, 2017. Disponível em: <https://doi.org/10.1016/j.compbiomed.2017.10.012>.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958. Disponível em: <https://doi.org/10.1037/h0042519>.

RUBERTO, C. D. et al. A feature learning framework for histology images classification. In: **Emerging trends in applications and infrastructures for computational biology, bioinformatics, and systems biology: systems and applications**. Elsevier Press, 2016. p. 37–48. Disponível em: <https://doi.org/10.1016/B978-0-12-804203-8.00003-1>.

RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. **International journal of computer vision**, Springer, v. 115, n. 3, p. 211–252, 2015. Disponível em: <https://doi.org/10.1007/s11263-015-0816-y>.

SAGI, O.; ROKACH, L. Ensemble learning: A survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Wiley Online Library, v. 8, n. 4, p. e1249, 2018. Disponível em: <https://doi.org/10.1002/widm.1249>.

SANTOS, L. F. S. dos et al. Multidimensional and fuzzy sample entropy (sampenmf) for quantifying h&e histological images of colorectal cancer. **Computers in biology and medicine**, Elsevier, v. 103, p. 148–160, 2018. Disponível em: <https://doi.org/10.1016/j.compbiomed.2018.10.013>.

SAÚDE, M. da. **Instituto Nacional de Câncer José Alencar Gomes da Silva. Estimativa 2020: Incidência de Câncer no Brasil. Rio de Janeiro**. [S.l.]: INCA, 2020.

SENA, P. et al. Deep learning techniques for detecting preneoplastic and neoplastic lesions in human colorectal histological images. **Oncology Letters**, Spandidos Publications, v. 18, n. 6, p. 6101–6107, 2019. Disponível em: <https://doi.org/10.3892/ol.2019.10928>.

SEVERYN, A.; MOSCHITTI, A. Twitter sentiment analysis with deep convolutional neural networks. In: ACM. **Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval**. 2015. p. 959–962. Disponível em: <https://doi.org/10.1145/2766462.2767830>.

SEZGIN, M.; SANKUR, B. Survey over image thresholding techniques and quantitative performance evaluation. **Journal of Electronic imaging**, International Society for Optics and Photonics, v. 13, n. 1, p. 146–166, 2004. Disponível em: <https://doi.org/10.1117/1.1631315>.

SHAMIR, L. et al. Iicbu 2008: a proposed benchmark suite for biological image analysis. **Medical & biological engineering & computing**, Springer, v. 46, n. 9, p. 943–947, 2008. Disponível em: <https://doi.org/10.1007/s11517-008-0380-5>.

SHANKAR, A. et al. Epileptic seizure classification based on gramian angular field transformation and deep learning. In: IEEE. **2020 IEEE Applied Signal Processing Conference (ASPCON)**. 2020. p. 147–151. Disponível em: <https://doi.org/10.1109/ASPCON49795.2020.9276717>.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, Springer, v. 6, n. 1, p. 60, 2019. Disponível em: <https://doi.org/10.1186/s40537-019-0197-0>.

SIMON, O. et al. Multi-radial lbp features as a tool for rapid glomerular detection and assessment in whole slide histopathology images. **Scientific reports**, Nature Publishing Group, v. 8, n. 1, p. 2032, 2018. Disponível em: <https://doi.org/10.1038/s41598-018-20453-7>.

SIRINUKUNWATTANA, K. et al. Gland segmentation in colon histology images: The glas challenge contest. **Medical image analysis**, Elsevier, v. 35, p. 489–502, 2017. Disponível em: <https://doi.org/10.1016/j.media.2016.08.008>.

SPANHOL, F. A. et al. A dataset for breast cancer histopathological image classification. **IEEE Transactions on Biomedical Engineering**, IEEE, v. 63, n. 7, p. 1455–1462, 2016. Disponível em: <https://doi.org/10.1109/TBME.2015.2496264>.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2016. p. 2818–2826. Disponível em: <https://doi.org/10.1109/CVPR.2016.308>.

TAINO, D. F. et al. A model based on genetic algorithm for colorectal cancer diagnosis. In: SPRINGER. **Iberoamerican Congress on Pattern Recognition**. 2019. p. 504–513. Disponível em: <https://doi.org/10.1007/978-3-030-33904-3_47>.

TAINO, D. F. et al. Analysis of cancer in histological images: employing an approach based on genetic algorithm. **Pattern Analysis and Applications**, Springer, v. 24, n. 2, p. 483–496, 2021. Disponível em: <https://doi.org/10.1007/s10044-020-00931-3>.

TAMBE, R. et al. Towards designing an automated classification of lymphoma subtypes using deep neural networks. In: **Proceedings of the ACM India Joint International Conference on Data Science and Management of Data**. [s.n.], 2019. p. 143–149. Disponível em: <https://doi.org/10.1145/3297001.3297019>.

TAN, C. et al. A survey on deep transfer learning. In: SPRINGER. **International conference on artificial neural networks**. 2018. p. 270–279. Disponível em: <https://doi.org/10.1007/978-3-030-01424-7_27>.

TAVOLARA, T. E. et al. A modular cgan classification framework: Application to colorectal tumor detection. **Scientific Reports**, Nature Publishing Group, v. 9, n. 1, p. 1–8, 2019. Disponível em: <https://doi.org/10.1038/s41598-019-55257-w>.

TENGUAM, J. J. et al. Multidimensional and multiscale higuchi dimension for the analysis of colorectal histological images. In: IEEE. **2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)**. 2020. p. 2833–2839. Disponível em: <https://doi.org/10.1109/BIBM49941.2020.9313575>.

TORREY, L.; SHAVLIK, J. Transfer learning. In: **Handbook of research on machine learning applications and trends: algorithms, methods, and techniques**. IGI global, 2010. p. 242–264. Disponível em: <https://doi.org/10.4018/978-1-60566-766-9.ch011>.

TOSTA, T. A. et al. Fitness functions evaluation for segmentation of lymphoma histological images using genetic algorithm. In: SPRINGER. **International Conference on the Applications of Evolutionary Computation**. 2018. p. 47–62. Disponível em: <https://doi.org/10.1007/978-3-319-77538-8_4>.

TOSTA, T. A. A. et al. Computational method for unsupervised segmentation of lymphoma histological images based on fuzzy 3-partition entropy and genetic algorithm. **Expert Systems with Applications**, Elsevier, v. 81, p. 223–243, 2017. Disponível em: <https://doi.org/10.1016/j.eswa.2017.03.051>.

TOSTA, T. A. A. et al. Unsupervised method for normalization of hematoxylin-eosin stain in histological images. **Computerized Medical Imaging and Graphics**, Elsevier, v. 77, p. 101646, 2019. Disponível em: <https://doi.org/10.1016/j.compmedimag.2019.101646>.

TRA, V. et al. Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 17, n. 12, p. 2834, 2017. Disponível em: <https://doi.org/10.3390/s17122834>.

TRIPATHI, S.; SINGH, S. K. Ensembling handcrafted features with deep features: an analytical study for classification of routine colon cancer histopathological nuclei images. **MULTIMEDIA TOOLS AND APPLICATIONS**, Springer, 2020. Disponível em: <https://doi.org/10.1007/s11042-020-08891-w>.

TZENG, E. et al. Adversarial discriminative domain adaptation. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition**. [s.n.], 2017. p. 7167–7176. Disponível em: <https://doi.org/10.1109/CVPR.2017.316>.

VAHADANE, A. et al. Structure-preserving color normalization and sparse stain separation for histological images. **IEEE transactions on medical imaging**, IEEE, v. 35, n. 8, p. 1962–1971, 2016. Disponível em: <https://doi.org/10.1109/TMI.2016.2529665>.

VETA, M. et al. Automatic nuclei segmentation in h&e stained breast cancer histopathology images. **PloS one**, Public Library of Science, v. 8, n. 7, p. e70221, 2013. Disponível em: <https://doi.org/10.1371/journal.pone.0070221>.

VETA, M. et al. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. In: IEEE. **2011 IEEE international symposium on biomedical imaging: from nano to macro**. 2011. p. 618–621. Disponível em: <https://doi.org/10.1109/ISBI.2011.5872483>.

WANG, C. F. A basic introduction to separable convolutions. 2018.

WANG, H. et al. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. **Journal of Medical Imaging**, International Society for Optics and Photonics, v. 1, n. 3, p. 034003, 2014. Disponível em: <https://doi.org/10.1117/1.JMI.1.3.034003>.

WANG, Y. et al. Breast cancer image classification via multi-level dual-network features and sparse multi-relation regularized learning. In: IEEE. **2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**. 2019. p. 7023–7026. Disponível em: <https://doi.org/10.1109/EMBC.2019.8857762>.

WANG, Y. et al. Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning. **IEEE Access**, IEEE, v. 8, p. 27779–27792, 2020. Disponível em: <https://doi.org/10.1109/ACCESS.2020.2964276>.

WANG, Z.; OATES, T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In: **Workshops at the twenty-ninth AAAI conference on artificial intelligence**. [S.l.: s.n.], 2015. v. 1.

WATANABE, K.; KOBAYASHI, T.; WADA, T. Semi-supervised feature transformation for tissue image classification. **PloS one**, Public Library of Science, v. 11, n. 12, 2016. Disponível em: <https://doi.org/10.1371/journal.pone.0166413>.

WINKENS, J. et al. Improved semantic segmentation for histopathology using rotation equivariant convolutional networks. 2018.

WOLFRAM, S. Wolfram research. **Inc., Mathematica, Version**, v. 8, p. 23, 2013.

XU, H. et al. Fractal dimension invariant filtering and its cnn-based implementation. In: **Proc. IEEE Conf. Comput. Vis. Pattern Recognit.** [s.n.], 2017. p. 3825–3833. Disponível em: <https://doi.org/10.1109/CVPR.2017.407>.

XU, J. et al. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. **Neurocomputing**, Elsevier, v. 191, p. 214–223, 2016. Disponível em: <https://doi.org/10.1016/j.neucom.2016.01.034>.

YAMASHITA, R. et al. Convolutional neural networks: an overview and application in radiology. **Insights into Imaging**, Springer, p. 1–19, 2018.

YANG, C.-L.; CHEN, Z.-X.; YANG, C.-Y. Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 20, n. 1, p. 168, 2020. Disponível em: <https://doi.org/10.3390/s20010168>.

YANG, X. et al. Concepts of artificial intelligence for computer-assisted drug discovery. **Chemical reviews**, ACS Publications, v. 119, n. 18, p. 10520–10594, 2019. Disponível em: <https://doi.org/10.1021/acs.chemrev.8b00728>.

YI, F. et al. Microvessel prediction in h&e stained pathology images using fully convolutional neural networks. **BMC bioinformatics**, BioMed Central, v. 19, n. 1, p. 64, 2018. Disponível em: <https://doi.org/10.1186/s12859-018-2055-z>.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: **Advances in neural information processing systems**. [S.l.: s.n.], 2014. p. 3320–3328.

YU, C. et al. Breast cancer classification in pathological images based on hybrid features. **Multimedia Tools and Applications**, Springer, p. 1–21, 2019.

ZHANG, H. et al. Recurrence plot-based approach for cardiac arrhythmia classification using inception-resnet-v2. **Frontiers in Physiology**, Frontiers Media SA, v. 12, 2021. Disponível em: <https://doi.org/10.3389/fphys.2021.648950>.

ZHANG, J. et al. Classification of digital pathological images of non-hodgkin's lymphoma subtypes based on the fusion of transfer learning and principal component analysis. **Medical Physics**, Wiley Online Library, v. 47, n. 9, p. 4241–4253, 2020. Disponível em: <https://doi.org/10.1002/mp.14357>.

ZIFF, R. M. Test of scaling exponents for percolation-cluster perimeters. **Physical review letters**, APS, v. 56, n. 6, p. 545, 1986. Disponível em: <https://doi.org/10.1103/PhysRevLett.56.545>.

APPENDIX A

Standard deviation values

The standard deviation values for the results described in Chapter 5 are presented in this appendix.

Table 30 – Standard deviation values for the results presented in Table 12.

Dataset	Sequential		Rec. Plot		GASF		GADF	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	1.96%	0.020	1.68%	0.017	1.88%	0.019	1.89%	0.019
LA	1.09%	0.011	0.94%	0.010	0.77%	0.008	0.84%	0.008
LG	1.39%	0.014	1.07%	0.011	0.89%	0.009	1.11%	0.011
NHL	1.77%	0.019	1.07%	0.011	2.39%	0.024	1.78%	0.018
UCSB	3.23%	0.032	3.79%	0.039	3.62%	0.036	4.81%	0.048
Average	1.89%	0.019	1.71%	0.018	1.91%	0.019	2.09%	0.021

Table 31 – Standard deviation values for the results presented in Table 14.

Dataset	DT		LOG		RaF		SVM		MLP	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	2.46%	0.026	1.12%	0.011	2.62%	0.027	1.81%	0.019	1.19%	0.012
LA	1.30%	0.013	0.66%	0.007	0.46%	0.004	0.27%	0.003	0.39%	0.004
LG	1.08%	0.011	0.59%	0.006	0.72%	0.007	0.44%	0.005	0.26%	0.003
NHL	1.88%	0.019	1.65%	0.017	1.18%	0.013	0.91%	0.010	1.95%	0.019
UCSB	4.24%	0.042	4.16%	0.042	2.73%	0.028	3.62%	0.039	3.68%	0.037
Average	2.19%	0.022	1.64%	0.017	1.54%	0.016	1.41%	0.015	1.49%	0.015

Table 32 – Standard deviation values for the results presented in Table 16.

Dataset	DT		LOG		RaF		SVM		MLP	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	2.70%	0.029	1.36%	0.014	1.57%	0.016	1.18%	0.012	2.19%	0.022
LA	1.13%	0.011	0.49%	0.005	0.35%	0.003	0.42%	0.004	0.47%	0.005
LG	1.74%	0.017	0.57%	0.006	0.62%	0.006	0.46%	0.005	0.36%	0.004
NHL	2.23%	0.023	1.96%	0.020	0.87%	0.009	1.27%	0.013	1.26%	0.013
UCSB	4.51%	0.045	1.78%	0.018	3.02%	0.031	3.09%	0.033	3.45%	0.036
Average	2.46%	0.025	1.23%	0.013	1.29%	0.013	1.29%	0.013	1.55%	0.016

Table 33 – Standard deviation values for the results presented in Table 18.

Dataset	Sequential		Rec. Plot		GASF		GADF	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	0.77%	0.008	0.65%	0.007	0.49%	0.005	0.51%	0.005
LA	0.43%	0.004	0.23%	0.002	0.29%	0.003	0.31%	0.003
LG	0.41%	0.004	0.58%	0.006	0.52%	0.005	0.73%	0.008
NHL	1.42%	0.014	0.92%	0.010	1.23%	0.013	0.88%	0.009
UCSB	2.47%	0.025	2.98%	0.030	3.15%	0.033	2.47%	0.025
Average	1.10%	0.011	1.07%	0.011	1.14%	0.012	0.98%	0.010

Table 34 – Standard deviation values for the results presented in Table 20.

Dataset	$\Phi + LOG(local) + ... + LOG(global)$		$\Phi + LOG(local) + ... + MLP(global)$		$\Phi + MLP(local) + ... + LOG(global)$		$\Phi + MLP(local) + ... + MLP(global)$	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	0.75%	0.008	0.63%	0.006	0.67%	0.007	0.63%	0.006
LA	0.33%	0.003	0.31%	0.003	0.36%	0.004	0.33%	0.003
LG	0.33%	0.003	0.28%	0.003	0.36%	0.004	0.39%	0.004
NHL	0.89%	0.009	0.94%	0.010	1.05%	0.011	0.88%	0.009
UCSB	1.83%	0.019	3.15%	0.033	2.16%	0.022	2.57%	0.028
Average	0.83%	0.008	1.06%	0.011	0.92%	0.010	0.96%	0.010

Table 35 – Standard deviation values for the results presented in Table 22.

Dataset	Proposed ensemble		ResNet-50		ADASYN		Data augmentation	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	0.67%	0.007	0.86%	0,009	0.49%	0,005	0.77%	0,008
LA	0.36%	0.004	0.48%	0,005	0.33%	0,003	0.88%	0,009
LG	0.36%	0.004	0.44%	0,004	0.50%	0,005	0.36%	0,004
NHL	1.05%	0.011	0.97%	0,010	1.57%	0,016	1.25%	0,013
UCSB	2.16%	0.022	4.47%	0,045	3.06%	0,031	4.51%	0,047
Average	0.92%	0.010	1.44%	0.015	1.19%	0.012	1.55%	0.016

Table 36 – Standard deviation values for the results presented in Table 25.

Dataset	Proposed ensemble		No F-CNN (LOG)		No F-CNN (MLP)		No weights	
	A_c	$F1$	A_c	$F1$	A_c	$F1$	A_c	$F1$
CR	0.67%	0.007	0.71%	0,007	0.48%	0,005	0.89%	0,009
LA	0.36%	0.004	0.20%	0,002	0.34%	0,003	0.28%	0,003
LG	0.36%	0.004	0.21%	0,002	0.32%	0,003	0.24%	0,002
NHL	1.05%	0.011	1.27%	0,013	0.85%	0,009	1.00%	0,010
UCSB	2.16%	0.022	2.45%	0,024	3.15%	0,034	2.00%	0,019
Average	0.92%	0.010	0.97%	0.010	1.03%	0.011	0.88%	0.009