University of Molise
Department of Medicine and Health Sciences
PhD Course in Clinical and Traslational Medicine

Radiomics and Formal Methods:
a dynamic duo for medicine in pixels

Doctoral Thesis

Tutor
Prof. Antonella Santone

PhD Candidate
Giulia Varriano

Coordinator of the PhD Program
Prof. Giovanni Scapagnini

Campobasso, 2024
Cycle XXXVI
SSD ING-INF/05

Ph.D. thesis reviewers

Prof. **Paul Tavolato**
Kolingasse 14-16, 1090 Wien
Austria
paul.tavolato@univie.ac.at

PostDoc. **Catarina Dinis Fernandes**
De Groene Loper 19 Flux 3.083, 5612 AZ Eindhoven
The Netherlands
c.dinis.fernandes@tue.nl

“Descrivere il passato, comprendere il presente, prevedere il futuro: questo è il compito della medicina” - Ippocrate

Resume

Questa tesi discute le tecnologie innovative che possono contribuire alla diagnosi precoce di diverse malattie. Nella nuova era della medicina personalizzata, l'assistenza sanitaria viene adattata ai singoli pazienti in base a fattori genetici, ambientali e di stile di vita per identificare la suscettibilità alle malattie, guidare le misure preventive e ottimizzare il trattamento, soprattutto in campi come l'oncologia. Le aree di interesse comprendono diverse discipline *-omiche*, tra cui l'analisi delle immagini radiologiche, chiamata *Radiomica*.
Passando all'Intelligenza Artificiale e alla Medicina, esploreremo i pro e i contro della sua applicazione in ambito sanitario. In particolare, proponiamo una nuova metodologia che utilizza i *Metodi Formali* e la Radiomica per risolvere le sfide presenti nella diagnosi delle malattie: la necessità di una grande quantità di dati, l'affidabilità della decisione finale e la complessa spiegabilità dei risultati. Vedremo la Radiomica con diverse modalità di *imaging* e poi come combinarla con i Metodi Formali per la diagnosi precoce, ottenendo prestazioni promettenti, interpretabilità dei risultati e maggiore fiducia nei sistemi esperti. La presente metodologia viene testata su varie parti anatomiche del corpo umano, tra cui cervello, seno, spalla, polmone, fegato e tessuti molli.
In sintesi, i progressi e le sfide della medicina personalizzata e l'integrazione dell'informatica possono concentrarsi sul ruolo dei metodi formali nell'affrontare alcune delle limitazioni dell'Intelligenza Artificiale. La speranza è che, con i progressi tecnologici, la Medicina Personalizzata trasformi l'assistenza sanitaria.

Parole chiave: Radiomica; Metodi formali; Imaging medico; Intelligenza Artificiale; Medicina personalizzata.

List of publications

List of the publications in Journals of the candidate

[1] Santone, A., Belfiore, M. P., Mercaldo, F., Varriano, G., and Brunese, L. (2021). On the adoption of radiomics and formal methods for covid-19 coronavirus diagnosis. *Diagnostics*, 11(2), 293.

[2] Rocca, A., Brunese, M. C., Santone, A., Avella, P., Bianco, P., ... and Brunese, L. (2021). Early diagnosis of liver metastases from colorectal cancer through CT radiomics and formal methods: a pilot study. *Journal of Clinical Medicine*, 11(1), 31.

[3] Varriano, G., Guerriero, P., Santone, A., Mercaldo, F., and Brunese, L. (2022). Explainability of radiomics through formal methods. *Computer Methods and Programs in Biomedicine*, 220, 106824.

[4] Natella, R., Varriano, G., Brunese, M. C., Zappia, M., Bruno, M., ... and Santone, A. (2023). Increasing differential diagnosis between lipoma and liposarcoma through radiomics: a narrative review. *Exploration of Targeted Anti-tumor Therapy*, 4(3), 498.

[5] Casale, R., Varriano, G., Santone, A., Messina, C., Casale, C., ... and Brunese, L. (2023). Predicting risk of metastases and recurrence in soft-tissue sarcomas via Radiomics and Formal Methods. *JAMIA open*, 6(2), ooad025.

[6] Varriano, G., Pavone, L., Santone, A., De Dominicis, M., Santella, E., ... and Brunese, L. Detecting Disease Activity from MRI Scans of Patients with Multiple Sclerosis Via Radiomics and Formal Methods. *Preprint available on: papers.ssrn.com.*

List of the publications in conferences of the candidate

[1] Varriano, G., D'Arienzo, D., Santone, A., Mercaldo, F., Valente, T., ... and De Rimini, M. L. (2022). Evaluating lung perfusion SPECT/CT imaging in patients with COVID-19 through radiomics and formal methods. *Clinical and Translational Imaging*, S97-S97.

[2] Varriano, G., Sorgente, V., Mercaldo, F., Santone, A., and Brunese, L. (2023, June). Computational cost of CT Radiomics workflow: a case study on COVID-19. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference* (COMPSAC) (pp. 1539-1544). IEEE.

List of papers still subject to submission and/or revision

[1] Varriano, G., Pavone, L., Di Muccio, L., Pasquale, F. Di Gennaro, G., ... and Brunese, L. *Radiomic imaging for the explicable study of Epilepsy by Formal Methods.*

[2] Mercaldo, F., Varriano, G., De Lucia, D., Caiazzo, C., Brunese, M.C., Santone, A., Brunese, L. *Early prediction of BI-RADS risk stratification through Radiomics and Formal Methods.*

[3] Varriano, G., Casolare, R., Mercaldo, F., Brunese, L., Santone, A. *Automatic and Explainable ROI on DICOM Imaging: a pilot study for COVID-19.*

[4] Varriano G., Cerminara A., Nardone V., Mercaldo, F., Brunese, L., Santone, A. *Radiomic assessment of capsulitis: Exploiting formal methods for precision diagnosis.*

[5] Rocca A., Brunese M.C., Bellifemine F., De Lucia D., Pacella G., Brunese, L., Santone, A. *Automated diagnosis of mild acute pancreatitis through Radiomics and Formal Methods.*

Abstract

This thesis discusses innovative technologies that can contribute to the early diagnosis of different diseases. In the new era of Personalized Medicine, healthcare is tailored to individual patients based on genetic, environmental and lifestyle factors to identify disease susceptibility, guide preventive measures and optimize treatment, especially in fields like oncology. The focus areas include several *-omics* disciplines, including the quantitative analysis of radiological images, called *Radiomics*.
Shifting on to Artificial Intelligence and Medicine, we will explore the pros and cons of its application in healthcare. In particular, we propose a new methodology using *Formal Methods* and Radiomics to solve the challenges present in disease diagnosis; the need for a large amount of data, the trustworthiness of the final decision and the complex explainability of the results. We will see Radiomics with different imaging modalities and then how to combine it with Formal Methods for early diagnosis, achieving promising performance, interpretability of results and increased confidence in expert systems. The present methodology is tested on various anatomical parts of the human body, including the brain, breast, shoulder, lung, liver and soft tissues.
In summary, the advancements and challenges in Personalized Medicine and the integration of Computer Science can focus on the role of Formal Methods in addressing some of the limitations of Artificial Intelligence. The hope is that, with technological advancements, Personalized Medicine will transform healthcare.

Keywords: Radiomics; Formal Methods; Medical Imaging; Artificial Intelligence; Personalized Medicine.

Contents

Resume **1**

List of publications **3**

Abstract **5**

1 Introduction **11**
1.1 Era of the Personalized Medicine 11
1.2 Artificial Intelligence and Medicine 12

2 Background **17**
2.1 Radiomics 17
2.1.1 Standards and software 20
2.1.2 PyRadiomics 21
2.2 Formal Methods 24
2.2.1 The Model Checking technique 25
2.3 Performance metrics 27

3 Methodology **29**
3.1 Image Segmentation 29
3.2 Feature Extraction 30
3.3 Feature Selection 31
3.4 Discretization 32
3.5 Formal Models 34
3.5.1 Formal Methods for single image 35
3.6 Formal Verification 38

3.7 Explainability of Radiomics through Formal Methods . . . 39
3.8 Related Work and Discussion . . . 42

4 Brain **43**
4.1 Detecting Disease Activity from MRI Scans of Patients with Multiple Sclerosis Via Radiomics and Formal Methods . . . 43
4.2 Radiomic imaging for the explicable study of Epilepsy by Formal Methods 48
4.3 Related Works and Discussion . . . 58

5 Breast **63**
5.1 Early prediction of BI-RADS risk stratification through Radiomics and Formal Methods . . . 63
5.2 Related Works and Discussion . . . 67

6 Shoulder **69**
6.1 Early Diagnosis of Adhesive Capsulitis: An Approach Leveraging Model Checking for Enhanced Precision . . . 69
6.2 Radiomics vs. Radiologists . . . 72
6.3 Related Works and Discussion . . . 73

7 Lungs and COVID-19 disease **75**
7.1 On the Adoption of Radiomics and Formal Methods for COVID-19 Coronavirus Diagnosis . . . 75
7.2 Computational cost of CT Radiomics workflow: a case study on COVID-19 . . . 80
7.3 Evaluating lung perfusion SPECT/CT imaging in patients with COVID-19 through Radiomics and Formal Methods . . . 84
7.4 Explainability in COVID-19 classification . . . 86
7.4.1 Differences in values distributions . . . 86
7.4.2 Explainability of the Minimum feature in this case study . . . 90
7.4.3 Explainable patterns . . . 92
7.4.4 Localization of the most important slices . . . 94
7.5 An Automatic Radiomic-Based Approach for Disease Localization: A Pilot Study on COVID-19 . . . 96
7.6 Related Works and Discussion . . . 102

8 Liver **109**
8.1 Early diagnosis of liver metastases from colorectal cancer through CT Radiomics and Formal Methods: a pilot study . . . 109
8.2 Related Works and Discussion . . . 114

9 Soft Tissue **115**
9.1 Increasing differential diagnosis between lipoma and liposarcoma through Radiomics: a narrative review . . . 115
9.1.1 Radiomics vs. Radiologists . . . 116
9.2 Predicting risk of metastases and recurrence in soft-tissue sarcomas via Radiomics and Formal Methods . . . 120
9.2.1 Statistical analysis and property verification . . . 122
9.3 Related Works and Discussion . . . 126

10 Discussion **127**
10.1 Never without Artificial Intelligence . . . 127
10.2 Limits of Radiomics . . . 128
10.3 Images are treasures . . . 129

11 Conclusion **131**

Acknowledgements **133**

References **136**

A Supplementary Material **157**
A.1 Epilepsy: results without filter . . . 157
A.2 Epilepsy: results on filtered images . . . 161

List of figures **167**

List of tables **171**

Chapter 1

Introduction

1.1 Era of the Personalized Medicine

Personalized medicine (or *precision medicine*) is an innovative approach to medical treatment and healthcare. It tailors medical decisions, practices, interventions and therapies to the individual patient based on his/her genetic, environmental and lifestyle characteristics. The goal of Personalized Medicine is to maximise the effectiveness of treatments while minimising potential side effects or harm. This added value makes it possible to provide more accurate and targeted healthcare for each patient and to save the resources of the national health system.
Personalized medicine is developing in several areas, one of which is *Genomics*. It stands for the study of a person's genetic composition, in particular through techniques such as DNA sequencing. Genetic information can be used to identify the risk of disease, to choose the most suitable treatments and thus predict the response to drugs (*Pharmacogenomics*). Healthcare professionals can prescribe drugs that are more likely to be effective and less likely to cause adverse reactions.
For nuclear medicine there is the *Dosiomics*, which is the measurement and assessment of radiation doses. It is critical because it ensures that the appropriate amount of radiation is delivered for therapeutic purposes, monitoring and minimising the radiation exposure of people in medical and industrial settings.
In the field of radiology, there is *Radiomics*, which involves the extraction and analysis of a large number of quantitative features from radiological images. These features include textural, shape and intensity metrics that go beyond what is visible to the

naked eye and provide detailed information on the spatial distribution, heterogeneity and other characteristics of the tissues or lesions being examined.
Personalized medicine can help identify an individual's susceptibility to certain diseases, which allows for targeted preventive measures such as early screening of cancer or monitoring of chronic diseases. For example, significant progress has been made in the field of oncology, characterising genetic mutations and biomarkers specific to a patient's tumour to determine the most effective therapies, such as immunotherapies.

In the realm of Personalized Medicine, sophisticated diagnostic tools are utilized to evaluate an individual's unique health status. These tools encompass genetic testing, the analysis of biomarkers, and various diagnostic methods to pinpoint diseases and uncover their root causes.
Almost every field of medicine has its corresponding *-omics* for the study of data related to Personalized Medicine. For example, *Immunomics* is an emerging field of biomedical research that focuses on the comprehensive study of the immune system and its interactions with various components of the human body (including genetics, genomics and *proteomics*), in order to gain a deeper understanding of the immune response and immune-related diseases for the development of Personalized immunotherapies.

In all fields of Personalized medicine, the management and analysis of large amounts of patient data is crucial. A first step has been taken with the implementation of Electronic Health Records (EHR) for collecting, storing and interpreting patient information to make informed medical decisions. For these reason, Medicine and Computer Science are strictly linked in a new multidiciplinary collaboration.
Personalized medicine can improve patients' clinical outcomes by optimising the use of health system resources. However, it also presents challenges related to data privacy, ethical considerations and the need for advanced infrastructure for data storage and analysis. The hope is that, in the near future, as technology and scientific knowledge advances, Personalized Medicine can transform the field of healthcare.

1.2 Artificial Intelligence and Medicine

Artificial intelligence (AI) was introduced in the medical sector in the mid-20th century, but only in recent decades has it gained considerable importance. Early applications focused on symbolic reasoning and rule-based systems: for example, Dendral is a program developed in the 1960s that marked the field of chemistry and is the precursor of diagnostics [1]. These are called "*expert systems*", which are knowledge-

based systems developed to assist healthcare professionals in diagnosis and decision-making. Different papers started to appear in the literature; think of MYCIN, the first expert system to diagnose bacterial infections and recommend antibiotic treatments [2]. Not all the community was convinced of these new innovative techniques, raising several problems about their use [3].
It is in the 2000s that Machine Learning (ML) algorithms have a rapid expansion: Neural networks and Decision trees begin to be applied in radiology and pathology. 2010 was the most significant year for AI in healthcare: Deep learning (DL), a subset of ML, shows outstanding results in medical imaging activities. Therefore, thanks to the amount of data collected in the scientific community, AI emerges in electronic health records, drug discovery, genomics and Personalized Medicine.

Currently, AI is widely integrated into various aspects of health care and often used in the daily clinical practice. It is used for the diagnosis of diseases, the prediction of patient outcomes, the discovery of drugs, the treatment of patients. However, the use of AI poses several problems [4], such as:

- needing of *large and high-quality data sets* for training and for validation, while in the healthcare industry, data can be fragmented and incomplete;
- ensuring data *privacy and security* and at the same time the sharing of medical data, is a major challenge;
- achieving *interoperability* between different health sectors and between different diagnostic or prognosis techniques;
- giving *interpretability* and transparency are not always easy to access in AI models because of the complex structure of models. On the contrary, in the healthcare system it is important to gain confidence and make decisions more understandable;
- determining *liability* in hazard: whenever mistakes or negative consequences occur, establishing fault is difficult.

Some of these limitations can be resolved by mathematical methods, i.e. transforming data into axioms and functions, guaranteeing the robustness of the decision and independence towards the case study dataset.
Among the most common mathematical methods in Computer Science are Formal Methods. These are based on a temporal logic that makes it possible to verify the correctness of a computer system, provided it satisfies certain properties established

by the requirements. The combination of Formal Methods and Radiomics gives rise to this thesis, in which an innovative methodology of disease diagnosis applicable to various fields is developed.

The background and the methodology will be discussed in Chapters 2 and 3. In particular, the thesis is structured by scanning different parts of the patient's body, as shown in Figure 1.1: starting with the brain, we can talk about Epilepsy and Multiple Sclerosis in Chapter 4. In the next Chapter 5 it will be analysed a very common disease, namely Breast Cancer. Consequently, in Chapter 6 we will look at Adhesive Capsulitis, which causes shoulder pain and is difficult to diagnose. At this point we will be in the lung part, in Chapter 7, where the methodology will be applied to the classification of COVID-19 patients and the recognition of Perfusion Deficit. Important results were found in the early prediction of Liver metastases from Colorectal Cancer, as explained in Chapter 8. For Soft Tissues, on the other hand, we will study sarcomas via a public database using this new methodology (Chapter 9).

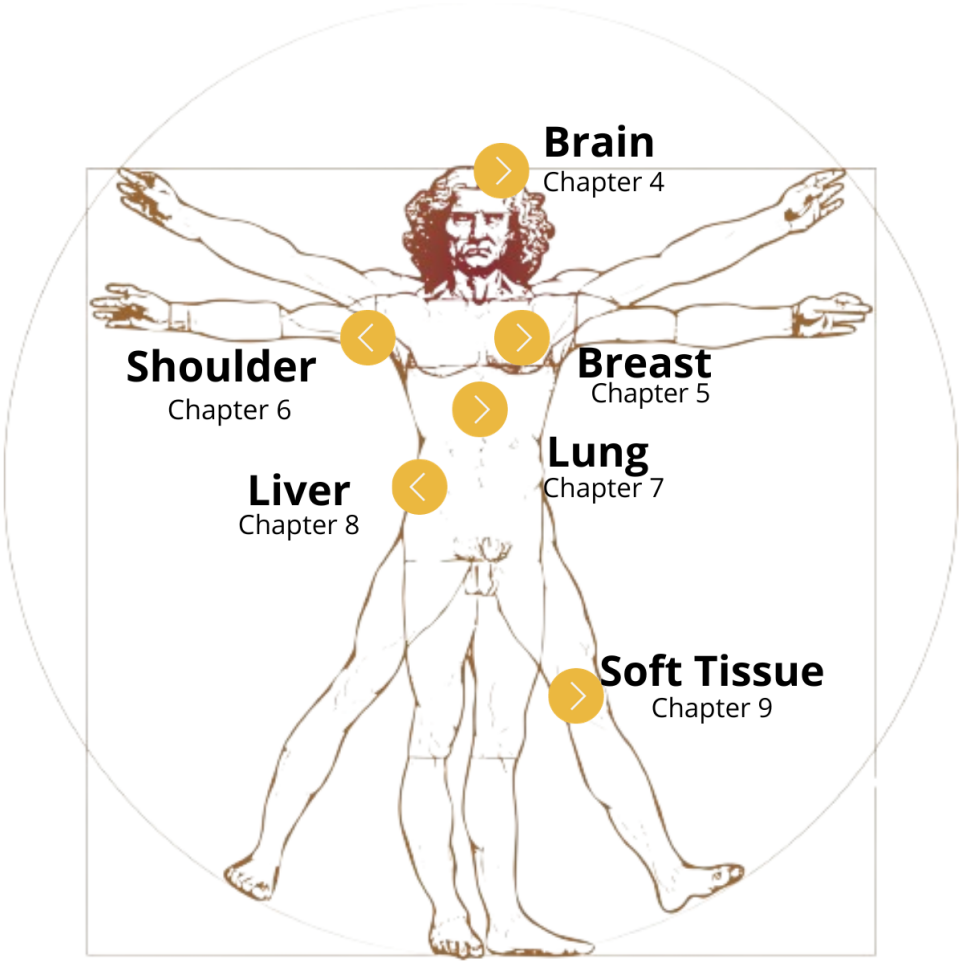


Figure 1.1: Structure of this thesis, where each Chapter is associated with an anatomical part of the human body to which the methodology has been applied.

Chapter 2

Background

2.1 Radiomics

In the first decade of the 2000s, a first application of the Personalized Medicine mentioned by Hippocrates as early as 460 BC was finally introduced. Furthermore, important discoveries were made after the possibility of sequencing the human genome [5]. The "-*omics*" disciplines were born starting with the need in molecular biology for a detailed description that would provide all useful information. This movement spread at incredible speed to all fields of medicine, moving from the study of genes, *Genomics*, to nuclear medicine, *Dosiomics* and finally to Radiology, *Radiomics*. This suffix indicates the handling of complex, multidimensional data [6].
Radiomics is the application of sophisticated techniques for the management and analysis of large digital images. We are talking about images from radiology scanners such as Computational Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US), but also from nuclear medicine such as Single Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography (PET).
The first to speak about Radiomics was R. Gillies [7], who recognised the importance of images not only as medical records, but as large data containers. These in fact, can accompany the analysis and diagnosis, making it more complete and robust.
The first software for this purpose in medicine were the Computer-Aided Diagnosis and Detection (CAD) systems, stand-alone processes that allow an answer to a problem [8]. Unfortunately, their use is limited to the single response of diagnosis or lesion identification; with Radiomics, on the other hand, the same data can be exploited at

the same time to execute multiple analysis, allowing the data-share.

This new discipline was conceived precisely as a development of CAD because it can be exploited standalone or, more powerfully, in combination with other data. This gives rise to new fields of research and application, such as *Radiogenomics* [9]: in this case, the fusion of Radiomics with Genomics data succeeds in putting Personalized Medicine into practice. Generally, radiomic data are expressed in relation to a tumour or lesion, which has its own genetic expression and may undergo mutation during a therapeutic treatment. Also in clinical practice, for example, there is an immediate advantage in characterising the consistency of the tumour detected to assess its malignancy or benignity.

The *practical process* of Radiomics involves several steps:

- acquisition of images;
- pre or post-processing of images (if required);
- identification of the volume of interest, from which data is to be extracted (e.g. the tumour/lesion);
- segmentation of the volume of interest using graphic tools;
- data extraction from medical images;
- feature selection or dimensionality reduction of data (if required);
- creation of useful models for classification.

Processing of images
To these steps can be added other activities for refining the analysis. For example, after image acquisition, image adjustment (in spatial terms, with *Resampling*) may be performed; *Filtering*, i.e. the implementation of filters on the image in order to reduce noise or highlight anatomical features, may also be performed.
In addition, before creating the model for the classification, a selection of data can be made to create successful predictive models in order to eliminate information redundancy and speed up performance.

Segmentation
The most complicated part of the radiomic process is the segmentation of the volume

of interest, also called *Region Of Interest* (ROI) or Volume of Interest (VOI). In fact, the contours of a tumour can be jagged, non-linear and therefore it is difficult to be accurate in manual tracking. Moreover, this step must be repeated for the entire database, which implies a high computational and human effort.
Manual segmentation by experienced readers is often considered the ground truth. However, this methodology suffers from high variability among readers and is labour intensive. Indeed, a ROI drawn by a doctor with 10 years of experience will be different from one drawn by a novice doctor (*inter-operator variability*); a ROI drawn on one patient will probably be different from one drawn on another patient (*intra-operator variability*) [10].
In conclusion, if manual segmentation is performed, it should be noted that this is operator-dependent, a variable that could led to bias during the classification. This is why some work on automatic segmentation has arrived in the literature, but without achieving strong effects in the real world activities [11, 12]. It can be said that the community has moved towards automatic segmentation but with manual validation [13] of medical experts. There is not a unique segmentation algorithm for all medical image applications: each segmentation, with appropriate parameter settings, can automatically or semi-automatically delineate the ROI and have different outcomes. Thus, it's crucial to establish standardized metrics for assessing segmentation algorithms.

Features extraction
The data extracted in this activity are called *radiomic features* and there are different types. The *semantic features* are those commonly used in the radiological lexicon (Size, Shape, Location, Vascularity, Spiculation, Necrosis) and consequently are easily linked to radiological evidences. In contrast, *agnostic* features are calculated using mathematical formulas based on information in the image, such as its histogram or matrices constructed based on the spatial relationship of grey levels in pixels. These are Haralick features [14], Laws [15], Wavelets [16], Laplace transforms, Minkowski functionals and Fractals dimensions.
In the category of agnostics, one can find first-, second- or higher-order features. *FIRST-Order features* are based on the histogram, i.e. the grey levels in the image without taking into account their spatial relationship. *Second-order features*, on the other hand, study the texture of the image using *voxels*, which are the three-dimensional version of pixels. Thus, similarity relationships between neighbouring voxels can be described. *Higher-order features*, on the other hand, apply image fil-

ters, such as fractals, to recognise patterns in the image.

All the advantages of Radiomics are, of course, accompanied by a number of disadvantages. Starting with the daily routine, it is sometimes difficult to make associations between the extracted data and the clinical and radiological evidences and this is a problem if we want to understand in the deep the automatic diagnosis.
Other limitations may be in the acquisition of data, as these are dependent on the type of scanner used in the hospital and the parameters set for the examination.

Feature selection
After completing the feature extraction, the features are utilized in a model to address clinical issues, like distinguishing between benign and malignant lesions/tumors. How do we determine which features are the most significant if there is no link to radiological evidence? In theory, all the extracted features serves as input for a predictive model, but this approach exponentially increases the model's size. Hence, a substantial number of candidate features need elimination or transformation through the *Feature selection* or *Dimensionality reduction*.
Radiomic features frequently display high correlations, indicating data redundancy, which allows for the discarding of some features and the grouping of others, replacing them with representative features, achieved through methods like Principal Component Analysis or Linear Discriminant Analysis. Within these representative features, those displaying the most informative characteristics take precedence and are considered most crucial. Parmar et al. [17] conducted a comparison of 14 distinct radiomic feature selection techniques, encompassing methods based on mutual information and 12 ML classifiers.

2.1.1 Standards and software

Most radiological images are stored in a special format, suitable for viewing, handling and transferring medical images: this main format is called *Digital Imaging and COmmunication in Medicine* (DICOM). This standard is deeply accurate: for example, it guarantees the same brightness variation on different devices with the same digital values, at 300 DPI and with 12 bit of gray levels (8 for normal digital devices) [18].
DICOM format is based on object-oriented programming, with a header and different tags for patients, radiological exam, scanner and printer, which together contribute to radiological final information. Medical images are represented by a series of tags, including the pixel value tag "image pixel module"; they can be compressed both in loseless or lossy way and can be visualized only with a DICOM viewer software that

is able to read all the metadata of the image. For a single exam, we will have a lot of DICOM (even 500 images) that create the radiological examination and each single image is called *slice*.

Very often, in addition to DICOM, one can find oneself working with NIFTI or NRRD files. NIFTI standard (*NeuroImaging Informatics Technology Initiative*) was introduced for brain images in neuroimaging to store new information, like the image orientation, which is commonly subject to changes in brain studies. This is done through voxel coordinates or linear transformations which describe the alignment of the volume according to a template-based system [19].

The NRRD format (*Nearly Raw Raster Data*) is suitable for the visualization and storage of N-dimensional raster data. It is a text file with two headers and different field about thicknesses, general space and orientation information.

With the advent of Radiomics, thousands of features have been developed to calculate radiomic parameters. In 2018, an international collaboration introduced the *Image Biomarkers Standardization Initiative* (IBSI) to create a radiomic feature extraction standard. In particular, they described the biomarker as "*a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention*" [20]. They established a comprehensive Radiomic Workflow, a formal definition of each radiomic feature and its extraction. The IBSI standard is implemented in a Python library called PyRadiomics [21]. As a graphical tool, PyRadiomics is included in the open-source software 3D Slicer, which allows visualisation and management of DICOM, ROI segmentation, feature extraction and so on [22].

2.1.2 PyRadiomics

"PyRadiomics" is a term that combines "Python" and "Radiomics": it is a library enabling the extraction and examination of various features from medical images like CT scans or MRI scans in 2D (pixel-based) or 3D (voxel-based) [21]. It is compliant to the IBSI standard, which hope to increase awareness of radiomic capabilities and expand the community.
Within it, there are FIRST-Order, second-order and high-order features. In particular, there are 7 *feature classes*, distinguished as follows:

- FIRST, which stands for *"FIRST-Order statistics"*. It describes the distribution of voxel intensities within the ROI (or mask) through commonly used and basic metrics like Mean, Median, Entropy;

- SHAPE features, which are high-order features, because they did not depend by the gray levels of the image, but only by the spatial relationship of the pixels/voxels. Features are computed on the approximated shape of the ROI defined by a triangular mesh generated through the application of a *Marching Cubes Algorithm* [23];
- GLCM, which stays for "*Gray Level Co-occurence Matrix*". It is the most known and used matrix and it calculates the number of times the combination of two levels i and j occurs in two adjacent pixels in the image, as shown in Figure 2.4;

1	0	2
2	1	1
1	0	0

Table 2.1: 3x3 image

	0	**1**	**2**
0	1	0	1
1	2	1	0
2	0	1	0

Table 2.2: GLCM Matrix

	0	**1**	**2**
0	0.16	0	0.16
1	0.33	0.16	0
2	0	0.16	0

Table 2.3: Normalized GLCM

Table 2.4: Example of how GLCM Matrix with distance $\delta = 2$ and $\sigma = 0°$ is built.

- GLDM is the "*Gray Level Dependence Matrix*" that quantifies gray level dependencies in the image. A gray level dependency is defined as the number of connected voxels within distance δ that are dependent on the center voxel;
- GLRLM is the "*Gray Level Run Length Matrix*" and it quantifies gray level runs. *Runs* are defined as the length, in number of pixels, of consecutive pixels that have the same gray level value;
- GLSZM is a "*Gray Level Size Zone*" quantifies gray level zones in an image. A *gray level zone* is defined as a the number of connected voxels that share the same gray level intensity. A voxel is considered connected if the distance is 1, according to the infinity norm;
- NGTDM, called "*Neighbouring Gray Tone Difference Matrix*" quantifies the difference between a gray value and the average gray value of its neighbours within distance δ. The sum of absolute differences for gray level i is stored in the matrix.

For each of these matrices, the standard distance measure is the Euclidean, but there are also other choices, like Manhattan distance or Infinity norm [24]. By default, the calculated GLCM is symmetrical. The list of all features contained in the seven feature classes is showed in table 2.5.

FIRST	10 Percentile; 90 Percentile; Energy; Entropy; Interquartile Range; Kurtosis; Maximum; Mean Absolute Deviation; Mean; Median; Minimum; Range; Robust Mean Absolute Deviation; Root Mean Squared; Skewness; Total Energy; Uniformity; Variance.
SHAPE 2D or 3D	Mesh surface; Pixel surface; Perimeter; Perimeter to Surface ratio; Sphericity; Spherical Disproportion; Maximum 2D diameter; Major Axis Length; Minor Axis Length; Elongation.
GLCM	Autocorrelation; Joint Average; Cluster Prominence; Cluster Shade; Cluster Tendency; Contrast; Correlation; Difference Average; Difference Entropy; Difference Variance; Joint Energy; Joint Entropy; Informational Measure of Correlation 1; Informational Measure of Correlation 2; Inverse Difference Moment; Maximal Correlation Coefficient; Inverse Difference Moment Normalized; Inverse Difference; Inverse Difference Normalized; Inverse Variance; Maximum Probability; Sum Average; Sum Entropy; Sum of squares.
GLDM	Small Dependence Emphasis; Large Dependence Emphasis; Gray Level Non-Uniformity; Dependence Non-Uniformity; Dependence Non-Uniformity Normalized; Gray Level Variance; Dependence Variance; Dependence Entropy; Low Gray Level Emphasis; High Gray Level Emphasis; Small Dependence Low Gray Level Emphasis; Small Dependence High Gray Level Emphasis; Large Dependence Low Gray Level Emphasis; Large Dependence High Gray Level Emphasis.
GLSZM	Small Area Emphasis; Large Area Emphasis; Gray Level Non-Uniformity; Gray Level Non-Uniformity Normalized; Size-Zone Non-Uniformity; Size-Zone Non-Uniformity Normalized; Zone Percentage; Gray Level Variance; Zone Variance; Zone Entropy; Low Gray Level Zone Emphasis; High Gray Level Zone Emphasis; Small Area Low Gray Level Emphasis; Small Area High Gray Level Emphasis; Large Area Low Gray Level Emphasis; Large Area High Gray Level Emphasis.
GLRLM	Short Run Emphasis; Long Run Emphasis; Gray Level Non-Uniformity; Gray Level Non-Uniformity Normalized; Run Length Non-Uniformity; Run Length Non-Uniformity Normalized; Run Percentage; Gray Level Variance; Run Variance; Run Entropy; Low Gray Level Run Emphasis; High Gray Level Run Emphasis; Short Run Low Gray Level Emphasis; Short Run High Gray Level Emphasis; Long Run Low Gray Level Emphasis; Long Run High Gray Level Emphasis.
NGTDM	Coarseness; Contrast; Busyness; Complexity; Strength.

Table 2.5: List of the radiomic features of the different classes available in PyRadiomics [21].

2.2 Formal Methods

A model is a simplified version of reality. We use models every day and in all fields: just think of a geographical map, a graphic, a biomedical project reproducing an artificial arm. The efficacy of the model is not depending by its value, but by the selection of the details included: the fewer superfluous details there are, the higher the quality of the model is.
A model can be expressed in different forms, verbal, written or by means of diagrams. *Formal Methods* allow a simplified representation of the connections of a system in a mathematical syntax [25]. They are often used in the area of critical systems, where there is a relative number of agents who have high stakes and who knows all the use cases of the system. In fact, this type of methods are used in software and hardware analysis to test if the implementation is aligned with the specification and vice versa [26, 27].
Specification is the process of describing the functional and non-functional requirements of a system that will be implemented. Formal Specification needs a defined syntax and semantics: one of the most used is the *Calculus of Communicating Systems* (CCS), proposed by Milner et. al. [28].

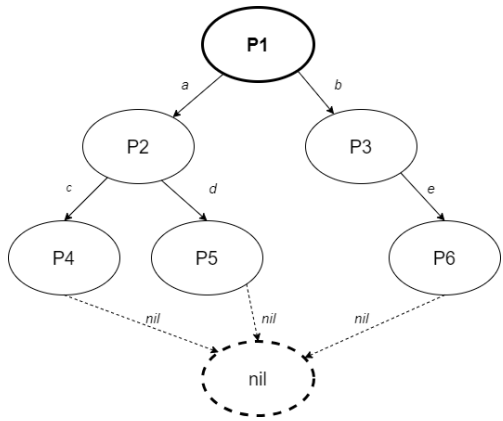


Figure 2.1: Graphical representation of a system behaviour. Circles are called *states* and are reachable only through the execution of *actions* (alphabet letters). $P1$ is the initial state, while *nil* is the state of termination.

The graphical representation of a system's behaviour is called *Labelled Transition System* (LTS) and it consists of states and transitions. The LTS begins from an initial state, serving as a predefined starting point; then, are illustrated the possible

movements from one state (referred to as "$P1$") to another state (designated as "$P2$") through a specific action (referred to as "a"). In the context of the CCS, a range of operators and an operational semantic are combined to automatically construct the LTS model of a system. This operational syntax has a great modelling power for the description of a system establishing axioms and transition rules. In Table 2.6 a subset of CCS language operators is illustrated, i.e. those exploited for modelling the case studies illustrated in the following chapters.

Operator	**Meaning**
proc $P1$	It establishes a new *Process*, namely a new state of the LTS, where $P1$ is the name of the state.
$a.P2$	It is the real definition of the system's behaviour: the system does an action a and then goes in the state $P2$. It is called "*Prefix*".
$a.P2 + b.P3$	The plus operator is called "*Choice*" and it represents the possibility of choosing whether to do an action a and then go to state $P2$ or to do b going to state $P3$.
$a.P2 \mid b.P2$	This operator gives the possibility of doing a and b at the same time, through the *interleaving*. It is called "*Parallel*".
nil	The *nil* operator is the unique that bring the LTS to a state of termination.

Table 2.6: Subset of CCS operators exploited in the following chapters.

2.2.1 The Model Checking technique

Model Checker is the agent that can test if the system respect the initial specification described in the LTS [29]. This is possible thanks to logical temporal properties that contain the behaviours that must be implemented to have a coherent specification.
Mu-calculus is the logic used in this thesis to check if the property is satisfied in a temporal and recursive way [30]. It exploits fixed-point operators are $\nu Z.\phi$ and $\mu Z.\phi$, where $\mu Z(\nu Z)$ binds free occurrences of Z in ϕ. The least fix-point of the recursive equation $Z = \phi$ is $\mu Z.\phi$, while $\nu Z.\phi$ is the greatest one. An occurrence of Z is free if it is not within the scope of a binder $\mu Z(\nu Z)$; a formula is closed if it do not contains free variables.
The syntax must reflect the logical operators used in the formal models and then other operators are added: below, there is the set of operations used to write a logical

property, where K is a set of actions.

$$\phi \quad ::= \quad \texttt{tt} \mid \texttt{ff} \mid Z \mid \phi \vee \phi \mid \phi \wedge \phi \mid [K]\phi \mid \ < K > \phi \mid \nu Z.\phi \mid \mu Z.\phi$$

The Model Checking problem is "given a formula ϕ and a LTS M and its generic state s, it must decide if $M \models \phi$". A LTS M satisfies a formula ϕ, denoted $M \models \phi$, if and only if $s_0 \models \phi$, where s_0 is the initial state of M. For the generic state s, the satisfaction of the formula ϕ is so defined:

- each state satisfies tt and no state satisfies ff;
- a state satisfies $\phi \wedge \psi$ (or $\phi \vee \psi$) if it satisfies ϕ and (or) ψ;
- $[K]\phi$ is a modal operator and is satisfied by a state which evolves in another state that meets ϕ for each performance of an action in K;
- $< K > \phi$ is a modal operator and is satisfied if exists a state which can evolve to another state that meets ϕ by performing an action in K.

Assuming that M is finite, the system will have a finite set of states, guaranteeing that computations will eventually get a conclusion. Nevertheless, with this method is impossible to discern if a transition includes an internal loop or if two distinct actions originate from the same state. Furthermore, it lacks the capacity to deduce an action count from a specific state or utilize quantifiers to evaluate the validity of a formula across all states. Fortunately, this logic offers simplicity, robust expressive capabilities and algorithmic properties, making it a valuable proposition for the verification of programs.
This checking process is possible under the Concurrency Workbench of the New Century (CWB-NC), a tool that can analyse the property on a finite state automaton [31].

The Model Checker agent evaluates whether the patient's model matches the model representing the disease. A possible formula can be, for example:

$$prop\ F_1 = (min\ X = \langle \alpha_1 \rangle \langle \alpha_2 \rangle \cdots \langle \alpha_n \rangle F_2 \vee \langle - \rangle X)$$

It tells the Model Checker to find the pattern $\gamma : \alpha_1 \alpha_2 \cdots \alpha_n$ followed by the pattern specified in F_2. The pattern γ could be present starting from the initial state of the LTS or starting from an intermediate state, possible thanks to the recursive operator, i.e., $\vee < - > X$.

2.3 Performance metrics

Different metrics allow to assess the classifier performance: one example is counting its *false positive* (all the true instances that are false in real world) and its *false negative* (all the false instances that are true in real world). Those instances can be collected in a *Confusion Matrix*, which explain how many correct classification are done and they can be used to calculate some performance indexes.
Accuracy measures the performances on the overall dataset, considering the correct instances on the whole number of patients (Equation 2.1).
In particular, *Precision* (Equation 2.2) (also called *positive predictive value*) is the fraction of relevant instances among the retrieved instances.
Recall (Equation 2.3) is the fraction of relevant instances that were retrieved. It is also called *Sensitivity* in the medical field. Both Precision and Recall are, therefore, based on relevance.
In the medical field is also used a metric called *Specificity*: it represents the likelihood of obtaining a negative test result when considering that the individual is really negative (Equation 2.4).

$$Accuracy = \frac{(truepositive + truenegative)}{(truepositive + truenegative + falsepositive + falsenegative)} \tag{2.1}$$

$$Precision = \frac{truepositive}{(truepositive + falsepositive)} \tag{2.2}$$

$$Recall = \frac{truepositive}{(truepositive + falsenegative)} \tag{2.3}$$

$$Specificity = \frac{truenegative}{(truenegative + falsepositive)} \tag{2.4}$$

Furthermore, the *Clinical Utility Indexes* (CUI) were calculated to take into account both occurrence and discrimination, as stated by Mitchell [32]. The value for CUI ranges from 0 to 1 and can be of:

- excellent utility (CUI $\geq$ 0.81);
- good utility (CUI $\geq$0.64);
- satisfactory/fair utility (CUI $\geq$ 0.49);

- poor utility (CUI < 0.49);
- very poor utility (CUI ≤ 0.36).

Chapter 3

Methodology

As part of this thesis, Formal Methods were integrated into the radiomic workflow for model creation and verification. An example of the union between medical and computer engineers, through Radiomics and Formal Methods, is provided in the Figure 3.2: each step of the diagram will be discussed in the following sections.

3.1 Image Segmentation

Image segmentation is a computer vision activity that involves the subdivision of an image into homogeneous parts, making analysis and understanding easier. Each segment typically represents regions with similar characteristics, such as colour. Think of the MRI of the thorax, in which the lungs and respiratory activity are seen: in this case, one can manually segment by colour (whiter for bone, darker for air) or automatically through a threshold (from *min_value* to *max_value*) or other edge-based methods.

Concerning the segmentation used in this thesis, my team and I state that the best ROI is the one that does not exclude any detail. This is done for brain, breast, shoulder and lungs. In some applications fields, such as liver (Chapter 8) and soft tissues (Section 9.2), it was necessary to apply the manual segmentation around the organ of interest from medical experts. Think of an MRI of the abdomen where to segment the liver: in a slice where there are several organs (liver, pancreas, gallbladder, etc.), the segmentation is difficult and hence almost necessarily manual.

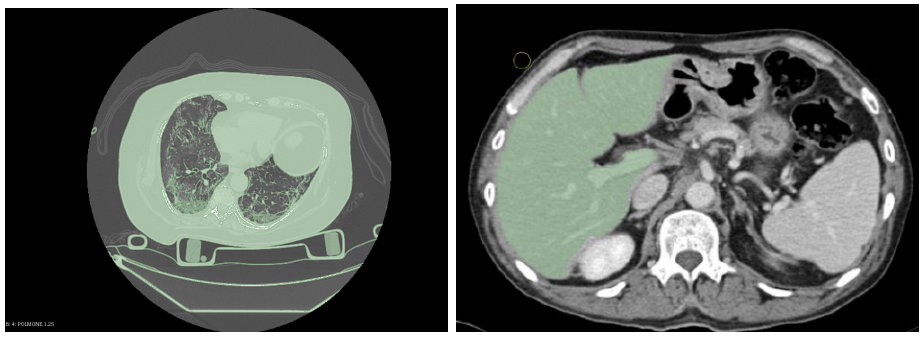


Figure 3.1: Two different type of segmentation. On the left an automatic segmentation based on threshold; on the right, a manual segmentation on the liver using 3DSlicer tool.

As illustrated on the left of Figure 3.1, the segmentation is automatic on the entire image using a threshold method on the grey value of the image: in particular, the segmentation is done taking all pixels between -600 and 600.
Instead, as on the right side in Figure 3.1, the segmentation is done manually by medical specialists accurately on the liver, excluding vena cava and other anatomical parts that can affect the analysis. To segment, in most cases an external graphics tool is required: in this thesis, *3DSlicer* [22] was always used, which includes various methods of both manual and automatic segmentation.

3.2 Feature Extraction

Feature extraction is the heart of the radiomic process. For each Chapter and then for each case study, the medical specialists with whom my team collaborated provided us with a dataset of patients undergoing medical imaging.
For all analyses, my team and I worked with DICOM images and NRRD masks: this allows to have hundreds of images per patient that can be modelled using Formal Methods. During the development of this thesis, we were also faced with examinations such as US, in which there is only one image. In that case, the formal model must be changed to allow communication between processes (because there are not slices). This second methodology is explored in more detail in Section 3.5.1.

From the ROIs segmented on the image, all the radiomic feature classes contained in the PyRadiomics library [21] are extracted. In almost all Chapters, with the extraction, each patient has multiple radiomic models, i.e. one per radiomic class of the IBSI standard (those of Section 2.1.2). Each patient is evaluated on all radiomic

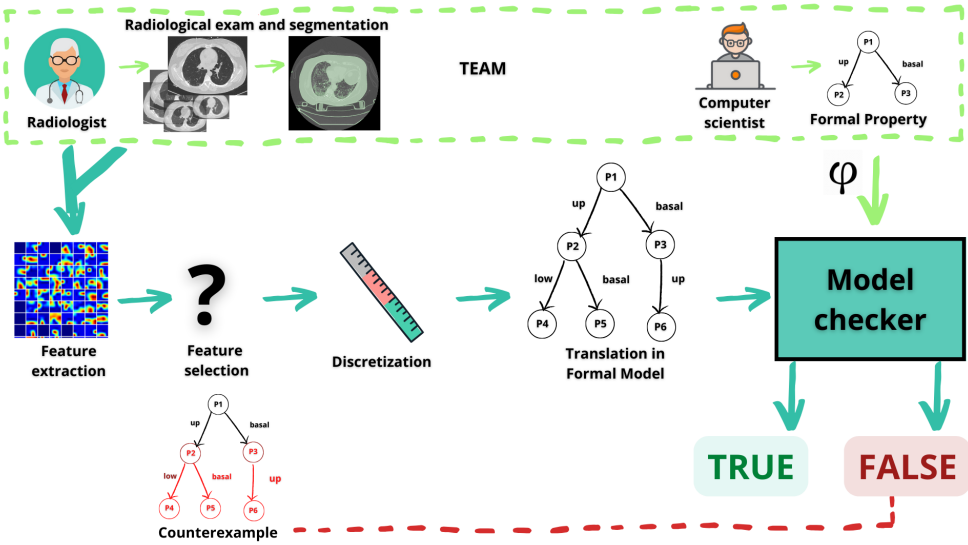


Figure 3.2: Methodology presented in this thesis, including collaboration between doctors and computer scientists, with image analysis through to modelling and results that become explainable thanks to the counterexample.

classes, but features are not mixed between different radiomic classes.

The methodology presented in this thesis is multidisciplinary, as it can include computer scientists and medical specialists. Bringing them together makes it possible to collect cases, analyse them and establish the formal property useful for classification. In the analysis part, the Feature Extraction and Selection process, the discretization into different levels and the creation of the formal model are carried out. Only as a last resort, the Model Checker allows the rule to be checked on the models and generates the classification. It is also possible to access the explanation of the decision by means of a counterexample.

3.3 Feature Selection

Here the first change in the workflow takes place. The integration of Formal Methods involves solving a mathematical problem that includes the combination of several features: this means that the problem must contain only a few features in order to achieve convergence. For the Feature Selection process, therefore, the aim is to choose between 2 and 5 features; data mining software are used. For instance, *Weka* tool [33]

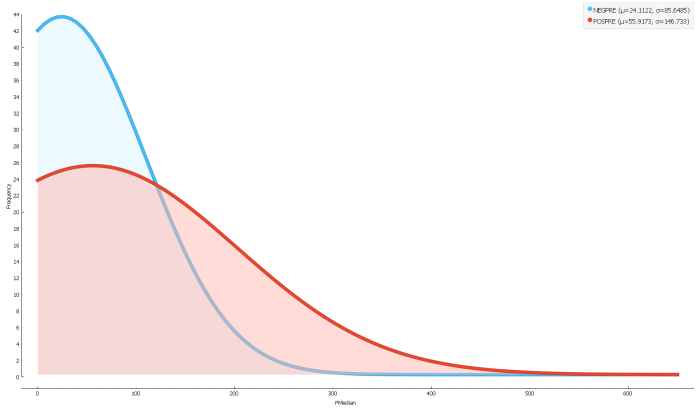


Figure 3.3: Display of the distribution of the feature Median of the FIRST class. In blue there is a category of patients, in red the other one; in the top right-hand corner there is also an indication about the mean of each curve.

uses algorithms and statistical techniques, such as decision trees and evaluators, to rank the most efficient attributes for classification. In addition, the *Orange* tool [34] is also able to graphically show the distribution of the feature value curves, allowing the user to make a guided selection of the best features for that case study (Figure 3.3). As explicated in the discussion of Section 3.7, it is not possible to completely renounce AI techniques, as in *real-world cases* the data to be handled involve enormous amounts. What this thesis succeeds in doing is the division between the AI used for Feature Selection and that used for the creation and verification of the patient model. However, the best selected features can be influenced by the exam settings, the database and the values; unfortunately, it is not certain that the chosen features are also valid for an external validation database as for the case study.

3.4 Discretization

In models it is not possible to track the exact value of features, hence a discretization step is mandatory prior to translation. In the context of data analysis, discretization refers to the process of converting continuous data into discrete intervals.
My team and I opted for a "3-level discretization", which means dividing the continuous range of values into three distinct intervals or categories. This process involves defining two threshold values that divide the value range into three intervals. Data points that fall within a specific interval are assigned to the corresponding interval.

For example:

- Low: Values below the first threshold;
- Medium: Values between the first and second threshold;
- High: Values above the second threshold.

The threshold chosen into this methodology is that of equidistributed discretization. For each feature, take the maximum value minus the minimum value and divide by the number of intervals to create three ranges of equal width.

$$threshold = (max - min)/nBins$$

For example, if a feature have values ranging from 0 to 30, the threshold would be:

$$threshold = (30 - 0)/3 = 10$$

This means there will be three bins: first for values greater than 0 and less than or equal to 10, second for values greater than 10 and less than or equal to 20, third for values greater than 20. In the translation into formal model, the discretized values are identified as "low" by $b1of3$, "medium" as $b2of3$ and "high" like $b3of3$.

This three-level discretisation simplifies data representation and facilitates analysis by classifying values into distinct groups. Discretized data may be easier to handle or interpret than a continuous range of values.
The choice of the number of intervals and threshold values can influence the results and interpretation of the analysis, widening or narrowing the intervals too much and often involves a certain level of optimisation according to the specific objectives of the analysis.

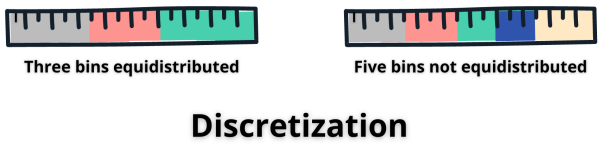


Figure 3.4: Different types of discretization are shown: on the left, an equidistributed method with 3 levels; on the right, a non-equidistributed method with 5 levels.

3.5 Formal Models

Once the features to be taken into account for the construction of the model have been chosen, automatic translation into formal language takes place thanks to a tool developed at the Radiomics Laboratory of the University of Molise.
As mentioned in the section 2.2, Formal Methods provide a model that reflects the specification of a computer system, namely its behaviour and the "rules of the game" that specify which input a certain output corresponds to. *For example, in the case of a vending machine, the behaviour includes different scenarios, such as: if the user inserts 1 coin, the vending machine will output tea; if the user inserts 2 coins the vending machine will output coffee.*
In the medical field, there is no standard behaviour of a patient. For this reason, my team and I decided that each slice of the radiological exam can be considered as the behaviour of the patient's state of health and therefore translated into a formal semantics to contribute to the construction of the model.

As explained in the Chapter 2 and Table 2.6, the CCS syntax has several powerful operators. The most significant one is certainly the Parallel, that allows *interleaving*, namely the simultaneous execution of several threads, processes, or events in a manner that makes them seem to occur simultaneously (i.e. multitasking). Understanding interleaving is essential to deal with condition-related problems and to ensure correct and predictable behaviour of programs. Indeed, managing interleaving requires effective synchronisation mechanisms and careful design.

$$\begin{aligned} proc\ P_1 &= (feature1.nil \mid feature2.nil).P2 \\ proc\ P_2 &= ... \end{aligned}$$

From the CCS syntax point of view, interleaving is guaranteed with the Parallel that is represented with the bar symbol | joining different features, as above. However, the syntax can be complicated: the example above is in fact wrong according to its semantics, because the Prefix operator is allowed only for Actions and not for Processes. Since we need this type of conceptual modelling because radiological examinations have hundreds of images, my team and I thought of implementing interleaving via the + operator, which is equivalent to Parallel, as explained by Milner [35].

$$\begin{aligned} proc\ P_1 &= feature1.feature2.P2 + feature2.feature1.P2 \\ proc\ P_2 &= ... \end{aligned}$$

To be equivalent, it is sufficient to write all combinations of radiomic features to

ensure their temporal and logic equivalence in the model at the moment of the formal checking, as illustrated above.

$prop\ P_1$	$=$	$b1of3\_inversevariance.b1of3\_clustershade.b3of3\_correlation.P2+$
		$b1of3\_clustershade.b3of3\_correlation.b1of3\_inversevariance.P2+$
		$b3of3\_correlation.b1of3\_clustershade.b1of3\_inversevariance.P2+$
		$b1of3\_clustershade.b1of3\_inversevariance.b3of3\_correlation.P2+$
		$b1of3\_inversevariance.b3of3\_correlation.b1of3\_clustershade.P2+$
		$b3of3\_correlation.b1of3\_inversevariance.b1of3\_clustershade.P2$
$prop\ P_2$	$=$	$b1of3\_inversevariance.b2of3\_clustershade.b3of3\_correlation.P3+$
		$b2of3\_clustershade.b3of3\_correlation.b1of3\_inversevariance.P3+$
		$b3of3\_correlation.b2of3\_clustershade.b1of3\_inversevariance.P3+$
		$b2of3\_clustershade.b1of3\_inversevariance.b3of3\_correlation.P3+$
		$b1of3\_inversevariance.b3of3\_correlation.b2of3\_clustershade.P3+$
		$b3of3\_correlation.b1of3\_inversevariance.b2of3\_clustershade.P3$
$prop\ P_3$	$=$	$...$
$prop\ P_{300}$	$=$	nil

Table 3.1: Creation of a Formal Model of a patient exam using the GLCM class and the radiomic features of Inverse Variance, Cluster Shade and Correlation.

For example, 3 features of GLCM class are taken: *Inverse Variance*, *Cluster Shade* and *Correlation*. As shown in Table 3.1, there are *Processes*, features and values. Each *Process* is a slice: if a MRI scan is composed of 300 slices, the model will have 300 processes. Each process contains the chosen radiomic features with the discretized values and all their combinations, to make them valid simultaneously during the formal checking. Each model can be traced back to an LTS and vice versa (Figure 2.1), which is its graphical representation.

3.5.1 Formal Methods for single image

In most of the papers, the databases used presented MRI, CT or PET images, so there are large sequences of images. If there are Us images, there is only one image per patient. This makes it impossible to use the methodology described above and new conceptual modeling is needed to create the formal models.
My team and I came up with a new methodology that would allow the Formal Methods to be used for classification and localization. By applying a *nxn* grid to the image,

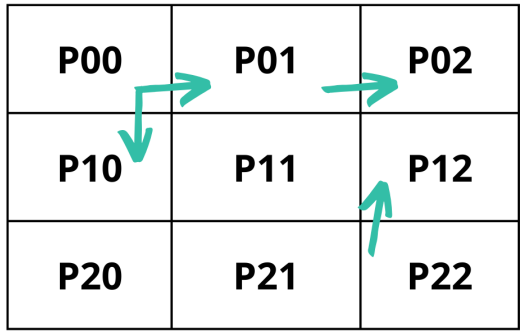


Figure 3.5: A different type of formal model using a single image: each process is a cell of a grid overlapped on the image. The syntax of this model can be seen in Table 3.2.

in fact, it is possible to divide the image into several cells, each of which will be combined with the others, by means of a model that allows one to understand which are its neighbours. As a result, the model is no longer linked between different slices, but between cell positions; it is possible to move by only one cell, except diagonally. This interleaving between cells allows classification and at the same time a localization of the signs of the disease, as in the "TRUE" cell a disease information will surely be present.

Referring back to Figures 3.5 and 7.16, one can see the construction of a $3x3$ grid on the image: always following the three-level discretisation, the formal model will be constructed using the positions of the other cells in relation to the one being considered. The first process will be the first cell, then it will use the features and then it will be able to go to the right or below by recalling the relative processes of the cells it moves to. The model is illustrated in Figure 3.2.

In this new method of modelling, the CCS model no longer contains time combinations of different features, but contains the same sequence of radiomic features with different targets: moving by one step, from cell $P00$ you can go only to the right from $P01$ or below from $P10$ because in other directions you would exit from the image. As a future development, this methodology could be incorporated into an iterative method that thickens the grid for better classification and localisation of even the smallest signs of disease.

$$
\begin{aligned}
prop\ P_{00} \quad = \quad & b1of3\_inversevariance.b1of3\_clustershade.b3of3\_correlation.sx.nil+ \\
& b1of3\_inversevariance.b1of3\_clustershade.b3of3\_correlation.up.nil+ \\
& b1of3\_inversevariance.b1of3\_clustershade.b3of3\_correlation.dx.P01+ \\
& b1of3\_inversevariance.b1of3\_clustershade.b3of3\_correlation.down.P10 \\
prop\ P_{01} \quad = \quad & b1of3\_inversevariance.b2of3\_clustershade.b3of3\_correlation.sx.P00+ \\
& b1of3\_inversevariance.b2of3\_clustershade.b3of3\_correlation.up.nil+ \\
& b1of3\_inversevariance.b2of3\_clustershade.b3of3\_correlation.dx.P02+ \\
& b1of3\_inversevariance.b2of3\_clustershade.b3of3\_correlation.down.P11 \\
prop\ P_{02} \quad = \quad & ... \\
prop\ P_{22} \quad = \quad & b3of3\_inversevariance.b2of3\_clustershade.b1of3\_correlation.sx.P12+ \\
& b3of3\_inversevariance.b2of3\_clustershade.b1of3\_correlation.up.P21+ \\
& b3of3\_inversevariance.b2of3\_clustershade.b1of3\_correlation.dx.nil+ \\
& b3of3\_inversevariance.b2of3\_clustershade.b1of3\_correlation.down.nil
\end{aligned}
$$

Table 3.2: Formal Model for single-image methodology (such as US scan), using the GLCM class and the radiomic features of Inverse Variance, Cluster Shade and Correlation.

3.6 Formal Verification

Once the database of formal models has been created, it is only necessary to check the presence of the disease and classify the patients. Thanks to the collaboration between medical engineers, radiologists and computer scientists, some more explicit cases of the disease were analysed and formal rules were created. These rules were then checked on the models, i.e. by asking the Model Checker if the combinations of the property were also in the models; thus the classification was carried out.
The rule almost always expresses the presence of the disease: therefore, if the Model Checker returns the value "TRUE" then the model includes the property and the patient is affected by the disease (namely, is sick); conversely, if it returns "FALSE" it means that the model does not include the property and consequently the patient shows no signs of the disease. These results can then be collected in a Confusion Matrix and used to calculate the metrics described in Section 2.3.

The advantage of Formal Methods lies in their explainability and interpretability. When the outcome is "FALSE", in fact, the Model Checker also clarifies the logic behind its decision by identifying the specific point at which the discrepancy between the LTS and the property occurs: this is called *Counterexample*. In Figure 3.6 there is LTS of a COVID-19 patient while a property associated with a healthy patient occurred. The generated counterexample includes the input LTS, along with clear annotations explaining the reasons for its failure to satisfy the property. In this specific instance in Figure 3.6, the property is searching for $b1of3\_variance$ and $b1of3\_fenergy$ whereas the input LTS starts with $b2of3\_variance$ and $b3of3\_fenergy$. Consequently, the initial state is marked in red and the ultimate output is "FALSE", signifying that the patient is not afflicted by COVID-19.

Figure 3.6: Taking a LTS and a formula to satisfy, the output is "False" and the initial state of the LTS is highlighted in red because from that state the property is no longer satisfied from the model.

3.7 Explainability of Radiomics through Formal Methods

Recent literature has established the effectiveness of AI in quantitative analysis, research, diagnosis and prognosis within the field of Radiomics [36–39]. A primary challenge in this context is to understand the *rationale behind model-specific predictions*: in this, AI-based solutions have not performed well in the real medical world due to the complexity of the models. These problems of interpretation are crucial for both patients and physicians, as the models could be enriched by their expertise and knowledge. *Explainability* remains an open problem within the scientific community and has already been addressed in several studies [40].
A solution lies in alternatives approaches of AI, such as Formal Methods [41, 42], which exploit the power of mathematical logic and offer a unique advantage. With Formal Methods, it becomes feasible to amalgamate the field of Computer Science with the expertise of radiologists to construct expert systems that aid in the diagnosis or prognosis of various diseases. This approach is automated, robust, simple, but it could be not completely disconnected from AI, which remains essential for the specific stages of handling large amounts of data.

I would like to emphasise that *algorithms can never replace the essential role of the radiologist*. The final automated result, i.e. the diagnosis, should be considered as a "secondary virtual opinion" complementing the radiologist's expertise; this could be an advantage in the midst of epidemiological emergencies, in the case of rare and difficult-to-diagnose diseases. The pipeline of Formal Method and Radiomics is extremely simple and in contrast to the "black boxes" of AI, which may appear more difficult to explain [43].

The main *objective* is to conduct a medical-oriented analysis of various radiomic features following the IBSI guidelines [20], trying to establish connections between these features and clinical or radiological visual evidences, effectively building upon previous diagnoses. For example, some key aspects can be:

1. Understand the range of feature values considered by the methodology for automated diagnosis;

2. Investigate these value ranges to discern patterns associated with different disease states;

3. Engage automatic pattern recognition to affirm and elucidate the automated

diagnosis;

4. Conduct localization of the most crucial image slices used in the diagnosis process.

The *slice localization process* allows these image slices to be provided to medical specialists as feedback, enabling them to affirm or dispute the visual qualitative diagnosis. This innovative approach offers doctors and their colleagues the opportunity to provide explanations for the automatic diagnosis: transparency is of fundamental importance in this translational field, where human expertise is indispensable.

FIRST-Order features are analysed: they describe the distribution of values of individual voxels independently of their spatial arrangement, indicating the presence of specific colours or values in an image by means of histograms, but not how these are arranged in the image. As a result, establishing a direct connection between numerical values and what a radiologist sees in a multi-slice CT examination can be quite challenging.
For example, FIRST features are unable to distinguish between a black-and-white chessboard pattern and a split image with one side black and the other white: in the case of Fig. 3.7, the histogram on the right shows a large number of pixels in only two columns of grey-level intensity, one for black pixels and one for white pixels. This limited information may be an impediment for researchers analysing the different patterns within the images. On the other hand, FIRST-Order statistics provide the general characteristics of an image and are currently the most widely used metrics for quantifying the phenotype of images [44].

FIRST-Order features provide features such as the mean, median, maximum and minimum values of voxel intensities. They also encompass parameters like Skewness (indicating asymmetry), Kurtosis (expressing flatness), Uniformity and entropy (for randomness). The basic features of Maximum, Minimum and Mean should correspond to the relative Hounsfield values [45] for the maximum, minimum and mean intensity levels, while the remaining features are essentially statistical measurements.

Section 7 can be considered as the basis for these links between numerical values and radiological evidences; in Section 7.4, the classification results will be interpreted regarding the existing literature. It is essential to note that in the methodology explained in the mentioned section, the authors chose a ROI which includes the entire CT image, ensuring that no part of the image is overlooked/underlooked in the calculation process. Radiomic features were calculated for each patient and variations

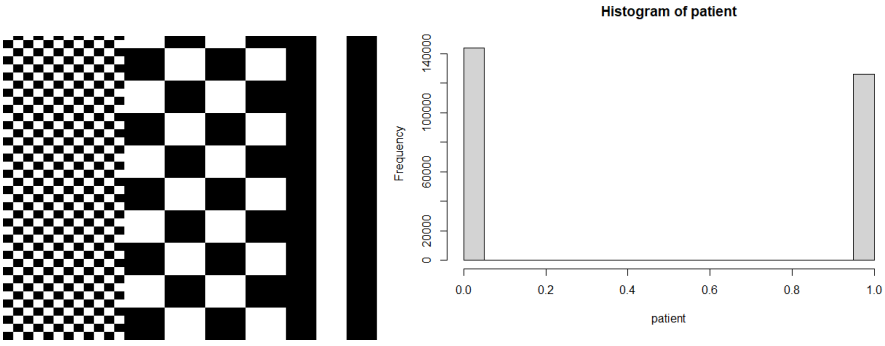


Figure 3.7: These structured images on the left all have the same FIRST-Order statistics values, well represented on the right by a histogram with two columns, one for black pixels and one for white pixels. This similarity is only apparent, because we can visualize the three images are different in terms of pattern combination and then, of meaning.

associated with different health states were highlighted: key features of the existing literature for lung studies were taken into account and efforts were made to provide meaning and interpretability to the entire process of using Formal Methods for COVID-19 classification.
A dataset of 35 patients may seem limited for AI techniques, which typically require a substantial volume of images to create robust models with high performance. However, this assumption does not apply to Formal Methods, which do not learn directly from images but rely on the knowledge and experience of radiologists, translating this expertise into properties of time logic. As a result, even a small data set allows for high performance. However, this method is not completely separate from AI: there are steps where AI is essential, such as the Feature Selection. Currently, Formal Methods stand out among the few that can address some of the limitations of AI, particularly the need for extended data to ensure model robustness.

This methodology will be applied in the following Chapters to different anatomical districts: in some works, modifications will be made in order to deepen or diversify the analysis. In each Chapter there will be an experimental part and one relating to discussion and state-of-the-art.

3.8 Related Work and Discussion

In the literature belonging to the field of Computer Science, Formal Methods were mainly used for testing software behaviour, to identify the malware attack or for other security issues. From 2010 onwards, Formal Methods were introduced also in the field of medical imaging: an example are the developed methodologies for image segmentation using Model Checking [46]. For example, the cited works by Massink et al. [47] used Spatial Logic and Model Checking for segmenting the images with the concept of similarity, achieved through a statistical similarity operator and from statistical texture analysis.

To the best of my team's knowledge, we were the first to introduce Formal Methods as a method of prediction, classification and localisation of diseases. We were also among the first to exploit these mathematical methods in conjunction with Radiomics.

To test the reliability of this *dynamic duo*, we tested the methodology on different anatomical districts of the human body, sometimes modifying the workflow according to the needs of the case study. For example, in the brain, a preprocessing step was added to handle the high structural complexity of the images; another example is in the COVID-19 Section, where DICOM sequences were modified to analyse only part of the patients. In each chapter of this thesis and for each anatomical part, there will be an in-depth discussion on the state-of-the-art and a comparison of the literature between AI and Formal Methods after the experiment.

Chapter 4

Brain

In this Chapter the Brain will be analysed: my team and I worked first on the prediction of Multiple Sclerosis using radiomic features and then on the prediction of active lesions of Epilepsy. In the latter, we modified the methodology by adding *pre-processing* on the images.

4.1 Detecting Disease Activity from MRI Scans of Patients with Multiple Sclerosis Via Radiomics and Formal Methods

Multiple sclerosis (MS) is a chronic autoimmune demyelinating disease that mainly affects the *Central nervous system* (CNS). Traditionally, the activity of the disease in MS is defined with the occurrence of new relapses or the appearance of new or enlarging lesions from the MRI. While clinical markers evaluating gait, dexterity and cognition have become more refined indicators of disease activity, regular monitoring through cerebral and spinal MRI with contrast agent remains essential, particularly in the early stages of the disease. *Gadolinium-enhancing* (Gd+) is a biomarker for acute inflammatory disease activity and is often employed as a measure of outcome in clinical trials for relapsing-remitting multiple sclerosis (RRMS) [48].

An open issue in the field of MS concerns potential alternatives to the repetitive gadolinium-enhanced MRI scans [49]. While the presence of contrast-enhanced lesions

Parameter	Value
Repetition Time (TR)	400 msec
Echo Time (TE)	12.76 msec
Slice thickness	5 mm
Matrix size	288x288
Flip angle	90°
Number of slices	24

Table 4.1: Parameters used for the axial 2D T1-weighted scan acquired before and after Gd contrast agent administration.

plays a vital role in diagnosis by demonstrating "spread over time" and distinguishing it from similar clinical conditions [50], there are concerns regarding the *safety and ethics* of using gadolinium-based contrast agents (GBCAs) [51]. These concerns include Gd accumulation in specific areas of the brain after repeated GBCAs administration [52–55] and the risk of nephrogenic systemic fibrosis (NSF) in patients with cited renal insufficiency [56–58]. In addition, the identification of Gd+ lesions in MRI images relies on manual interpretation by experienced neuroradiologists, a time-consuming treatment prone to variability [59].

In recent years, there have been numerous efforts to develop automated techniques for the detection of Gd+ lesions in MRI scans of patients with multiple sclerosis. Many of these approaches use ML or AI methods [60,61], specifically applied to MRI images taken after the administration of Gd+ [61,62]. Notably, Radiomics has demonstrated the ability of certain features to estimate the age of lesions in patients with MS [63], as well as distinguish between MS and neuromyelitis optica spectrum disorders based on spinal MRI images [64, 65].

The primary focus of my research revolves around the development of an automated method for the *detection of active lesions*. This detection will be carried out using pre-contrast T1-weighted (T1w) MRI images through a combination of Radiomics and Formal Methods. A total of 96 patients (comprising 66 females and 30 males, with a mean age of 35.2 years) diagnosed with RRMS and receiving treatment at the Department of Neurology of IRCCS Neuromed were selected. Among these patients, 48 out of the 96 displayed radiological activity in their T1w post-contrast MRI scans, while the other 48 did not exhibit any radiological activity on their MRI scans. The study received approval from the Ethics Committee of the Neuromed Research Institute *(code 06-17)* in compliance with the Declaration of Helsinki and all patients

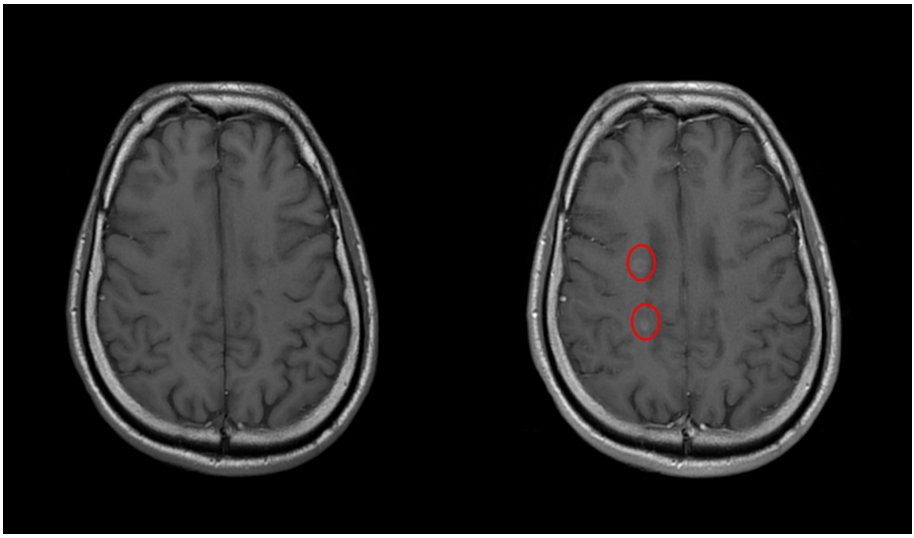


Figure 4.1: On the left side, a view of a pre-contrast T1 weighted MRI. On the right side, the T1w post-contrast MRI of the same patient, which showed radiological activity (highlighted by two red circles).

provided written informed consent to participate in the research. All patients underwent imaging using a 1.5T MRI scanner, specifically a General Electric HDxt system, equipped with an 8-channel head coil. The MRI protocol included both pre-contrast and post-contrast axial 2D T1w scans, with the scan parameters detailed in Table 4.1.

In the case of patients exhibiting radiological activity, the *identification of Gd+ lesions* in the post-contrast T1w MRI scans was conducted by two experienced neurologists; they noted the specific slice numbers where these lesions were observed, as in Figure 4.1, following the well-established guidelines [66]. The acquisition of pre-contrast T1w MRI scans was essential to differentiate between T1-hyperintensities potentially attributed to lipid- and iron-laden microglia/macrophages [67] and genuine contrast-enhancing T1-hyperintensities.

Acquiring T1w images, allowed me to extract all the radiomic features as per the IBSI standards [20]. Subsequently, Orange [34] is used to perform *feature selection*, choosing only 4 features for each class to ensure that the mathematical model could reach a solution in a timely manner. Unfortunately, not all the feature classes exhibited a distinct separation between patients with and without Gd+ lesions: in Table 4.2 are listed the selected features for this case study.

Since the aim is to evaluate the prognostic value of this method in detecting ra-

Feature Class	Feature
FIRST	10 Percentile; Mean Absolute Deviation; Mean; Median; Robust Mean Absolute Deviation.
GLCM	Idmn; Cluster Prominence; Cluster Shade; Correlation.
GLDM	Small Dependence Low Gray Level Emphasis; High Gray Level Emphasis; Large Dependence High Gray Level Emphasis; Small Dependence High Gray Level Emphasis.
GLRLM	Short Run Low Gray Level Emphasis; Gray Level Variance; High Gray Level Run Emphasis; Long Run High Gray Level Emphasis.
GLSZM	Zone Percentage; Gray Level Variance; High Gray Level Zone Emphasis; Small Area High Gray Level Emphasis.
NGTDM	Strength; Complexity; Contrast.

Table 4.2: List of the selected features for models construction.

diological activity without the administration of Gd, only the pre-contrast 2D T1w images were used, all in DICOM format. The produced property was created to classify patients with disease activity and the performance were evaluated through the *Confusion Matrix*. The analysis of these performance metrics showed that the most suitable feature class was the GLDM: as shown in Table 4.3, the methodology got a classification Accuracy of 66%, with a Precision of 67%, a Recall of 60% and a Specificity of 70%.

Another advantage of Formal Methods is that they are easy to explain. For instance, the property applied by the classifier is clearly showed in Figure 4.2. The features under consideration are *Small Dependence Low Gray Level Emphasis* and *High Gray Level Emphasis* from the GLDM feature class. The combination of these two features values are also enriched by a notation, for example "$x2$", indicating that the action is repeated two times in a row. The number of actions and their consecutive presence can significantly influence the classifier's decision.
In this specific case, the critical actions to consider are a combination of *medium-high* Small Dependence Low Gray Level Emphasis and *medium-high* High Gray Level Emphasis values, or *low values* for both features in the combination. This information is valuable for understanding the classification process, linking radiomic features to radiological evidence and indirectly aiding in lesion localization. When the feature

```
prop F1 =  <b1of3_dsmalldependencelowgraylevelemphasis> <b3of3_dhighgraylevelemphasis> F2
prop F2 =  <b2of3_dsmalldependencelowgraylevelemphasis> <b3of3_dhighgraylevelemphasis> F3 - x2
prop F3 =  <b3of3_dsmalldependencelowgraylevelemphasis> <b2of3_dhighgraylevelemphasis> F4 - x2
prop F4 =  <b1of3_dsmalldependencelowgraylevelemphasis> <b2of3_dhighgraylevelemphasis> F5
prop F5 =  <b1of3_dsmalldependencelowgraylevelemphasis> <b1of3_dhighgraylevelemphasis> tt - x3
```

Figure 4.2: Actions described in the property to find in the models, which lead the classifier to make a decision about the state of health of the patients. On the right side there are some annotation on how many times the action is repeated in a row.

values closely resemble the repeated pattern, it suggests the presence of Gd+ lesions.

The literature on AI for Multiple Sclerosis have shown promise in the field of neurological diseases [68, 69]. However, there is still no precise description of the number or type of radiomic features that unequivocally characterize the pathology. In future research, the incorporation of Formal Methods can enhance the explainability of these studies. With Formal Methods, it will be feasible to investigate the localization of disease activity and understand the reasons behind it. This approach can provide insights into which values are associated with neurological evidence, strengthening the relationship between quantitative and qualitative data in radiology.

Accuracy Statistics	**Value**
Accuracy	0.66%
Precision	0.67%
Recall	0.60%
Specificity	0.70%

Table 4.3: Accuracy and other indexes to gain more information about the clinical usefulness of the methodology.

4.2 Radiomic imaging for the explicable study of Epilepsy by Formal Methods

Epilepsy is a prevalent and chronic neurological disorder, which, if not effectively managed, can significantly diminish the quality of life. Unfortunately, a substantial portion of epilepsy patients does not respond well to therapies, resulting in an elevated risk of injury and potential cognitive impairment [70].
Temporal lobe epilepsy (TLE) is the most frequent form of focal epilepsy, constituting approximately 30 percent of all epilepsy cases [71]. This type of epilepsy is characterized by seizures originating in the temporal lobe of the brain. MRI is a standard diagnostic tool employed to identify TLE and uncover any underlying structural abnormalities in the brain that may serve as the *epileptogenic focus*.
However, in clinical practice, MRI is still insufficient in about one third of patients, as it may appear normal on visual examination, i.e. no lesions or abnormalities are detected. This subgroup of cases is commonly referred to as "TLE MRI-negative" [72] and for these patients further investigations are required to identify the exact location of the *epileptogenic focus*.

The use of PET with 18F-fluorodeoxyglucose (18FDG-PET) is a widely adopted approach for the localization of seizure foci and can provide valuable lateralization or

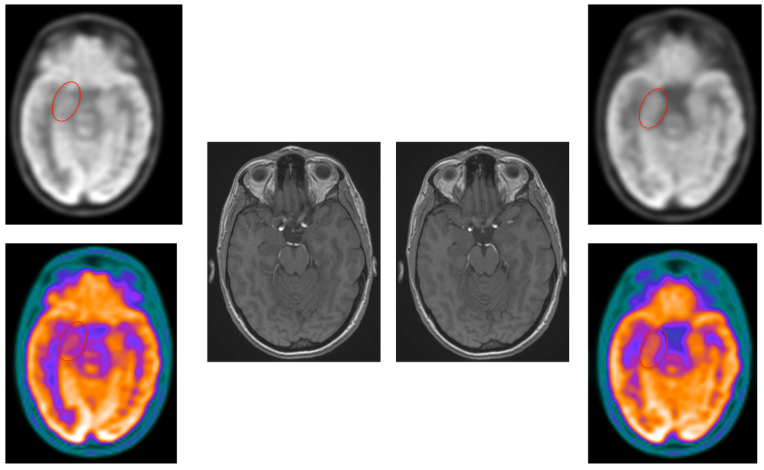


Figure 4.3: An example of a patient who is MRI-negative but PET-positive (on the left) and another patient who is MRI-negative and PET-negative (on the right). The red circle highlights the area of hypometabolism, which is clearly visible on the PET images of the PET-positive patient (on the left) but absent in the PET-negative patient (on the right).

localization information in 60-90% of patients with TLE. In focal epilepsy, a typical FDG-PET observation involves a decrease in interictal glucose metabolism in the temporal lobe [73, 74], as shown in figure 4.3.

Although functional nuclear imaging techniques are advantageous in MRI-negative TLE patients, it is important to note that these techniques are not universally accessible, especially in developing countries. Alternatively, patients with MRI-negative TLEs may opt for intracranial Electroencephalography (EEG), a procedure that carries similar risks as resective surgery and high costs [75].
Unconventional MRI sequences, such as double inversion recovery (DIR) or enhanced diffusion MRI (dMRI), have been investigated to analyse brain abnormalities in MRI-negative TLE patients. However, these sequences are time-consuming, are not standardised on all MR scanners and are not available in all hospitals [76]. Therefore, the community is looking for alternative time- and cost-saving approaches to identify disease-specific markers from MR-negative scans of patients with TLE. These approaches exploit AI techniques such as ML to detect TLE-negative MRI status using neuroimaging data; the limitation of these methods is access to large datasets for training classifiers [77].

In the standard diagnostic procedure for patients with suspected epilepsy, routine EEG monitoring is followed by a brain MRI. If the MRI results are negative, a PET scan is often requested to assess whether abnormalities attributable to epilepsy are

Patient	**No. Slices**	**Slice Thickness** (mm)	**Repetition time** ($msec$)	**Echo time** ($msec$)	**Slice gap** (mm)
001	146	1.0	6.99	2.85	1.0
003	172	1.8	7.99	2.17	0.9
005	176	1.8	6.96	2.16	0.9
006	172	1.8	7.07	2.19	0.9
007	154	1.0	7.22	2.95	1.0
008	164	1.8	9.51	3.81	0.9
009	192	1.6	8.00	2.44	0.8
010	172	1.8	9.56	3.83	0.9
011	156	1.0	7.34	2.98	1.0
016	208	1.6	8.46	2.73	0.8
017	168	1.0	7.05	2.84	1.0
019	172	1.8	9.58	3.83	0.9
025	164	1.8	6.96	2.87	0.9
026	156	1.0	7.34	2.93	1.0

Table 4.4: List of MRI parameters of the 3d T1w SPGR MRI for patients included in the dataset.

detectable, thus confirming the diagnosis of epilepsy. In this Section, the *primary objective* is to predict PET positivity or negativity on the basis of a negative MRI using Radiomics and Formal Methods using only their negative MRI scans.
Each patient underwent a brain MRI, including a T1-weighted 3D SPGR and subsequently an 18F-FDG PET scan to confirm or exclude the diagnosis of TLE. All MRI scans were negative for TLE, while the PET scan was discordantly positive with the MRI findings in nine of the fourteen patients. It is essential to note that the imaging protocols were not consistent between all patients due to the retrospective nature of the study, as listed in Table 4.4.

Radiomics has been used in recent years for tasks such as identifying laterality in MRI-negative TLE patients and distinguishing medial TLE patients from healthy controls, but to the best of my knowledge, this is the first combination of Radiomics and Formal Methods for the purpose of identifying epilepsy in negative MRI scans [78, 79]. Differently from the classic methodology, as MRI images can be prone to noise, *a pre-processing step* involving the application of a Gaussian filter to the MRI images was planned. This allowed to compare the performance of the original images with those subjected to pre-processing, determining whether filtering improves the results:

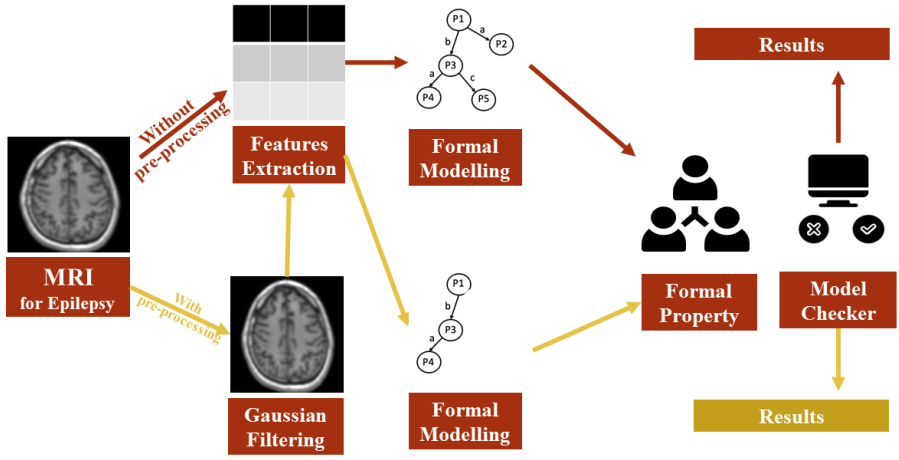
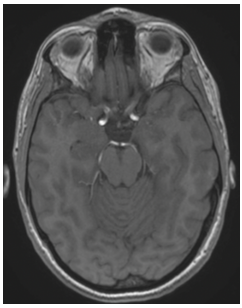
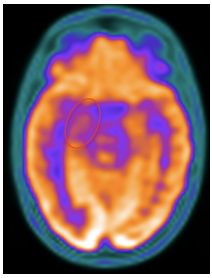


Figure 4.4: The proposed model involves two distinct steps. The first step is performed without any image pre-processing, while the second step includes image pre-processing. In both steps, MRI series of the patient are modelled for both the original images and the filtered ones. Utilizing a Formal Property and the Model Checker agent, let obtain the results of the automatic diagnosis.



(a) MRI image of a patient with epilepsy.

(b) PET image of a patient with epilepsy.



(c) Application of a color map to the PET image of a patient with epilepsy.

Figure 4.5: Three different medical images of a patient with epilepsy.

the methodology is described in Figure 4.4. This retrospective study looked at 14 patients with MRI-negative scans, consisting of 9 individuals with negative PET scans and 5 with positive PET scans, collecting demographic and clinical characteristics. These patients were recruited from the Epilepsy Surgery Unit of the IRCCS Istituto Neurologico Mediterraneo Neuromed, located in Pozzilli, Italy, between January 2011 and February 2022. An illustrative example of a patient with negative MRI and positive PET scan (left), as well as a patient with negative MRI and negative PET scan (right), can be seen in Figure 4.3. All participants underwent high-resolution 3D T1w Spoiled gradient echo (SPGR) MRI scans with a GE Signa HDxt 3.0T MRI scanner equipped with an 8-channel cranial coil, as well as 18FDG-PET interictal scans. For the purpose of this study, only 3D T1w MRI scans were considered for analysis; a thorough visual inspection of all MRI and FDG-PET scans was conducted by experienced board-certified professionals, including a neuroradiologist and nuclear physicians. This visual examination aimed to identify and exclude any potentially epileptogenic lesions within the MRI scans. For example, Figure 4.5a displays the MRI scan of a patient with negative MRI results. In Figure 4.5b, the corresponding PET image for the same patient is presented. The PET image exhibits slight blurring (highlighted with a red circle) typically located in the grey matter, indicating PET positivity. This characteristic on the PET image becomes even more evident when a color map is applied (as depicted in Figure 4.5c), with the blurring circled in red.

As far as *Feature Selection* is concerned, the Orange software has always been used, looking at the distribution of values for each feature: the following Table 4.5 shows the list of selected features for the models.

Filter application, formal modelling and verification

```
Folder ='../006'
filesname = dir(fullfile(Folder,'*.dcm'));
total_Im = numel(filesname);
for i = 1:total_Im
    f = fullfile(Folder, filesname (i).name)
    dicom = dicomread (f);
    dicom1 = imgaussfilt3(dicom, 2);
    montage ({dicom, dicom1});
    title('Originals vs Gaussian');
    tit = sprintf('/dicom0sd.dcm',i);
    dicomwrite (dicom1, tit);
end
```

Listing 4.1: Script used for applying the Gaussian filter to MRI.

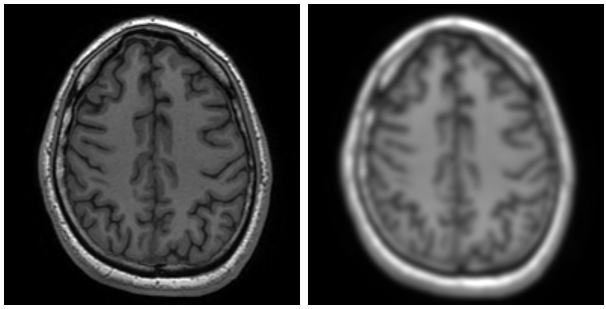


Figure 4.6: Comparison between the original MRI image (on the left) and the filtered MRI image (on the right).

Regarding the workflow with pre-processing, a *3D Gaussian filter* was employed on medical images using the "imgaussfilt3" function from MATLAB *(The Mathworks, Inc., R2023a)*[1] via the script presented in Listing 4.1. The Gaussian filter is a linear smoothing filter that relies on the Gaussian function, functioning as a low-pass filter and is primarily employed to eliminate noise that adheres to a normal distribution. The width of the Gaussian filter is determined by the parameter σ: a larger value results in a broader frequency band for the Gaussian filter, as discussed in Wang [80]. In the Listing script, this parameter is set at value 2.
For instance, Figure 4.6 displays the same slice of patient 26 both before and after the application of the filter. It is evident how the image on the right, which has undergone

[1] https://it.mathworks.com/help/images/ref/imgaussfilt3.html

the Gaussian filter, differs from the original image because of the smoothing.
For both the original medical images and their filtered counterparts, the values are *discretized* into three classes with equal amplitude: low, medium and high. The models were constructed using the CCS syntax, which involves creating a process for each slice of the medical image and encompassing all the combinations of discretized feature values that must simultaneously hold.
To perform the *classification* between PET-positive and PET-negative patients, it is essential to define the property that characterizes PET-positive patients. Eight rules were developed, incorporating input from both medical experts and computer scientists. These outline the varying levels of features that typify a TLE patient: for example, one property highlights medium-high values for the *Variance* feature, while the *Kurtosis* feature consistently remains at low level.

Feature Class	**Features**
FIRST	Kurtosis, Variance, Skewness
GLCM	Cluster Prominence, Cluster Shade, Icm1, MCC
GLDM	Large Dependence High Gray Level Emphasis, Gray Level Variance, High Gray Level Emphasis, Small Dependence High Gray Level Emphasis
GLRLM	Gray Level Variance, High Gray Level Run Emphasis, Long Run High Gray Level Emphasis, Short Run High Gray Level Emphasis
GLSZM	Large Area Emphasis, Large Area Low Gray Level Emphasis, Large Area High Gray Level Emphasis, Zone Variance
NGTDM	Coarseness, Complexity, Busyness

Table 4.5: These are the features selected from the feature classes of the IBSI standard: these features are used for the analysis in the Radiomic workflow.

Results on original image

The performance metrics of the developed mathematical model are presented in detail as Supplementary Material in the *Appendix* A.1 below. For the *FIRST class*, as shown in Table A.1, the results indicate that the Accuracy hovered around 50%, with Precision ranging from 60% to 100%. Recall was generally low in all cases, while Specificity remained consistently at 60% or higher.

In the *GLCM class*, the performance metrics results are displayed in Table A.2. The Accuracy ranges from 57% to 86% and the Precision consistently remains above 80% in all eight cases. Recall exhibits substantial variability, with a minimum of 33% in two instances and a maximum of 100% in two other instances. Specificity, on the other hand, maintains an average of approximately 80%.

In the *GLDM class*, the performance metrics results can be found in Table A.3. Accuracy surpasses 50% in only two instances, while Precision varies from 67% to 100%. Recall consistently remains below 90% and Specificity is consistently 40% or higher.

In the *GLRLM class*, the performance metrics results are presented in Table A.4. The best case achieved an Accuracy of 71%, while the two worst cases reached 43%. Precision consistently remained above 57%, while Recall showed significant values around 90-100% in two instances. Specificity varied uniformly from 20% to 100%.

The results for the *GLSZM class* are summarized in Table A.5: Accuracy ranges from a minimum of 36% in the worst case to a maximum of 71% in two instances. Precision reaches 100% in only one case, while Recall does not exceed 67% in any case. Specificity is at least 80% in half of the cases.

The results for the *NGTDM class*: are presented in Table A.6. Accuracy is 50% or higher in six cases, while Precision reaches 100% in three cases. Recall is consistently below 67% and Specificity reaches 100% in three instances.

The best property for each feature class is presented below in Table 4.6, highlighting the one with the most favorable metrics and minimal heterogeneity among them. The property “GLCM7-16” stands out as the best performer, while “FIRST5-25” exhibits the least impressive results. In general, except for the FIRST class, Accuracy exceeds 60%. Precision ranges between 70% and 86%, with Recall hitting 100% in two cases; Specificity attains 80% in just one case.

Results without filter				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
ruleFIRST5-25	0.57	0.71	0.56	0.60
ruleGLCM7-16	**0.86**	0.82	**1.0**	0.60
ruleGLDM5-25	0.71	0.73	0.89	0.40
ruleGLRLM5-25	0.71	0.70	**1.0**	0.20
ruleGLSZM5-25	0.71	**0.86**	0.67	**0.80**
ruleNGTDM5-25	0.64	0.75	0.67	0.60

Table 4.6: Overview table of the best metrics for the feature classes under study.

Results on filtered image

To assess the impact of image filtering, a Δ variation value is calculated by comparing the average metrics for the filtered and original images. A negative value indicates a decline in performance, while a positive value suggests an improvement. Detailed tables are provided in *Appendix* A.2.

Results for the *FIRST class* in Table A.7 exhibit generally low values for both Accuracy (never exceeding 57%) and Recall (never surpassing 55%). Applying the filter did not yield significant enhancements, except for two rules, which led to increased Precision and Specificity while reducing Recall in both cases.
For the *GLCM class*, the results obtained are good for more than one rule, with Precision always greater than or equal to 80%. The application of the filter, for each rule, led to a deterioration, as displayed in Table A.8.
The results of the features belonging to the *GLDM class* in Table A.9 show medium-high values for Accuracy, Precision and Specificity while lower values for Recall. The application of the filter resulted in a Δ greater than zero only for rule3-11 (Precision and Specificity increase while Recall decreases), furthermore with the application of rule11-19 the values do not change.
The results obtained from the *GLRLM class* and listed in Table A.10, show Accuracy around 50%, Precision around 60 per cent, rather low Recall values while for Specificity both high (80%) and low (20%) values. The results obtained from the filtered images show a general improvement for three rules in which Precision and Specificity increase while Recall values decrease.
For the *GLSZM class*, all four metrics have medium to high values. For only two rules are improvements obtained (Precision and Specificity increase) as a result of applying the filter (Table A.11).
Observing the results obtained from the features of the *NGTDM class*, we see average values for Accuracy, Precision and Recall while higher values for Specificity. For five out of eight rules, there is a general improvement in the results from the filtered medical images: specifically, there is an increase in Precision and Specificity, as shown in Table A.12.

In Table 4.7, the chosen rules for classification based on the highest metrics from the original medical images are presented, along with the results obtained from the same rules for the filtered images. For the feature classes FIRST, GLRLM and NGTDM a positive Δ indicates an improvement. In these cases, there is an increase

Results						
Feature Class	**Property**	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
FIRST	rule5-25	0.57	0.71	0.55	0.60	
FIRST	filtered	0.50	**1**	0.22	**1**	**+0.07**
GLCM	rule7-16	0.86	0.82	**1**	0.60	
GLCM	filtered	0.64	0.83	0.56	0.80	**-0.11**
GLDM	rule5-25	0.71	0.73	0.89	0.40	
GLDM	filtered	0.64	0.75	0.67	0.60	**-0.01**
GLRLM	rule5-25	0.71	0.69	**1**	0.20	
GLRLM	filtered	0.64	0.83	0.56	0.80	**+0.06**
GLSZM	rule5-25	0.71	0.86	0.67	0.80	
GLSZM	filtered	0.21	0	0	0.60	**-0.56**
NGTDM	rule5-25	0.64	0.75	0.67	0.60	
NGTDM	filtered	0.64	**1**	0.44	**1**	**+0.10**

Table 4.7: A comparison of the results obtained from the original and filtered medical images is presented below, indicating the chosen property based on the highest metrics for the original images.

in both Precision and Specificity, but Recall worsens. The Accuracy either decreases (FIRST and GLRLM) or remains the same (NGTDM). A high Precision suggests that a significant percentage of the cases classified as positive are indeed positive, while high Specificity indicates a low number of false positives and accurate classification of negative patients. However, a decrease in Recall suggests an increase in false positives, leading to less accurate predictions when Accuracy also decreases.
The GLSZM class exhibits the lowest Δ value, with a Precision and Recall of 0% and an Accuracy of 22%; these results are attributed to the classifier identifying no true positives.

4.3 Related Works and Discussion

In the medical community, there is a growing need of focused research on applying Radiomics to neurological disease, like the detection of radiological activity in patients with MS. An example is the study conducted by Pontillo et al. [68], aimed to address the *clinicoradiologic paradox*, where MRI alone explains only a small portion of the variability in clinical outcomes for MS patients [81]. To overcome this challenge, they combined Radiomics with ML techniques and a larger dataset than the one used in my study, achieving results that are comparable to mine, with an average performance index of around 70%. In general, this suggests the promise of Radiomics and AI in enhancing the understanding and diagnosis of complex medical conditions like MS.
The distinction between my approach, which employs Formal Methods and the one by Pontillo et al. [68], which uses AI techniques, is significant; Formal Methods do not rely on learning from images, which can be an advantage when working with small datasets. This minimizes the potential for biases stemming from unbalanced datasets, often encountered in medical imaging. Additionally, the use of Formal Methods enhances the *explainability* and understandability of results. These methods are often referred to as "white-boxes" because they provide transparent, structured and interpretable models [43]. This characteristic facilitates communication and collaboration between radiologists, their colleagues and the data, fostering a deeper understanding of the diagnostic process. The interconnection between medical experts and data through Formal Methods can lead to improved decision-making and more effective clinical outcomes.

Another interesting example is the study by Peng et al. [69], which aimed to predict unenhanced lesion evolution on MRI scans. They focused on MRI images from MS patients with radiological activity, manually defining ROIs of the lesions. Surprisingly, my results achieved similar levels of performance as their but without the need for manual lesion segmentation. My method automatically generates ROIs encompassing the entire image, which allows to explore every detail of the image and conduct localization effectively, thanks to the Formal Methods technique. In addition, there is not preprocessing of the MRI images, but future developments might involve them to improve the interpretability of results.
Preprocessing can help to understand the differences between patients with specific conditions, such as MS and healthy individuals. The study by Ardakani et al. [82] is a prime example of how filters can be used to strengthen the association between numerical values derived from radiomic features and the visual radiological evidence within medical images. This process aids in making the diagnostic and decision-

making process more transparent, fostering a deeper understanding of the data and its implications for clinical practice.

Another widely studied disease in the field of Radiomics is *Epilepsy*. My team and I tried to predict the positivity of PET scans in 14 patients based on their initially MRI-negative scans using a combination of Radiomics and Formal Methods. This can be turned into an advantage for patients who do not have to perform multimodal imaging exam. In contrast to our approach, certain studies have focused on utilizing different MRI sequences on the premise that abnormalities in the signal can be correlated with the presence of brain lesions. An example is the work conducted by Beheshti et al. [77], where they investigated the utility Fluid-Attenuated Inversion Recovery (FLAIR) signals obtained from specific brain ROIs in conjunction with ML algorithms. Their primary aim was to distinguish between patients with MRI-negative TLE and healthy controls, while also achieving lateralization in MRI-negative TLE patients. The results demonstrated a 75% correct classification rate, with impressive prediction accuracies for tasks such as discriminating between healthy controls and patients with right TLE (87.71%), healthy controls and patients with left TLE (83.01%) and distinguishing between left TLE and right TLE patients (76.19%). This strategy underscores the potential of the FLAIR sequence as a robust tool for the classification and lateralization of MRI-negative patients with TLE.
In this case study, I do not pinpoint the specific location of the lesion. However, for lateralization Formal Methods can locate the precise *slice* in which the verification agent finds the sign of the disease achieving with similar performance results.
As shown in Table 4.7, the use of MRI produces a number of false positives that is 0, while the number of false negatives is higher. This approach ensures a high level of specificity in the diagnosis, even at the cost of reduced sensitivity.
Formal Methods enabled improvements in classification and localization performance, making them more explainable to healthcare professionals and their colleagues. This increased clarity in the diagnostic process can be immensely beneficial in the field of Radiomics. Although the results achieved are still far from making the method applicable in the clinical practice and despite the limitations about the complexity of the images and the feature selection phase, the use of image preprocessing techniques and the use of segmented MRI images of *White matter* and *Grey matter* instead of raw MRI images, could greatly improve the performance of the methods.

In the Epilepsy field, to the best of my knowledge, the leverage Radiomics and Formal Methods in predicting PET scan results based solely on initially negative

MRI scans in patients with TLE is a pioneering attempt. Furthermore, this approach contributes to the refinement of classification techniques, as highlighted in the Section 3.7. For example, I investigated the impact of applying a filter to the MRI images on the results of our method. The filter led to positive changes (Δ), indicating an improvement in performance. Specifically, I observed an increase in Specificity and Precision for each class in 5 out of 6 cases, primarily because there were no false positives. However, at the same time the classifier returned more false negatives, as indicated by decreased Recall values.
The core idea behind this study was to develop a method that relies solely on the analysis of negative MRI images to predict the results of PET 18F-FDG scans, which are considered the gold standard for diagnosing TLE in MRI-negative patients. Since PET scans can be expensive and not always accessible, having a method that depends only on MRI scans could potentially reduce costs for national health systems and expedite the diagnostic process.
In literature, the DL framework proposed by Zhang et al. [83] analysed 201 pediatric patients with TLE and achieved accurate detection of *epileptic foci* from 18F-FDG sequences, with an AUC for the classification of 0.93. They showed the effectiveness of the DL framework in accurately identifying *epileptic foci* and metabolic levels, surpassing the performance of other existing methods. While Zhang et al. use a Gaussian kernel filter of 8 mm, the script described in Section 4.2 has $\sigma = 2$, but both show positive impacts on performances. Unlike my study, it is not described whether IBSI-compliant features are used and how segmentation is performed.

Limitations about this study is the relative *small number* of patients included in the study, which may not be sufficient to generalize the results. A larger and multicentric dataset is required to validate and strengthen the proposed method.
Secondly, while the results are promising, they may not meet the standards necessary for *clinical adoption*. There is a need for further research to enhance the performance of the method: for example, the exploration of different filtering approaches to address the issue of image noise, which can significantly impact the method's results. In summary, future studies should focus on expanding the patient dataset and refining the method to make it more robust and clinically applicable. These efforts will contribute to the development of a more effective diagnostic tool for TLE in MRI-negative patients.

Formal Methods offer a significant advantage in the field of Radiomics by providing a mathematical and explainable framework for analyzing medical images; they can help identify repeated actions, as illustrated in Figure 4.2, that can be linked

either directly to radiological evidence or to the presence of a particular disease. For example, if a ROI is segmented along the edges of an organ, the association between the feature values and the ROI precisely aids in the localization of the disease within the image.

Chapter 5

Breast

This chapter focuses on the classification of Breast Cancer using the BI-RADS system with the methodology of Radiomics and Formal Methods.

5.1 Early prediction of BI-RADS risk stratification through Radiomics and Formal Methods

Breast cancer (BC) stands as the most prevalent neoplasia among women across the globe [84] and it is the leading cause of cancer-related deaths in Europe. In Italy, the year 2021 witnessed a tragic balance of 12.500 lives claimed by this disease; surprisingly, the year 2022 brought an estimate of 55.700 new BC diagnoses among women, contributing to an overall prevalence of 834.200 women struggling with this condition. This increase in BC prevalence can be attributed to advancements in early detection, evolving risk factors and the aging of the population [85].
The timely prediction of prognosis and treatment response remains a central focus in both clinical practice and research, especially for BC since it often follows a protracted clinical course, with various therapeutic options. Effective prognosis and treatment response prediction can play a crucial role in substantially reducing BC mortality. Traditional screening methods rely on mammography, complemented by conventional US.

In 2003, the fourth edition of the *Breast Imaging Reporting and Data System* (BI-RADS) atlas, which standardized the diagnostic characterization of US-detected

breast lesions, was published by the American College of Radiology. The most recent edition of this atlas was released in 2013 and comprises:

- BI-RADS 0 is a diagnosis that needs to be combined with other imaging;
- BI-RADS 1 is defined as no lesions or negative findings;
- BI-RADS 2 is defined as a benign lesion without suspicious characteristics;
- BI-RADS 3 is possible benign lesion with less than 2% malignant probability;
- BI-RADS 4 is defined as a suspicious lesion with 2% to 95% malignant probability that is recommended for biopsy;
- BI-RADS 5 is a highly suspected malignancy, with more than 95% malignant probability;
- BI-RADS 6 is a known malignancy or pathologically proven to be malignant.

Due to the wide range of malignancy probability, category 4 is further divided into three subcategories: 4A, 4B and 4C, with malignancy probabilities of 2–10%, 10–50% and 50–95%, respectively [84]. MRI offers the advantage of 3D breast imaging without radiation exposure and without the need for breast compression, which can be uncomfortable for patients. It also provides coverage of breast areas near the chest wall and armpit. However, MRI is not currently a part of BC standard screening programs, but is a supplementary screening in high-risk women that enhances early detection and ultimately improves BC prognosis. A valuable aspect to consider is the evaluation of the relative quantity of fibroglandular tissue and, simultaneously, Breast Parenchymal Enhancement (BPE) using MRI. One potential approach is to utilize breast density measurements obtained through MRI to identify women who should commence early annual MRI screening.
Historically, the high costs associated with MRI have limited its widespread use. However, the development of more concise protocols has the potential to mitigate these costs and make MRI screening more accessible, particularly for medium-risk women with dense breast tissue. Recent studies have underscored the limitations of subjective visual estimations of breast density, which exhibit considerable variability among different observers and even within the same reader. These challenges have prompted a search for objective methods, similar to mammography, to improve the accuracy and consistency of breast density assessment [86].
Radiomics in breast MRI images can assist less experienced radiologists in accurately

assessing breast lesions, thereby reducing the risk of both underestimation and overestimation. Subsequently, I devised a Formal Method that integrates Radiomics to predict the BI-RADS score, which is indicative of the malignancy level of breast lesions.

Early diagnosis refers to the identification of a disease in its initial stages, ideally before it reaches an incurable state. This is particularly crucial in BC prevention [85], where detecting a potential tumor in time is essential for initiating less invasive and effective treatments, ultimately leading to successful outcomes. Preventive measures typically involve specific examinations, such as breast examinations [87]. While these clinical assessments are essential for evaluating possible issues, they may not provide a precise diagnosis. To achieve a more accurate understanding of a patient's clinical status, additional diagnostic tools are employed. Mammography [88], a non-invasive X-ray procedure, plays a significant role in investigating potential breast pathologies, allowing the detection of tumor-like lesions characterized by irregular margins or distorted areas. In cases of diagnostic uncertainty, healthcare providers may perform needle biopsy or percutaneous biopsy to obtain a more comprehensive clinical picture.

In this case study, 169 women underwent *breast MRI examinations* at the conventional radiological center "Potito" in Campobasso, Italy, between January 2020 and July 2022. These patients all underwent standardized breast MRI examinations using a 1.5T MRI Philips Ingenia scanner, which were reviewed and analyzed by Dr. Caiazzo, a highly experienced breast radiologist with over 35 years of expertise. The selection of patients comprised individuals diagnosed with breast pathologies based on the results of their MRI examinations conducted in different phases of sinus compression. The MRI examinations involved both pre-contrast and post-contrast imaging, with the administration of a contrast medium for enhanced visualization. The dataset is composed by patients with different diagnoses by characterisation and tumour stage, classified into 4 distinct categories of the BI-RADS system (from B2 to B5-B6 unified). The dataset was reduced from 169 patients to the number of 99 due to problems during the Feature Extraction.
After a thorough analysis of the value distributions, categories B2 and B3 were merged by labelling them as the *"Low Risk" category* and categories B4 and B5-B6 as the *"High Risk" category*. This consolidation aims to differentiate healthy patients or those with a low risk of developing cancer from those with a high risk.
During the feature selection, only a few features exhibited significant differences; *Busyness* and *Strength* from the NGTDM class, *Sphericity* and *Maximum 2D DiameterscColumn* from the SHAPE class, *Difference Average* from the GLCM class. In

this case, the models created are mixed and are interclass, i.e. several features of different classes are grouped in the same model. With these chosen features, we were then able to create models for each patient and to conduct formal verification in order to obtain the diagnostic results.

In collaboration with Dr. Caiazzo at the Radiomics Laboratory of University of Molise, the formal property for the classification was formulated signifying the presence of a tumor mass within the breast. Namely the property is:

- *Satisfied*: Signifying a high risk of tumor presence, including patients categorized as B4 or B5-B6;
- *Not Satisfied*: Denoting a low risk of tumor presence, encompassing patients classified as B2 or B3.

The qualitative assessment of the model's classification performance is carried out using the *Confusion Matrix*, obtaining on one side 23 True Positives and 10 False Negatives, on the other side 38 True Negatives and 6 False Positives. The calculated metrics were applied to a total of 77 patients, a reduction from the initial 99 by randomly excluding 22 patients of the B2 score, with the sole purpose of creating a balanced dataset for each category. The Confusion Matrix yielded positive results with excellent indexes, especially in terms of Precision and Recall, as visible in Table 5.1.

Accuracy Statistics	**Value**
Accuracy	0.792
Precision	0.793
Recall	0.697
Specifity	0.863

Table 5.1: Accuracy and other indexes to gain more information about the clinical usefulness of the methodology.

5.2 Related Works and Discussion

The BI-RADS classification system plays a significant role in assessing the probability of breast malignancy. Its importance lies in optimizing patient care by ensuring that follow-up examinations and procedures are neither overused nor underused [89]. To achieve this goal, it is also important for women to be well-informed about the interpretation and management strategies associated with each BI-RADS category.
The introduction of AI into medical diagnostics, while promising, can pose challenges related to the familiar and explainable nature of the diagnostic process. Many AI models are inherently complex, making it challenging to provide straightforward explanations for the diagnostic outcomes they produce.
In essence, while AI has the potential to enhance medical diagnostics, it also raises important considerations about transparency, interpretability and ensuring that patients can understand and trust the diagnostic process, especially in fields such as BC diagnosis [4].

In the past, CAD software often provided unsatisfactory results in the field of medical imaging; now, with the emergence of Radiomics, the approach is transforming itself. An example is the study conducted by Fleury et al. [90], where they manually segmented 206 lesions and analyzed them using ML. While their work did not adhere to the IBSI standard, as opposed to my work, they achieved notable results including a 0.84 AUC, a 71.4% Recall and a 76% Specificity. My methodology achieved similar results without the need for manual segmentation and using only a few features per patient, showing to be on par with published data from other studies focusing on BI-RADS categories in mammography or US. For these reasons, the potential of Radiomics and Formal Methods is to deliver comparable performance to traditional methods.

Radiomics and BI-RADS classification system can indeed be considered equivalent in terms of performance. In the study conducted by Perre et al. [91], which involved 117 MRIs following the IBSI standard and manual segmentation, 15 features were analyzed, resulting in an AUC of 0.88, in particular thanks to the significance of texture-related features in the analysis.
My methodology achieved similar levels of performance without the need for manual segmentation. Furthermore, through the use of Formal Methods, the model was reduced to just 5 features, providing a more straightforward understanding of the connection between numerical values and clinical practice. This approach aims to make medical experts, including physicians and radiologists, more comfortable with the implementation of radiomic techniques.

The use of MRI images and textural features in this study is consistent with findings in the literature, which underscore the importance of texture features in various studies for the detection of BC. These results not only validate the effectiveness of Radiomics but also highlight its potential to simplify and enhance the diagnostic process [92,93].

Chapter 6

Shoulder

This chapter delves into the problem of Adhesive Capsulitis in radiology, for which an experienced radiologist is required. The methodology will be applied, plus a comparison with the performance of radiologists with different types of experience.

6.1 Early Diagnosis of Adhesive Capsulitis: An Approach Leveraging Model Checking for Enhanced Precision

Adhesive capsulitis (AC), more commonly referred to as *frozen shoulder*, is a condition marked by inflammation and thickening of the capsule surrounding the shoulder joint [94]. This condition causes pain and restricted movement, limiting the shoulder's range of motion. While the precise cause of AC is frequently unclear, it may be linked to factors like diabetes, prior shoulder injuries or extended periods of immobilization. The usual approach to managing this condition includes physical therapy, pain relief measures and, in certain instances, medical imaging and interventions to relieve symptoms and regain shoulder functionality.

AC can be difficult to identify for several reasons. Firstly, the onset is gradual and symptoms may only be felt when the AC has already progressed and the severity of symptoms and speed of progression may vary between individuals. Unfortunately, the symptoms are similar to other conditions and may therefore be underestimated: people may not seek medical attention for mild shoulder discomfort, especially if they

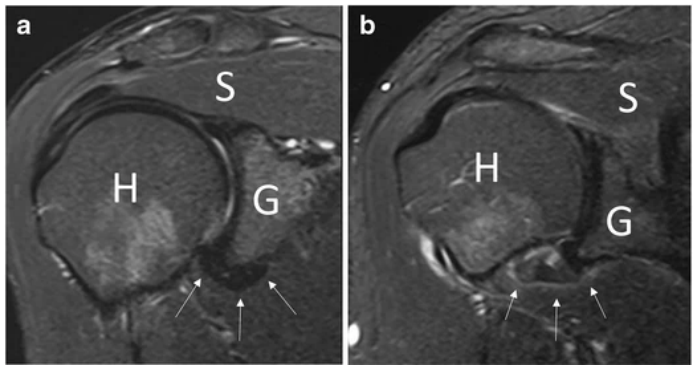


Figure 6.1: Coronal MRI image weighed in T2 saturated with fat. Image *b* shows thickening (up to 10mm) and hyperintensity of the signal of the glen-humeral ligament complex of the axillary sac with hyperintensity/edema of adjacent soft tissues [97].

attribute it to normal ageing or overuse. All of these reasons contribute to delaying the diagnosis of AC.

Radiological evidence is also difficult for an inexperienced radiologist to detect. With MRI (shown in Figure 6.1), images with excellent contrast between soft tissues and high spatial resolution can be generated in every direction, which is why this technique is one of the key methods for differentiating shoulder diseases [95]. The most obvious signs from MRI are the increase in the signal strength of the lower glen-humeral ligament; the thickness of the humeral coraco ligament and the appearance of the subcorachoid fat triangle [95, 96].

The first research question for this study *"Is it possible to diagnose AC through the use of Model Checking and radiomic features?"* Explore the feasibility of diagnosing AC through an automated methodology, allows for investigating the usefulness of radiomic features in musculoskeletal radiology. This research question seeks to develop an automatic approach that employs radiomic features extracted from MRI scans to create a formal model, which is then utilized to verify if a patient is afflicted by AC using Model Checking and formal properties.

"To what extent does the proposed approach for AC diagnosis perform in terms of Accuracy when compared to traditional diagnostic modalities?" This second question aims to assess the effectiveness of the proposed diagnostic tool concerning its Accuracy compared to traditional diagnostic methods, such as those performed by radiologists with varying levels of expertise. The objective is to evaluate how well the tool performs and to identify potential applications.

This study applies my innovative methodology for the automatic classification of AC patients; subsequently, through a survey, radiologists evaluate the MRI images to determine whether patients have AC or not and the results are compared with those obtained from the automated tool to assess its performance and possible use cases. In particular, the dataset was provided by Professor Marcello Zappia and includes DICOM images from 55 patients who underwent MRI scans between 2014 and 2022 at the "Istituto Diagnostico Varelli SRL", among whom 32 have been diagnosed with AC, while 23 have not. Although there were variations in MRI protocols among patients, the selection process focused on MRI images in both coronal and sagittal planes. For this study, the ROI analysis was performed automatically on the entire slice. On average, 22 slices per patient were analyzed, covering both coronal and sagittal acquisitions.

Feature Selection is done with Orange [34], selecting 5 features per radiomic class: Tables 6.1 and 6.2 provide a comprehensive list of all features for each class that was part of our experimentation.

FIRST	**GLCM**	**GLDM**
Energy	Cluster Prominence	Dependence Variance
Median	Contrast	Gray Level Non Uniformity
Minimum	Difference Average	Gray Level Variance
Total Energy	Difference Variance	High Gray Level Emphasis
Variance	Id	Large Dependence Low Gray Level Emphasis

Table 6.1: Most relevant features for FIRST, GLCM and GLDM classes.

In the final step, the Model Checker is employed to automatically validate the formal properties against the patient models: a "TRUE" signifies that the patient is afflicted with AC; conversely, a "FALSE" indicates the absence of the disease.

GLRLM	**GLSZM**	**SHAPE**
Gray Level Non Uniformity	Gray Level Non Uniformity	Flatness
Gray Level Variance	Large Area Emphasis	Least Axis Length
High Gray Level Run Emphasis	Large Area High Gray Level Emphasis	Mesh Volume
Long Run High Gray Level Emphasis	Large Area Low Gray Level Emphasis	Surface Area
Run Length Non Uniformity	Zone Variance	Voxel Volume

Table 6.2: Most relevant features for GLRLM, GLSZM and SHAPE classes.

6.2 Radiomics vs. Radiologists

To estimate the effectiveness of the automatic diagnosis tool, we approached a total of 11 professional radiologists from the Department of Medicine and Surgery at the University of Molise to participate in this study. The invitations to these radiologists are accompanied by the explanation of the objectives and emphasizing that (i) their involvement was entirely voluntary, (ii) any personal data would be treated with the utmost confidentiality and (iii) the estimated time required for patient categorization was approximately five minutes, with the option for participants to withdraw at any point. The survey was conducted via email and the entire process, from selecting participants to sending invitations and gathering responses, spanned approximately two months.

Among the participants, 4 were categorized as having a Beginner Level (comprising Master's Degree students), 3 held an Intermediate Level (early professionals) and 4 were classified as having an Expert Level (professional radiologists). Therefore, to address my research question, we included three groups of radiologists categorized by their level of experience and all participants were asked to evaluate, based on MRI images, whether patients were affected by AC.

Below in the Table 6.3 there are summarized the results of the different experts and the proposed tool; the tool has superior performance to the group of beginners and intermediates, but comparable to an expert radiologist.

	Accuracy	**Precision**	**Recall**	**Specificity**
Beginners	0.54	0.60	0.56	0.51
Intermediates	0.53	0.61	0.58	0.45
Experts	0.72	0.77	**0.76**	0.65
Tool	**0.74**	**0.82**	0.72	**0.78**

Table 6.3: Summary of participant performances in comparison with the methodology.

6.3 Related Works and Discussion

In the current literature, there is little research focusing on the application of Radiomics in diagnosing AC. The majority of studies primarily concentrate on establishing correlations between AC and shoulder pain among cancer patients. As a matter of fact, AC frequently manifests in cancer patients, often occurring during chemotherapy or following surgical procedures (e.g. breast cancer). In a study by Hayashi et al. [98], a cohort of patients underwent ^{18}FDG PET-CT scans, during which SUV parameters were measured within manually delineated ROIs on transaxial images. The analysis of radiotracer uptake revealed that symptomatic patients exhibited higher SUVmax and mean SUV values compared to asymptomatic patients, highlighting the significance of considering AC as a noteworthy factor in cancer patients, particularly when heightened metabolic activity is observed, as noted in prior research by Yang [99].
These imaging techniques are employed because CT scans are effective in identifying inflammatory diseases, allowing for the detection of hypermetabolic activity (active inflammation) in the shoulder joint through PET-CT scans. In fact, AC patterns have been characterized in a study by Kim et al. [100], involving 21 patients, where SUVs were assessed within four distinct ROIs, revealing four distinct patterns associated with AC. However, it is important to acknowledge that a limitation of using PET-CT is that these patterns are not entirely unique, as other medical conditions can sometimes exhibit similar radiographic findings, as discussed by Salem [101].

In my work, I studied 55 patients to diagnose the presence of AC, without the information about the presence of cancer. The aim is in fact to efficiently detect radiological evidence of capsulitis, which is sometimes invisible even to experienced radiologists (Table 6.3).
To the best of my knowledge, my team and I are the first in this field to provide a technique for automatically classifying ACS ("Frozen Shoulder"). Triveni et al. explored image Feature Extraction for assessing shoulder pain [102]. It is important to note that shoulder pain does not denote "frozen shoulder", however this cited work demonstrates the potential effectiveness of GLCM features. They propose a segmentation method based on region and watershed applied to MRI images for discerning shoulder disorders. Their investigation overlaid feature extraction and deep feature extraction techniques to comprehensively analyze shoulder disorders and the nature of shoulder pain. The GLCM was identified as being adept at identifying the specific area of pain onset and its type. However, their study did not classify AC patients; rather, it determined which algorithm produced a more accurate segmentation of the

shoulder. While they explored various types of features, our study relies solely on textual GLCM values without the need for artificial intelligence techniques.
Moreover, the effectiveness of this class of features is confirmed by Scott et al. [103]. The experts state that GLCM-based segmentation showed promising results for tendons in ultrasound images of patients with shoulder pain. The discrepancy in observed values between healthy and tendinopathic tendons emerges as a potential tool to objectively assess damage in tendinopathy. Although our study is focused on AC rather than tendinopathy, the examination of GLCM in MRI confirms the efficacy of this class of features, which is equally applicable in MRI as demonstrated in US imaging.

Chapter 7

Lungs and COVID-19 disease

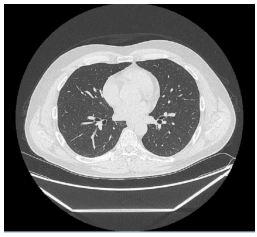
This chapter studies the problem of COVID in depth, both by applying the classical and matrix methodology on DICOM without external graphic software and then reducing the radiological examination in the number of slices. We also studied how some recently COVID-positive patients could present Perfusion Deficit detectable on SPECT, applying the methodology in the field of nuclear medicine. While conducting the studies, there was an opportunity to delve into a case related to the interpretability of the results concerning certain features.

7.1 On the Adoption of Radiomics and Formal Methods for COVID-19 Coronavirus Diagnosis

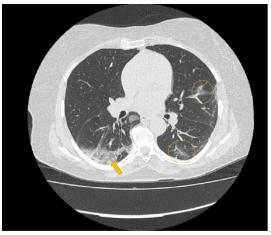
Due to the huge pandemic caused by the novel *Coronavirus*, in the 2021 there was a pressing demand for rapid and automated diagnostic methods. In the Radiomics Laboratory of the University of Molise, my team and I introduced Formal Methods in a two-tiered approach, which have not been previously employed in this context.
It is well-known that every medical image is a combination of grey-levels that produce the morphology of the image. In this specific case, the attention is focused on *High Resolution CT images* (HRCT) of lungs: the external part of a healthy lung would be white, while the inner part would be black. In this black part there could be different movements that display normal respiration and the performance of vital functions. Through these images, in the black section of the lung, radiologists can note some

white spots that can be marked as disease symptoms, but they would not know to which disease the marks refer. For example, in Figure 7.1 there are three different HRCT belonging to three different patients: one image has "very black" lungs, while the others have some white spots. As a matter of fact, the Figure 7.1a is an HRCT of a healthy patient and it does not show white spots or pneumonia signs. Instead the Figures 7.1b and 7.1c are similar, because they both show some disease marks: the first is from a COVID-19 patient, while the second is from a lung disease patient. Thanks to this, it is clear just how difficult it is to see the differences between a general pneumonia and Coronavirus disease with the naked eye.

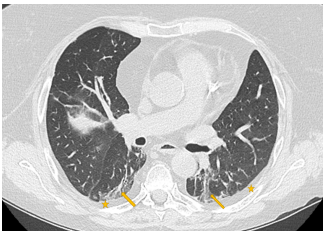
This approach aimed to accomplish the *following objectives*: (i) determine whether a patient's lungs are healthy or show signs of a general pulmonary infection; (ii) if the previous tier detects a general pulmonary disease, identify whether the patient is afflicted by the novel Coronavirus disease.
The experiment uses a dataset of HRCT images collected by the authors, achieving an Accuracy rate exceeding 81% in disease detection. To better facilitate the classification between healthy and sick patients, the methodology is split into two steps, as shown in Figure 7.2. If, in the first phase, the classification is about healthy or sick patients (affected by any lung disease), in the second phase, the task is to classify COVID-19 or lung disease patients. Consequently, there is a different model and property for each step. In the first phase, if the patient is sick, it pass to the second step, otherwise the patient is healthy and the method can be stopped because there are not Coronavirus signs. In the second step, hence, if the property is true, the patient is affected by COVID-19, otherwise the patient is a lung disease patient.



(a) HRCT of a healthy patient: disease marks are absent in the image.



(b) HRCT of a COVID patient: yellow arrows indicate the main signs of the disease.



(c) HRCT of a lung disease patient. Yellow arrows indicate the main disease marks.

Figure 7.1: Three different medical images of three different patients: a healthy one, a COVID-19 one and a lung disease patient.

FIRST	GLDM	GLRLM	GLSZM
Energy	Dependence non uniformity normalized	Gray level variance	Gray level variance
Interquartile range	Gray level variance	Run percentage	Large area emphasis
Skewness	Large dependence emphasis		Size zone non uniformity normalized
Total energy	Large dependence low gray level emphasis		
Variance			Zone variance

Table 7.1: Radiomic feature set for the first tier (or step 1).

FIRST	GLDM	GLRLM
90 Percentile	Dependence non uniformity	Gray level non uniformity
Kurtosis	Gray level non uniformity	Long run emphasis
Minimum	High gray level emphasis	Run length non uniformity
Skewness	Large dependence emphasis	Short run high gray level emphasis
Variance	Small dependence high gray level emphasis	

Table 7.2: Radiomic feature set for the second tier (or step 2).

The Feature Selection was done manually to choose the best 5 features to solve this classification problem, as shown in Table 7.1 and 7.2. In Radiomics, there is no standard value to identify healthy or sick patients. For reasons of coherence and clarity, I show in Table 7.3 an idea of the average values gained from PyRadiomics in each medical status category.

Each HRCT exam has a number of slices of around 200–600 images. In the database used for this study, there are 35 patients divided into three categories according to their medical status: nine healthy patients, six lung disease patients and 20 Coronavirus disease patients. Each COVID-19 patient had a positive RT-PCR test; on the contrary, the other patients have a negative RT-PCR test.
How is natural there is more similarity between two sick patients than between a healthy one and a sick one, consequently in the first step the differences between pa-

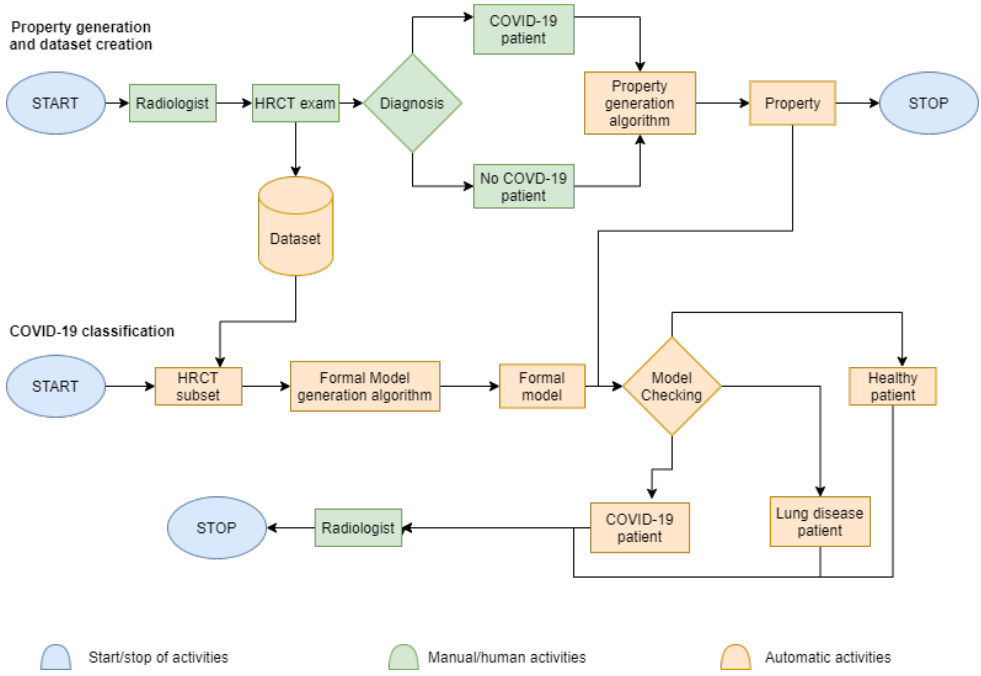


Figure 7.2: Diagram of the manual and automatic activities for the earlier identification of COVID-19 or Lung disease patients. The initial process outlines the conventional procedure that a patient must follow to undergo an HRCT exam and receive a diagnosis. In contrast, the second process utilizes the insights obtained from the first process (comprehensive dataset and the expert knowledge of the radiologist) in an attempt to replicate the accurate diagnosis.

tients are really discriminant (Table 7.4); otherwise, in the second step, the differences between medical status categories decrease (Table 7.5).

As summarized in Table 7.4, in the first tier, for the classification between healthy and sick patients, an Accuracy greater than 82% for the FIRST-Order and for the GLRLM and GLSZM matrices is achieved. In Table 7.5, the classification concerns COVID-19 and non-COVID-19 patients; therefore, the results do not exceed those of the first tier, but there is an improvement in terms of number of false negatives (Recall or Sensitivity). In this latter case, the most important radiomic classes are the GLDM and GLRLM matrices.

Features	**Healthy**	**COVID-19**	**Other Disease**
90 Percentile	-	−144.22	−15.00
Kurtosis	-	5.88	3.84
Minimum	-	−1146.39	−1023.00
Skewness	0.45	1.17	1.09
Variance	260 000.20	201 752.40	216 780.40

Table 7.3: Average values for healthy, COVID and lung disease patients. For healthy patients, there are only two parameters because these are the most discriminant for distinction by sick patients. In fact, these five features belong to the second-tier classification (see Tables 7.1 and 7.2).

Feature Class	False Positive	**Precision %**	False Negative	**Recall %**	**Accuracy %**
FIRST	0	100	6	81.25	82.86
GLDM	1	96.15	6	80.64	80
GLRLM	1	96.15	5	83.33	82.86
GLSZM	2	92.31	4	85.71	82.86

Table 7.4: Step 1: Results of the comparison between healthy and lung disease patients.

Feature Class	**False Positive**	**Precision %**	**False Negative**	**Recall %**	**Accuracy %**
FIRST	4	80.95	3	85	78.12
GLDM	3	85	3	85	81.25
GLRLM	5	79.16	1	95	81.25

Table 7.5: Step 2: Results of the comparison between general lung disease and COVID-19 patients.

7.2 Computational cost of CT Radiomics workflow: a case study on COVID-19

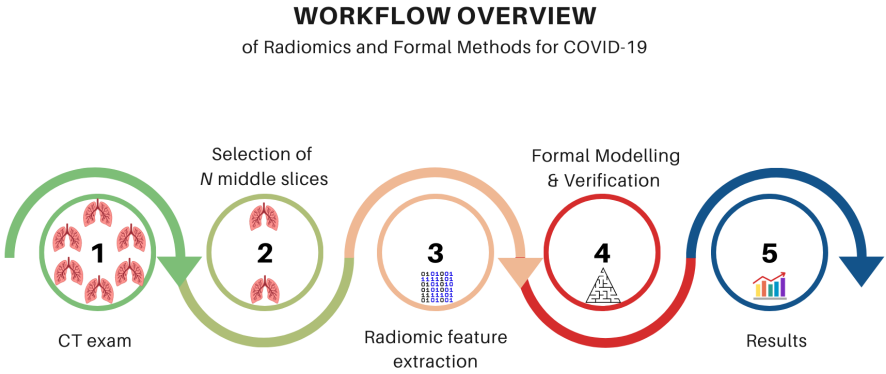


Figure 7.3: Starting from the CT exam, the 30 middle slices are selected, from which the radiomic features are extracted. Through temporal logic, models are created and checked. For the "Full Exam" study, the step 2 will not be performed.

Once studied the entire radiological exam, with 600 slices per patient, I had different questions about the diagnosis: "*is it possible to discriminate between healthy and COVID-19 patients only by using any central CT images of the radiological exam?*". This question is well-founded when using the methodology of Formal Methods, because they do not need training and testing steps, but they describes the disease in a rigorous formal way. Indeed, during the classification process, I asked what number of images the classifier needs to correctly recognize a COVID-19 patient.

The structure of this methodology is similar to the classic radiomic workflow described in Section 3, as shown in Figure 7.3. In addition, for each patient, only some center slices of the CT scan are chosen as the basis from which the radiomic features are retrieved. The "Full Exam" study will not include step 2 in its methodology.

The dataset consists of 53 patients of three different categories: 36 patients are COVID-19 patients, 9 healthy patients and 8 patients with different lung pneumonitis. The imaging modality refers to HRCT scan of each patient in the axial plane. The dataset is not publicly available due to privacy restrictions.
To answer the previous research question, only *30 intermediate slices* were considered for each patient. The slices were selected looking for those where is a better visualization of the lungs (the central slices, where the lung is completely open), going to

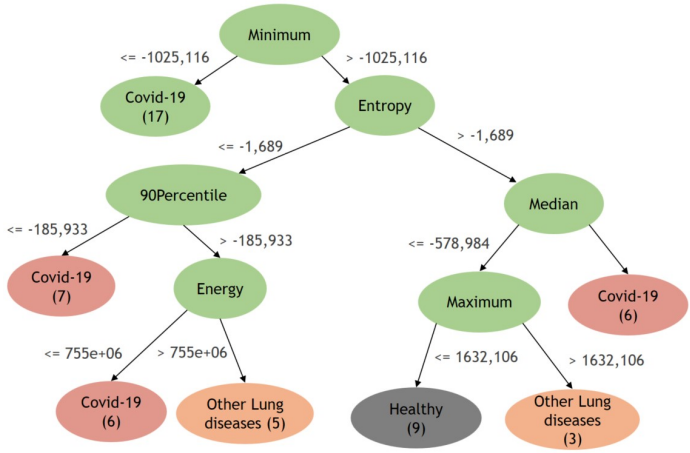


Figure 7.4: This is the visualisation of a tree decision algorithm. In this case of Selection with Weka, the feature selected are those closest to the root: Minimum, Entropy, Median, Energy.

evaluate whether the method is able to correctly diagnose the disease. The method proposes to choose only a few central slices because on the one hand, the radiologist would be favoured to check the outcome of the automatic diagnosis by studying a reduced set of the total radiological examination; I know that, on the other hand, the diagnosis could be wrong or incomplete due to the reduced information.

Firstly, I exploit the classic methodology on the entire radiological exam (with 400 slices on average). In a second moment, I modified the CT exam maintaining only 30 consequent slices in the middle of the axial series. After having created this sort of "*reduced exam*", the diagnoses are repeated maintaining the same dataset and features.

Exploiting PyRadiomics, the attention is focused on FIRST-Order Features and the Feature Selection Step was possible thanks to Weka [104]: using a decision tree, the "Best first" method was used to evaluate the attributes, which returned the "Minimum" feature as the most relevant feature (Fig. 7.4).

Model Checker agent [29] must answer to two different questions: the first question is "*Is the patient healthy or sick*?", while in the second step I ask if the patient has radiological evidences typical of COVID-19. In this way, the designed idea of classification is a double approach: first, the classifier decides if the patient is healthy or not. Consequently, if the patient is healthy, it stops the execution; otherwise, it tries to diagnose if it is a COVID-19 patient or a lung disease patient. This cascade

methodology is described in Section 7.1.
The comparison is done between the *Full Exam* of average 400 slices and the *Reduced Exam* obly composed by the middle 30 slices. The results of the different diagnosis are displayed in percentage in the following Tables 7.6 and 7.7, only for demonstration purposes of the performances with one complete examination and one only focused on the central slices.

Regarding the first step of classification between sick and healthy patients, the performances remain more or less stable: there is an actual decrease in Precision and a stability in Recall values. This means that by decreasing images, the rate of false positives rises while false negatives remain unchanged. Thus more healthy patients being classified as sick: thus the computational problem becomes a problem for the national health system, which has to delve into the exams of more patients and incurring more expenses.
In the second step, then in the classification between patients with COVID-19 and with general lung pneumonia, it can be said that the performance reverses. In this case, all indexes drop as many percentage points, while Precision and Specificity are the only stable indexes. Probably, reducing the CT exam, the mathematical rule created by computer scientists and physicians turns out to be more generic and is not able to encompass several radiological cases of COVID-19.

	Accuracy	**Precision**	**Recall**	**Specificity**
Full Exam	86.8	100	22.2	100
Reduced Exam	84.9	66.7	22.2	97.7

Table 7.6: Results for step 1, in which it is possible to distinguish between healthy and sick patients (Covid-19 and Lung disease patients).

	Accuracy	**Precision**	**Recall**	**Specificity**
Full Exam	45.1	100.0	34.8	100.0
Reduced Exam	20.0	100.0	4.7	100.0

Table 7.7: Results for step 2, in which I distinguish between COVID and lung disease patients.

I have answered the initial research question: focusing the radiomic model only on one part of the CT study does not lead to better performances; on the contrary, it may significantly lower performance. This is again an affirmation that Radiomics sees "what the human eye does not see"; evidently, there is important information in the other images of the CT sequence for the recognition of COVID-19 patients.

In addition, my team and I state the performance stability is guaranteed by Formal Methods, which do not learn from the data but merely describe its behavior: this is why Formal Methods are more robust than other AI techniques.
The verification of properties is done by a Model Checker, which offers the additional advantage of providing a counter example in case the property is not verified, making the model more understandable and explainable. Numerous studies have been conducted in the fields of Cybersecurity and Bioengineering domain [26, 105], where it has been noted that the use of Formal Methods leads to a net increase in performance.

7.3 Evaluating lung perfusion SPECT/CT imaging in patients with COVID-19 through Radiomics and Formal Methods

Figure 7.5: The assumption is that different health states can result in different LTS.

The sequence of inflammatory reactions observed in individuals with COVID-19 could potentially result in *pulmonary embolism* (PE), which may worsen the overall prognosis. Lung perfusion SPECT/CT (Q-scan) scans performed on symptomatic patients who have been discharged following a COVID-19 infection can be used to verify the presence or absence of pulmonary vascular complications. The objective is to explore the novel approach that combines Radiomics and Formal Methods, offering a virtual secondary assessment capable of identifying perfusion irregularities, thereby enhancing the precision of patient-specific diagnostic and therapeutic strategies.
A group of 23 patients, all of whom had recently recovered from COVID-19 and had no prior lung-related conditions (such as lung cancer or abnormal CT scan findings like lung bullae), were included in the Q-scan study due to persistent shortness of breath one month after their discharge. Through visual and semiquantitative analysis of their SPECT exams, they were categorized into two groups: those with no signs of lung perfusion abnormalities (14 patients) and those with evident abnormalities (9 patients). The evaluation of pulmonary lobar perfusion (given as cts/volume % for each lobe) was conducted using GE Healthcare's Q-Lung software.
In a comparative assessment alongside the *Q-Scan findings*, the model exhibited congruent characteristics in 13 out of 23 patients. It successfully detected perfusion abnormalities in 8 out of 9 patients who had a positive Q-Scan and it ruled out perfusion abnormalities in 5 out of 14 patients with a negative Q-Scan. Discordant outcomes were observed in the remaining 10 out of 23 patients, especially in those who initially tested negative. However, in this subgroup, the *semiquantitative analysis* using

Q-Lung revealed perfusion defects measuring less than 10% per lobe, which was considered within the normal range but may warrant further evaluation.
While data is still preliminary and based on a limited patient sample, this Formal methodology demonstrated promising agreement with the Q-Scan results. Further analyses, including the incorporation of co-registered CT data, are needed to enhance the robustness of the approach.

7.4 Explainability in COVID-19 classification

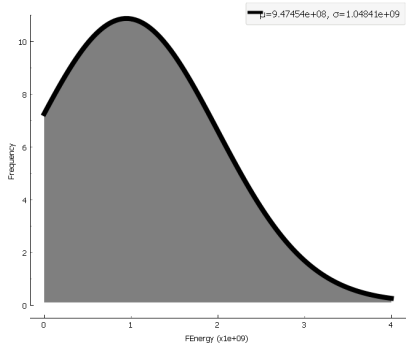
7.4.1 Differences in values distributions

Formal Methods are understandable to both medical professionals and Computer scientists or Radiomics experts, also thanks to the incorporation of human activities in the workflow. By examining the diagnostic information derived from the application of Formal Methods to COVID-19 patients and the analysis of their corresponding CT scans, the goal is to *establish connections between numerical values and radiological evidence*, with the hope of establishing the basis for the implementation of a Radiomic standard to improve the understanding of lung diseases.

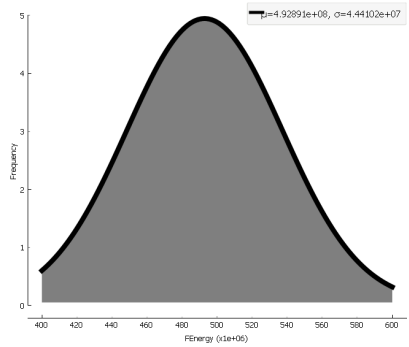
Starting from some features of the FIRST-Order used for the automated diagnosis of COVID-19 patients (Energy, Interquartile Range, Skewness, Total energy, Variance, 90 Percentile, Kurtosis and Minimum) as in Section 7.1, their values distributions were studied as Gaussian graphs with annotations regarding the mean and the corresponding standard deviation. In the analysis between healthy and COVID-19 values, it can be seen that *Energy*, *Minimum* and *Variance* values in COVID-19 patients were lower than in healthy patients. On the contrary, the remaining features like *Skewness* and *Kurtosis*, were higher in COVID-19 patients than in healthy patients.

Energy [20] is a measure of the size of voxel values and is often associated with volume measurement, but in this case the formula does not consider spatial dimensions. Energy is also an indicator of the color, brightness and size change rate of gray values in local areas, which essentially represents the uniformity of an image. In this context there are different value distributions between sick and healthy classes, as illustrated in Figure 7.6. The figure shows the distribution curves and includes annotations for mu, representing the average value and $sigma$, which stands for the standard deviation; the name of the function is plotted on the x axis, while the y axis represents the frequency of the values.

Kurtosis and *Skewness* measurements are statistical estimations that assess the shape of a distribution and quantify how much it deviates from a Gaussian (normal) distribution, as demonstrated in Figure 7.7.
In particular, Kurtosis indicates whether the distribution is higher or flatter than a normal distribution; namely, it evaluates the uniformity of the image through the shape of the distribution curve. In lung studies [107], a high Kurtosis value suggests a higher degree of tissue heterogeneity and microstructural complexity, while a low Kurtosis value indirectly reflects a higher degree of cross-linking and tissue uniformity,



(a) Energy value distribution for COVID-19 patients.



(b) Energy value distribution for healthy patients.

Figure 7.6: These histograms show the total distribution of Energy values for COVID-19 patients (on the left side) and for healthy patients (on the right side). The curve of healthy patients is more similar to a Gaussian curve than the curve on the left; in addition, the values of healthy patients are higher than ones of COVID-19 patients.

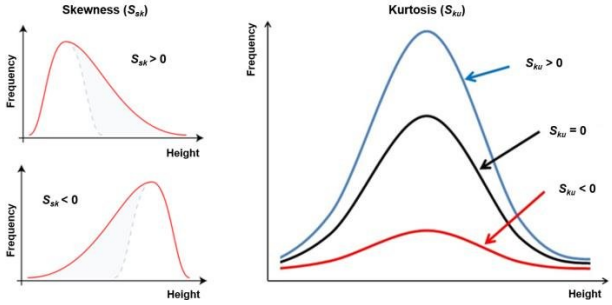
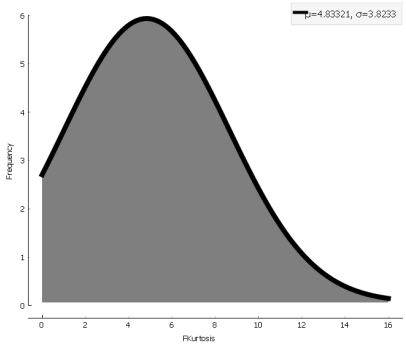


Figure 7.7: Value distribution measured with skewness and kurtosis [106].

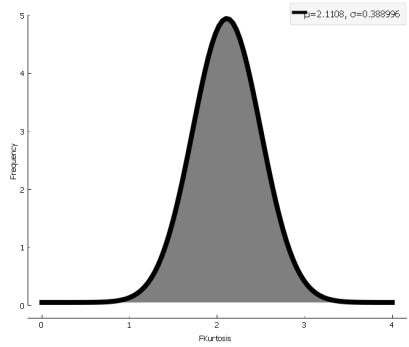
like a honeycomb pattern. These variations are directly evident in histograms, which show distinct behaviours between different categories of patients, as in Figure 7.8.

Skewness measures the degree of asymmetry in the histogram distribution. A positive Skewness value indicates a positively distorted distribution, where many values are concentrated on the bottom, while a negative Skewness value represents a negatively distorted distribution with many values concentrated on the top, as shown in Figure 7.7. When the distribution is symmetrical and follows a normal Gaussian pattern, the Skewness is zero. In this case study, the Skewness distribution curves are represented in the figure 7.9.

Two significant features used in the diagnosis that consistently showed higher



(a) Kurtosis value distribution for COVID-19 patients.

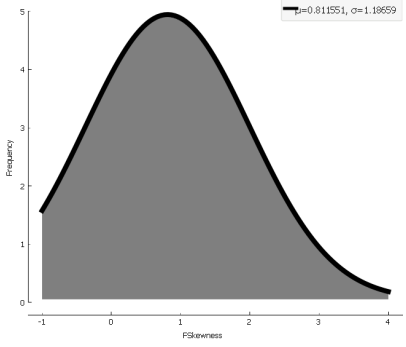


(b) Kurtosis value distribution for healthy patients.

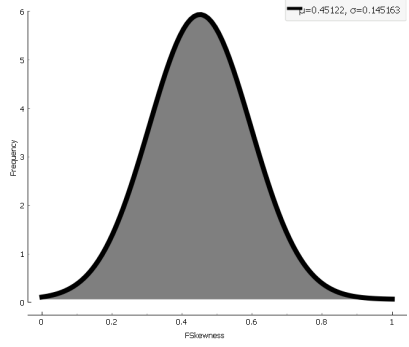
Figure 7.8: These histograms show the total distribution of Kurtosis values for COVID-19 patients (on the left side) and for healthy patients (on the right side). The differences about the form of the curve are greater; then, healthy values are lower than the others.

values in COVID-19 patients are Variance and Minimum. The *Variance* index assesses the average of the squared distances of each intensity value from the mean value, quantifying the spread of the distribution around the mean, as shown in Figure 7.10.

The *Minimum* measurement provides an estimate of the minimum intensity, which corresponds to the lowest intensity level within the image: intuitively, the Minimum feature represents the density of air within the lungs on the CT scan (Figure 7.11).

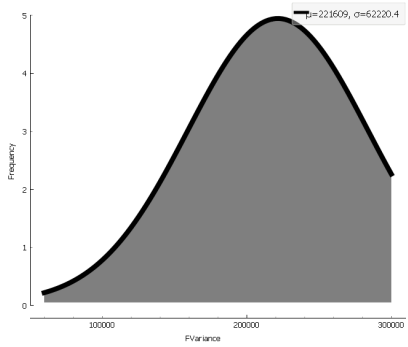


(a) Skewness value distribution for COVID-19 patients.

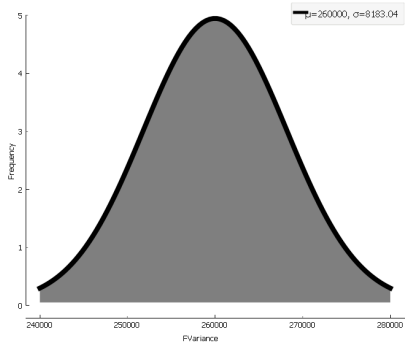


(b) Skewness value distribution for healthy patients.

Figure 7.9: Skewness values for COVID-19 patients and for healthy patients. The distribution of healthy patients is on average less than the one of sick patients.



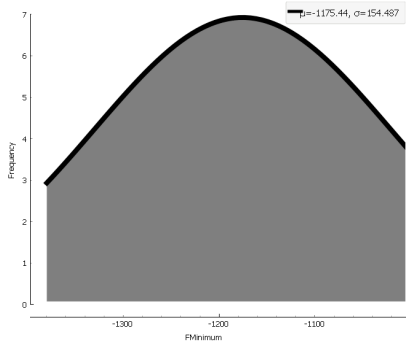
(a) Variance value distribution for COVID-19 patients.



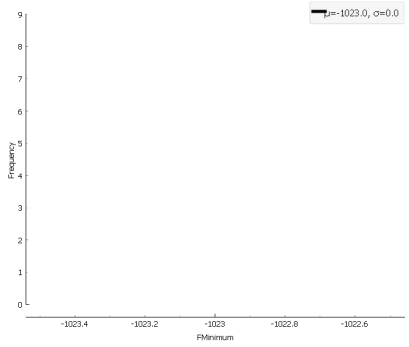
(b) Variance value distribution for healthy patients.

Figure 7.10: Variance values for COVID-19 patients and for healthy patients: healthy values are higher and better distributed around the mean, as a normal distribution.

These last features appear to be crucial in a lung study, given the observed variations in the distribution of values. These objective data align with some radiological results, such as those of lung function tests given by lung density measurements: these show how lung tissue attenuates X-rays, reflecting structural abnormalities within the lungs [108].



(a) Minimum value distribution for COVID-19 patients.



(b) Minimum value distribution for healthy patients.

Figure 7.11: Minimum intensity values for COVID-19 patients; the graph for healthy patients is absent because the values are fixed on -1023,00 with a standard deviation of 0,00. This is the major difference between the two categories.

7.4.2 Explainability of the Minimum feature in this case study

During the Feature Selection phase, some of the features mentioned in the previous Section 7 were chosen using Weka [104]. The results are depicted in the tree structure shown in Figure 7.4, with the most important feature at the root of the tree. *Minimum* feature is in the root of the tree and, as defined by IBSI, it represents the minimum intensity in the image. In the field of radiology and lung studies, this minimum value should accurately correspond to the value of air captured by the CT examination [109, 110]. However, since air is present in both healthy and diseased patients and is unaffected by the presence of the disease, it raised questions about why it holds significance for the AI tool.

Deepening why Weka chose the "Minimum" feature as the most discriminating one, it can be seen that for healthy patients, these values were consistently at -1023.0, resulting in an empty distribution chart as shown in Figure 7.11b. In contrast, the *Minimum* values for some COVID-19 patients have occasionally exceeded -1023.0, as in Fig. 7.11a. The hypothesis that the fixed value -1023.0 correctly represents air in the lungs led to consider this as a potential bias related to radiological machine settings for COVID-19 patients. This assumption, regardless of the number of patients, was later confirmed through the use of formal methodology.
For the classification task, as shown in Table 7.8, there are different FIRST-Order features according to the two levels of the methodology. The first part of the table includes the features used for the classification of healthy and sick patients, while the second part concerns the classification of COVID-19 and patients with lung diseases. Initially, the *Minimum* feature was included in the automated diagnosis of COVID-19. As a result, the mathematical formulas used for second-level diagnosis incorporated *Minimum* with the *Skewness* feature. Based on the above radiological assumptions, the combinations were modified by eliminating the *Minimum* feature. After re-evaluating the diagnosis, the results remained consistent: this confirms that the *Minimum* feature is not a significant factor in the classification between healthy patients and COVID-19 and may not be suitable for lung studies, contrary to what other authors suggest in various documents relating to the lungs [111, 112].

In conclusion, the explanation offered by Formal Methods makes it possible to validate or refute the results of the AI tool, as demonstrated in this case. This in-depth analysis provides the means to interpret and explain the results, addressing both medical specialists and their collaborators. It serves as a practical illustration of a feedback mechanism for physicians, allowing them to contribute to the development

FIRST 1st part		**FIRST 2nd part**
Energy Variance		Skewness
Precision	**Specificity**	**Accuracy**
1.00	1.00	0.79

Table 7.8: List of features used for the property generation and their performance metrics. In the second part, the feature Minimum was removed.

and use of medical decision support systems in collaboration with Data and Computer scientists.

7.4.3 Explainable patterns

Automatic diagnosis is possible thanks to a mathematical model known as "property" or "formula" or "rule", designed to represent the disease independently, without the patient's characteristics. In addition to enabling diagnosis, these properties can also provide insight and information about patient models and thus improve the interpretation of results. In this example 7.12, only *Variance* and *Energy* features have been

```
prop F1 =(min X = <b1of3_fvariance> <b1of3_fenergy> F2 \/ < - > X)
prop F2 =(min X = <b1of3_fvariance> <b1of3_fenergy> F3 \/ < - > X)
prop F3 =(min X = <b1of3_fvariance> <b1of3_fenergy> F4 \/ < - > X)
prop F4 =(min X = <b1of3_fvariance> <b1of3_fenergy> F5 \/ < - > X)
prop F5 =(min X = <b1of3_fvariance> <b1of3_fenergy> F6 \/ < - > X)
prop F6 =(min X = <b1of3_fvariance> <b1of3_fenergy> F7 \/ < - > X)
prop F7 =(min X = <b1of3_fvariance> <b1of3_fenergy> F8 \/ < - > X)
prop F8 =(min X = <b1of3_fvariance> <b1of3_fenergy> F9 \/ < - > X)
prop F9 =(min X = <b1of3_fvariance> <b1of3_fenergy> tt \/ < - > X)
```

Figure 7.12: Example of logical "formula" that contains discretized values for two features, Variance and Energy.

identified as decisive factors in determining the disease and will be used to classify the entire patient dataset. However, within the formal properties there can be the repetition of some discretized values, which helps in the classification process, because they can be recognised as a *pattern*. In Figure 7.13 there is an example of pattern in COVID-19 and healthy categories, noting how many times a particular value is repeated consecutively. This pattern suggests that the combination of *low Variance values* and *low Energy values* for nine times in a row can reliably identify COVID-19 and distinguish it from healthy patients.

```
prop F1 =  <b1of3_fvariance> <b1of3_fenergy> x9
```

Figure 7.13: Example of pattern recognition on a formula used in classification of COVID-19 patients. In this case, the formula contains a unique pattern: the repetition of *low variance* and *low energy* for 9 times in a row.

To investigate and support these hypotheses, the structure of formal models for each patients is deepening with pattern recognition technique. In the case of healthy patients, the model tends to exhibit *medium variance* and *medium-low energy* values. This technique not only provided information on the distribution of values (as illustrated in figures 7.10b and 7.6b), but also highlights the fact that this property may not be universally applicable to all patients. As shown in Table 7.8, this property does not achieve 100% performance in terms of metrics.
The study of models allows researchers to gain better control over models, which can

have a significant impact on performance metrics such as Accuracy and Specificity. Moreover, the pattern extraction process is neither long nor expensive, making it an economical application in the medical industry. This, in turn, increases the overall explainability of the results, a benefit for both specialist doctors and patients. For medical specialists, this information allows them to understand the logic behind automatic diagnosis of the patient; to refine the settings if the diagnosis does not align with the real state of health; to enjoy a more robust decision support system that offers greater confidence.

```
prop F1 =  <b2of3_fvariance> <b3of3_fenergy>
prop F2 =  <b2of3_fvariance> <b1of3_fenergy> x3
prop F4 =  <b2of3_fvariance> <b3of3_fenergy>
prop F4 =  <b2of3_fvariance> <b1of3_fenergy> x2
prop F6 =  <b2of3_fvariance> <b3of3_fenergy> x2
prop F7 =  <b2of3_fvariance> <b1of3_fenergy>
prop F7 =  <b1of3_fvariance> <b1of3_fenergy> x2
prop F9 =  <b2of3_fvariance> <b1of3_fenergy> x23
prop F10 =  <b2of3_fvariance> <b2of3_fenergy>
prop F10 =  <b3of3_fvariance> <b2of3_fenergy>
prop F11 =  <b2of3_fvariance> <b2of3_fenergy>
```

Figure 7.14: The pattern recognition searching returns how many times in a row different combination of features are found. This is an example of pattern recognition on a formula used in classification of healthy patients. Healthy patients are characterised by different patterns, which the most recurrent is the $mediumVariance$ together to $lowEnergy$.

7.4.4 Localization of the most important slices

As demonstrated, FIRST-Order features show variations between different health states, effectively capturing radiological evidences and translating it into a range of values. This concept highlights a great challenge faced by radiologists in the analysis of CT examinations, which consists in the management of hundreds of images: in fact, the human effort required by radiologists intensifies and could increase errors, cases of neglected details or cases of divergent automatic diagnosis, where the radiologist is uncertain as to which image led to the classification discrepancy.

With the algorithm developed for the use of Formal Methods, a simple but powerful method can locate the most crucial slices within a multi-slice CT exam. This algorithm identifies slices where specific patterns are detected, revealing which slices are evaluated by the diagnostic algorithm. In Figure 7.9, the formula is enriched by annotations on the significant slices; for example, "SLICE 246" refers to the 246^{th} slice in the series of CT exams. In this image, the disease is clearly visible, often appearing as grounded glass opacity (GGO), a distinctive sign of COVID-19 [113].

```
prop F1 =(min X = <b1of3_fvariance> <b1of3_fenergy> F2 \/ < - > X)
SLICE 246

prop F2 =(min X = <b1of3_fvariance> <b1of3_fenergy> F3 \/ < - > X)
SLICE 251

prop F3 =(min X = <b1of3_fvariance> <b1of3_fenergy> F4 \/ < - > X)
SLICE 253

prop F4 =(min X = <b1of3_fvariance> <b1of3_fenergy> F5 \/ < - > X)
SLICE 255

prop F5 =(min X = <b1of3_fvariance> <b1of3_fenergy> F6 \/ < - > X)
SLICE 261
```

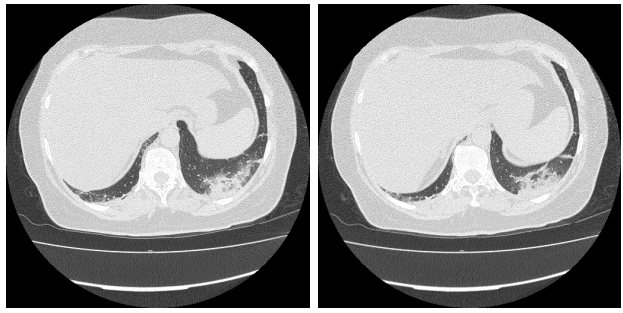


Table 7.9: On the upper bound there is an example of a "explainable" formula with indications on the localization of the most important slices to be reviewed by the specialists. On the lower bound, on the left there is the $246th$ slice, relative to the first value of the property. On the right there is the slice relative to the second property, namely the slice number 251.

This result has significant implications for the explanation and practical use of the method. Medical specialists can now focus their attention on a reduced subset of slices (e.g., 30 slices) instead of navigating through the totality of a multi-slice CT exam. While this approach retains the essence of a multi-slice rating, it is simplified to focus on a limited number of slices. Nevertheless, this outcome can only highlight the crucial elements detected by the automated diagnosis. A thorough and in-depth interpretation of the lung's condition is a task that remains within the purview of medical specialists.

7.5 An Automatic Radiomic-Based Approach for Disease Localization: A Pilot Study on COVID-19

In recent years, *Radiomics* has gained increasing popularity across various realms of radiology, spanning from oncology to forensic medicine. Radiomics serves a multi-faceted purpose, encompassing diagnostic, prognostic and a generalized extraction of valuable data from radiological images [7]. This innovative approach to medical practice not only facilitates Personalized Medicine but also offers numerous advantages, including a wealth of information that enhances both diagnosis and patient treatment. However, as with any pioneering technique, there remain several unresolved issues within the literature. These include the absence of standardization and clinical evidence associated with Radiomics data, concerns related to ethical regulation and the need for automated standardized segmentation tools [114]. These challenges underscore the importance of establishing standards and foster a collective anticipation for a future integration of Radiomics into routine radiological practices.

In the radiomic workflow, the *pivotal stage involves defining the region of interest* (ROI) through segmentation, which is the delineation of the specific area from which data is extracted. Presently, identifying and outlining a ROI is typically performed manually, making it an exceedingly time- and resource-intensive process [115]. My team and I propose an automated approach that enables the classification and localization of a specific disease without the need for manual ROI delineation. Specifically,

FIRST	Total Energy, 90 Percentile, Mean Absolute Deviation;
GLCM	Sum Average, Autocorrelation, Contrast, Joint Average;
GLDM	Large Dependence High Gray Level Emphasis, Dependence Non Uniformity, Gray Level Non Uniformity, High Gray Level Emphasis;
GLRLM	Run Length Non Uniformity, Gray Level Non Uniformity, High Gray Level Run Emphasis, Long Run High Gray Level Emphasis;
GLSZM	Small Area High Gray Level Emphasis, Gray Level Non Uniformity, High Gray Level Zone Emphasis, Large Area Low Gray Level Emphasis;
NGTDM	Strength, Complexity.

Table 7.10: List of radiomic features taken into account for the final diagnosis.

the methodology applies an $n \times n$ grid to the medical image, dividing it into a matrix that generates multiple regions corresponding to the cells in the matrix. The *objective* is to establish a methodology for detecting and localizing diseases using radiomic features to study and early recognize diseased patients from healthy ones. The ultimate goal is to develop tools for future works that can automatically identify and localize diseases. This method demonstrates promising performance while offering several advantages: (i) it automatically analyzes all parts of the image, ensuring no detail is overlooked; (ii) it is not resource-intensive; (iii) manual segmentation is not required; (iv) it is scalable for different diseases through the parameter n.

In Figure 7.15, the workflow of the proposed methodology is presented. It commences with a database provided by radiologists, consisting of DICOM images from both healthy and diseased patients. Prior to the Feature Extraction process, the segmentation of the image under analysis is done by dividing each DICOM slice into n equal sections by applying an $n \times n$ grid to each image, where each cell of the matrix represents a candidate ROI. Radiomic features are then extracted from each cell of the matrix to build formal models and effectively localize a disease. In contrast to the classical approach described in Section 7.1, models are constructed relating features to each other based on the position of the cell in the matrix.

A ROI involves identifying and outlining a specific area within an image. In medical images, this allows for a focused analysis in a localized region, yielding more precise feature values and enhancing the model's accuracy. The notion of a *region* can be likened to the geometric concept of *bounding boxes*. Given a set of points P in N

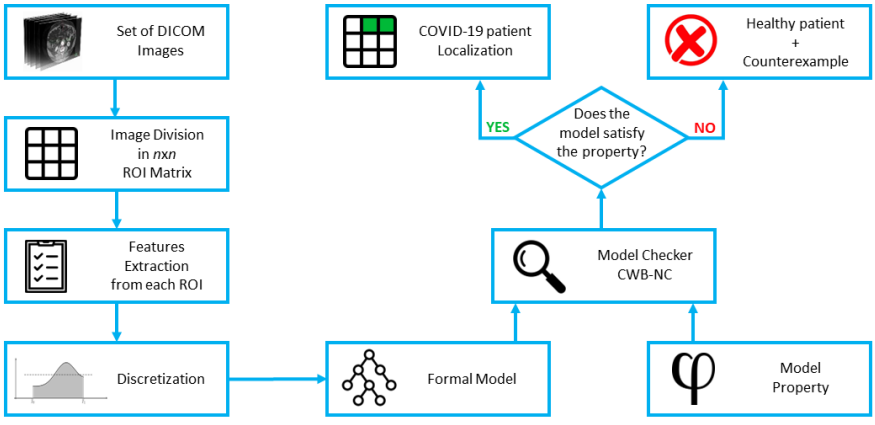


Figure 7.15: Schema of the developed method.

dimensions, a bounding box is the smallest measure box (in terms of area or volume) that encompasses all the points. For the creation of bounding boxes, a Python method called *Quokka* was used and it facilitates the generation of axes-aligned rectangles on an image by creating a grid, if the grid's dimension is parametrically determined by n. When provided with a number n, it generates an $n \times n$ grid on the image, resulting in $n \times n$ distinct boxes. Consequently, each square is generated by considering the minimum coordinates (upper left corner of the region) and the maximum coordinates (lower right corner of the region). As the illustration in Figure 7.16, in a DICOM image with dimensions of 512x512 pixels and a $3x3$ matrix, the boundary values for columns/rows would be 0, 170, 340 and 512. Taking the R1 region (the first box in Figure 7.16), its minimum coordinates would be $x = 0$ and $y = 0$, while the maximum coordinates would be $X = 170$ and $Y = 170$.

For the case study, the focus revolves around the classification and localization of COVID-19 disease. The database involves examining 40 DICOM patients acquired through a high-resolution CT scanner, of which 31 patients were diagnosed with COVID-19 and 9 healthy patients. The Feature Selection process results are highlighted in Table 7.10. For the experiment, a $3x3$ grid ($n = 3$) is set as it proved sufficient to localize the COVID-19 pathology, particularly in the lung, which occupies a substantial area in the image.

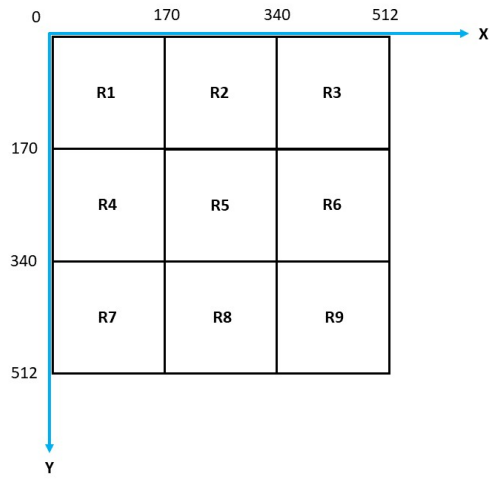


Figure 7.16: ROI matrix represented on the axes of the system considered by the DICOM type images.

Once the nxn matrix structure is established on DICOM images, features are extracted for each DICOM series and possible ROI (e.g., R1), treating each of the nxn boxes as the effective ROI each time. For a given region, such as $R1$, the comprehensive formal model concatenates the corresponding region (i.e., $R1$) from each slice. I implemented the sequential conjunction of each region, with each one containing the exhaustive interleaving of all features. For example, `proc P1` in Table 3.1 represents the modeling of the first slice of the ROI under analysis, while `proc P2` mirrors the second slice of the same ROI under analysis and so forth. Upon completion of this model construction step, each patient will possess nxn formal models, one for each region identified by the grid. In collaboration with radiology experts, a subset of 4 patients was selected to formulate a formal "property" encapsulating information about the presence of COVID-19 disease.

For the formal verification, a comprehensive evaluation is conducted considering two distinct types of ROIs. Initially, the **ROI is the entire image**. The Model Checker tool is employed as explained in Section 2.2 and it returns a single truth value ('TRUE' or 'FALSE') for each patient. Table 7.11 presents the classification results when applying the ROI to the entire image.

Next, the **matrix-algorithm** with $n = 3$ is applied maintaining the same radiomic features outlined in Table 7.10 on the DICOM images, formally and iteratively checking each candidate ROI. This $3x3$ grid proved sufficient for localizing the COVID-19 pathology, particularly as it manifests in the lungs wich occupy a substantial area in the image. In this scenario, following the execution of the complete workflow using the $3x3$ grid, results are gathered for each cell of the grid. The outcome is TRUE if the formal property is satisfied in the cell and FALSE otherwise. To be classified as having COVID-19, a patient must exhibit 3 TRUE cells at a distance of 1 horizontally or vertically (not diagonally). Both for entire than for candidate ROI, a *True Positive* denotes a COVID-19 patient correctly classified as such, while a *False Positive* indicates a healthy patient erroneously classified as having COVID-19. The performance indices of classification for GLRLM class achieve an Accuracy of 0.72, a Precision of 0.78, a Recall of 0.9.

Conversely, for **disease localization**, the matrix is conceptually divided into 3 vertical bands: the first representing the right lung, the middle one depicting the mediastinum and the third signifying the left lung. If a patient has 1 TRUE cell, the corresponding band is considered localized. In this case, a *True Positive* corresponds to a COVID-19 patient correctly localized, while a *False Positive* represents a healthy patient where a sign of COVID-19 disease is localized. All patients clas-

Table 7.11: Classification performances with the segmentation on the whole DICOM image (above) and with the 3x3 regions matrix put on the DICOM images (below).

Radiomic class	**Accuracy**	**Precision**	**Recall**
GLDM	0.75	0.83	0.83
GLDM $3x3$	0.65	0.76	0.81
GLRLM	0.77	0.86	0.83
GLRLM $3x3$	0.72	0.78	0.90
NGTDM	0.75	0.78	0.93
NGTDM $3x3$	0.77	0.77	1

sified as COVID-19 also underwent localization, while patients classified as healthy did not undergo localization. Table 7.12 illustrates the performance achieved by this approach using the localization grid for each radiomic class. Compared to the Gold Standard, 21 out of 28 patients (0.75%) were localized for the right lung band, 20 out of 31 patients (0.64%) for the middle band and 23 out of 28 patients for the left lung (0.82%).

As observed in this particular case of COVID-19 disease, employing a grid on the image yields positive results when utilizing classes such as NGTDM and GLRLM. In contrast, there is a decline in performance for the GLDM class. This decline could be attributed to the class's reliance on voxel dependencies with respect to a central voxel. With smaller ROIs, these dependencies may become weaker, leading to more classification errors. Another factor contributing to this deterioration could be related to the features within the GLDM class. The Feature Selection was based on the entire ROI, where variations in intensity are more pronounced. Future studies will aim to shed light on these results by correlating them with clinical evidence for improved interpretability.
This methodology is flexible thanks to its parameterization by the variable n allow-

Table 7.12: Localization performances with the 3x3 regions matrix put on the DICOM images.

Radiomic class	**Accuracy**	**Precision**	**Recall**
GLDM	0.70	0.77	0.87
GLRLM	0.77	0.79	0.97
NGTDM	0.77	0.77	1

ing its application across various levels of granularity. This adaptability extends to different anatomical contexts, facilitating its effective utilization across a spectrum of scenarios with varying degrees of detail. Opting for a larger n leads to a denser grid, directly proportional to the increased magnitude of n. This means that by selecting a higher value for n, it is possible to achieve finer granularity in disease localization. Moreover, this method provided insights into manipulating the coordinates of the axis system, enabling an understanding of the DICOM coordinates to derive the n boxes from the nxn matrix, each serving as a distinct region. Segmenting a DICOM through coordinates is challenging due to the complex structure of this file. Notably, the DICOM system is oriented with the Y axis downwards, as depicted in Figure 7.16. This innovative methodology not only enables classification but also localization and visualization simultaneously. The newly generated result matrix can be superimposed on the HRCT DICOM of the patient, as displayed in Figure 7.17. This facilitates the localization of disease markers within the radiological images of the patients.

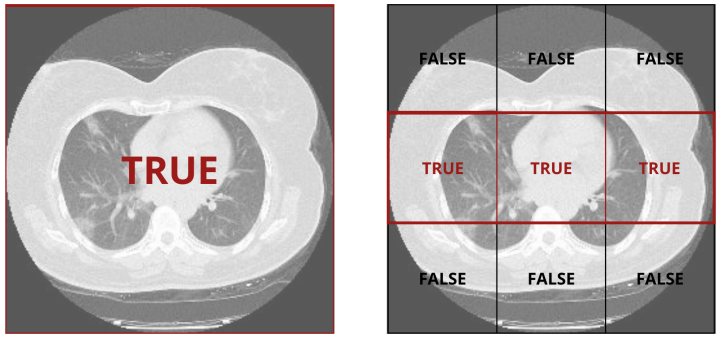


Figure 7.17: The ROI matrix can be superimposed on the DICOM of the patient to facilitate the localization of disease markers within the radiological images.

7.6 Related Works and Discussion

In the experimental analysis of COVID-19 patients, I observed a different *performance* in the first step compared to the second step of classification. This difference reflects the real-world scenario, where there is typically a clearer distinction between patients with healthy lungs and those with lung diseases (Table 7.4), compared to the differentiation between lung disease and COVID-19 patients (Table 7.5). The latter two groups may share more similarities due to the overlapping symptoms and clinical evidences.
However, when examining the results in terms of Accuracy, they are on par with recent AI techniques. The advantage of using this method is the ability to obtain an accurate, automated, quick and easily accessible "*second virtual opinion*" for radiologists. This approach can complement the expertise of healthcare professionals, enhancing diagnostic processes and decision-making.
Instead of using *HRCT*, Afat et al. [116] opted for Dual-Energy Computed Tomography (DECT) of the lung in their study involving 14 COVID-19 patients. They aimed to investigate the relationship between pulmonary perfusion patterns and opacities using 3D segmentation. In contrast to my approach, this study required the expertise of two radiologists, along with the utilization of the Linkert scale and statistical analysis. However, as stated by other researchers [117], HRCT is generally the preferred imaging modality for COVID-19 examinations and classifying patients with Coronavirus disease.

The use of *AI in the medical field* for the detection of COVID-19 patients is explored in various studies, including the one presented by Rajaraman et al. [38]. They presented a method involving weakly labeled data augmentation for deep learning in COVID-19 detection using chest X-rays to expand the training data by recognizing COVID-19-related lung opacities. The underlying premise of their work is that COVID-19 manifestations are related to those of other types of pneumonitis, including viral or bacterial ones. While we concur with this assumption, it is worth noting that COVID-19 and pneumonia can sometimes exhibit striking similarities, potentially leading to misdiagnoses. Their method, unlike mine, uses chest X-ray images for training a convolutional neural network-based algorithm, achieving an Accuracy of 88.89%. However, their approach heavily relies on extensive datasets and entails complex processes within the DL model. These processes include image segmentation to generate ROIs, various preprocessing steps to adjust images and standardization of images to ensure computational efficiency. Such extensive preprocessing and dataset requirements are not necessary when using Formal Methods, which still manage to

obtain results similar to those in the literature.

Loey et al. [118] proposed a novel detection model to address the lack of images, particularly in chest X-ray images, for COVID-19 detection. Their approach involves using Generative Adversarial Networks (GANs) and Deep Transfer Learning to mitigate the dataset shortage. The concept is to aggregate as many COVID-19 images as possible and employ GAN networks to generate additional images and enhance the Accuracy of COVID-19 detection, to overcome the limited availability of X-ray images related to COVID-19. While they focus on using GANs and transfer learning to bolster the dataset, my work utilizes HRCT images and Formal Methods for detection, minimizing the dependency on large datasets and advanced ML techniques. The application of Formal Methods and Model Checking in the context of COVID-19 diagnosis, as far as the authors are aware, is relatively novel and not widely explored. This approach differentiates the presented study from other research in the field: while other studies focus on similar patient categories in their databases and employ different techniques, such as Radiomics, the application of Formal Methods for diagnosis and validation adds a unique dimension to this work. This approach aims to provide a rigorous and explainable methodology for COVID-19 detection, which can be important for medical applications, especially when large datasets or advanced ML techniques are not readily available or suitable.

The *critical role of radiologists* is also underscored in the study by Buttner et al. [123]. They introduced a semi-quantitative method to identify COVID-19 patients in need of intensive care unit (ICU) treatment or intubation. In contrast to my method, their approach utilized a dataset of 28 COVID-19 patients for a statistical study, which involved methods like Receiver Operating Characteristic (ROC)

Table 7.13: State-of-the-art comparison for COVID-19 classification with AI.

Method	**Images**	**Healthy**	**COVID-19**	**Others**	**Accuracy**
Brunese et al. [119]	X-ray	✓	✓	✓	0.97%
Hemdan et al. [120]	X-ray	✓	✓	-	0.90%
Rajaraman et al. [38]	X-ray	✓	✓	✓	0.89%
Loey et al. [118]	X-ray	✓	✓	✓	0.85%
Civit-Masot et al. [37]	X-ray	✓	✓	✓	0.85%
Wang et al. [121]	CT	-	✓	✓	0.82%
Song et al. [122]	CT	✓	✓	-	0.86%
This method	CT	✓	✓	✓	**0.83**%

analysis and AUC calculations; they established a hazard ratio of 95%. Additionally, their method requires image segmentation at three different levels and manual ROI labeling. However, their approach exhibits the capability to predict ICU admission with an Accuracy rate of 85.6% and 71.9% for intubation.

Thinking about all the images that the doctor has to look at in order to ascertain the presence of a COVID-19 symptom, I thought about whether it might not be possible to rely on the robustness of Formal Methods to *decrease the number of images in the radiological examination* and focus only on the central ones. I performed an experiment maintaining the same type of cascade classification.
Regarding the first step of classification between sick and healthy patients, the performances remain more or less stable: there is an actual decrease in Precision and a stability in Recall values. This means that by decreasing images, the rate of false positives rises while false negatives remain unchanged. Thus more healthy patients being classified as sick: thus the computational problem becomes a problem for the national health system, which has to delve into the exams of more patients and incurring more expenses.
In the second step, then in the classification between patients with COVID-19 and with general lung pneumonia, it can be said that the performance reverses. In this case, all indexes drop as many percentage points, while Precision and Specificity are the only stable indexes. Probably, reducing the CT exam, the mathematical rule created by computer scientists and physicians turns out to be more generic and is not able to encompass several radiological cases of COVID-19.
One can think that reducing the complexity of medical images, such as CT scans, may indeed speed up the diagnostic process for radiologists. However, this simplification might not be suitable for radiomic analysis, which relies on detailed image features to provide valuable information for diagnosis and treatment planning. Researchers are actively working on developing methods for automatic segmentation and delineation of ROIs within medical images to address this challenge and ensure the integrity of radiomic analysis.

Subsequently, we tried to evaluate the effects of Long COVID: patients who still presented with breathlessness and fatigue months after COVID were subjected to SPECT scans to assess the presence of perfusion deficits. Baeza et al. [124] studied thousands of images of patients with pulmonary embolism, pneumonia and healthy patients in order to perform a classification on the basis of radiomic features. When

differentiating images with PE, they obtained recall, specificity and precision of 75.1%, 98.2%, 88.9%. After performing an automatic ROI on the basis of threshold segmentation, as in my work they used PyRadiomics for Feature Extraction.

Researchers in radiomic studies [44] have emphasized that the size of the ROI area is a critical confounding factor in the analysis. In my automated diagnosing approach there is not a specific ROI and the computations are performed on the entire image. Consequently, this aspect can be seen as an initial advantage for the application of Formal Methods in Radiomics.
Regarding the radiomic features employed in studies on lung functions, the literature includes various investigations that correlate numerical values with radiological evidence. Many studies commonly use parameters such as Mean Lung Attenuation (MLA) and other histogram-based indices for quantifying emphysema and pulmonary fibrosis with densitometry techniques [108]. Instead, emphysema is primarily assessed using percentiles, while fibrosis is predominantly assessed using *Skewness* and *Kurtosis*, as incorporated into previous methodology (Section 3.7). For instance, studies related to Idiopathic pulmonary fibrosis (IPF) have highlighted the predictive importance of Kurtosis in the diagnosis of this condition [107]. Much like my methodology based on FIRST-Order features, these researchers have emphasized the role of histogram features, including Kurtosis, Skewness, Mean Lung Density, Median and Variance in assisting healthcare professionals in predicting factors such as mortality in IPF patients.
Moreover, histogram measures are better suited for capturing disease-related patterns because the interstitial patterns of reticulation and honeycomb exhibit extremely low or high intensities, which are situated in both tails of the histogram [125]. Therefore, a *low Kurtosis value* indirectly corresponds to a high extent of reticulation and honeycomb, while a *higher Kurtosis value* suggests mild IPF. In the previous case study of Section 7.1, the dataset shows a higher Kurtosis in COVID-19 patients: this association between numerical values and radiological evidence may be linked to the presence of hypoxemia (a low level of oxygen in the blood), which is often associated with COVID-19 pneumonia.

Similarly, *Skewness* is a relevant feature in lung density studies and it is closely related to the Kurtosis feature. Most published articles on Kurtosis and Skewness for CT densitometry of the lungs focus on Interstitial lung disease (ILD) [107]. These studies typically find that lower Kurtosis and Skewness values are observed in patients with advanced disease and more compromised lung functions. These findings suggest

heterogeneity in lung density and, consequently, a deterioration in lung functions.
In my case study, the hypothesis presented in the literature is confirmed, even though the Skewness and Kurtosis values of COVID-19 patients are higher than those of healthy patients. We believe this discrepancy is due to the important values being found in the tails of the distribution, leading to extreme low or high values, as stated in [125].
In various lung studies [126], researchers have suggested that the *Variance* feature may be an indicator associated with the non-uniformity levels of the image. Another study on fungal pneumonia used this feature in the diagnosis [127]. In my study, the *Variance feature* was useful for classifying COVID-19 patients with high-performance metrics.

With Formal Methods can be matched different types of segmentation and based on it, build different logical models. For example, Formal Methods allows to ideally draw a new type of ROI only using image coordinates and creating on it a *3x3* matrix. Here, each image is associated with 9 ROIs, each of which corresponds to a formal model, with improvements in performance indices for 2 out of 3 radiomic classes.
Thanks to automatic segmentation, there is no need to calculate variability indices such as the *Intraclass Correlation Coefficient* (ICC) [128], as is done in other studies [129]. In addition, currently DICOM segmentation is typically carried out using external Radiomics software, which allows manual drawing of ROIs on the image or the use of seed-growing methods. However, high-quality ROI segmentation software is not always freely available; with my work, DICOM segmentation can be performed automatically and at no cost using the PyRadiomics script, eliminating the need for other external software.
In studies like [130, 131], researchers have employed a hybrid approach to automatically detect suspicious lesions using an adaptive fuzzy region-growing algorithm. This segmentation method initiates from a random seed point that expands across slices to encompass the entire tumor region. However, my method propose that these seed-based segmentation methods can be enhanced by leveraging the concept of a ROI matrix. This approach provides more control over seed point selection and can be supported by the matrix's structure, even though the initial seed selection remains manual. In prior research, segmentation typically begins with an operator manually identifying the ROI and then proceeds in an unsupervised manner, relying on exploring a patient's examination to locate the tumor. In contrast, my method integrates the classification and localization tasks simultaneously, significantly reducing the time

required for diagnosis and providing valuable assistance to medical specialists.

Chapter 8

Liver

This chapter delves into the prediction of liver metastases caused by Colon-Rectal Cancer.

8.1 Early diagnosis of liver metastases from colorectal cancer through CT Radiomics and Formal Methods: a pilot study

Liver metastases present a significant challenge in the management of *colorectal carcinoma* (CRC) and remain a primary cause of mortality. A comprehensive, multidisciplinary approach to CRC treatment is now established, involving a collaborative effort among surgeons, oncologists, radiotherapists and radiologists [132, 133].
As of today, achieving an R0 surgical resection (complete resection with no residual tumor) of liver metastases remains the gold standard for curative treatment. Recent advances in liver surgery have introduced parenchyma-sparing techniques and precision US-guided resections, expanding the possibilities even in cases of recurrent metastases. However, it is crucial to emphasize the importance of precise and early detection of all intrahepatic lesions prior to surgical resection, as this is essential for optimizing the overall oncological management [134].
Despite significant advancements in radiological technologies, the accurate diagnosis of *liver metastases* originating from *colorectal cancer* (CRCLM) continues to be a pivotal concern. Current clinical practice primarily relies on tri/quadraphasic CT

scans, with MRI used selectively. Notably, the most challenging scenario involves the detection of liver micrometastases, as they often elude identification through standard radiological protocols during their early stages. Therefore, an earlier and more precise diagnosis of hepatic micrometastases holds the potential to significantly enhance the appropriateness of treatment and surveillance in CRC patients [135, 136].
Previous investigations have explored the utility of AI models for early diagnosis of liver metastases. However, these studies were primarily based on ML approaches and were constrained by relatively short follow-up periods [137]. These studies underscored a critical limitation of AI and ML techniques: the requirement for a substantial cohort of cases to attain a satisfactory level of sensitivity and specificity.

The *objective* of this case study is to identify liver metastases in their early stages by leveraging non-invasive radiomic features based on shape analysis of CT images. This approach aims to detect metastases before they become discernible through conventional imaging protocols.
CT scan data retrospectively reviewed of 30 patients collected between January 2013 and June 2021 at the Pineta Grande Hospital Castel Volturno, Caserta, Italy. All patients included in the study underwent CT scan at our centre using SOMATOM® Definition Flash, Siemens Healthcare, Erlangen, Germany. A total of 21 patients were recruited for a *case-control group* (CCG), consisting of "healthy" individuals who had undergone triphasic CT scans, showing no signs of cancer or chronic illnesses in their radiological or medical history. The *case group* (CG) comprised 9 patients with colorectal cancer, who either developed liver lesions within a time frame spanning from 4 months to 8 years or they were individuals who developed liver metastases following their initial staging, prior to any liver surgery. Additionally, 2 patients were included in the CG after undergoing R0 liver resection. The presence of any liver lesion detected in the initial CT scan served as an exclusion criterion; patients with any indications of underlying liver disease were not included in the CCG; the presence of benign liver lesions, such as angiomas and hepatic cysts, did not disqualify individuals from CCG. The exclusion criteria were evidence of sinchronous liver lesion at first CT scan; underlying liver disease in both groups; CT scan performed in other centers or with other type of scan settings. The inclusion criteria were:

1. Evidence of CRC diagnosed at CT scan, histopathologically confirmed;

2. Patients who performed the first CT scan and follow-up exams at our centre, in order to set the protocol on the same scan;

3. No evidence of liver lesion at the moment of primary diagnosis or at follow up after surgery;

4. Findings of previous liver surgery in metachronous patients already treated with surgical approach.

In this case, manual segmentation is included in the methodology, as the image at liver level has too much noise and the analysis could be affected by the other organs in the image. Trained physicians manually outlined and meticulously reviewed the ROIs using the 3D Slicer software [22]. Notably, the *ROI segmentation* was carried out on each individual slice throughout the entire series of portal phase images, excluding specific anatomical structures like the portal vein, branches of the inferior vena cava and the gallbladder; this method ensured that the entire liver parenchyma and biliary structures were included in our analysis.

FIRST	**GLDM**	**GLCM**	**GLRLM**	**GLSZM**
Entropy	Dependence Entropy	Autocorrelation	High Gray-Level Run Emphasis	High Gray-Level Zone Emphasis
Interquartile Range	High Gray-Level Emphasis	Joint Average	Long Run Low Gray-Level Emphasis	Low Gray-Level Zone Emphasis
Mean Absolute Deviation	Large Dependence Low Gray-Level Emphasis	Joint Entropy	Low Gray-Level Run Emphasis	Small Area Low Gray-Level Emphasis
Robust Mean Absolute Deviation	Low Gray-Level Emphasis	Sum Average	Short Run Low Gray-Level Emphasis	
Uniformity	Small Dependance Low Gray-Level Emphasis	Sum Entropy		

Table 8.1: List of features selected as "significant" in this specific case study of the liver.

CONFUSION MATRIX		**Actual Values**	
		sick	*healthy*
Predicted Values	*sick*	TP = 7	FP = 0
	healthy	FN = 2	TN = 21

Table 8.2: The Confusion Matrix helps to understand how many and which patients are misclassified.

Feature extraction was performed on an average of 70 ROI per patient (56–92 CI). Subsequently, the Feature Selection process was carried out: radiomic features were analyzed using Weka [104] and below in Table 8.1, there is the list of features chosen as "significant" without being considered redundant.
As results, FIRST, GLDM and GLSZM feature classes were the most efficient considering their distribution between "healthy" and "metastatic" livers. The Precision index reached 100%, which means that patients were only categorized as positive if they would indeed be affected by colorectal liver metastases, resulting in an overall Accuracy of 93.3%; Recall or Sensitivity, stood at 77.8%. The Confusion Matrix in Figure 8.2 and the Table of performances 8.3 provide the values obtained from this automated diagnostic process. Furthermore, additional statistics by Mitchell et al. [32] are included.

Accuracy Statistics	**Value**	
Recall (Sensitivity)	77.8%	
Specificity	100.0%	
Precision (Positive Predictive Value)	100.0%	
Negative Predictive Value	91.3%	
Positive Likelihood Ratio (+Ve)	Inf	
Negative Likelihood Ratio (-Ve)	0.222	
Test Score (or fraction correct) %	93.3%	
Prevalence	30.0%	
Utility statistics	**Rating**	**95% Confidence Interval**
Clinical Utility (+Ve)	Good	0.778
Clinical Utility (-Ve)	Excellent	0.913

Table 8.3: Accuracy and utility statistics according to Mitchell statistics [32] to gain more information about the clinical usefulness of the methodology.

While our method exhibits a high degree of selectivity, resulting in a relatively lower Recall of 77.8%, the achieving goal is a Precision of 100%. This means that the classifier will classify a patient into the positive class only when it is absolutely certain, even if it may not identify all patients in advance. As a matter of fact, in cases where the tumor is not yet detectable by radiologists, it is preferable to prioritize a high level of confidence in patient predictions. Healthy individuals are not unnecessarily subjected to preventive care and frequent visits, thus conserving healthcare resources for any potential false negatives that might develop the disease later on.
This approach can represent a significant stride toward the utilization of Radiomics as a tool for predicting chronic diseases.

8.2 Related Works and Discussion

The studies by Granata et al. [138] and Wesdorp et al. [139] provide a valuable overview of the current state of *Radiomics in literature about cancer management*: the focus is predominantly on the assessment of various aspects, such as response to chemotherapy, interventional approaches, long-term survival and the prediction of genetic mutations in the primary tumor. These studies have commonly used pre-treatment CT imaging to derive radiomic features.
In contrast, my study introduces a novel application of Radiomics by predicting the development of liver metastases. It exploits a radiomic model based on CT imaging to analyze the liver parenchyma, aiming to enhance early detection of lesions. Radiomic features are derived from pre-treatment imaging to predict the first metastasis and post-hepatobiliary surgery imaging to predict recurrence. This research expands the scope of radiomics applications, particularly in the context of liver metastasis, offering the potential for earlier and more accurate detection and improved patient outcomes.

As a matter of fact, the study by Lee et al. [136] explored the prediction of metachronous liver metastases using texture features extracted from CT imaging. However, their findings did not yield significant differences between cases and controls and their results suggest that a multi-slice analysis, particularly one derived from whole liver segmentation, might be more effective in predicting the development of liver metastases. My study, focused on segmentation of the whole CT liver imaging for predicting the development of liver metastases, may offer a new perspective and approach to address the limitations and challenges identified in previous research.

The decision to exclude clinical data from models is aligned with the findings of Taghavi et al. [140], who demonstrated that a radiomic model alone can yield comparable performance to combined models that incorporate clinical features. This supports the notion that radiomic features extracted from imaging data can be highly informative and have diagnostic value on their own. In my study, there is an improvement compared to the ML methods applied in the cited study: while their Accuracy decreased from 93.3% to 86% when clinical features were excluded, Formal Methods maintains high diagnostic Accuracy, offering a robust model for predicting liver metastases. Moreover, the formal model's ability to achieve these results with a smaller dataset and fewer features is advantageous, as it can significantly reduce computational time and resource requirements, making it more practical for real-world clinical applications where time and resources are often limited.

Chapter 9

Soft Tissue

In this chapter, after reviewing the literature on the potential of Radiomics compared to conventional radiology, we used the methodology to predict the risk of metastases due to the risk of Soft Tissue sarcomas.

9.1 Increasing differential diagnosis between lipoma and liposarcoma through Radiomics: a narrative review

Soft tissue sarcomas (STS) are uncommon and diverse diseases, without frequently exhibiting noticeable symptoms. Diagnosing them is of fundamental importance, as is determining the level of malignancy, which can be high, moderate or low. Guidelines from the Italian Medical Oncology Association and the European Society of Medical Oncology advocate for the use of MRI since clinical examinations typically prove ineffective [141].

Diagnosing STS using AI techniques encounters limitations due to the availability of limited datasets, resulting in less robust methods. Nevertheless, the integration of AI techniques with Radiomics offers a novel approach to the diagnosis of rare diseases: the outcomes reported in the literature are promising in terms of performance and data interpretability.
As a novelty, the combination of Radiomics and Formal Methods enables tumor classification, site localization and the prediction of metastatic risk; in addition, the col-

laboration between computer scientists and radiologists, allows linking quantitative features to radiological evidence with exceptional performance and advancements in the diagnosis of rare diseases.

Lipomas and other *benign adipocytic tumors* rank as the most prevalent soft-tissue tumors, with lipomas alone accounting for a minimum of 30% of all benign soft tissue tumors. However, obtaining precise incidence data can be challenging, as they are frequently underreported due to the fact that many cases remain asymptomatic and do not seek medical attention.
Liposarcomas represent a category of intermediate to malignant soft-tissue tumors with lipomatous characteristics, constituting roughly 20% of all STS. In contrast, lipomas do not pose a risk of local progression or metastasis. The intermediate locally aggressive tumors are Atypical lipomatous tumor (ALT) and Well-differentiated liposarcoma (WD-LPS). The malignant ones are: Myxoid liposarcoma (M-LPS), Dedifferentiated liposarcoma (DD-LPS), Pleomorphic liposarcoma (PM-LPS) and Myxoidpleomorphic liposarcoma (MPM-LPS). Because of the similarity of ALT, WD-LPS and DD-LPS with lipoma at imaging, the gold standard used for classification is the genetic testing.

Usually, US is used for superficial lipomas, presenting homogenous hyperechogenic mass with well-defined margins. Consistently, when MRI examinations are conducted on lipomas, they consistently display a uniform isointense signal to fat across all sequences, with a consistent low signal in sequences designed to saturate fat signals.
In contrast, using US for liposarcoma tends to present as a heterogeneous, multilobulated mass, often well-defined. However, distinguishing liposarcoma from lipoma through US can be challenging; in CT scans, WD-LPS generally appears as predominantly fatty masses that contain non-lipomatous elements.

9.1.1 Radiomics vs. Radiologists

Is it possible for Radiomics to differentiate between lipomas and ALT, WD liposarcomas? By leveraging a database that includes histopathological data related to the diagnosis and MDM2 amplification, radiologists have the potential to develop a model that can distinguish between lipomas and ALT/WDLPS using the T1W TSE MRI series [142].

The *clinical visual examination* conducted by radiologists involves assessing the presence of septae thicker than 2 mm, nodules, patches of non-adipose tissue and determining whether the tumor is located deep within the tissue or near the surface. In contrast, in the radiomic analysis, images need to be segmented by selecting a ROI

where it is possible to identify disease-related markers. This *segmentation* process can be carried out either manually or automatically. When performed manually, it may be influenced by the operator and is subject to operator-dependent variations. Various software options are available for image segmentation: among these, the feature definitions in the IBEX software [143] do not adhere to the IBSI standard [20] due to the possibility of non-negligible non-conformities arising from its preprocessing steps with the primary radiomic standard.

Within the community, the most interesting findings revolve around the connections established between Radiomics and clinical data. Specifically for FIRST-Order features, the *Energy* and *Total Energy* values are inefficient because lipomas typically exhibit a mass that does not produce a substantial negative signal. Additionally, those features were less significant due to their limited specificity and AUC.
Among the second-order features, *Run Length Non Uniformity* from the GLRLM class emerged as the most effective performer: a lower value indicates a greater level of homogeneity. *Gray Level Non Uniformity* also proved to be a strong performer, with a lower value signifying a higher degree of similarity in grey levels across the image. Additionally, the *Energy* of the GLCM exhibits higher values when the image texture is more chaotic, reflecting increased scattering and greater variability among voxels. Those features can be fundamental for placing specific emphasis on the MDM2 gene [144], which is the primary objective to differentiate well-differentiated liposarcomas from lipomas using MRI images exclusively.

As a matter of fact, to determine the most suitable medical imaging modality for identifying STSs, the performance of T1w images is compared to that of T1w combined with T2w images, along with the assessments made by three expert radiologists [145]. It is observed that incorporating T2w images into the analysis can result in improved classification performance when evaluating patients.
The performance of radiologists exhibits a significant improvement in terms of sensitivity but a corresponding decrease in specificity, while their performance in terms of the AUC remains relatively consistent. Undoubtedly, the reliability of a medical radiologist with expertise can be considered stronger when complemented by Radiomics software, which can help compensate for the decrease in specificity.
The true challenge in Radiomics is to outperform musculoskeletal radiologists in differentiating Soft-Tissue Lipoma and Liposarcoma, as highlighted in literature [146]. This can be achieved by developing a predictive model using Radiomics and ML, allowing the classification using preoperative axial T1w MRI images and incorporating histological data. The MRI T1w images is robust, which allows for conducting new

experiments and deepening the understanding of these diseases.

Discussing about imaging, within the literature there are studies that developed and validated a model that relies on Radiomics, MRI images and ML to distinguish between atypical LPS (ALT) and WDLPS [147]. When working with histological data and enhanced T1-weighted MRI images obtained from various medical centers, the radiomic analysis necessitates the discretization of these images, to either enhance or reduce variations in the color channel. The impact of discretization is measured by the *Coefficient of Variation,* which indicates:

- 0-5% means absence of variations;
- 5-25% acceptable variations;
- >25% non acceptable variation.

Only 20% of the total extracted features exhibit acceptable variation. This unique characteristic warrants further in-depth investigation to determine the impact on the final diagnosis or the stability of the features, in conjunction with considerations about image resolution. Notably, patients with lipomas show variations between machines from different vendors, while patients with liposarcomas only display differences in GLCM features. This observation underscores the robustness and clinical utility of Radiomics in terms of interoperability, emphasizing its potential for widespread use in everyday clinical practice.

Radiomics has surpassed individual radiologists both in terms of accuracy and *understandability.* According to existing literature, a radiologist achieves an Accuracy rate of only 69%. The prevalent source of error stems from atypical spindle cells occasionally found in lipomas, which may lead musculoskeletal radiologists to misinterpret them as LPS, whereas ML algorithms correctly identify and differentiate them [146]. The competition between the performance of Radiomics and expert radiologists comes into focus, for example, when distinguishing between ALT and LPS using axial T1w and FS T2w MRI images [148]. On one hand, when three distinct models are constructed (one based on T1w features, another on FS T2w features and a third using T1w combined with T2w features) there appear to be no significant differences in the diagnostic efficacy of T1w or T2w models. On the other hand, considering the ability of three musculoskeletal radiologists with over a decade of experience to make diagnoses, the question arises whether a ML model might be better. Generally, the results obtained were impressive for both radiologists and Radiomics models, but distinctions emerge among them. Following the cited study [148], AUC, Specificity and

Aim	**Information Data**	**Radiomic performances**
WDLPS vs Lipoma Vos et al. [145]	MDM2 detection	Sensitivity: Radiologist > Radiomics Specificity: Radiologist < Radiomics Accuracy: Radiologist = Radiomics
LPS vs Lipoma Malinauskaite et al. [146]	Histologic	Sensitivity: Radiologist < Radiomics Specificity: Radiologist < Radiomics Accuracy: Radiologist < Radiomics
ALT vs Lipoma Tang et al. [148]	Histologic	Sensitivity: Radiologist > Radiomics Specificity: Radiologist < Radiomics Accuracy: Radiologist < Radiomics

Table 9.1: Performances comparison of Radiologists and Radiomics.

Accuracy values of radiologists are lower than those achieved by Radiomics; however, radiologists excel in terms of Sensitivity. This underscores the notion that the use of Radiomics can be a valuable asset in clinical practice, even in cases involving rare diseases, all without the need for invasive interventions or contrast-enhanced imaging.

The *ability* to diagnose STSs utilizing Radiomics and automated technologies has the potential to be more precise and effective. There are, however, a number of issues that must be resolved, including the requirement for sizable and varied datasets, the standardization of imaging methods and the validation of AI models in clinical settings. AI combined with Radiomics has the potential to enhance STS risk classification, treatment planning and monitoring. However, to fully exploit the potential of AI with Radiomics in identifying and treating STSs, more study and collaboration between physicians and data scientists is required.

9.2 Predicting risk of metastases and recurrence in soft-tissue sarcomas via Radiomics and Formal Methods

STSs are a group of *rare malignancies* which represents about 1% of all cancers [149]. Despite their low incidence, they have a metastatic percentage that can reach about 50% for high-grade STSs [150, 151]. Generally, the prognosis of patients who develop metastases is: three-year survival rate lower than 50% for those undergoing surgical metastasectomy and lower than 20% in those who are not candidate to surgery [150]. Thus, an imaging method which potentially enables the *prediction of metastases occurrence* in this set of patients might be of high benefit [151].

Assessing *tumor heterogeneity* in pathological samples obtained from biopsies can be a complex task and the information gathered may heavily rely on the specific part of the tumor sampled [152]. Recent reports have shown promise in the application of Radiomics for *musculoskeletal sarcomas* [153]; nonetheless, a common challenge is the substantial number of imaging features evaluated by the software [154].
The introduction of Formal Methods into the Radiomics process allows for a reduced dataset of patients and/or images to be used for model computation, resulting in a model that is both interpretable and comprehensible while preserving its robustness. The primary *objective* was to introduce a Formal Method for predicting distant metastases and local recurrence in extremity STSs, which has not been explored first in existing literature. Notably, this technique was non-invasive and eliminated the need for biopsy, as radiomic features were derived from non-enhanced magnetic resonance images.
An open-source de-identified database[+] was used as source of data [155, 156] with 51 histologically confirmed cases of extremity STSs. All of these patients underwent pre-treatment FDG-PET and MRI scans during the period from November 2004 to November 2011. The MRI protocols were not uniform and only T2FS or short tau inversion recovery (STIR) sequences were chosen. Among the selected exams, one was obtained in the sagittal plane, five were acquired in the coronal plane and the remaining 41 exams were captured in the axial plane. Further specifics regarding the MRI acquisition protocols can be found in the "MRIAcquisition" file available online with the public database.

[+]http://doi.org/10.7937/K9/TCIA.2015.7GO2GSKS

	Age [mean(min-max)]	Sex	Volume cm3 [median(min-max)]	N. Patients
Group A	48.5 (16-82)	13 F - 8 M	295.8 (28.8-2937.6)	21
Group B	59.9 (34-78)	11 D - 15 M	539.9 (109.3-3958.2)	26

Table 9.2: Some clinical data for Group A and Group B.

Following these criteria, the analysis includes a total of 47 patients out of the initial 51 (comprising 23 men and 24 women, with a median age of 69 years, ranging from 16 to 82 years); these patients were categorized into two groups based on clinical outcomes: "metastases/local recurrence" (group B) and "no metastases/no local recurrence" (group A). During the follow-up period, 21 patients did not experience the development of metastases or local recurrence (group A), while 26 patients did (group B). Specifically, 23 patients developed metastases and 3 patients experienced local recurrence.
In *group A*, the median duration between diagnosis and the last follow-up was 790 days, with a range of 458 to 2121 days. Meanwhile, in *group B*, the median period between diagnosis and the occurrence of metastases or local recurrence was 216.5 days, spanning from 66 to 1196 days. The median volume of segmentations, inclusive of visible edema, was 295.8 cm^3 for group A, with a range of 28.8 to 2937.6 cm^3, while group B had a median volume of 539.9 cm^3, ranging from 109.3 to 3958.2 cm^3.
In terms of histological grade, 25 patients were diagnosed with high-grade sarcoma, with 8 patients in group A and 17 in group B. Additionally, 15 patients had intermediate-grade sarcoma, consisting of 8 patients in group A and 7 in group B. Four patients were identified as having low-grade sarcoma and all of them were in group A. Histological grade information was unavailable for 3 patients, with 1 in group A and 2 in group B. For modeling purposes, two distinct MRI exams (from two different patients) were used, while the remaining 45 exams (from the remaining patients) were designated as the testing dataset.

Segmentations were derived from the previously mentioned public database and these segmentations included visible edema. Each segmentation underwent a visual assessment and adjustement by a radiologist with 7 years of experience. The 3DSlicer software (version 4.13) was employed for this process [22] and all the radiomic features were computed using PyRadiomics version 3.0.1 with the hyper-parameters detailed in the following Table 9.3.

From the segmentation of a left thigh pleomorphic sarcoma, a total of 102 fea-

Parameter	Option Name	Value
fixed bin count	binCount	50
image normalization	normalize	True
outliers to remove	removeOutliers	3
interpolator for resampling	interpolator	sitk.sitkBSpline
voxel size for resampling	resampledPixelSpacing	[0.6,0.6,0]
forcing to texture calculation	force2D	True

Table 9.3: Hyper-parameters settings for the feature extraction with PyRadiomics: other parameters were set to default. For each exam, features were extracted from each image separately.

tures were extracted. The features employed for the Formal Method classifier were selected with the Correlation Attribute Evaluation method using Weka software version 3.8.5 [33,104]. The Feature Selection process narrowed it down to 2 FIRST-Order features and 3 Shape 2D features, that were discretized in low, basal and up with the equal-width partitioning.

At the moment of the Formal Verification, the application of Formal Methods can also facilitate the localization of the *sarcoma site*. Since it is a mathematical approach, it enables the verification of the regions in the model where the property holds true: in radiological context, this is the pinpointing the location of the disease in the radiological examination. Determining the exact site of the sarcoma would further enhance the radiologist's confidence in the diagnosis made.

9.2.1 Statistical analysis and property verification

Following visual assessment, 43 segmentations were retained and 4 segmentations were manually adjusted to enhance the delineation of the lesion's profile. From these segmentations, a total of 102 Radiomics features were extracted, originating from T2FS or STIR images. However, for the formal modeling, only 5 features were taken into account:

1. *Kurtosis* (FIRST-Order feature) measures the "peakedness" of the value distribution within the ROI in the image;

2. *Skewness* (FIRST.Order feature) quantifies the asymmetry of the value distribution concerning the mean value;

3. *Elongation* (Shape feature 2D) reflects the relationship between the two principal components of the ROI shape;

4. *Sphericity* (Shape feature 2D) assesses the roundness of the tumor region's shape in relation to a circle;

5. *Mesh Surface* (Shape feature 2D) computed based on the approximated shape defined by the mesh circumference.

Once the CCS models were generated for all 47 patients, the disease property was computed by the analysis of 2 exams. Following this, the phase of Formal Model Verification examined whether the property held true when applied to the remaining 45 CCS models. The property itself was collaboratively composed by a radiologist and two computer science researchers and it is structured as follows:

$$
\begin{array}{lcl}
prop\ F_0 & = & F_1 \vee F_{10} \\
prop\ F_1 & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis\rangle F_2\ \vee \langle -\rangle X) \\
prop\ F_2 & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis\rangle F_3\ \vee \langle -\rangle X) \\
prop\ F_3 & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis\rangle F_4 \vee \\
 & & \langle -b3of3sphericity, b1of3kurtosis\rangle X) \\
prop\ F_4 & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis\rangle tt \vee \\
 & & \langle -b3of3sphericity, b1of3kurtosis\rangle X) \\
prop\ F_{10} & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis, b2of3kurtosis\rangle \\
 & & \langle b3of3_s meshsurface, b3of3elongation\rangle F_{11} \vee \langle -\rangle X) \\
prop\ F_{11} & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis, b2of3kurtosis\rangle \\
 & & \langle b3of3_s meshsurface, b2of3meshsurface\rangle F_{12} \vee \langle -\rangle X) \\
prop\ F_{12} & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis, b2of3kurtosis\rangle \\
 & & \langle b3of3_s meshsurface, b3of3elongation\rangle F_{13} \vee \langle -\rangle X) \\
prop\ F_{13} & = & (minX = \langle b3of3sphericity\rangle\langle b1of3kurtosis, b2of3kurtosis\rangle \\
 & & \langle b3of3elongation, b3of3meshsurface\rangle F_{14} \vee \langle -\rangle X) \\
prop\ F_{14} & = & (minX = \langle b2of3meshsurface, b2of3elongation\rangle [-] ff \vee \langle -\rangle X)
\end{array}
$$

Table 9.4: Formal Property to verify on the patient's model the presence of the disease.

To measure the *effectiveness* of the property in predicting the occurrence or absence of metastases/local recurrence (comparing group B to group A), various metrics were taken into consideration: Specificity, Sensitivity (also known as Recall), Accuracy, Positive predictive value and Negative predictive value. Additionally, the intercorrelation among the chosen features was computed using the Spearman correlation coefficient in Figure 9.1.

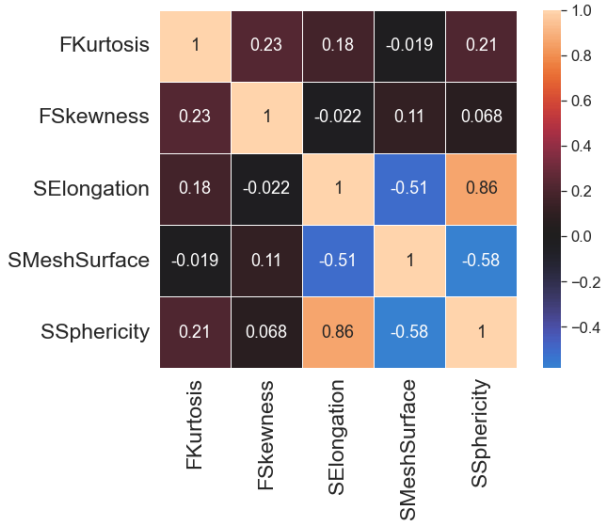


Figure 9.1: Inter-correlation among selected features calculated with Spearman correlation coefficient.

After confirming the validity of the property, the subsequent step involved the computation of the following metrics of Table 9.5. The positive CUI calculated was 0.606 and the negative CUI was 0.491. Both of them demonstrate a satisfactory or fair utility level. When constructing the Confusion Matrix for the classification, in the Actual Group A of 21 patients, 14 were accurately identified, while 7 were misclassified as Group B. In the Actual Group B of 26 patients, 21 were correctly classified and 5 were mistakenly labeled as Group A.

Accuracy Statistics	**Value**	
Recall (Sensitivity)	0.81%	
Specificity	0.61%	
Accuracy	0.74%	
Precision (Positive Predictive Value)	0.75%	
Negative Predictive Value	0.74%	
Utility statistics	**Rating**	**95% Confidence Interval**
Clinical Utility (+Ve)	Fair	0.606
Clinical Utility (-Ve)	Fair	0.491

Table 9.5: Accuracy and utility indexes according to Mitchell statistics [32] to gain more information about the clinical usefulness of the methodology.

9.3 Related Works and Discussion

Radiomics has the potential in providing decision support for diagnosing complex diseases, for example distinguishing between the grade of LPS to early predict metastatic risk in patients with high-grade neoplasms. This assessment could have significant implications for determining appropriate treatment strategies and reducing the likelihood of recurrence. It also has the potential to aid in selecting the most effective therapeutic interventions for tumors, ultimately leading to better patient outcomes.

Regarding *radiomic features for classification of STS*, Kurtosis, Skewness, Elongation and Sphericity have been associated with specific characteristics and prognostic factors in various tumors. For example, a positive *Skewness* and higher *Kurtosis* have been linked to increased heterogeneity in tumors [157, 158]. Heterogeneity refers to variations in tissue characteristics within the tumor and higher heterogeneity can be an indicator of a poorer prognosis. Indeed, the presence of at least two out of three characteristics, including heterogeneity, necrosis and peritumoral enhancement, was shown to predict overall survival and metastasis-free survival in STSs [159].
Researchers have also used features like *Elongation* and *Sphericity* to differentiate between low-grade and high-grade STSs: these features may capture structural characteristics of tumors that are indicative of their grade, which is an important factor in cancer diagnosis and prognosis. However, the Spearman correlation coefficient of 0.86 between Elongation and Sphericity suggests a strong positive correlation between these two radiomic features; I decided to retain both features based on the article of Peeken et al. [160].
In the study by Crombe et al. [161], authors developed radiomic-based models to predict metastatic relapse within a 2-year timeframe. They utilized a training dataset consisting of 50 patients and tested their best supervised model on a cohort of 20 patients, achieving an Accuracy of 0.75. Similarly, in the work conducted by Tian et al. [162], they employed a training dataset comprising 54 sarcomas and a testing dataset with 23 sarcomas achieving AUC and Accuracy values of 0.902 and 0.913, respectively.

The aforementioned studies focused on specific histotypes or utilized the same MRI scanner; in contrast, my study encompassed various histologies and involved different MRI scanners. A *limitation* of my study stems from variations in histological characteristics; as reported in sources like [163], the likelihood of distant metastasis in STSs varies widely, ranging from 20% to nearly 100%, depending on the tumor grade and histological type. This study did not delve into the impact of histological variations, hoping for further exploration in future research.

Chapter 10

Discussion

Radiomics is not only accompanied by AI techniques, which have shown promise in the automatic extraction of radiomic features and image analysis [164]. Radiomic models have exhibited the capability to forecast various tumor features such as histology, genetic traits and treatment responses like pathological reactions, recurrence, lymph node involvement and survival across different cancer types. But, are Radiomics and other techniques really implementable for daily practice?

10.1 Never without Artificial Intelligence

Artificial Intelligence has become an integral part of our modern world and there are different reasons why we might find it challenging to imagine life without AI. For example, AI technologies have automated many tasks, *increasing efficiency and productivity* in almost all industries: think of repetitive, dangerous or high-risk works. From manufacturing to data analysis, AI is able to *perform repetitive and time-consuming tasks* faster and more accurately than humans and this is happening because ML can analyse large data sets to discover valuable insights and patterns. As a matter of fact, in the new medicine, even AI is starting to be used to perform surgeries remotely [165] and to revolutionise healthcare by aiding disease diagnosis, drug discovery and personalised treatment plans.

While, on the one hand, the availability of data to be analysed grows, on the other hand, AI can become a *threat to cyber and medical security* [166]. Since the introduction of these techniques into the real world, several issues have been raised

concerning privacy rights and the handling of special data [167].
In summary, AI has already become deeply embedded in our daily lives, leading to greater efficiency, innovation and improved quality of life. Although there are legitimate concerns about the ethical and social implications, it is clear that many of the benefits of progress that seem obvious are not natural, but due to the use of AI.

In this thesis, efforts were made to minimize the intervention of AI during the analysis workflow. However, it is not possible to eliminate AI altogether: the amount of data created by Radiomics requires methods that can manage all the information. In this context, AI and statistics are closely related [168]. Statistics is a fundamental component of AI and its methods are fundamental to the development of learning systems that recognize models and make decisions without explicit programming. While Statistics is primarily concerned with the analysis of underlying data and probabilities, AI goes beyond statistical methods, incorporating broader concepts such as knowledge representation, problem solving and decision making. However, the two fields intersect significantly, with AI often using statistical principles.

10.2 Limits of Radiomics

The main problem with Radiomics lies in its *integration into clinical practice*. The literature says that Radiomics helps in diagnosis, treatment planning and monitoring of treatment response, but in the real world it is still difficult to find a system with radiomic analysis integration, partly because different diseases are not yet correlated with different features.

Another problem is the *variability of the features*, namely for each disease and for each study, there can be different *suitable features*. Researchers are deepening this open issue through radiomic study with multimodal and multicentric radiological cases, e.g. combining CT and MRI of 4 different hospitals, because this may provide more comprehensive information on a patient's condition. In addition, some state-of-the-art centres are able to supplement image information with *histological and clinical data* from patients: this allows for a more robust analysis and probably a better classification.
Fortunately, the IBSI standard [20] is essential to ensure reproducibility of results between different institutions. Perhaps in the coming years, thanks to a rich literature on the features used, researchers will be able to establish which parameters are best for each type of disease and imaging. Note that Radiomics is also used for neurological disorders, cardiovascular diseases and other non-oncological applications.

Like any medical technology, this innovative technique today raises *ethical and regulatory considerations*, especially with regard to patient privacy and data security, which are stored in the radiological examinations. The use of patient imaging data for research purposes must comply with strict privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the US [169] and the *General Data Protection Regulation* (GDPR) in Europe [170].
Radiomics is fascinating as it involves radiologists, data scientists, bioinformaticians and physicians. Multidisciplinary teams work together to develop and validate radiomic models, analyse data and enrich the literature with scientific evidences.

10.3 Images are treasures

Gillies opened up a new view on data analysis to the world [7]. Just as the images taken from mobile phone or those found on the internet can be processed, analysed and manipulated by computers, so too can medical images be subject to analysis. Everything is based on the concept that digital images are made up of individual elements in 2D, called *pixels* or in 3D, called *voxels*. Each pixel represents a data point with specific values of colour, brightness and other attributes.

From the image study was born the *Computer vision* [171] is a subfield of AI that focuses on the development of algorithms, models and systems for the extraction, processing and interpretation of visual data from the physical world. In general, *image processing* techniques are employed to identify objects, extract relevant features and comprehend visual information. Conversely, in ML, images serve as input data for training artificial neural networks, which are employed in tasks such as image classification and object detection.
In the field of *healthcare* [172], medical imaging play a vital role in diagnosing and monitoring medical conditions. These images represent a structured form of data, with values that provide descriptions of their visual content. As in other domains, medical images undergo various processing steps, such as image enhancement, object segmentation and feature extraction, utilizing specific techniques and algorithms. For instance, there are techniques like "*Filtering*" to enhance or suppress particular features within an image (e.g., Gaussian and Sobel filters) and "*Transformation*" methods for frequency domain analysis (e.g., Wavelet transforms). These processes enable healthcare professionals to extract valuable information from medical images, aiding in diagnosis and patient care. *Edge detection* is a critical technique in the realm of medical images, as it enables the identification of boundaries within an area

by detecting rapid changes in pixel intensity. This capability is of paramount importance in the detection of tumors and other medical conditions, because it is the basis for the definition of the *Region of Interest* [173]. Recent advancements in this field have led to the development of new methods for the automatic delineation of ROI through a process known as "*Thresholding*". This technique involves separating the foreground from the background to extract specific features within a particular area of interest.

Real-world medical images, however, often present *challenges* such as variations in lighting and artifacts resulting from patient positions or movements during the examination. *Interpreting images* transcends pixel-level processing and encompasses higher-level tasks such as object recognition, anomaly detection and understanding the relationships between objects in a scene. These are tasks that closely align with the responsibilities of a radiologist [174]. Therefore, it is essential to emphasize that Radiomics cannot replace the role of a medical expert. While Radiomics and advanced technologies play a valuable role in medical image analysis and diagnosis, the expertise and clinical judgment of trained professionals remain irreplaceable in the field of healthcare.

Chapter 11

Conclusion

Radiomics, often combined with AI techniques, is promising for medical image analysis and can really make a difference in the diagnosis of disease or the prediction of treatment for various diseases. Radiomic models built through Formal Methods can also add interpretability and reliability to the results, which are currently lacking in some works in the literature. Indeed, thanks to the methodology presented in this thesis, it is possible to understand the connection between the classifier's decision and the radiomic numerical characteristics, while maintaining high performance.
In our case, the *most promising study* was on the liver: we were able to predict patients who would have liver metastases due to colorectal cancer with a *level of performance* around 93%. The *most difficult work*, however, was on the brain: our methodology tried to predict Multiple Sclerosis with a performance level around 67%. We know that brain images are very complex structurally as they include even more information. Nevertheless, we managed to maintain a satisfactory performance as Formal Methods are knowledge-based, where the domain expert is needed, which makes them more robust and reliable.
However, it remains the fundamental role of AI in efficiency and productivity in medicine, with an impact on surgery, disease diagnosis, drug discovery and personalized treatment plans. In future work, one might think new type to model radiological examination, compare different discretization of features and deepen the connections between numerical values and radiomic features thanks to the explainability provided by Formal Methods.

The challenges of integrating Radiomics into clinical practice and of feature variability between various diseases are often raised in the scientific community. However, Radiomics remains a fascinating multidisciplinary field involving radiologists, data and computer scientists, bioinformaticians and physicians. Everything is based on Gilles' statement: "*Radiomics: images are more than images, they are data*".

Acknowledgements

Abbreviations

The following abbreviations are used in this manuscript:

EHR	Electronic Health Records
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
US	Ultrasound
SPECT	Single Photon Emission Computed Tomography
PET	Positron Emission Tomography
CAD	Computer-Aided Diagnosis and Detection system
ROI	Region of interest
VOI	Volume of Interest
DICOM	Digital Imaging and COmmunications in Medicine
NIFTI	NeuroImaging Informatics Technology Initiative
NRRD	Nearly Raw Raster Data
IBSI	Image Biomarker Standardisation Initiative
CCS	Calculus of Communicating System
LTS	Labelled Transition System
CWB-NC	Concurrency Workbench - New Century
CUI	Clinical Utility Index
MS	Multiple sclerosis
CNS	Central nervous system
Gd+	Gadolinium-enhancing
RRMS	Relapsing-remitting multiple sclerosis

GBCAs	gadolinium-based contrast agents
NSF	Nephrogenic systemic fibrosis
TR	Repetition Time
ET	Echo Time
T1w	T1-weighted
TLE	Temporal Lobe Epilepsy
18FDG-PET	18F-fluorodeoxyglucose PET
EEG	electroencephalography
DIR	double inversion recovery
dMRI	enhanced diffusion MRI
SPGR	Spoiled Gradient Echo
FLAIR	Fluid-Attenuated Inversion Recovery
AUC	Area Under the Curve
FCD	Focal cortical dysplasia
BC	Breast cancer
BI-RADS	Breast Imaging Reporting and Data System
BPE	Breast parenchymal enhancement
AC	Adhesive Capsulitis
HRCT	High Resolution Computed Tomography
COVID-19	SARS-CoV2 or Coronavirus
PE	Pulmonary Embolism
GGO	Ground glass opacity
GAN	Generative Adversarial Networks
ICU	Intensive care unit
ROC	Receiver Operating Characteristic
MLA	Mean Lung Attenuation
IPF	Idiopathic pulmonary fibrosis
ILD	Interstitial lung disease
ICC	Intraclass Correlation Coefficient
CRC	Colorectal cancer
CRCLM	Liver metastases originating from colorectal cancer
CCG	Case control group
CG	Case group
STS	Soft tissue sarcoma
ALT	Atypical lipomatous tumor

WD-LPS	Well-differentiated liposarcoma
M-LPS	Myxoid liposarcoma
DD-LPS	Dedifferentiated liposarcoma
PM-LPS	Pleomorphic liposarcoma
MPM-LPS	Myxoidpleomorphic liposarcoma
STIR	Short tau inversion recovery
HIPAA	Health Insurance Portability and Accountability Act
GDPR	General Data Protection Regulation

Funding

These work has been partially supported by MUR - REASONING: foRmal mEthods for computAtional analySis for diagnOsis and progNosis in imagING - PRIN.
These work has been partially supported by e-DAI (Digital ecosystem for integrated analysis of heterogeneous health data related to high-impact diseases: innovative model of care and research), Health Operational Plan, FSC 2014-2020, PRIN-MUR-Ministry of Health.
These work has been partially supported by National Plan for NRRP Complementary Investments D^3 4 Health: Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care.

References

[1] E. A. Feigenbaum and B. G. Buchanan, "Dendral and meta-dendral," *Artificial Intelligence*, vol. 59, pp. 233–240, 1993.

[2] E. H. Shortliffe, "Mycin: A rule-based computer program for advising physicians regarding antimicrobial therapy selection," Ph.D. dissertation, Stanford University Ph. D. dissertation, 1974.

[3] K. S. Metaxiotis and J.-E. Samouilidis, "Expert systems in medicine: academic illusion or real power?" *Information management & computer security*, vol. 8, no. 2, pp. 75–79, 2000.

[4] W. J. von Eschenbach, "Transparency and the black box problem: Why we do not trust ai," *Philosophy & Technology*, vol. 34, no. 4, pp. 1607–1622, 2021.

[5] E. Abrahams and M. Silver, "The history of personalized medicine," *Integrative neuroscience and personalized medicine*, pp. 3–16, 2010.

[6] C. M. Micheel, S. J. Nass, G. S. Omenn, *et al.*, "Committee on the review of omics-based tests for predicting patient outcomes in clinical trials; board on health care services; board on health sciences policy; institute of medicine. evolution of translational omics: Lessons learned and the path forward," *Evolution of translational omics: lessons learned and the path forward. Washington, DC: National Academies Press*, 2012.

[7] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[8] G. S. Omenn, S. J. Nass, C. M. Micheel, *et al.*, *Evolution of translational omics: lessons learned and the path forward.* National Academies Press, 2012.

[9] M. A. Mazurowski, "Radiogenomics: what it is and why it is important," *Journal of the American College of Radiology*, vol. 12, no. 8, pp. 862–866, 2015.

[10] B. Zhao, "Understanding sources of variation to improve the reproducibility of radiomics," *Frontiers in oncology*, vol. 11, p. 826, 2021.

[11] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "Computer-aided grading of gliomas combining automatic segmentation and radiomics," *International journal of biomedical imaging*, vol. 2018, 2018.

[12] A. Stefano, A. Comelli, V. Bravatà, S. Barone, I. Daskalovski, G. Savoca, M. G. Sabini, M. Ippolito, and G. Russo, "A preliminary pet radiomics study of brain metastases using a fully automatic segmentation method," *BMC bioinformatics*, vol. 21, no. 8, pp. 1–14, 2020.

[13] C. Gu and M.-C. Lee, "Semiautomatic segmentation and tracking of semantic video objects," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 572–584, 1998.

[14] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.

[15] K. I. Laws, "Texture energy measures," in *Proc. Image understanding workshop*, 1979, pp. 47–51.

[16] P. Grossmann, O. Grove, N. El-Hachem, *et al.*, "Identification of molecular phenotypes in lung cancer by integrating radiomics and genomics," *Sci Transl Med*, 2017.

[17] C. Parmar, P. Grossmann, J. Bussink, and P. Lambin, "Machine learning methods for quantitative radiomic biomarkers," *Machine learning applications for Radiomics*, vol. 5, p. 125, 2017.

[18] M. Coriasco, N. Balossino, O. Rampado, and S. Rabellino, *L'immagine digitale in diagnostica per immagini.* Springer, 2013.

[19] M. Larobina and L. Murino, "Medical image file formats," *Journal of digital imaging*, vol. 27, pp. 200–206, 2014.

[20] M. Hatt, M. Vallieres, D. Visvikis, and A. Zwanenburg, "Ibsi: an international community radiomics standardization initiative," 2018.

[21] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.

[22] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

[23] G. M. Nielson, "On marching cubes," *IEEE Transactions on visualization and computer graphics*, vol. 9, no. 3, pp. 283–297, 2003.

[24] S. Salzberg, "Distance metrics for instance-based learning," in *International Symposium on Methodologies for Intelligent Systems*. Springer, 1991, pp. 399–408.

[25] D. F. Sprinz and Y. Wolinsky-Nahmias, *Models, numbers, and cases: methods for studying international relations*. University of Michigan Press, 2004.

[26] F. Mercaldo, R. Casolare, G. Ciaramella, G. Iadarola, F. Martinelli, F. Ranieri, and A. Santone, "A real-time method for can bus intrusion detection by means of supervised machine learning," *Proceedings of the 19th International Conference on Security and Cryptography (SECRYPT 2022)*, 2022.

[27] L. Coppolino, R. Nardone, A. Petruolo, L. Romano, and A. Souvent, "Exploiting digital twin technology for cybersecurity monitoring in smart grids," in *Proceedings of the 18th International Conference on Availability, Reliability and Security*, 2023, pp. 1–10.

[28] R. Milner, *A calculus of communicating systems*. Springer, 1980.

[29] E. M. Clarke, "Model checking," in *Foundations of Software Technology and Theoretical Computer Science: 17th Conference Kharagpur, India, December 18–20, 1997 Proceedings 17*. Springer, 1997, pp. 54–56.

[30] E. A. Emerson, "Model checking and the mu-calculus." *Descriptive Complexity and Finite Models*, vol. 31, pp. 185–214, 1996.

[31] R. Cleaveland, J. Parrow, and B. Steffen, "The concurrency workbench: A semantics-based tool for the verification of concurrent systems," *ACM Transactions on Programming Languages and Systems (TOPLAS)*, vol. 15, no. 1, pp. 36–72, 1993.

[32] A. J. Mitchell, "Sensitivity× ppv is a recognized test called the clinical utility index (cui+)," *European journal of epidemiology*, vol. 26, pp. 251–252, 2011.

[33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *ACM SIGKDD explorations newsletter*, vol. 11, no. 1, pp. 10–18, 2009.

[34] J. Demšar, T. Curk, A. Erjavec, Črt Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, and B. Zupan, "Orange: Data mining toolbox in python," *Journal of Machine Learning Research*, vol. 14, pp. 2349–2353, 2013.

[35] R. Milner, "Four combinators for concurrency," in *Proceedings of the first ACM SIGACT-SIGOPS symposium on Principles of distributed computing*, 1982, pp. 104–110.

[36] P. Zhang, Y. Zhong, Y. Deng, X. Tang, and X. Li, "Cosingan: learning covid-19 infection segmentation from a single radiological image," *Diagnostics*, vol. 10, no. 11, p. 901, 2020.

[37] J. Civit-Masot, F. Luna-Perejón, M. Domínguez Morales, and A. Civit, "Deep learning system for covid-19 diagnosis aid using x-ray pulmonary images," *Applied Sciences*, vol. 10, no. 13, p. 4640, 2020.

[38] S. Rajaraman and S. Antani, "Weakly labeled data augmentation for deep learning: a study on covid-19 detection in chest x-rays," *Diagnostics*, vol. 10, no. 6, p. 358, 2020.

[39] L. Brunese, F. Martinelli, F. Mercaldo, and A. Santone, "Machine learning for coronavirus covid-19 detection from chest x-rays," *Procedia Computer Science*, vol. 176, pp. 2212–2221, 2020.

[40] E. Sorantin, M. G. Grasser, A. Hemmelmayr, S. Tschauner, F. Hrzic, V. Weiss, J. Lacekova, and A. Holzinger, "The augmented radiologist: Artificial intelligence in the practice of radiology," *Pediatric Radiology*, pp. 1–13, 2021.

[41] A. Santone, G. Vaglini, and M. L. Villani, "Incremental construction of systems: An efficient characterization of the lacking sub-system," *Science of Computer Programming*, vol. 78, no. 9, pp. 1346–1367, 2013.

[42] S. Gradara, A. Santone, and M. Villani, "Delfin+: An efficient deadlock detection tool for ccs processes," *Journal of Computer and System Sciences*, vol. 72, no. 8, pp. 1397–1412, 2006.

[43] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys (CSUR)*, vol. 51, no. 5, p. 93, 2018.

[44] L. Dercle, S. Ammari, M. Bateson, P. B. Durand, E. Haspinger, C. Massard, C. Jaudet, A. Varga, E. Deutsch, J.-C. Soria, *et al.*, "Limits of radiomic-based entropy as a surrogate of tumor heterogeneity: Roi-area, acquisition protocol and tissue site exert substantial influence," *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[45] S. Zannos, *Godfrey Hounsfield and the Invention of Cat Scans (Unlocking the Secrets of Science)*. Mitchell Lane Pub Inc, August 1, 2002.

[46] A. A. Taha, A. Hanbury, and O. A. J. del Toro, "A formal method for selecting evaluation metrics for image segmentation," in *2014 IEEE international conference on image processing (ICIP)*. IEEE, 2014, pp. 932–936.

[47] F. Banci Buonamici, G. Belmonte, V. Ciancia, D. Latella, and M. Massink, "Spatial logics and model checking for medical imaging," *International Journal on Software Tools for Technology Transfer*, vol. 22, pp. 195–217, 2020.

[48] D. Li, M. J. Li, A. Traboulsee, G. Zhao, A. Riddehough, and D. Paty, "The use of mri as an outcome measure in clinical trials." *Advances in neurology*, vol. 98, pp. 203–226, 2006.

[49] F. Bagnato, "Gadolinium should always be used to assess disease activity in ms–commentary," *Multiple Sclerosis Journal*, vol. 26, no. 7, pp. 769–770, 2020.

[50] C. Granziera and D. S. Reich, "Gadolinium should always be used to assess disease activity in ms–yes," *Multiple Sclerosis Journal*, vol. 26, no. 7, pp. 765–766, 2020.

[51] À. Rovira and M. P. Wattjes, "Gadolinium should always be used to assess disease activity in ms–no," *Multiple Sclerosis Journal*, vol. 26, no. 7, pp. 767–769, 2020.

[52] T. Kanda, K. Ishii, H. Kawaguchi, K. Kitajima, and D. Takenaka, "High signal intensity in the dentate nucleus and globus pallidus on unenhanced t1-weighted mr images: relationship with increasing cumulative dose of a gadolinium-based contrast material," *Radiology*, vol. 270, no. 3, pp. 834–841, 2014.

[53] Y. Errante, V. Cirimele, C. A. Mallio, V. Di Lazzaro, B. B. Zobel, and C. C. Quattrocchi, "Progressive increase of t1 signal intensity of the dentate nucleus on unenhanced magnetic resonance images is associated with cumulative doses of intravenously administered gadodiamide in patients with normal renal function, suggesting dechelation," *Investigative radiology*, vol. 49, no. 10, pp. 685–690, 2014.

[54] L. D. Weberling, P. J. Kieslich, P. Kickingereder, W. Wick, M. Bendszus, H.-P. Schlemmer, and A. Radbruch, "Increased signal intensity in the dentate nucleus on unenhanced t1-weighted images after gadobenate dimeglumine administration," *Investigative radiology*, vol. 50, no. 11, pp. 743–748, 2015.

[55] M. Adin, L. Kleinberg, D. Vaidya, E. Zan, S. Mirbagheri, and D. Yousem, "Hyperintense dentate nuclei on t1-weighted mri: relation to repeat gadolinium administration," *American Journal of Neuroradiology*, vol. 36, no. 10, pp. 1859–1865, 2015.

[56] M. P. Sormani, B. Stubinski, P. Cornelisse, S. Rocak, D. Li, and N. D. Stefano, "Magnetic resonance active lesions as individual-level surrogate for relapses in multiple sclerosis," *Multiple Sclerosis Journal*, vol. 17, no. 5, pp. 541–549, 2011.

[57] R. A. Bermel, X. You, P. Foulds, R. Hyde, J. H. Simon, E. Fisher, and R. A. Rudick, "Predictors of long-term outcome in multiple sclerosis patients treated with interferon beta," *Annals of neurology*, vol. 73, no. 1, pp. 95–103, 2013.

[58] D. Miller, F. Barkhof, and J. Nauta, "Gadolinium enhancement increases the sensitivity of mri in detecting disease activity in multiple sclerosis," *Brain*, vol. 116, no. 5, pp. 1077–1094, 1993.

[59] B. J. Bedell and P. A. Narayana, "Automatic segmentation of gadolinium-enhanced multiple sclerosis lesions," *Magnetic resonance in medicine*, vol. 39, no. 6, pp. 935–940, 1998.

[60] Z. Karimaghaloo, M. Shah, S. J. Francis, D. L. Arnold, D. L. Collins, and T. Arbel, "Automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain mri using conditional random fields," *IEEE transactions on medical imaging*, vol. 31, no. 6, pp. 1181–1194, 2012.

[61] R. He and P. A. Narayana, "Automatic delineation of gd enhancements on magnetic resonance images in multiple sclerosis," *Medical physics*, vol. 29, no. 7, pp. 1536–1546, 2002.

[62] S. Datta, B. R. Sajja, R. He, R. K. Gupta, J. S. Wolinsky, and P. A. Narayana, "Segmentation of gadolinium-enhanced lesions on mri in multiple sclerosis," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 25, no. 5, pp. 932–937, 2007.

[63] E. M. Sweeney, T. D. Nguyen, A. Kuceyeski, S. M. Ryan, S. Zhang, L. Zexter, Y. Wang, and S. A. Gauthier, "Estimation of multiple sclerosis lesion age on magnetic resonance imaging," *Neuroimage*, vol. 225, p. 117451, 2021.

[64] Y. Liu, D. Dong, L. Zhang, Y. Zang, Y. Duan, X. Qiu, J. Huang, H. Dong, F. Barkhof, C. Hu, *et al.*, "Radiomics in multiple sclerosis and neuromyelitis optica spectrum disorder," *European radiology*, vol. 29, no. 9, pp. 4670–4677, 2019.

[65] X. Ma, L. Zhang, D. Huang, J. Lyu, M. Fang, J. Hu, Y. Zang, D. Zhang, H. Shao, L. Ma, *et al.*, "Quantitative radiomic biomarkers for discrimination between neuromyelitis optica spectrum disorder and multiple sclerosis," *Journal of Magnetic Resonance Imaging*, vol. 49, no. 4, pp. 1113–1121, 2019.

[66] M. Filippi, P. Preziosa, B. L. Banwell, F. Barkhof, O. Ciccarelli, N. De Stefano, J. J. Geurts, F. Paul, D. S. Reich, A. T. Toosy, *et al.*, "Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines," *Brain*, vol. 142, no. 7, pp. 1858–1875, 2019.

[67] S. Cakirer, E. Karaarslan, and A. Arslan, "Spontaneously t1-hyperintense lesions of the brain on mri: a pictorial review," *Current Problems in Diagnostic Radiology*, vol. 32, no. 5, pp. 194–217, 2003.

[68] G. Pontillo, S. Tommasin, R. Cuocolo, M. Petracca, N. Petsas, L. Ugga, A. Carotenuto, C. Pozzilli, R. Iodice, R. Lanzillo, *et al.*, "A combined radiomics

and machine learning approach to overcome the clinicoradiologic paradox in multiple sclerosis," *American Journal of Neuroradiology*, vol. 42, no. 11, pp. 1927–1933, 2021.

[69] Y. Peng, Y. Zheng, Z. Tan, J. Liu, Y. Xiang, H. Liu, L. Dai, Y. Xie, J. Wang, C. Zeng, *et al.*, "Prediction of unenhanced lesion evolution in multiple sclerosis using radiomics-based models: a machine learning approach," *Multiple Sclerosis and Related Disorders*, vol. 53, p. 102989, 2021.

[70] S. Rastogi, C. Lee, and N. Salamon, "Neuroimaging in pediatric epilepsy: a multimodality approach," *Radiographics*, vol. 28, no. 4, pp. 1079–1095, 2008.

[71] W. Muhlhofer, Y.-L. Tan, S. G. Mueller, and R. Knowlton, "Mri-negative temporal lobe epilepsy—what do we know?" *Epilepsia*, vol. 58, no. 5, pp. 727–742, 2017.

[72] W. H. Theodore, S. Sato, C. Kufta, M. Balish, E. Bromfield, and D. B. Leiderman, "Temporal lobectomy for uncontrolled seizures: the role of positron emission tomography," *Annals of neurology*, vol. 32, no. 6, pp. 789–794, 1992.

[73] W. H. Theodore, M. E. Newmark, S. Sato, R. Brooks, N. Patronas, R. de la Paz, G. Dichiro, R. M. Kessler, R. Margolin, R. G. Manning, *et al.*, "{18F} fluorodeoxyglucose positron emission tomography in refractory complex partial seizures," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 14, no. 4, pp. 429–437, 1983.

[74] B. W. Abou-Khalil, G. J. Siegel, J. Chris Sackellares, S. Gilman, R. Hichwa, and R. Marshall, "Positron emission tomography studies of cerebral glucose metabolism in chronic partial epilepsy," *Annals of neurology*, vol. 22, no. 4, pp. 480–486, 1987.

[75] T. Tanriverdi, A. Ajlan, N. Poulin, and A. Olivier, "Morbidity in epilepsy surgery: an experience based on 2449 epilepsy surgery procedures from a single institution," *Journal of neurosurgery*, vol. 110, no. 6, pp. 1111–1123, 2009.

[76] D. Sone, "Making the invisible visible: Advanced neuroimaging techniques in focal epilepsy," *Frontiers in Neuroscience*, vol. 15, p. 699176, 2021.

[77] I. Beheshti, D. Sone, N. Maikusa, Y. Kimura, Y. Shigemoto, N. Sato, and H. Matsuda, "Flair-wise machine-learning classification and lateralization of

mri-negative 18f-fdg pet-positive temporal lobe epilepsy," *Frontiers in neurology*, vol. 11, p. 580713, 2020.

[78] E.-N. Cheong, J. E. Park, D. E. Jung, and W. H. Shim, "Extrahippocampal radiomics analysis can potentially identify laterality in patients with mri-negative temporal lobe epilepsy," *Frontiers in neurology*, vol. 12, p. 706576, 2021.

[79] Y. W. Park, Y. S. Choi, S. E. Kim, D. Choi, K. Han, H. Kim, S. S. Ahn, S.-A. Kim, H. J. Kim, S.-K. Lee, *et al.*, "Radiomics features of hippocampal regions in magnetic resonance imaging can differentiate medial temporal lobe epilepsy patients from healthy controls," *Scientific reports*, vol. 10, no. 1, p. 19567, 2020.

[80] M. Wang, S. Zheng, X. Li, and X. Qin, "A new image denoising method based on gaussian filter," in *2014 International Conference on information science, electronics and electrical engineering*, vol. 1. IEEE, 2014, pp. 163–167.

[81] D. Chard and S. A. Trip, "Resolving the clinico-radiological paradox in multiple sclerosis," *F1000Research*, vol. 6, 2017.

[82] A. A. Ardakani, N. J. Bureau, E. J. Ciaccio, and U. R. Acharya, "Interpretation of radiomics features: a pictorial review," *Computer Methods and Programs in Biomedicine*, p. 106609, 2021.

[83] Q. Zhang, Y. Liao, X. Wang, T. Zhang, J. Feng, J. Deng, K. Shi, L. Chen, L. Feng, M. Ma, *et al.*, "A deep learning framework for 18 f-fdg pet imaging diagnosis in pediatric patients with temporal lobe epilepsy," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 48, pp. 2476–2485, 2021.

[84] F. Galati, G. Moffa, and F. Pediconi, "Breast imaging: Beyond the detection," *European Journal of Radiology*, vol. 146, p. 110051, 2022.

[85] A. S. Tagliafico, M. Piana, D. Schenone, R. Lai, A. M. Massone, and N. Houssami, "Overview of radiomics in breast cancer diagnosis and prognostication," *The Breast*, vol. 49, pp. 74–80, 2020.

[86] J. Seo, E. Ko, B.-K. Han, E. Y. Ko, J. H. Shin, and S. Y. Hahn, "Automated volumetric breast density estimation: a comparison with visual assessment," *Clinical radiology*, vol. 68, no. 7, pp. 690–695, 2013.

[87] D. Saslow, J. Hannan, J. Osuch, M. H. Alciati, C. Baines, M. Barton, J. K. Bobo, C. Coleman, M. Dolan, G. Gaumer, *et al.*, "Clinical breast examination:

practical recommendations for optimizing performance and reporting," *CA: a cancer journal for clinicians*, vol. 54, no. 6, pp. 327–344, 2004.

[88] C. Balleyguier, S. Ayadi, K. Van Nguyen, D. Vanel, C. Dromain, and R. Sigal, "Birads™ classification in mammography," *European journal of radiology*, vol. 61, no. 2, pp. 192–194, 2007.

[89] M. M. Eberl, C. H. Fox, S. B. Edge, C. A. Carter, and M. C. Mahoney, "Bi-rads classification for management of abnormal mammograms," *The Journal of the American Board of Family Medicine*, vol. 19, no. 2, pp. 161–164, 2006.

[90] E. Fleury and K. Marcomini, "Performance of machine learning software to classify breast lesions using bi-rads radiomic features on ultrasound images," *European Radiology Experimental*, vol. 3, pp. 1–8, 2019.

[91] S. V. Perre, L. Duron, A. Milon, A. Bekhouche, D. Balvay, F. H. Cornelis, L. Fournier, and I. Thomassin-Naggara, "Radiomic analysis of htr-dce mr sequences improves diagnostic performance compared to bi-rads analysis of breast mr lesions," *European radiology*, vol. 31, pp. 4848–4859, 2021.

[92] F. Angelone, C. Ricciardi, G. Gatta, M. Sansone, A. M. Ponsiglione, M. P. Belfiore, F. Amato, and R. Grassi, "Breast density analysis on mammograms: Application of machine learning with textural features," in *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*. IEEE, 2022, pp. 295–300.

[93] M. Sansone, R. Grassi, M. P. Belfiore, G. Gatta, F. Grassi, F. Pinto, G. V. La Casella, R. Fusco, S. Cappabianca, V. Granata, *et al.*, "Radiomic features of breast parenchyma: assessing differences between for processing and for presentation digital mammography," *Insights Into Imaging*, vol. 12, pp. 1–13, 2021.

[94] J. P. Tasto and D. W. Elias, "Adhesive capsulitis," *Sports medicine and arthroscopy review*, vol. 15, no. 4, pp. 216–221, 2007.

[95] M. Zappia, F. Di Pietto, A. Aliprandi, S. Pozza, P. De Petro, A. Muda, and L. M. Sconfienza, "Multi-modal imaging of adhesive capsulitis of the shoulder," *Insights into imaging*, vol. 7, no. 3, pp. 365–371, 2016.

[96] P. A. G. Teixeira, C. Balaj, A. Chanson, S. Lecocq, M. Louis, and A. Blum, "Adhesive capsulitis of the shoulder: value of inferior glenohumeral ligament signal

changes on t2-weighted fat-saturated images," *American Journal of Roentgenology*, vol. 198, no. 6, pp. W589–W596, 2012.

[97] W. Zhao, X. Zheng, Y. Liu, W. Yang, V. Amirbekian, L. E. Diaz, and X. Huang, "An mri study of symptomatic adhesive capsulitis," *PLoS One*, 2012.

[98] D. Hayashi, E. Gould, R. Shroyer, E. van Staalduinen, J. Yang, M. Mufti, and M. Huang, "Shoulder adhesive capsulitis in cancer patients undergoing positron emission tomography-computed tomography and the association with shoulder pain," *World Journal of Radiology*, vol. 13, no. 10, p. 344, 2021.

[99] S. Yang, D. H. Park, S. H. Ahn, J. Kim, J. W. Lee, J. Y. Han, D. K. Kim, J. Y. Jeon, K. H. Choi, and W. Kim, "Prevalence and risk factors of adhesive capsulitis of the shoulder after breast cancer treatment," *Supportive Care in Cancer*, vol. 25, pp. 1317–1322, 2017.

[100] D. H. Kim, D. H. Sung, H. Y. Ga, and J. Y. Choi, "Metabolic patterns of the shoulder joint on 18 f-fluorodeoxyglucose positron emission tomography/computed tomography in adhesive capsulitis," *Annals of nuclear medicine*, vol. 28, pp. 136–144, 2014.

[101] U. Salem, L. Zhang, J. L. Jorgensen, R. Kumar, and B. Amini, "Adhesive capsulitis mimicking metastasis on 18f-fdg-pet/ct," *Clinical nuclear medicine*, vol. 40, no. 2, pp. e145–e147, 2015.

[102] B. Triveni, P. Bhargavi, and S. Jyothi, "Deep feature extraction of mri image–a reliable tool for shoulder pain analysis," *INFORMATION TECHNOLOGY IN INDUSTRY*, vol. 9, no. 3, pp. 1016–1030, 2021.

[103] I. Scott, D. Connell, D. Moulton, S. Waters, A. Namburete, A. Arnab, and P. Malliaras, "An automated method for tendon image segmentation on ultrasound using grey-level co-occurrence matrix features and hidden gaussian markov random fields," *Computers in Biology and Medicine*, vol. 169, p. 107872, 2024.

[104] I. H. Witten, E. Frank, M. A. Hall, C. Pal, and M. DATA, "Practical machine learning tools and techniques," in *DATA MINING*, vol. 2, 2005, p. 4.

[105] A. Santone, M. P. Belfiore, F. Mercaldo, G. Varriano, and L. Brunese, "On the adoption of radiomics and formal methods for covid-19 coronavirus diagnosis," *Diagnostics*, vol. 11, no. 2, p. 293, 2021.

[106] A. Bonyár, "Application of localization factor for the detection of tin oxidation with afm," in *2015 IEEE 21st International Symposium for Design and Technology in Electronic Packaging (SIITME)*. IEEE, 2015, pp. 25–30.

[107] A. Stefano, M. Gioè, G. Russo, S. Palmucci, S. E. Torrisi, S. Bignardi, A. Basile, A. Comelli, V. Benfante, G. Sambataro, *et al.*, "Performance of radiomics features in the quantification of idiopathic pulmonary fibrosis from hrct," *Diagnostics*, vol. 10, no. 5, p. 306, 2020.

[108] M. Mascalchi, G. Camiciottoli, and S. Diciotti, "Lung densitometry: why, how and when," *Journal of thoracic disease*, vol. 9, no. 9, p. 3319, 2017.

[109] I. Fornacon-Wood, H. Mistry, C. J. Ackermann, F. Blackhall, A. McPartlin, C. Faivre-Finn, G. J. Price, and J. P. O'Connor, "Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform," *European radiology*, vol. 30, pp. 6241–6250, 2020.

[110] P. Brown, J. Zhong, R. Frood, S. Currie, A. Gilbert, A. Appelt, D. Sebag-Montefiore, and A. Scarsbrook, "Prediction of outcome in anal squamous cell carcinoma using radiomic feature analysis of pre-treatment fdg pet-ct," *European journal of nuclear medicine and molecular imaging*, vol. 46, no. 13, pp. 2790–2799, 2019.

[111] L. Moura, C. Dartora, C. Mattjie, R. Barros, and A. M. M. da Silva, "Texture-based feature extraction for covid-19 pneumonia classification using chest radiography," *EAI Endorsed Transactions on Bioengineering and Bioinformatics*, 2021.

[112] Q. Wu, S. Wang, L. Li, Q. Wu, W. Qian, Y. Hu, L. Li, X. Zhou, H. Ma, H. Li, *et al.*, "Radiomics analysis of computed tomography helps predict poor prognostic outcome in covid-19," *Theranostics*, vol. 10, no. 16, p. 7231, 2020.

[113] W. Schmitt and E. Marchiori, "Covid-19: round and oval areas of ground-glass opacity," *Pulmonology*, vol. 26, no. 4, p. 246, 2020.

[114] J. E. Van Timmeren, D. Cester, S. Tanadini-Lang, H. Alkadhi, and B. Baessler, "Radiomics in medical imaging—"how-to" guide and critical reflection," *Insights into imaging*, vol. 11, no. 1, pp. 1–16, 2020.

[115] K. Tingelhoff, K. W. Eichhorn, I. Wagner, M. E. Kunkel, A. I. Moral, M. E. Rilk, F. M. Wahl, and F. Bootz, "Analysis of manual segmentation in paranasal

ct images," *European archives of oto-rhino-laryngology*, vol. 265, pp. 1061–1070, 2008.

[116] S. Afat, A. E. Othman, K. Nikolaou, and S. Gassenmaier, "Dual-energy computed tomography of the lung in covid-19 patients: Mismatch of perfusion defects and pulmonary opacities," *Diagnostics*, vol. 10, no. 11, 2020.

[117] M.-Y. Ng, E. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M. M.-s. Lui, C. S.-Y. Lo, B. Leung, P.-L. Khong, *et al.*, "Imaging profile of the covid-19 infection: radiologic findings and literature review." *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, 2020.

[118] M. Loey, F. Smarandache, and N. E. M. Khalifa, "Within the lack of chest covid-19 x-ray dataset: A novel detection model based on gan and deep transfer learning," *Symmetry*, vol. 12, no. 4, 2020.

[119] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus covid-19 detection from x-rays," *Computer Methods and Programs in Biomedicine*, 2020.

[120] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[121] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, and B. Xu, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/24/2020.02.14.20023028

[122] S. Ying, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, R. Wang, Y. Chong, J. Shen, Y. Zha, and Y. Yang, "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/02/25/2020.02.23.20026930

[123] L. Büttner, A. Aigner, F. N. Fleckenstein, C. M. Hamper, M. Jonczyk, B. Hamm, O. Scholz, and G. Böning, "Diagnostic value of initial chest ct findings for the need of icu treatment/intubation in patients with covid-19," *Diagnostics*, vol. 10, no. 11, 2020.

[124] S. Baeza, D. Gil, I. Garcia-Olivé, M. Salcedo-Pujantell, J. Deportós, C. Sanchez, G. Torres, G. Moragas, and A. Rosell, "A novel intelligent radiomic analysis of perfusion spect/ct images to optimize pulmonary embolism diagnosis in covid-19 patients," *EJNMMI physics*, vol. 9, no. 1, p. 84, 2022.

[125] H. J. Kim, M. S. Brown, D. Chong, D. W. Gjertson, P. Lu, H. J. Kim, H. Coy, and J. G. Goldin, "Comparison of the quantitative ct imaging biomarkers of idiopathic pulmonary fibrosis at baseline and early change with an interval of 7 months," *Academic radiology*, vol. 22, no. 1, pp. 70–80, 2015.

[126] Y. Li, X. Liu, K. Xu, Z. Qian, K. Wang, X. Fan, S. Li, Y. Wang, and T. Jiang, "Mri features can predict egfr expression in lower grade gliomas: a voxel-based radiomic analysis," *European radiology*, vol. 28, no. 1, pp. 356–362, 2018.

[127] D. Kim, L. J. Jensen, T. Elgeti, I. G. Steffen, B. Hamm, and S. N. Nagel, "Radiomics for everyone: A new tool simplifies creating parametric maps for the visualization and quantification of radiomics features," *Tomography*, vol. 7, no. 3, pp. 477–487, 2021.

[128] F. Sardanelli and G. Di Leo, "Reproducibility: intraobserver and interobserver variability," *Biostatistics for radiologists: planning, performing, and writing a radiologic study*, pp. 125–140, 2009.

[129] S. Roy, T. D. Whitehead, J. D. Quirk, A. Salter, F. O. Ademuyiwa, S. Li, H. An, and K. I. Shoghi, "Optimal co-clinical radiomics: Sensitivity of radiomic features to tumour volume, image noise and resolution in co-clinical t1-weighted and t2-weighted magnetic resonance imaging," *EBioMedicine*, vol. 59, p. 102963, 2020.

[130] M. Donelli, G. Espa, and P. Feraco, "A semi-unsupervised segmentation methodology based on texture recognition for radiomics: A preliminary study on brain tumours," *Electronics*, vol. 11, no. 10, p. 1573, 2022.

[131] S. G. Sapate, A. Mahajan, S. N. Talbar, N. Sable, S. Desai, and M. Thakur, "Radiomics based detection and characterization of suspicious lesions on full field digital mammograms," *Computer methods and programs in biomedicine*, vol. 163, pp. 1–20, 2018.

[132] A. I. Valderrama-Treviño, B. Barrera-Mera, J. C. Ceballos-Villalva, and E. E. Montalvo-Javé, "Hepatic metastasis from colorectal cancer," *Euroasian journal of hepato-gastroenterology*, vol. 7, no. 2, p. 166, 2017.

[133] F. A. Haggar and R. P. Boushey, "Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors," *Clinics in colon and rectal surgery*, vol. 22, no. 04, pp. 191–197, 2009.

[134] G. Torzilli, M. Montorsi, M. Donadon, A. Palmisano, D. Del Fabbro, A. Gambetti, N. Olivari, and M. Makuuchi, ""radical but conservative" is the main goal for ultrasonography-guided liver resection: prospective validation of this approach," *Journal of the American College of surgeons*, vol. 201, no. 4, pp. 517–528, 2005.

[135] S.-X. Rao, D. M. Lambregts, R. S. Schnerr, W. van Ommen, T. J. van Nijnatten, M. H. Martens, L. A. Heijnen, W. H. Backes, C. Verhoef, M.-S. Zeng, *et al.*, "Whole-liver ct texture analysis in colorectal cancer: does the presence of liver metastases affect the texture of the remaining liver?" *United European gastroenterology journal*, vol. 2, no. 6, pp. 530–538, 2014.

[136] S. J. Lee, R. Zea, D. H. Kim, M. G. Lubner, D. A. Deming, and P. J. Pickhardt, "Ct texture features of liver parenchyma for predicting development of metastatic disease and overall survival in patients with colorectal cancer," *European radiology*, vol. 28, pp. 1520–1528, 2018.

[137] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Artificial intelligence in radiology," *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[138] V. Granata, R. Fusco, M. L. Barretta, C. Picone, A. Avallone, A. Belli, R. Patrone, M. Ferrante, D. Cozzi, R. Grassi, *et al.*, "Radiomics in hepatic metastasis by colorectal cancer," *Infectious Agents and Cancer*, vol. 16, no. 1, pp. 1–9, 2021.

[139] N. Wesdorp, V. van Goor, R. Kemna, E. Jansma, J. van Waesberghe, R. Swijnenburg, C. Punt, J. Huiskens, and G. Kazemier, "Advanced image analytics predicting clinical outcomes in patients with colorectal liver metastases: a systematic review of the literature," *Surgical Oncology*, vol. 38, p. 101578, 2021.

[140] M. Taghavi, S. Trebeschi, R. Simões, D. B. Meek, R. C. Beckers, D. M. Lambregts, C. Verhoef, J. B. Houwers, U. A. van der Heide, R. G. Beets-Tan, *et al.*, "Machine learning-based analysis of ct radiomics model for prediction of colorectal metachronous liver metastases," *Abdominal Radiology*, vol. 46, no. 1, pp. 249–256, 2021.

[141] A. Saifuddin, V. Andrei, R. Rajakulasingam, I. Oliveira, and B. Seddon, "Magnetic resonance imaging of trunk and extremity myxoid liposarcoma: diagnosis, staging, and response to treatment," *Skeletal Radiology*, vol. 50, pp. 1963–1980, 2021.

[142] A. T. J. Lee, K. Thway, P. H. Huang, and R. L. Jones, "Clinical and molecular spectrum of liposarcoma," *Journal of Clinical Oncology*, vol. 36, no. 2, p. 151, 2018.

[143] L. Zhang, D. V. Fried, X. J. Fave, L. A. Hunter, J. Yang, and L. E. Court, "Ibex: an open infrastructure software platform to facilitate collaborative work in radiomics," *Medical physics*, vol. 42, no. 3, pp. 1341–1353, 2015.

[144] A. F. Mavrogenis, J. Lesensky, C. Romagnoli, M. Alberghini, G. D. Letson, and P. Ruggieri, "Atypical lipomatous tumors/well-differentiated liposarcomas: clinical outcome of 67 patients," *Orthopedics*, vol. 34, no. 12, pp. e893–e898, 2011.

[145] M. Vos, M. Starmans, M. Timbergen, S. van der Voort, G. Padmos, W. Kessels, W. Niessen, G. Van Leenders, D. Grünhagen, S. Sleijfer, *et al.*, "Radiomics approach to distinguish between well differentiated liposarcomas and lipomas on mri," *Journal of British Surgery*, vol. 106, no. 13, pp. 1800–1809, 2019.

[146] I. Malinauskaite, J. Hofmeister, S. Burgermeister, A. Neroladaki, M. Hamard, X. Montet, S. Boudabbous, *et al.*, "Radiomics and machine learning differentiate soft-tissue lipoma and liposarcoma better than musculoskeletal radiologists," *Sarcoma*, vol. 2020, 2020.

[147] B. Leporq, A. Bouhamama, F. Pilleul, F. Lame, C. Bihane, M. Sdika, J.-Y. Blay, and O. Beuf, "Mri-based radiomics to predict lipomatous soft tissue tumors malignancy: a pilot study," *Cancer Imaging*, vol. 20, no. 1, pp. 1–8, 2020.

[148] Y. Tang, J. Cui, J. Zhu, and G. Fan, "Differentiation between lipomas and atypical lipomatous tumors of the extremities using radiomics," *Journal of Magnetic Resonance Imaging*, vol. 56, no. 6, pp. 1746–1754, 2022.

[149] M. Kransdorf, "Malignant soft-tissue tumors in a large referral population: distribution of diagnoses by age, sex, and location." *AJR. American journal of roentgenology*, vol. 164, no. 1, pp. 129–134, 1995.

[150] K. G. Billingsley, J. J. Lewis, D. H. Leung, E. S. Casper, J. M. Woodruff, and M. F. Brennan, "Multifactorial analysis of the survival of patients with distant metastasis arising from primary extremity sarcoma," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 85, no. 2, pp. 389–395, 1999.

[151] M. Brennan, "Soft tissue sarcoma: Advances in understanding and management," *The Surgeon*, vol. 3, no. 3, pp. 216–223, 2005.

[152] D. L. Longo *et al.*, "Tumor heterogeneity and personalized medicine," *N Engl J Med*, vol. 366, no. 10, pp. 956–7, 2012.

[153] S. Gitto, R. Cuocolo, D. Albano, F. Morelli, L. C. Pescatori, C. Messina, M. Imbriaco, and L. M. Sconfienza, "Ct and mri radiomics of bone and soft-tissue sarcomas: a systematic review of reproducibility and validation strategies," *Insights into Imaging*, vol. 12, no. 1, pp. 1–14, 2021.

[154] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, *et al.*, "Radiomics: the process and the challenges," *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1234–1248, 2012.

[155] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, p. 5471, 2015.

[156] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[157] F. Davnall, C. S. Yip, G. Ljungqvist, M. Selmi, F. Ng, B. Sanghera, B. Ganeshan, K. A. Miles, G. J. Cook, and V. Goh, "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?" *Insights into imaging*, vol. 3, no. 6, pp. 573–589, 2012.

[158] Z. Yang, L. H. Tang, and D. S. Klimstra, "Effect of tumor heterogeneity on the assessment of ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification," *The American journal of surgical pathology*, vol. 35, no. 6, pp. 853–860, 2011.

[159] A. Crombé, P.-J. Marcellin, X. Buy, E. Stoeckle, V. Brouste, A. Italiano, F. Le Loarer, and M. Kind, "Soft-tissue sarcomas: assessment of mri features correlating with histologic grade and patient outcome," *Radiology*, vol. 291, no. 3, pp. 710–721, 2019.

[160] J. C. Peeken, M. B. Spraker, C. Knebel, H. Dapper, D. Pfeiffer, M. Devecka, A. Thamer, M. A. Shouman, A. Ott, R. von Eisenhart-Rothe, *et al.*, "Tumor grading of soft tissue sarcomas using mri-based radiomics," *EBioMedicine*, vol. 48, pp. 332–340, 2019.

[161] A. Crombé, M. Kind, D. Fadli, F. Le Loarer, A. Italiano, X. Buy, and O. Saut, "Intensity harmonization techniques influence radiomics features and radiomics-based predictions in sarcoma patients," *Scientific reports*, vol. 10, no. 1, pp. 1–13, 2020.

[162] L. Tian, D. Zhang, S. Bao, P. Nie, D. Hao, Y. Liu, J. Zhang, and H. Wang, "Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas," *Clinical Radiology*, vol. 76, no. 2, pp. 158–e19, 2021.

[163] C. D. Fletcher, K. Unni, and F. Mertens, *World Health Organization classification of tumours. Pathology and genetics of tumours of soft tissue and bone.* IARC press, 2002.

[164] R. Forghani, P. Savadjiev, A. Chatterjee, N. Muthukrishnan, C. Reinhold, and B. Forghani, "Radiomics and artificial intelligence for biomarker and prediction model development in oncology," *Computational and structural biotechnology journal*, vol. 17, p. 995, 2019.

[165] K. Kakhi, R. Alizadehsani, H. D. Kabir, A. Khosravi, S. Nahavandi, and U. R. Acharya, "The internet of medical things and artificial intelligence: trends, challenges, and opportunities," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 3, pp. 749–771, 2022.

[166] M. Taddeo, T. McCutcheon, and L. Floridi, "Trusting artificial intelligence in cybersecurity is a double-edged sword," *Nature Machine Intelligence*, vol. 1, no. 12, pp. 557–560, 2019.

[167] W. Guidance, "Ethics and governance of artificial intelligence for health," *World Health Organization*, 2021.

[168] D. J. Hunter and C. Holmes, "Where medical statistics meets artificial intelligence," *New England Journal of Medicine*, vol. 389, no. 13, pp. 1211–1219, 2023.

[169] P. F. Edemekong, P. Annamaraju, and M. J. Haydel, "Health insurance portability and accountability act," *StatPearls*, 2018.

[170] G. D. P. Regulation, "General data protection regulation (gdpr)," *Intersoft Consulting, Accessed in October*, vol. 24, no. 1, 2018.

[171] G. Stockman and L. G. Shapiro, *Computer vision*. Prentice Hall PTR, 2001.

[172] J. Gao, Y. Yang, P. Lin, D. S. Park, *et al.*, "Computer vision in healthcare applications," 2018.

[173] M. Gudmundsson, E. A. El-Kwae, and M. R. Kabuka, "Edge detection in medical images using a genetic algorithm," *IEEE transactions on medical imaging*, vol. 17, no. 3, pp. 469–474, 1998.

[174] D. T. Huff, A. J. Weisman, and R. Jeraj, "Interpretation and visualization techniques for deep learning models in medical imaging," *Physics in Medicine & Biology*, vol. 66, no. 4, p. 04TR01, 2021.

Appendix A

Supplementary Material

A.1 Epilepsy: results without filter

FIRST Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	**0.57**	**1.0**	0.33	**1.0**
rule3-9	0.50	**1.0**	0.22	**1.0**
rule5-25	**0.57**	0.71	**0.56**	0.60
rule7-16	0.50	0.75	0.33	0.80
rule19-25	0.50	**1.0**	0.22	**1.0**
rule3-11	0.43	0.60	0.33	0.60
rule5-6	0.50	**1.0**	0.22	**1.0**
rule9-16	0.50	**1.0**	0.22	**1.0**

Table A.1: Summary table of metrics for the FIRST feature class.

GLCM Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	0.57	0.80	0.44	0.80
rule3-9	0.57	**1.0**	0.33	**1.0**
rule5-25	0.71	0.86	0.67	0.80
rule7-16	**0.86**	0.82	**1.0**	0.60
rule19-25	**0.86**	0.82	**1.0**	0.60
rule3-11	0.79	0.88	0.78	0.80
rule5-6	0.79	**1.0**	0.67	**1.0**
rule9-16	0.57	**1.0**	0.33	**1.0**

Table A.2: Summary table of metrics for the GLCM feature class.

GLDM Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	0.50	**1.0**	0.22	**1.0**
rule3-9	**0.71**	0.73	**0.89**	0.40
rule5-25	**0.71**	0.73	0.89	0.40
rule7-16	0.50	0.67	0.44	0.60
rule19-25	0.50	0.75	0.33	0.80
rule3-11	0.50	0.75	0.33	0.80
rule5-6	0.50	**1.0**	0.22	**1.0**
rule9-16	0.50	**1.0**	0.22	**1.0**

Table A.3: Summary table of metrics for the GLDM feature class.

GLRLM Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	0.57	0.67	0.67	0.40
rule3-9	0.64	0.67	0.89	0.20
rule5-25	**0.71**	0.70	**1.0**	0.20
rule7-16	0.43	0.57	0.44	0.40
rule19-25	0.43	0.67	0.22	0.80
rule3-11	0.43	0.67	0.22	0.80
rule5-6	0.50	**1.0**	0.22	**1.0**
rule9-16	0.50	**1.0**	0.22	**1.0**

Table A.4: Summary table of metrics for the GLRLM feature class.

GLSZM Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	0.57	0.67	**0.67**	0.40
rule3-9	0.36	0.50	0.22	0.60
rule5-25	**0.71**	0.86	0.67	0.80
rule7-16	0.64	0.75	**0.67**	0.60
rule19-25	**0.71**	0.86	**0.67**	0.80
rule3-11	0.57	0.71	0.56	0.60
rule5-6	0.57	**1.0**	0.33	**1.0**
rule9-16	0.64	0.83	0.56	0.80

Table A.5: Summary table of metrics for the GLSZM feature class.

NGTDM Class				
Property	*Accuracy*	*Precision*	*Recall*	*Specificity*
rule11-19	0.43	0.56	0.56	0.20
rule3-9	0.57	**1.0**	0.33	**1.0**
rule5-25	**0.64**	0.75	**0.67**	0.60
rule7-16	0.50	0.63	0.56	0.40
rule19-25	0.50	**1.0**	0.22	**1.0**
rule3-11	0.50	0.63	0.56	0.40
rule5-6	0.43	0.57	0.44	0.40
rule9-16	0.50	**1.0**	0.22	**1.0**

Table A.6: Summary table of metrics for the NGTDM feature class.

A.2 Epilepsy: results on filtered images

FIRST Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.50	1	0.22	1	
rule19-25 (filtered)	0.36	imp	0	1	**-0.34**
rule3-11	0.43	0.60	0.33	0.60	
rule3-11 (filtered)	0.43	1	0.11	1	**+0.15**
rule5-6	0.50	1	0.22	1	
rule5-6 (filtered)	0.36	imp	0	1	**-0.34**
rule9-16	0.50	1	0.22	1	
rule9-16 (filtered)	0.36	imp	0	1	**-0.34**
rule11-19	0.57	1	0.33	1	
rule11-19 (filtered)	0.43	0.67	0.22	0.80	**-0.2**
rule3-9	0.50	1	0.22	1	
rule3-9 (filtered)	0.43	1	0.11	1	**-0.04**
rule5-25	0.57	0.71	0.55	0.60	
rule5-25 (filtered)	0.50	1	0.22	1	**+0.07**
rule7-16	0.50	0.75	0.33	0.80	
rule7-16 (filtered)	0.36	imp	0	1	**-0.26**

Table A.7: Results obtained from the original medical images ceived in terms of Accuracy, Precision, Recall, Specificity and Δ for the FIRST class.

GLCM Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.86	0.81	1	0.60	
rule19-25 (filtered)	0.79	0.88	0.78	0.80	**-0.01**
rule3-11	0.79	0.88	0.78	0.80	
rule3-11 (filtered)	0.50	0.67	0.44	0.60	**-0.26**
rule5-6	0.79	1	0.67	1	
rule5-6 (filtered)	0.43	0.60	0.33	0.60	**-0.38**
rule9-16	0.57	1	0.33	1	
rule9-16 (filtered)	0.36	imp	0	1	**-0.39**
rule11-19	0.57	0.80	0.44	0.80	
rule11-19 (filtered)	0.36	imp	0	1	**-0.31**
rule3-9	0.57	1	0.33	1	
rule3-9 (filtered)	0.43	1	0.11	1	**-0.09**
rule5-25	0.71	0.86	0.67	0.80	
rule5-25 (filtered)	0.57	1	0.33	1	**-0.03**
rule7-16	0.86	0.82	1	0.60	
rule7-16 (filtered)	0.64	0.83	0.56	0.80	**-0.11**

Table A.8: Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLCM class.

GLDM Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.50	0.75	0.33	0.80	
rule19-25 (filtered)	0.36	imp	0	1	**-0.26**
rule3-11	0.50	0.75	0.33	0.80	
rule3-11 (filtered)	0.43	1	0.11	1	**+0.04**
rule5-6	0.50	1	0.22	1	
rule5-6 (filtered)	0.36	imp	0	1	**-0.34**
rule9-16	0.50	1	0.22	1	
rule9-16 (filtered)	0.36	imp	0	1	**-0.34**
rule11-19	0.50	1	0.22	1	
rule11-19 (filtered)	0.50	1	0.22	1	**0**
rule3-9	0.71	0.73	0.89	0.40	
rule3-9 (filtered)	0.50	0.75	0.33	0.80	**-0.08**
rule5-25	0.71	0.73	0.89	0.40	
rule5-25 (filtered)	0.64	0.75	0.67	0.60	**-0.01**
rule7-16	0.50	0.67	0.44	0.60	
rule7-16 (filtered)	0.43	0.60	0.33	0.60	**-0.06**

Table A.9: Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLDM class.

GLRLM Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.43	0.67	0.22	0.80	
rule19-25 (filtered)	0.36	imp	0	1	**-0.19**
rule3-11	0.43	0.67	0.22	0.80	
rule3-11 (filtered)	0.36	imp	0	1	**-0.19**
rule5-6	0.50	1	0.22	1	
rule5-6 (filtered)	0.36	imp	0	1	**-0.34**
rule9-16	0.50	1	0.22	1	
rule9-16 (filtered)	0.36	imp	0	1	**-0.34**
rule11-19	0.57	0.67	0.67	0.40	
rule11-19 (filtered)	0.50	1	0.22	1	**+0.10**
rule3-9	0.64	0.67	0.89	0.20	
rule3-9 (filtered)	0.50	0.67	0.44	0.60	**-0.05**
rule5-25	0.71	0.69	1	0.20	
rule5-25 (filtered)	0.64	0.83	0.56	0.80	**+0.06**
rule7-16	0.43	0.57	0.44	0.40	
rule7-16 (filtered)	0.43	1	0.11	1	**+0.18**

Table A.10: Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLRLM class.

GLSZM Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.71	0.86	0.67	0.80	
rule19-25 (filtered)	0.57	0.80	0.44	0.80	**-0.11**
rule3-11	0.57	0.71	0.56	0.60	
rule3-11 (filtered)	0.57	0.80	0.44	0.80	**+0.04**
rule5-6	0.57	1	0.33	1	
rule5-6 (filtered)	0.43	1	0.11	1	**-0.09**
rule9-16	0.64	0.83	0.56	0.80	
rule9-16 (filtered)	0.43	0.67	0.22	0.80	**-0.18**
rule11-19	0.57	0.67	0.67	0.40	
rule11-19 (filtered)	0.43	0.57	0.44	0.40	**-0.12**
rule3-9	0.36	0.50	0.22	0.60	
rule3-9 (filtered)	0.43	1	0.11	1	**+0.22**
rule5-25	0.71	0.86	0.67	0.80	
rule5-25 (filtered)	0.21	0	0	0.60	**-0.56**
rule7-16	0.64	0.75	0.67	0.60	
rule7-16 (filtered)	0.50	0.67	0.44	0.60	**-0.12**

Table A.11: Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLSZM class.

NGTDM Class					
Rule	*Accuracy*	*Precision*	*Recall*	*Specificity*	Δ
rule19-25	0.50	1	0.22	1	
rule19-25 (filtered)	0.36	imp	0	1	**-0.34**
rule3-11	0.50	0.63	0.56	0.40	
rule3-11 (filtered)	0.50	0.75	0.33	0.80	**+0.08**
rule5-6	0.43	0.58	0.44	0.40	
rule5-6 (filtered)	0.79	1	0.67	1	**+0.41**
rule9-16	0.50	1	0.22	1	
rule9-16 (filtered)	0.36	imp	0	1	**-0.34**
rule11-19	0.43	0.56	0.56	0.20	
rule11-19 (filtered)	0.50	0.63	0.56	0.40	**+0.08**
rule3-9	0.57	1	0.33	1	
rule3-9 (filtered)	0.36	imp	0	1	**-0.39**
rule5-25	0.64	0.75	0.67	0.60	
rule5-25 (filtered)	0.64	1	0.44	1	**+0.10**
rule7-16	0.50	0.63	0.56	0.40	
rule7-16 (filtered)	0.71	0.78	0.78	0.60	**+0.20**

Table A.12: Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the NGTDM class.

List of Figures

1.1 Structure of this thesis, where each Chapter is associated with an anatomical part of the human body to which the methodology has been applied. 15

2.1 Graphical representation of a system behaviour. Circles are called *states* and are reachable only through the execution of *actions* (alphabet letters). $P1$ is the initial state, while *nil* is the state of termination. 24

3.1 Two different type of segmentation. On the left an automatic segmentation based on threshold; on the right, a manual segmentation on the liver using 3DSlicer tool. 30
3.2 Methodology presented in this thesis, including collaboration between doctors and computer scientists, with image analysis through to modelling and results that become explainable thanks to the counterexample. 31
3.3 Display of the distribution of the feature Median of the FIRST class. In blue there is a category of patients, in red the other one; in the top right-hand corner there is also an indication about the mean of each curve. 32
3.4 Different types of discretization are shown: on the left, an equidistributed method with 3 levels; on the right, a non-equidistributed method with 5 levels. 33
3.5 A different type of formal model using a single image: each process is a cell of a grid overlapped on the image. The syntax of this model can be seen in Table 3.2. 36
3.6 Taking a LTS and a formula to satisfy, the output is “False” and the initial state of the LTS is highlighted in red because from that state the property is no longer satisfied from the model. 38

3.7 These structured images on the left all have the same FIRST-Order statistics values, well represented on the right by a histogram with two columns, one for black pixels and one for white pixels. This similarity is only apparent, because we can visualize the three images are different in terms of pattern combination and then, of meaning. 41

4.1 On the left side, a view of a pre-contrast T1 weighted MRI. On the right side, the T1w post-contrast MRI of the same patient, which showed radiological activity (highlighted by two red circles). 45

4.2 Actions described in the property to find in the models, which lead the classifier to make a decision about the state of health of the patients. On the right side there are some annotation on how many times the action is repeated in a row. 47

4.3 An example of a patient who is MRI-negative but PET-positive (on the left) and another patient who is MRI-negative and PET-negative (on the right). The red circle highlights the area of hypometabolism, which is clearly visible on the PET images of the PET-positive patient (on the left) but absent in the PET-negative patient (on the right). . . 48

4.4 The proposed model involves two distinct steps. The first step is performed without any image pre-processing, while the second step includes image pre-processing. In both steps, MRI series of the patient are modelled for both the original images and the filtered ones. Utilizing a Formal Property and the Model Checker agent, let obtain the results of the automatic diagnosis. 50

4.5 Three different medical images of a patient with epilepsy. 51

4.6 Comparison between the original MRI image (on the left) and the filtered MRI image (on the right). 52

6.1 Coronal MRI image weighed in T2 saturated with fat. Image *b* shows thickening (up to 10mm) and hyperintensity of the signal of the glenhumeral ligament complex of the axillary sac with hyperintensity/edema of adjacent soft tissues [97]. 70

7.1 Three different medical images of three different patients: a healthy one, a COVID-19 one and a lung disease patient. 76

7.2 Diagram of the manual and automatic activities for the earlier identification of COVID-19 or Lung disease patients. The initial process outlines the conventional procedure that a patient must follow to undergo an HRCT exam and receive a diagnosis. In contrast, the second process utilizes the insights obtained from the first process (comprehensive dataset and the expert knowledge of the radiologist) in an attempt to replicate the accurate diagnosis. 78

7.3 Starting from the CT exam, the 30 middle slices are selected, from which the radiomic features are extracted. Through temporal logic, models are created and checked. For the "Full Exam" study, the step 2 will not be performed. 80

7.4 This is the visualisation of a tree decision algorithm. In this case of Selection with Weka, the feature selected are those closest to the root: Minimum, Entropy, Median, Energy. 81

7.5 The assumption is that different health states can result in different LTS. 84

7.6 These histograms show the total distribution of Energy values for COVID-19 patients (on the left side) and for healthy patients (on the right side). The curve of healthy patients is more similar to a Gaussian curve than the curve on the left; in addition, the values of healthy patients are higher than ones of COVID-19 patients. 87

7.7 Value distribution measured with skewness and kurtosis [106]. 87

7.8 These histograms show the total distribution of Kurtosis values for COVID-19 patients (on the left side) and for healthy patients (on the right side). The differences about the form of the curve are greater; then, healthy values are lower than the others. 88

7.9 Skewness values for COVID-19 patients and for healthy patients. The distribution of healthy patients is on average less than the one of sick patients. 88

7.10 Variance values for COVID-19 patients and for healthy patients: healthy values are higher and better distributed around the mean, as a normal distribution. 89

7.11 Minimum intensity values for COVID-19 patients; the graph for healthy patients is absent because the values are fixed on -1023,00 with a standard deviation of 0,00. This is the major difference between the two categories. 89

7.12 Example of logical "formula" that contains discretized values for two features, Variance and Energy. 92
7.13 Example of pattern recognition on a formula used in classification of COVID-19 patients. In this case, the formula contains a unique pattern: the repetition of *low variance* and *low energy* for 9 times in a row. 92
7.14 The pattern recognition searching returns how many times in a row different combination of features are found. This is an example of pattern recognition on a formula used in classification of healthy patients. Healthy patients are characterised by different patterns, which the most recurrent is the $mediumVariance$ together to $lowEnergy$. 93
7.15 Schema of the developed method. 97
7.16 ROI matrix represented on the axes of the system considered by the DICOM type images. 98
7.17 The ROI matrix can be superimposed on the DICOM of the patient to facilitate the localization of disease markers within the radiological images. 101

9.1 Inter-correlation among selected features calculated with Spearman correlation coefficient. 124

List of Tables

2.1 3x3 image 22
2.2 GLCM Matrix 22
2.3 Normalized GLCM 22
2.4 Example of how GLCM Matrix with distance $\delta = 2$ and $\sigma = 0°$ is built. 22
2.5 List of the radiomic features of the different classes available in PyRadiomics [21]. 23
2.6 Subset of CCS operators exploited in the following chapters. 25

3.1 Creation of a Formal Model of a patient exam using the GLCM class and the radiomic features of Inverse Variance, Cluster Shade and Correlation. 35
3.2 Formal Model for single-image methodology (such as US scan), using the GLCM class and the radiomic features of Inverse Variance, Cluster Shade and Correlation. 37

4.1 Parameters used for the axial 2D T1-weighted scan acquired before and after Gd contrast agent administration. 44
4.2 List of the selected features for models construction. 46
4.3 Accuracy and other indexes to gain more information about the clinical usefulness of the methodology. 47
4.4 List of MRI parameters of the 3d T1w SPGR MRI for patients included in the dataset. 49
4.5 These are the features selected from the feature classes of the IBSI standard: these features are used for the analysis in the Radiomic workflow. 53
4.6 Overview table of the best metrics for the feature classes under study. 55

4.7 A comparison of the results obtained from the original and filtered medical images is presented below, indicating the chosen property based on the highest metrics for the original images. 57

5.1 Accuracy and other indexes to gain more information about the clinical usefulness of the methodology. 66

6.1 Most relevant features for FIRST, GLCM and GLDM classes. 71
6.2 Most relevant features for GLRLM, GLSZM and SHAPE classes. . . . 71
6.3 Summary of participant performances in comparison with the methodology. 72

7.1 Radiomic feature set for the first tier (or step 1). 77
7.2 Radiomic feature set for the second tier (or step 2). 77
7.3 Average values for healthy, COVID and lung disease patients. For healthy patients, there are only two parameters because these are the most discriminant for distinction by sick patients. In fact, these five features belong to the second-tier classification (see Tables 7.1 and 7.2). . . . 79
7.4 Step 1: Results of the comparison between healthy and lung disease patients. 79
7.5 Step 2: Results of the comparison between general lung disease and COVID-19 patients. 79
7.6 Results for step 1, in which it is possible to distinguish between healthy and sick patients (Covid-19 and Lung disease patients). 82
7.7 Results for step 2, in which I distinguish between COVID and lung disease patients. 82
7.8 List of features used for the property generation and their performance metrics. In the second part, the feature Minimum was removed. 91
7.9 On the upper bound there is an example of a "explainable" formula with indications on the localization of the most important slices to be reviewed by the specialists. On the lower bound, on the left there is the $246th$ slice, relative to the first value of the property. On the right there is the slice relative to the second property, namely the slice number 251. 94
7.10 List of radiomic features taken into account for the final diagnosis. . . 96
7.11 Classification performances with the segmentation on the whole DICOM image (above) and with the 3x3 regions matrix put on the DICOM images (below). 100

7.12 Localization performances with the 3x3 regions matrix put on the DICOM images. 100
7.13 State-of-the-art comparison for COVID-19 classification with AI. . . . 103

8.1 List of features selected as "significant" in this specific case study of the liver. 111
8.2 The Confusion Matrix helps to understand how many and which patients are misclassified. 112
8.3 Accuracy and utility statistics according to Mitchell statistics [32] to gain more information about the clinical usefulness of the methodology. 112

9.1 Performances comparison of Radiologists and Radiomics. 119
9.2 Some clinical data for Group A and Group B. 121
9.3 Hyper-parameters settings for the feature extraction with PyRadiomics: other parameters were set to default. For each exam, features were extracted from each image separately. 122
9.4 Formal Property to verify on the patient's model the presence of the disease. 123
9.5 Accuracy and utility indexes according to Mitchell statistics [32] to gain more information about the clinical usefulness of the methodology. 125

A.1 Summary table of metrics for the FIRST feature class. 157
A.2 Summary table of metrics for the GLCM feature class. 158
A.3 Summary table of metrics for the GLDM feature class. 158
A.4 Summary table of metrics for the GLRLM feature class. 159
A.5 Summary table of metrics for the GLSZM feature class. 159
A.6 Summary table of metrics for the NGTDM feature class. 160
A.7 Results obtained from the original medical images ceived in terms of Accuracy, Precision, Recall, Specificity and Δ for the FIRST class. . . 161
A.8 Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLCM class. 162
A.9 Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLDM class. 163

A.10 Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLRLM class. 164
A.11 Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the GLSZM class. 165
A.12 Results obtained from original and pre-processed medical images in terms of Accuracy, Precision, Recall, Specificity and Δ for the NGTDM class. 166