FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Chest Radiography Content-Based Image Retrieval

Maria Francisca Fontes da Silva

Mestrado em Engenharia Eletrotécnica e de Computadores

Supervisor: João Manuel Patrício Pedrosa

Second Supervisor: Ana Maria Rodrigues de Sousa Faria de Mendonça

July 31, 2023

Chest Radiography Content-Based Image Retrieval

Maria Francisca Fontes da Silva

Mestrado em Engenharia Eletrotécnica e de Computadores

Faculdade de Engenharia da Universidade do Porto

July 31, 2023

Resumo

A radiografia torácica desempenha um papel vital no diagnóstico médico e monitorização da evolução de várias situações patológicas. Quando confrontados com casos complexos, os radiologistas frequentemente recorrem à comparação com imagens anteriores, tornando a pesquisa e análise de bases de dados de imagens médicas demoradas e desafiadoras. Para abordar esta questão, os sistemas de Recuperação de Imagens Baseada em Conteúdo (CBIR) oferecem uma solução eficiente, encontrando automaticamente casos semelhantes em grandes bases de dados, para auxiliar os radiologistas no processo de tomada de decisão. No domínio médico, os sistemas CBIR tradicionais podem não ser os mais recomendados para lidar com grandes bases de dados e analisar imagens mais complexas. Alguns sistemas CBIR existentes aplicados a bases de dados de imagens de raios-X torácico (CXR) falharam em recuperar patologias semelhantes devido ao manuseio inadequado das características das imagens e processos de similaridade insuficientemente orientados. Esta dissertação apresenta uma análise abrangente de um sistema CBIR, chamado LXIR (Sistema de Recuperação de Imagens de CXR Baseado em Lesões), especificamente adaptado para lesões presentes em imagens de CXR. Os principais objetivos do trabalho desenvolvido foram avaliar a generalização, robustez e eficácia do sistema na recuperação de lesões em diversas bases de dados. A arquitetura YOLOv5x foi utilizada para deteção e extração de características de lesões. As bases de dados VinDr-CXR e ChestXray-14 foram utilizadas como coleções de busca e o conjunto de teste pertencente à base de dados VinDr-CXR foi utilizado como conjunto de consulta. Foram realizadas três experiências para avaliar o desempenho do sistema, testando a sua generalização e robustez, investigando a relação entre distância e similaridade visual à lesão pesquisada e avaliando a dissimilaridade entre diferentes classes e a dissimilaridade entre as subclasses dentro da classe de Lesões Parenquimatosas (PaL). Ao longo destas experiências, a importância da seleção do limiar para otimizar o desempenho da recuperação foi identificada. Além disso, enfatizou-se a necessidade de existirem mais dados anotados e métodos de anotação melhorados para aumentar a precisão na discriminação de subclasses. Assim, este estudo demonstra o potencial do sistema LXIR para aplicações médicas, enfatizando a importância de extração de características personalizadas e estratégias de recuperação. Os resultados contribuem com informações valiosas para otimizar sistemas CBIR e guiar pesquisas futuras para aprimorar o desempenho do sistema de recuperação no domínio médico.

Palavras-chave: Radiografia torácica, Sistemas de Recuperação de Imagens Baseados em Conteúdo, Sistema de Recuperação de Imagens de CXR Baseado em Lesões.

Abstract

Chest radiography plays a vital role in medical diagnosis and monitoring the progression of various conditions. When faced with complex cases, radiologists often rely on comparisons with previous images, making the search and analysis of medical imaging databases time-consuming and challenging. To address this, Content-Based Image Retrieval (CBIR) systems offer an efficient solution by automatically finding similar cases in large databases, assisting radiologists in their decision-making process. In the medical domain, traditional CBIR systems may not be optimal for handling large databases and fine-grained image analysis. Some existing CBIR systems applied to Chest X-ray (CXR) image databases have failed to retrieve similar pathologies due to improper handling of image features and insufficiently guided similarity processes. This study presents a comprehensive analysis of a CBIR system, the Lesion-based CXR Image Retrieval (LXIR) system, specifically customised to lesions in CXR images. The main objectives were to assess the system's generalisation, robustness, and effectiveness in retrieving lesions from diverse databases. YOLOv5x architecture was employed for lesion detection and feature extraction, with the VinDr-CXR and ChestXray-14 databases serving as search collections and the VinDr-CXR test set as the query/test set. Three experiments were conducted to evaluate the system's performance by testing generalisation and robustness, investigating the relationship between distance and visual similarity to the query, and assessing dissimilarity between different classes and the dissimilarity between subclasses within the Parenchymal Lesion (PaL) class. Throughout the experiments, the significance of threshold settings to optimise retrieval performance was identified. Additionally, the need for more annotated data and refined annotation methods to improve subclass discrimination accuracy was emphasised. Thus, this study demonstrates the potential of the LXIR system for medical applications, emphasising the importance of personalised feature extraction and retrieval strategies. The findings contribute valuable insights into optimising CBIR systems and guiding future research to enhance retrieval performance in the medical domain.

Keywords: Chest radiography, Content-Based Image Retrieval systems, Lesion-based CXR Image Retrieval.

Agradecimentos

Gostaria de expressar a minha sincera gratidão a todas as pessoas que contribuíram para o desenvolvimento desta dissertação. O seu apoio, orientação e encorajamento foram absolutamente essenciais para a concretização deste trabalho.

Em primeiro lugar, quero manifestar o meu profundo agradecimento ao meu orientador, o Professor Doutor João Pedrosa, e à minha coorientadora, a Professora Doutora Ana Maria Mendonça, pela qualidade da sua orientação e pelo empenho constantemente demonstrado. Além disso, sou grata pela oportunidade de integrar um grupo de investigação, que me permitiu crescer academicamente, e também pelas valiosas e numerosas contribuições que generosamente me ofereceram ao longo de todo este processo.

Em segundo lugar, estendo o meu apreço a todos os professores que fizeram parte da minha jornada académica, contribuindo para a expansão do meu conhecimento e aperfeiçoamento das minhas capacidades intelectuais.

De forma especial, gostaria de expressar um profundo e emocionado obrigado aos meus pais. Foram eles que me proporcionaram a oportunidade de estudar na faculdade que sempre ansiei frequentar, um sonho tornado realidade graças ao seu incansável apoio e encorajamento em cada passo e decisão que tomei.

Gostaria também de agradecer aos meus amigos e colegas de faculdade que sempre me apoiaram e ajudaram ao longo destes 5 anos.

A todos aqueles que contribuíram de forma direta ou indireta para a realização desta dissertação, o meu sincero obrigada.

Francisca Silva

"The noblest pleasure is the joy of understanding."

Leonardo da Vinci

Contents

List of Figures

List of Tables

Abbreviations

ACG	Average Cumulative Gain
ALARA	As Low As Reasonably Achievable
AP	Average Precision
ATH	Attention-based Triplet Hashing
BoW	Bag-of-Words
CBIR	Content-Based Image Retrieval
CNN	Convolutional Neural Network
CR	Computed Radiography
CSPNet	Cross-Stage Partial Network
CT	Computed Tomography
CXR	Chest X-Ray
DCG	Discounted Cumulative Gain
DCMH	Deep Cross-Modal Hashing
DL	Deep Learning
DR	Digital/Direct Radiography
DRBM	Deep Restricted Boltzmann Machine
FNAB	Fine-Needle Aspiration Biopsy
FR	Feature Representation
GP	General Practitioner
HOG	Histogram of Oriented Gradient
IDCG	Ideal Discounted Cumulative Gain
IoU	Intersection over Union
IR	Image Receptor
KDE	Kernel Density Estimation
kVp	Kilovoltage Peak
L_2 norm	Euclidean Distance
LDA	Linear Discriminant Analysis
LBP	Local Binary Pattern
LSH	Locality-Sensitive Hashing
LXIR	Lesion-based CXR Image Retrieval
mAP	mean Average Precision
mAs	Milliampere Seconds
Mbps	Megabits per second
ML	Machine Learning
nDCG	Normalised Discounted Cumulative Gain
NHS	National Health Service
NMS	Non-Maximum Suppression

OD	Optical Density
OID	Object/Image Distance
PA	Posteroanterior
PACS	Picture Archiving and Communication System
PaL	Parenchymal Lesion
PANet	Path Aggregation Network
PCA	Principal Component Analysis
RBM	Restricted Boltzmann Machine
ROI	Region of Interest
SAE	Stacked Auto-Encoder
SID	Source/Image Distance
SIFT	Scale-Invariant Feature Transform
SPP	Spatial Pyramid Pooling
SURF	Speeded-Up Robust Features
YOLO	You Only Look Once

Chapter 1

Introduction

1.1 Context

Chest radiography (CXR) is a very affordable and effective diagnostic tool used by doctors as it provides a large amount of information about the thoracic anatomy and still is a mainstay in the diagnosis of many pulmonary diseases. With these exams, different pathologies can be discovered and as they are used to detect many conditions, they require plenty of time, attention, and understanding from physicians.

In dubious cases, it is common for radiologists to look back on previous and memorised similar CXR images to support their decision about the case that they are analysing. Naturally, the experts do not remember all the previous cases from a certain pathology and if they try to search for them in a large database of CXR images, this process is extremely time-consuming and inefficient.

With the help of an automatic CXR image retrieval system that searches and finds all the cases with similar features (i.e., the same pathology, in the same location, with the same size, etc.), this procedure would be faster and would allow junior radiologists to learn the properties of the pathology's classes through case comparison. Systems that can make the search and comparison process automatic normally use Content-Based Image Retrieval (CBIR) methods that offer a more consistent and reliable medical image analysis, as presented in [7]. CBIR methods have a vital role in indexing and finding images with visual similarities like shape, size, position, etc. These methods are mainly divided into two steps named feature representation and feature indexing.

In the CXR domain, some CBIR methods have been already applied that can offer a good similarity between medical images of the same condition. When these methods extract features from the whole image to represent a condition, they dismiss the possibility of the existence of multiple lesions in the same image. As discussed in reference [13], a solution to this is the use of CBIR methods that are focused on the lesion. These methods do not extract the features from the whole image, discarding the factors of the image that are not relevant for the radiologist, like the anatomical similarities, the position of the patient's chest, the contrast, etc. However, when a lesion retrieval process has a limited amount of data, the diversity of cases of each class may be reduced.

Thus, it is important to optimise these methods in order to improve image retrieval and overcome the limitations of the system.

1.2 Goals

In this study, the primary research objective is to comprehensively evaluate the performance and effectiveness of the CBIR system called LXIR (Lesion-based CXR Image Retrieval), presented in reference [12], that is applied to the medical domain with a specific focus on lesions present in Chest X-ray (CXR) images. The main aim is to investigate the generalisation and robustness of the LXIR system in retrieving lesions from diverse databases, namely VinDr-CXR and ChestX-ray14. Additionally, this work aims to explore the relationship between distance measurements and visual similarity in the retrieval process and to assess the dissimilarity between different classes and the dissimilarity between the subclasses within the Parenchymal Lesion (PaL) class.

The significance of this research lies in its potential to address the challenges faced in medical image retrieval, particularly in the context of CXR databases. By applying the LXIR system and evaluating its performance through a series of experiments, valuable insights can be gained into the system's generalisation capabilities, robustness, and overall effectiveness in retrieving lesions from diverse databases. This research contributes to the advancement of CBIR methods in the medical domain, where traditional approaches may fall short in handling large databases and fine-grained analysis.

Furthermore, the findings from this study have practical implications for radiologists and medical professionals, as a robust and efficient CBIR system can significantly assist in their decision-making processes and expedite the search for similar cases in large medical image repositories. The results of this research can also guide the optimisation of CBIR systems and further research in the medical domain to improve retrieval performance, especially when dealing with subclass discrimination and outlier challenges in medical image databases.

1.3 Structure of the report

The following chapters will be dedicated to:

- Chapter 2: Fundamental concepts of Radiography and, specifically, of Chest Radiography;
- Chapter 3: State-of-the-art of Content-Based Image Retrieval in medical applications, with a specific focus on its application in Chest Radiography;
- Chapter 4: Description of the databases and the methodology applied;
- Chapter 5: Description of the performed experiments, the obtained results and the conclusions drawn;
- Chapter 6: Conclusions drawn from this dissertation and future work.

Chapter 2

Fundamental Concepts

This chapter serves as an introduction to the fundamental concepts of Chest Radiography, as it constitutes the primary imaging modality under analysis and investigation in this research project.

2.1 Radiography

In 1879, Sir William Crookes observed that the passage of electric current through a high vacuum tube caused a "greenish fluorescence of the walls" [14]. He named this new type of radiant energy cathode rays.

In 1895, a German physicist and mathematician called Wilhelm Conrad Roentgen observed that when a Crookes tube was involved in a black paper, he could see fluorescence in a nearby barium platinocyanide-coated paper. He concluded that a hitherto unknown force that was made at the moment of impact when the cathode rays collided with a solid object caused this phenomenon. He called this type of energy "*X-rays*", because the *X* mathematical symbol represents the unknown.

In a reduced amount of time, Roentgen managed to identify the properties of X-rays that are known today. In 1901, he received the first Nobel Prize in Physics as an award for the enormous impact that his discovery had on the development of science.

Although this was revolutionary, when individuals were exposed to significant doses of X-rays, their health suffered injuries like *erythema*, a skin burn. Through the years, more serious reactions were noticed, like acute radiation sickness, the development of cancer and DNA damage.

However, with proper management of radiation, such as the use of radiation-protection procedures, these risks can be minimised and the protection of the patient and the radiologist can be safeguarded.

2.1.1 Physical Concepts of X-rays

The first tubes used by Roentgen for the study of X-rays were a design made by Crookes for the study of the "cathode rays". These tubes were also known as gas tubes, and they started the release of the electrons by ionisation provoked by the impact of the positively charged molecules of residual gas (present at a reduced level of pressure in the glass enclosure) with the cathode structure

[1]. The production of X-rays was performed when the electrons were abruptly decelerated on a platinum anode target or on the glass wall itself. At that time, the radiation produced was dependent on numerous things like the environmental conditions (that cause variation in temperature and humidity), the residual gas inside the tube, the voltage waveform applied and the current characteristics of the power source. The penetrating power of radiation was dependent on its intensity, and the electrical characteristics of multiple gas tubes differed among them.

In 1913, a significant success in X-ray tube technology occurred when William Coolidge developed the hot cathode electron source. He denominated this tube as "Coolidge X-ray tube", figure 2.1, that was a very high vacuum tube with an embedded solid tungsten disk target in a completely evacuated glass envelope, a spiral tungsten cathode filament and an anode built with copper. This kind of tube mitigated the problems of the previous one and could continuously operate for several hours, providing an intensity and penetrating power of radiation approximately constant [15].

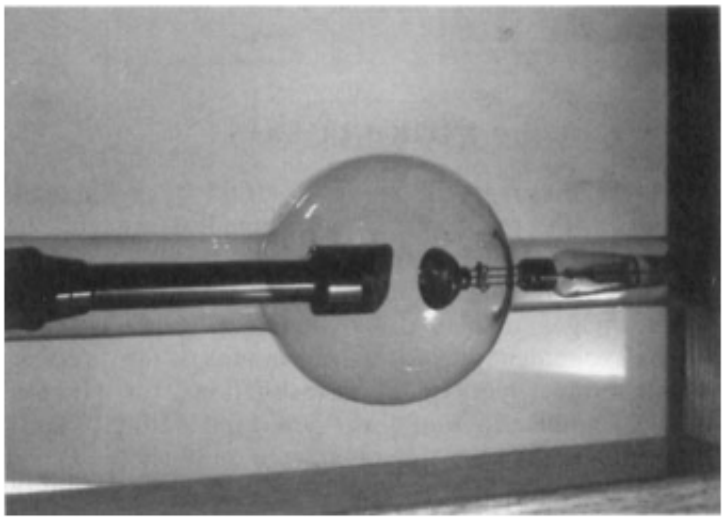


Figure 2.1: The Coolidge X-ray Tube [1].

The X-ray tube has been considered one of the most important components for a successful X-ray examination. The better quality it has, the better the final result will be [16]. The main components of the X-ray tube are shown below in figure 2.2, and they consist of the cathode assembly, the anode assembly, the tube envelope, the rotor and stator, and the tube housing [17].

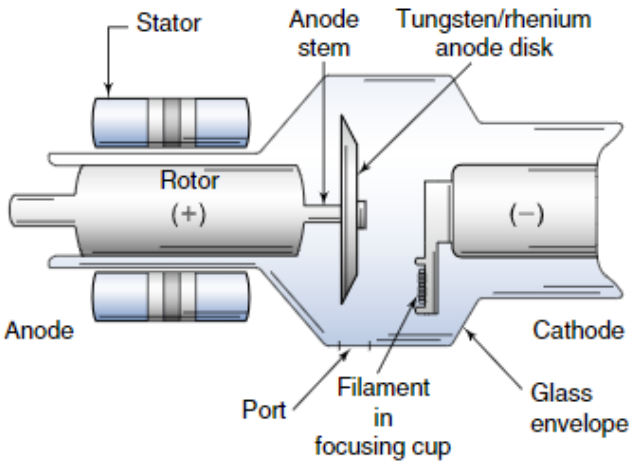


Figure 2.2: The basic components of the X-ray Tube [2].

Cathode Assembly

The cathode of an X-ray tube is the source of the negatively charged electrons and, generally, it presents a larger and a smaller filament within a focusing cup. An electric field is produced when there is a high voltage between the cathode and the anode. This electric field, combined with the shape of the filament in use, creates an electrostatic lens that directs the electrons to a target, and their impact forms an area called the focal spot. A smaller focal spot, created by the smaller filament, leads to fine detail radiography. However, as the local heat on the target becomes greater, the maximum power allowed must diminish, leading to a decrease in X-ray intensity.

The size of the filament is normally determined by a balance between the necessity of X-ray intensity production and the duration of radiation exposure. Wide exposure times may cause motion blur in the image. When, in a practical application, this phenomenon is not a problem or when resolving capability is essential, it is common to use a shorter filament. The larger filament is used when the application requires a high X-ray intensity in a short time interval of exposure [17].

Anode Assembly

At an early age, a fixed anode was constructed with a fixed copper block and a tungsten target. This type of anode has a low heat capacity, causing a limitation on the X-ray intensity produced. In applications that do not require high intensity, like dental radiography, stationary anodes are still used.

With the appearance of a rotating anode, the heat capacity of the tube improved, leading to a higher X-ray intensity, and the effective area of the target became larger. This happens because, during the exposure time, the target is expeditiously rotated, permitting an increase in the instantaneous heat load on the target.

In both designs of anode tubes, the target is departing from the perpendicular of the axis defined by the anode and the cathode by a small angle. The hide angle formed by the X-rays that come from the target is restrained by the collimation of the exit port of the housing tube in one way and by the self-absorption of the anode material target in the other way. This last referred limitation is called *heel effect* and makes the intensity of the X-ray beam diminish through the anode-cathode axis. At the perpendicular axis, this effect does not occur. However, as the cathode side has a higher X-ray intensity, in some cases, the operator positions the patient in a way where his body part with more attenuation is irradiated by the cathode side of the X-ray field.

The effective focal spot size is originated from the projection on the target done by the electrons' path. Second, the line-focus principle, a larger target angle leads to a larger field size and effective focal spot size. For a finer spatial resolution, a diminished effective focal spot size is desirable.

X-ray Tube Envelope

Typically, the structure of an X-ray tube contains glass or, more commonly, a metal envelope because metal has better electrical properties and extends the life of the tube. This component can keep the required environment constant. In the production of X-rays, the heat, which is most of the energy that is not transformed in X-rays, is transferred to the envelope and then to the insulating oil that surrounds the tube envelope. Also, some X-ray tube assemblies have a fan that helps to dissipate the heat by blowing air over the tube [2].

Rotor and Stator

During the exposure to X-rays, the components of the electric induction motor used for the rotation of the anode are the rotor, made of copper, and the stator. The stator is an electric motor external to the tube envelope and when it is practised an alternating current on its windings, inside the area of the rotor is induced a change in the magnetic field. This leads to a turn of the rotor and consequently, a rapid rotation of the anode because the rotor and the anode are strongly attached through the anode stem. This stem is typically built with molybdenum or stainless steel, and its function is to protect the high-strength ball bearings in the rotor from high heat. They allow a smooth rotation of the anode at high speeds.

X-ray Tube Housing

X-ray tube housing is the external composition of the X-ray tube, and it is used to protect the patient by shielding the leakage radiation, radiation that does not belong to the primary beam defined by the exit port of the housing. It also provides electric isolation and structural support.

Before the exit port, the leakage radiation has already been "hardened" by an inherent filtration that attenuates the low-energy radiations.

After the exit port, it is common to harden again the higher effective beam with an additional filter, typically made with aluminium. These processes eliminate the low-energy X-rays that only affect the patient's dose.

A set of collimators is used outside the tube housing, allowing the operator to limit the X-ray beam to the region of interest.

2.1.2 X-rays Properties

X-rays have electrical and magnetic properties and are a form of ionising radiation. *Ionising radiation* is a type of radiation that produces ions using a certain quantity of energy to separate tightly bound electrons from atoms. When these high-energy electrons are accelerated by an electric field and hit the metal target, they release X-ray photons. These photons have enough energy to ionise atoms and molecules in the body, which can damage DNA and cause cancer.

In radiography, ionising radiation is used to produce images of the body by passing a beam of ionising radiation through the body and capturing the resulting shadows on a special film or digital detector.

X-rays are an electromagnetic type of radiation that is invisible, electrically neutral, has no mass and travels at the speed of light, $3 \times 10^8 m/s$, in a vacuum. The beams produced by X-rays are heterogeneous and poly energetic, various photons with different energies produce each one of them. Each photon in a divergent X-ray beam travels in a straight line, and the X-rays cannot be reflected or refracted. By making the air behave like a conductor, X-rays can consequently discharge a body that was electrically charged.

The maximum energy of a beam is expressed in kilovoltage peak (kVp), and it can determine the quality and penetration of each X-ray beam. For medical proposes, the X-ray range of energies applied is between [30,150] kVp and the energies are selected depending on the type of examination that is being performed, the patient's age, the pathology at hand and the patient's condition. Commonly, this measure is combined with other factors like the amount of X-ray exposure (measured in milliampere seconds, (mAs)) and the focus-to-skin distance measured between the skin of the patient and the X-ray tube. When it is used with higher energy, generally, the X-ray beam is more penetrating and the contrast in the image is lower.

The radiation dose used in a patient must comply with the "As Low As Reasonably Achievable" (ALARA) principle. To minimise the dose of radiation, the amount of time of exposure to this type of radiation must be reduced, a safe distance must be guaranteed and the use of radiopaque materials that reduce the exposure of the neighbouring areas to radiation, that is, the use of shielding materials, must be maximised. Other protection measures that must be considered are to avoid replicating the same exam in a patient and, in the case of a much-need exam on a pregnant patient, the radiation dose must be decreased, using a precise collimator and shielding materials to reduce the exposure of the developing fetus to the radiation. [2].

2.1.3 Acquisition System and Image Formation

X-rays can cause chemical changes in radiographic and photographic films, leading to the appearance of images. Image receptors are substances used in diagnostic radiography that produce light when X-rays collide with them. By being absorbed or scattered, X-rays can create a secondary photon as a consequence of the interaction made with the matter (human body), *photoelectric effect*. This depends on the energy of each photon and the composition and thickness of the tissues and structures being exposed to this type of radiation.

When an X-ray beam passes through a patient's body and interacts with an *image receptor* (IR), like a digital-imaging system, a radiographic image is produced. The anatomic area of interest is structurally represented by the variations in the transmission and absorption of the X-ray beams, as shown in figure 2.3.

The primary X-ray beam is the radiation that comes from the tube and is attenuated by the air. When the radiation encounters matter, a portion of the energy is absorbed and scatter radiation is produced.

Scatter radiation is more difficult to control and it is commonly less energetic than the primary X-ray beam. This kind of radiation causes unwanted exposure across the image receptor to the radiation, creating image noise, a layer of "fog", that leads to the loss of the image contrast. The intensity pattern of the radiation that exits the matter (Remnant/Exit Radiation), that is, the remains of the primary X-ray beam, conducts to the creation of the radiographic image.

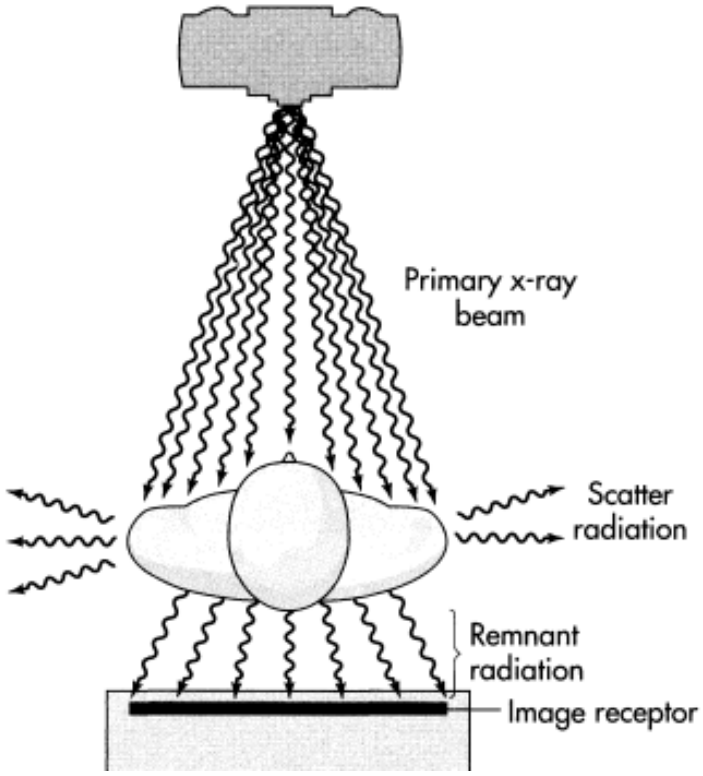


Figure 2.3: The phenomenon of X-ray beam attenuation [3].

This variation of intensities makes visible the brightness levels of the image. Dense tissues, such as bones, absorb more radiation and appear whiter in the image. Less dense tissues, like soft tissues and fluids, scatter more radiation and appear grey or black in the image.

The *Radiographic Contrast* is defined as the difference of the blackness (also called *Optical Density* (OD) in consecutive structures inside the image. For different anatomic tissues, the radiograph may exhibit different levels of brightness, as shown in image B in figure 2.4. A low-contrast image is an image with an overall grey appearance. The desired contrast image is an image where all areas show detailed information, but the less relevant parts are less noticeable.

The main contrast control factor is kilovoltage. A higher kilovoltage leads to a more penetrating X-ray beam, which consequently reduces the white areas of the image. Although, a high kilovoltage does not necessarily lead to the darkness of the easily penetrating subjects because, for an optimum contrast of the image, less intensity of X-rays during the exposure time is needed. Thus, for a higher contrast, the kilovoltage is reduced, resulting in beams with less penetration, and the amount of radiation is increased for a higher X-ray intensity.

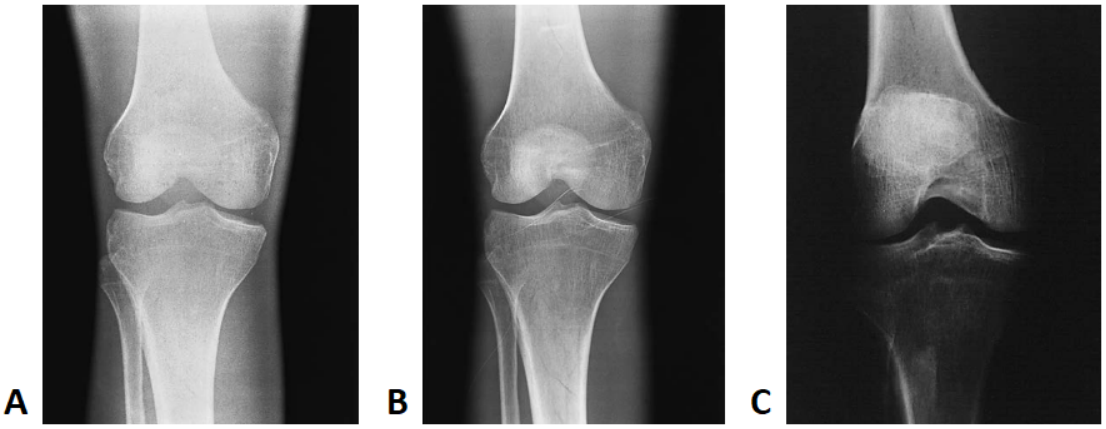


Figure 2.4: Example of a radiograph with: excessive brightness (A), sufficient brightness (B) and insufficient brightness (C) [2].

Another element that influences the image quality is the *Image Detail*. An image with higher detail presents precise edges and lines, oppositely, a low-detail image has blurred (*"out of focus"*) edges. Image detail is affected by the focal spot size, by a short movement of the patient during the exposure, by the source/image distance (SID, distance between the source of radiation and the Image Receptor) and by the object/image distance (OID), the distance between the object and the IR.

The last element that affects the quality of the image is the *Distortion* that is defined by the incoherence between the size/shape of the produced image and the real size of the object. The size distortion is also a consequence of the SID and the OID. The shape distortion is a result of different magnifications of components of the object. this distortion is the lowest when the IR is parallel to the position of the object.

Radiographic Film

For many decades, before the creation of digital image receptors, film-screen imaging was used for the acquisition of radiographic imaging, although it presents many deficiencies.

During this radiographic procedure, cassettes function as film holders that generally accommodate two intensifying screens and protect the films with their light-tight and rigid structure. The intensifying screens are plates coated with fluorescent crystals, and phosphors, that, when exposed to X-rays, emit light. Their role is to reduce the quantity of exposure needed to produce an image. If they have damaged areas, dirt or stains, the light that exposes the film is restricted, resulting in the appearance of artefacts on the image [3]. Also, the image contrast is limited, the intensity screens are expensive and the occurrence of damaged areas is usual.

This method requires long film processing times and the utilisation of chemicals, making the images harder to manipulate, duplicate or store digitally. The IR detects a limited dynamic range (the range of exposure intensities) making the film sensitive to overexposure and underexposure and the visibility of the tissues that vary considerably in X-ray attenuation is restricted.

Filmless Radiography

Although filmless systems are expensive, their necessity for less space, time, and processing chemicals conjugated with the vantages of digital electronic images results in an advantageous conversion. Digital imaging allows a multiplanar image reconstruction and a higher spatial frequency that results in improved spatial resolution.

Computed Radiography (CR) uses an imaging plate made of photostimulable phosphors that is exposed in a special cassette. This special cassette is then introduced to a special processor that with a small beam of its high-intensity laser stimulates the phosphor particles on the plate, causing the latent image to be converted to a visible image and captured by a photomultiplier tube. This tube emits an electronic signal that is processed by a computer to create a digital image.

This image can be viewed on a high-resolution monitor, stored electronically, and even with a laser film printer, duplicates can be created. CR is also less sensitive to motion and can be used to image patients who are unable to remain still for traditional radiography.

Digital/Direct Radiography (DR) uses special radiographic tables and vertical cabinets with radiation receptors that react to variations of the remnant radiation. These sensors are placed in the same position as in film radiography and, instead of using cassettes to capture the image, they convert the radiation into a digital signal that is directly transferred to a computer, where an image can instantaneously be seen and manipulated. This technique does not involve any processing.

These techniques automatically correct the visual image quality and allow faster image acquisition and the ability to manipulate the images electronically to enhance visibility.

2.1.4 Chest Radiography

In 2021 in England, plain Chest Radiography was the most commonly required test by General practitioners (GPs) with approximately 1.8 million tests, and through all sources, approximately

7.9 million tests. In the figure 2.5 presented below, we can observe that Plain Radiography remains the most common imaging test when compared to Ultrasound and Computed Tomography (CT) scans from April 2021 to March 2022.

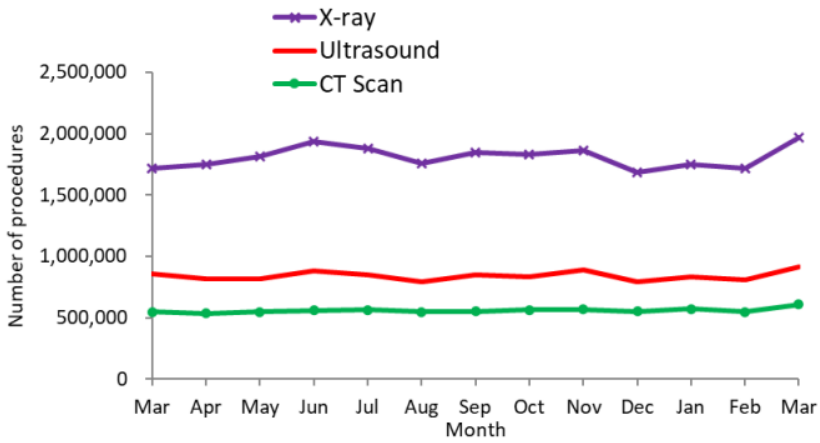


Figure 2.5: National Health Service (NHS) imaging activity in England from April 2021 to March 2022 [4].

Chest radiographs are commonly used in clinical practice for a diversity of diagnostic and therapeutic purposes. They are frequently the first line of imaging for patients presenting with respiratory symptoms, like coughing, shortness of breath, and chest pain. CXR images can help to identify a wide range of medical conditions, including pneumonia, lung cancer, tuberculosis, and heart disease. More precisely, they help to identify pathologies, like Aortic Enlargement, Atelectasis, Calcification, Cardiomegaly, Consolidation, Infiltration, Interstitial Lung Disease, Lung Opacity, Nodule/Mass, Pleural Effusion, Pleural Thickening, Pneumothorax and Pulmonary Fibrosis. They can also be used to assess the function and health of the lungs, heart, and other structures in the chest. In addition, Chest radiographs may be used to guide certain medical procedures, such as percutaneous Fine-Needle Aspiration Biopsy (FNAB) of pulmonary nodules and drainage of the pleural fluid. Overall, CXR images are a valuable tool in clinical practice and can provide essential information for the diagnosis and treatment of many medical conditions.

The frontal CXR image is observed as if looking at the patient, with the patient's right side on the viewer's left. In a normal CXR image, all the components that can be identified are [5]:

- Lungs: right and left; right upper, middle, and lower lobes; left upper (including lingula) and lower lobes
- Pulmonary arteries: main, right, left, right interlobar, left lower lobe
- Airway: trachea, carina, main bronchi;
- Fissures: minor, superior accessory, inferior accessory, azygos;
- Aorta: ascending, arch ("knob"), descending;
- Veins: superior vena cava, azygos, left superior intercostal ("aortic nipple");
- Aortopulmonary window;
- Right paratracheal stripe;

- Junction lines—anterior, posterior;
- Azygoesophageal recess;
- Paraspinal lines;
- Left subclavian artery;
- Heart: right atrium, left atrial appendage, left ventricle, locations of the four cardiac valves;
- Bones: spine, ribs, clavicles, scapulae, humeri.

The Hilar corresponds to the main bronchi and the main pulmonary vasculature. The left hilum is often represented above the right hilum, but it can vary from patient to patient. The two hilar normally have the same size. In the two lungs, some asymmetry between them is normal. In comparison to the left diaphragm, the right diaphragm is commonly 1.5 to 2.0 cm higher. The stomach bubble can be observed below the left hemidiaphragm. The angle formed by the lateral chest wall and the dome of each hemidiaphragm is called the costophrenic angle, and this angle must be visible. In a posteroanterior (PA) Chest radiograph, the size of the heart is generally $\leq 50\%$ of the largest diameter of the thoracic cage and the lugs' opacity must be equal. Also, the left pulmonary artery is located 3 cm down the left main bronchus and then goes up and out at around 45°. Its width must be shorter than the aortic knob.

Some anatomic structures in a PA CXR image are presented in figure 2.6.

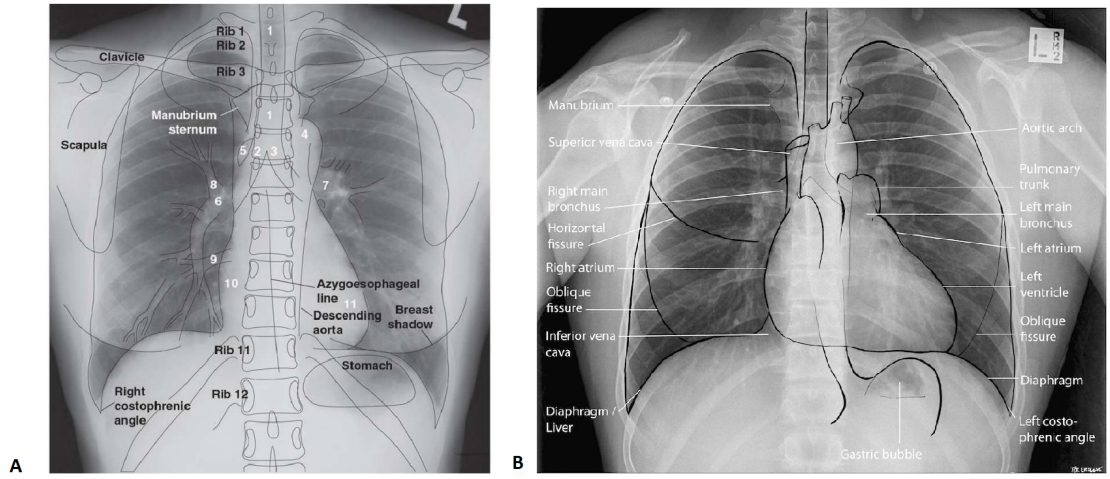


Figure 2.6: Normal anatomic structures presented in a Posteroanterior Chest Radiograph:**A**-(1) trachea, (2) right mainstem bronchus, (3) left mainstem bronchus, (4) aortic "knob" or arch, (5) azygos vein emptying into superior vena cava, (6) right interlobar pulmonary artery, (7) left pulmonary artery, (8) right upper lobe pulmonary artery (truncus anterior), (9) right inferior pulmonary vein, (10) right atrium, (11) left ventricle, and the other labelled structures [5]; **B**- Another example of a labelled PA CXR image with normal anatomy [6].

2.2 Towards Automatic Chest Radiography Image Retrieval

Digital Radiography was revolutionary, but with the conjugation of Picture Archiving and Communication (PACS), images started to be remotely observed and manipulated in different locations by clinicians.

Nowadays, automatic systems that interpret and analyse CXR images are in constant development. They use advanced algorithms and techniques involving the areas of machine learning, deep learning and computer vision. These systems allow more efficient and accurate detection and diagnosis of pathologies from various medical branches.

Recent advances in this area make use of deep learning approaches that assist radiologists in the interpretation of CXR images, managing to identify abnormalities and also monitoring and tracking their progress.

With the appearance of communication systems and picture archiving, image collections of different modalities and pathologies lead to the opportunity to find similarities between different cases and increasingly contribute to the evolution of image search techniques like CBIR systems that, by making use of visual features as search criteria, like low-level features (p.e. colour, texture, and shape) complement traditional text-based image retrievals. These methods eliminate uneconomic and subjective manual labelling by directly extracting in an automatic or semi-automatic way these features. In the following chapter, it is better explained the utilisation of CBIR in the medical domain.

Chapter 3

Content-Based Image Retrieval in Medical Imaging

A CBIR system serves as an effective decision support tool, employing image content and non-image metadata, such as image labels, to extract objective and quantifiable imaging features. By applying a case-based search approach, CBIR systems can identify visual similarities among images, helping in the retrieval of relevant cases.

In contrast to models solely reliant on pathology labels provided by specialists, CBIR systems utilize a search criterion based on visual similarities. This approach helps address potential incoherence arising from the comparison of images with the same pathology but accompanied by different labels, thus mitigating retrieval inaccuracies. CBIR models consist of two main steps: feature representation and feature indexing.

Feature representation in a CBIR system involves the extraction of relevant and discriminative features from medical images. These features aim to capture essential characteristics of the images, such as texture, shape, or intensity, which are then used to represent the content of the images in a more compact and meaningful manner.

On the other hand, feature indexing is the process of organising and storing the extracted features in a way that facilitates efficient and fast retrieval. This step typically involves creating an index or database of feature vectors, allowing for quick comparisons and similarity searches when a query image is presented to the CBIR system. Efficient feature indexing is crucial for reducing retrieval times and ensuring the system can handle large-scale medical image databases effectively.

In this chapter, we present state-of-the-art CBIR techniques applied to medical imaging data, encompassing domains such as machine learning (ML), computer vision, and deep learning (DL). These techniques offer valuable insights into the advancement of CBIR systems in the medical domain, enabling more accurate and efficient retrieval of relevant medical cases.

3.1 Feature Representation

Feature representation is an essential step in the processing of medical images. With the advances in ML and DL techniques, it is possible to increase the number of features extracted, however, the search for an outstanding solution becomes more difficult when high-dimensional features are used [18].

In feature representation, the low-level content of each image is represented with feature vectors. It is possible to link these vectors to high-level perceptions of the images. High-dimensional data is harder to analyse and holds redundant and irrelevant features that can reduce the system's performance. It requires a larger amount of storage space for the feature vector and it can affect the retrieval accuracy and lead to higher computational complexity [19].

A good feature representation is essential to achieve superior performance in medical image retrieval. This process can be classified about whether the features are acquired through an expert-driven procedure, i.e., handcrafted features designed through domain expert knowledge, or entirely by a data-driven procedure, i.e., learned features designed through an automatic process based on data points [20].

3.1.1 Handcrafted

Prior to the widespread adoption of deep learning, feature extraction in medical retrieval systems primarily relied on handcrafted methods. Many of these systems still employ such techniques for feature representation, often combining multiple handcrafted feature methods to enhance retrieval performance. However, when confronted with extensive medical data, the resulting combination of features may become excessively large, impeding scalable retrievals and adversely affecting retrieval efficiency [7].

In general, according to algorithms based on expert knowledge, in each image the features are extracted and correspond to specific information, as low-level visual information (e.g. colour, texture, shape, density, statistical, morphological, geometric) [20]. In medical image retrieval, most of the handcrafted methods are based on generic features such as a local texture feature called *Scale-Invariant Feature Transform* (SIFT), proposed by Lowe et al. [21], that, by finding local extrema in the Difference-of-Gaussian space, detects the scale-invariant key points. SIFT descriptors use a *Bag-of-Words* (BoW), first applied by Sivic and Zisserman in 2003 [22], to be modelled/quantized. This computes the local features by counting the frequency of the generated visual words in each image. Other techniques used as generic features are the Local Binary Pattern (LBP) [23], the Histogram of Oriented Gradient (HOG) [24], Gabor [25] and the Speeded-Up Robust Features (SURF) [26], [27].

Another category of features adopted by medical image retrieval was the *holistic features* that represent the global information of the full image [7].

Regardless, despite their wide application, handcrafted features have shown to be time-consuming and computationally expensive, particularly when large-scale image repositories are used. The

methods applied cannot be generalised as they are created for specific medical data and, even in each case, the features may differ visually when multiple dimensions and modalities are used.

3.1.2 Data-driven (ML/DL)

In deep learning methods, features are learned from multiple layers of abstraction, where the higher levels are formed by a set of lower-level features. In the higher levels of abstraction, where humans do not know how to explicitly specify the features that characterise a certain class, the capacity to learn features automatically allows the system to be more independent of user-defined features and to learn complex functions that map, directly from data, the inputs to the outputs. As the amount of data presented in the databases continues to grow, there is an increasing need to automatically learn features from images to optimise this step [28].

Supervised deep neural networks require a massive amount of labelled images, annotated by domain experts, to train the parameters in each layer so, in the medical area where the databases have a limited amount of labelled images, it is common that this supervised feature learning methods cannot be employed as it may result in over-fitting. A common framework of a supervised deep network is the Convolutional Neural Network (CNN) [29], [30].

Multiple *unsupervised deep neural networks* have been proposed for feature representation, like an *Auto-Encoder* that learns the feature representations by minimising the reconstruction error between the input and the output. Ideally, the input and the output, that is a reconstructed version of the input, are similar. Originally, the auto-encoder only possesses one hidden layer and can be viewed as a combination of an encoder and a decoder [31]. Because the single-layer auto-encoder is often too shallow to learn features, the representation power improved significantly when several auto-encoders were stacked to form deep *Stacked Auto-Encoders* (SAEs), like in [32].

In addition to auto-encoders, *Restricted Boltzmann Machines* (RBM) proposed by [33] and *Deep Restricted Boltzmann Machines* (DRBM) that can be seen as multiple RBMs stacked together, can also construct unsupervised deep neural networks and tackle medical feature representations and other tasks [34].

3.2 Feature Indexing

After the process of feature extraction, each image is represented by its corresponding feature vector. Feature indexing, one of the main steps in a CBIR system, presents unique challenges as it involves creating an efficient approach to organise and index the region within the medical image that contains relevant information for analysis or diagnosis, thereby guiding an effective search and retrieval process. This specific region of interest (ROI) typically possesses intrinsic characteristics that necessitate careful consideration.

In the context of feature indexing, a common strategy is to perform a k-nearest neighbour search in the feature space. This method involves computing and ranking the distance between the query image's feature vector and those of the images in the database. The k most similar images

determined by their ranking are then retrieved as potential matches to the query image. This process enables an efficient and accurate retrieval of relevant medical images from the database.

3.2.1 Similarity/Distance Measures

After completing the offline/training phase of the CBIR system, during which rich content from each image is extracted and features are learned, in the online phase, the features from the query image are normally extracted and the image class is determined, allowing for the elimination of semantically irrelevant classes from the retrieval process. By discarding such classes, the system can focus on retrieving images that are visually similar and conceptually related to the query.

The online phase continues with the estimation of the similarity between the query image and the images in the database from a human visual perspective. The goal is to retrieve database images that share a visual resemblance with the query.

The simplest way of calculating this visual similarity is by using distance measures. In general, a lower distance corresponds to a higher similarity between the query image and the database image. Each similarity measure can impact the system's retrieval performance differently, and the selection of an appropriate measure is crucial to achieving accurate and efficient retrieval. Commonly used similarity measures include:

Minkowski-Form Distance

Given the feature vectors of two images I and J, the distance L_p between them is given by:

$$D(I,J) = (\sum_i |f_i(I) - f_i(J)|^p)^{\frac{1}{p}} \tag{3.1}$$

In the equation 3.1 the $D(I,J)$ is the distance L_p and the f_i is the feature i from the feature vector of the image I and J.

For $p = 1$, the distance is the L_1 norm distance function, also known as the Manhattan/ City Block Distance. For $p = 2$, this distance corresponds to the well-known Euclidean distance (L_2 norm) and for $p = \infty$, the distance is called the Chebyshev distance or L_∞ [35].

Cosine Similarity

This metric measures the similarity of the images through the angle that the query feature vector, z_q, and a database image feature vector, z_i, form [36], [37]. The expression of this metric is presented below:

$$cos(z_q, z_i) = \frac{z_q . z_i}{||z_q|| ||z_i||} \tag{3.2}$$

In $z_q.z_i$ of the equation 3.2 it is applied the scalar product of the vectors.

Mahalanobis Distance

This distance is called quadratic distance, and it is defined by the following formula:

$$D(I,J) = \sqrt{(F_I - F_J)^T C^{-1} (F_I - F_J)} \tag{3.3}$$

In equation 3.3, C represents the covariance matrix of the feature vectors.

This distance can also be defined for each dimension of the feature vector independently:

$$D(I,J) = \sum_{i=1}^{N} \frac{(F_I - F_J)^2}{c_i} \tag{3.4}$$

c_i, in the equation 3.4, corresponds to the variance of the feature component i.

Hamming Distance

When this distance is applied to two feature vectors, x and y, it represents the number of times that the elements of the vectors, for the same position i, are different [38], [39]:

$$D(x,y) = \sum_{i=1}^{n} \sim (x_i = y_i) \tag{3.5}$$

Where n corresponds to the dimension of the feature vector x and y. Also in equation 3.5, $\sim$ corresponds to the negation of the value of the expression that follows, having the value of 1 if x_i is different from y_i or the value of 0 if they are equal.

3.2.2 Vocabulary Tree

Proposed by Nistér and Stewenius in 2006 [40], Vocabulary Tree methods can be applied for large-scale medical image retrieval, as shown in figure 3.1. These methods employ a structure based on a tree that accelerates the similarity indexing. Compared to other traditional retrieval methods that use an exhaustive search, vocabulary tree methods are more accurate. When applied to large databases, they improve the computational efficiency. Some variations of this technique have also been used in medical image retrieval, for example, when a weighted version of a framework based on a vocabulary tree was used in a mammogram image retrieval where a higher weight was given to the features with lower frequencies in the mammogram. This is because they present more significant information than the higher frequency features in the mammogram and it is a way of escaping overcount [41].

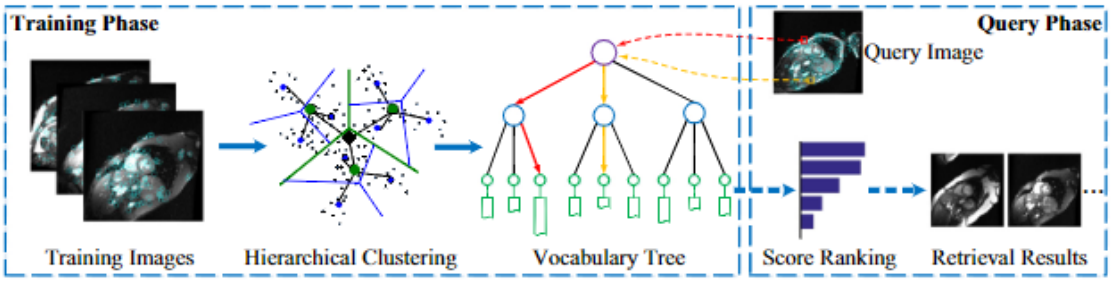


Figure 3.1: Example of a framework for a Vocabulary Tree based image retrieval [7].

Vocabulary tree based methods can be applied to different types of medical images as they use local features instead of global ones, although, in some cases, only using local features may not be enough for the image representation step. Also when these methods are employed in a larger image database, the training phase may be computationally expensive.

3.2.3 Hashing

Hashing methods have been much requested in the fields of ML and computer vision for indexing massive data, as mentioned in Wang et al.survey [42]. The original data is first compressed into short binary codes that are derived from the hashing functions that were defined and then, by computing the similarity distances in binary Hamming space, the nearest-neighbour search is employed, as shown in figure 3.2. The hashing methods can be data-independent or data-dependent.

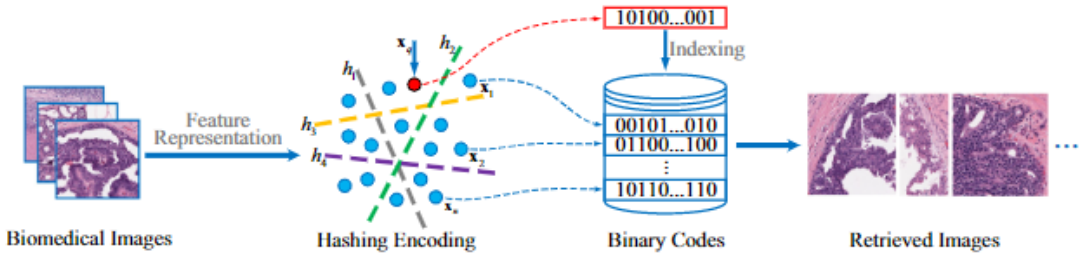


Figure 3.2: Example of a framework for a Hashing based image retrieval [7].

Data-independent methods

These methods create generalised hashing functions that compact any kind of data into binary codes. A well-known method from this category is called Locality-Sensitive Hashing (LSH). This method is based on arbitrary vectors from a particular distribution in order to maximise the probability of collision of identical points when they are mapped into the Hamming space with high probability [43].

Data-dependent/Learning to hash methods

From a given training image repository, data-dependent methods create hashing functions. They can achieve similar or even better retrieval accuracy with short binary codes than data-independent methods.

These methods can be based on whether the training data carries labels or not, being divided into supervised methods, unsupervised methods and semi-supervised methods. Supervised learning is the most common form. It uses labelled data divided into categories and aims to find the best input-output function, trained with a set of data, that can be generalised when tested with other sets of data. The error between the output score and the ideal pattern can be measured with the help of an objective function that generally has adjustable parameters. Unsupervised learning has the goal of uncovering the hidden structure of the data or its distribution and is normally used for classification and regression. As in some cases data labels are difficult to acquire, unsupervised learning is a better option because it is based on deductions from datasets and is independent of label information. Some examples of unsupervised learning techniques are clustering and blind signal separation, such as the Principal Component Analysis (PCA) and the Independent Component Analysis. Semi-supervised learning has the property of training a small-sized labelled dataset and slowly labelling unlabelled data with the purpose of augmenting the training data [18].

Data-dependent methods can also be based on the form of hashing functions that can be linear or nonlinear. With simple projections, linear hashing functions map and split up the original feature space. They showed to be computationally efficient and easy to optimise. However, they cannot take care of the cases where the difference in the data is subtle and linearly inseparable.

To override such limitations, nonlinear hashing methods were created. They learn hashing functions based on manifold structures and kernel matrices, insert the intrinsic structure in a high-dimensional space, and, non-linearly, plot the feature vectors into binary codes.

3.2.4 Other methods

Other methods can accelerate the similarity search, improve retrieval accuracy and reduce the dimension of the feature vectors while preserving the most relevant knowledge.

In addition to hashing, one example is *Feature Compression* which compresses extensive image features into smaller sizes. A popular method of feature compression is the *Principal Components Analysis*, PCA. The retrieved images are computed based on a distance measure after the similarity indexing can be reordered through *re-ranking methods* that move the most suitable images to the top or optimise the diversity in the best results. Another example of feature compression is Linear Discriminant Analysis (LDA) [44], also used for classification, is a technique that has the purpose of maximising the divergence between classes (inter-classes variation) and minimising the variation inside the same class (intra-class variation) by finding a linear combination of the features.

Thus, as re-ranking methods just need to process a few images, they can be very efficient, and the retrieval precision can be improved as they use multiple information sources to consider and compare the similarity [7].

3.3 Evaluation Metrics

The evaluation of the system is necessary to demonstrate its usefulness and the possible impact that the implemented system can have. Generally, the concerns in large-scale medical retrieval systems can be reduced to a balance between efficiency and accuracy. The main task is to evaluate the retrieval performance levels of the entire retrieval system.

In large-scale medical image retrieval, as in generic information retrieval, the evaluation measures provide a comparison and validation of the retrieval methods applied and a quantitative analysis.

One of the measures that can be applied to image retrieval systems is called the *top precision* which is defined as the number of the top-ranked relevant retrievals before the top irrelevant dataset image [18].

In the medical domain and dealing with massive medical data, large intra-class variation and small inter-class variation may be more challenging to overcome when noisy images are included in the data, influencing the retrieval performance of the system. To evaluate the class-consistency, a major indicator used is the *precision*, [45]:

$$precision = \frac{|relevant images \cap retrieved images|}{|retrieved images|} \tag{3.6}$$

The task of deciding the similarity/relevance is not dismissed. When only the class labels that determine the similarity of the images and the image classes are identified, *annotation-based evaluation criteria* is used. In *user-based criteria*, the domain experts can provide a more fine-grained retrieval evaluation. In many medical image retrieval systems, this criterion is employed for performance evaluation [46].

Another measure is the *Average Cumulative Gain* (ACG).

$$ACG = \frac{\sum_{n=1}^{k} s_n}{k} \tag{3.7}$$

For k retrieved similar images, s_n is the graded similarity of the n retrieved image. The s_n value is given by the ratio of the common positive labels between the n^{th} retrieved image and the query divided by the total number of the positive labels in the query image [9]

The *Normalised Discounted Cumulative Gain* (nDCG) can be applied in learning to rank tasks by measuring the ranking quality [47]. This standard metric is a normalization of the *Discounted Cumulative Gain* (DCG) function. At nDCG, the DCG is normalized by the maximum/ideal value

of DCG, also called *Ideal Discounted Cumulative Gain* (IDCG) [10].

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i}}{log_2(i+1)} \tag{3.8}$$

For p retrieved images, rel_i refers to the relevance value.

The relevance values are given by the radiologist that evaluates and organises the retrieval images attributing them a value of similarity. The most similar will have the highest score and the less similar will have the lowest score. This allows the retrieval with a higher similarity will be placed in positions of more importance.

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{3.9}$$

For p retrieved images.

The efficiency of the system can be measured based on the *feature computation time* which is the time that was spent to extract the features from all images; the *retrieval time* represents the time that was needed to retrieve the images after applying a precise number of queries; the *memory cost* that is the number of Megabits consumed per second (Mbps) on the online search [8].

Also, a way of evaluating the performance of the system is by evaluating the relevance of retrieved images using the mean Average Precision (mAP) metric [48] which is defined as the mean of the Average Precision (AP) for each query.

$$AP(q_j) = \frac{1}{m_j} \sum_{1<k<m_j} Precision(R_{jk}) \tag{3.10}$$

R_{jk} is the set of retrieved images and m_j is the number of relevant images for the image query j.

$$MAP(Q) = \frac{1}{N} \sum_{1<j<N} AP(q_j) \tag{3.11}$$

Q is the set of the query images and N is their total number. $AP(q_j)$ is the Average Precision of the query image q_j.

3.4 Content-Based Image Retrieval in CXR

In this section, CBIR techniques applied to plain CXRs are presented, as well as their advantages and disadvantages. In general, CBIR methods that are focused on the lesion and do not extract the features from the whole image, they do not discard factors of the image that are not relevant to the radiologist, such as the anatomical similarities, the position of the patient chest, the contrast, etc.

In 2021 Fang et al. reported a network called Attention-based Triplet Hashing (ATH) [8] that compared to other deep hashing methods achieves better medical image retrieval performance. The system receives three images as input (a query image, a positive image with the same pathology

and visual aspect and a negative image from another pathology and different visual aspect) and with the help of a spatial-attention module, it captures their ROI information and represents each image by a hash code.

During the training phase, the system makes use of the triplet cross-entropy loss in order to achieve the maximum hash code discriminability and class separability, i.e. the classification and the similarity information in the hash codes remains by penalising the classification losses and the similarity losses. This method also uses as input triplet labels from classification datasets to expand the use of small-sample information with the objective of overcoming the problem of imbalanced samples due to the scarce representation of disease cases. These triplet labels give a notion of relative similarity between the images. The system used the triplet labels and the ground-truth labels in the learning of the hash codes and the learning of the classification likelihood. In the training phase with the purpose of minimising the distance between visually similar images and maximising the distance between dissimilar images, a hinge ranking loss is created. This loss is composed of two losses called the triplet loss and the cross-entropy loss.

This system makes use of the Euclidean distance as a part of the triplet loss, which objective is to measure if the learned hash codes can satisfy the given triplet labels. For that, the triplet loss computes the likelihood of the given triplet labels. The cross-entropy loss is applied in order to penalise the classification loss when the predicted classes are compared to the ground-truth classes. The triplet loss leads to an improved classification performance and, by making use of the triplet labels, the small-sample information can contribute to solving the problem of imbalanced-sample. The cross-entropy loss helps in the preservation of the classification data inside the hash codes. The retrieved images are ranked from the lowest hash-code distance to the highest hash-code distance. This ATH system shows state-of-the-art performance results when applied to medical databases.

An example of medical image retrieval when the system is applied to a database of ophthalmic images and when applied to a dataset of CXR images is shown in figure 3.3.

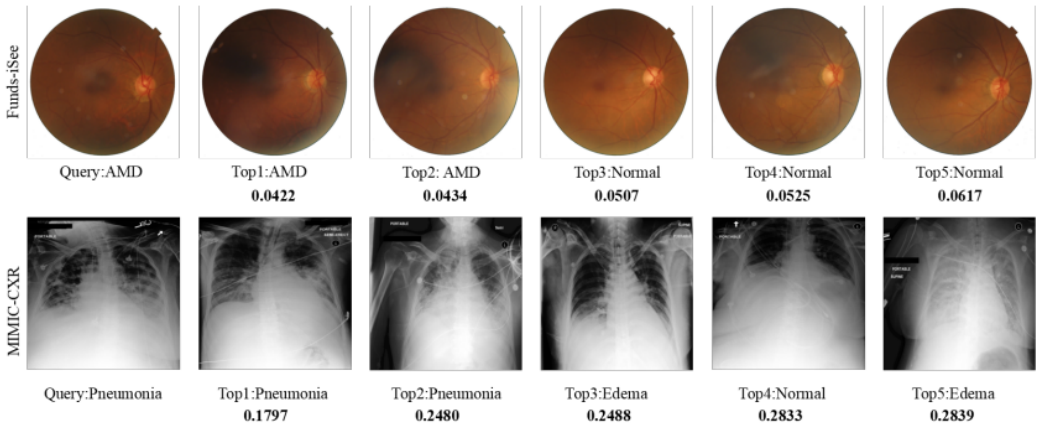


Figure 3.3: Example of retrieved images when two query medical images are tested [8].

Also in 2021, Haq et al. presented a framework based on deep learning [9] that showed better performance when compared to other state-of-the-art image retrieval systems. This medical image retrieval system is composed of a deep neural network that is based on a colour generator trained with CXR image datasets, a graph network based on code-similarity values and, in order to find the different communities formed by similar images, a detection scheme is applied after. The retrieval of a defined number k, of images was defined in two steps. Firstly a region that only includes the node of the query image is created and, based on weighted modularity, a search for the closest image community, is done. Secondly, after finding that community, the k most similar images are retrieved, which correspond to the nodes inside the region of the query image that generates the maximum gain in weighted modularity. Two examples of retrieved images, when the system is tested with two queries from a chest x-ray database, are presented in figure 3.4.

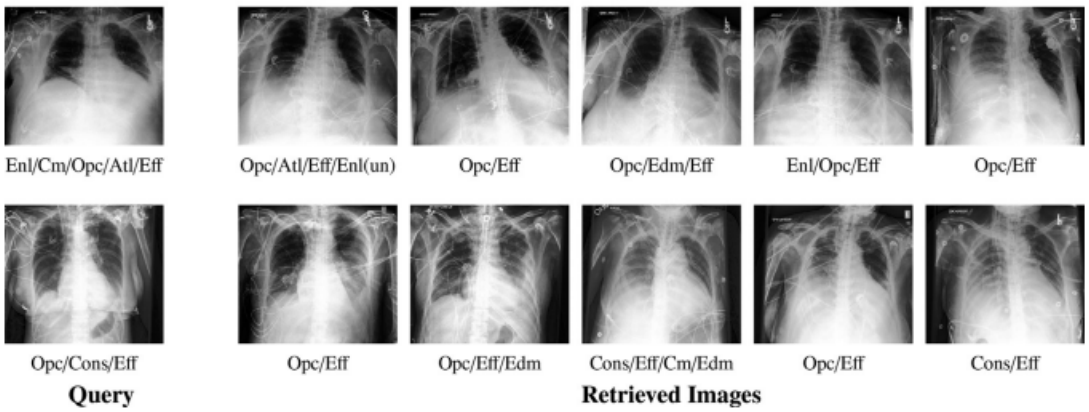


Figure 3.4: Example of retrieved images when two query images are tested [9].

This system was used as a CXR image retrieval system, and only the image content was considered for the extraction of the features. However, this system can also be applied to other medical image retrieval tasks. One of its limitations is that, in the decision-making process, when the clinical information was not considered the performance on some clinical diagnosis-based disease labels was lower.

Another framework proposed in 2022 by Zhang et al. was a category-supervised cross-modal hashing retrieval system applied to a CXR image database [49] that is composed of a category-supervised hashing network that learns the hashing code for each category and applies it as supervised information to guide the learning of the images modality and the text's modality hashes. Also, it uses a union hashing network that has the ability to learn the correlation between the existing modalities. With these networks, it is possible to effectively guide the medical cross-modal hashing retrieval. Compared to Deep Cross-Modal Hashing (DCMH), this method achieved better mAP values.

In 2020, a method proposed by Silva et al., used saliency maps for feature representation, focusing the generated features on the pathological characteristics. As illustrated in figure 3.5, the training process was divided into two different steps.

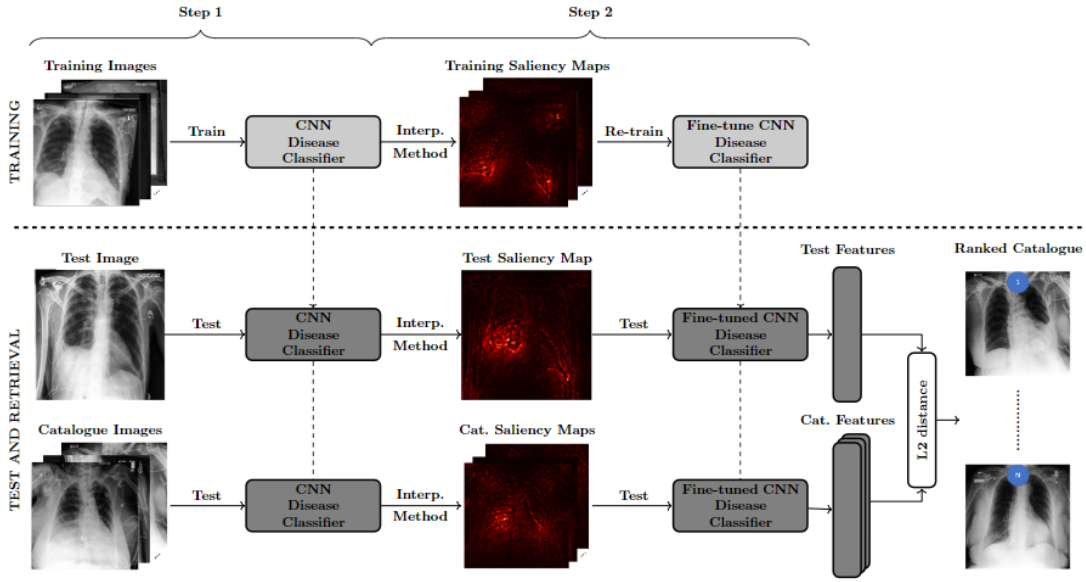


Figure 3.5: Overview of the approach proposed by Silva et al. [10].

In the first one, a pre-trained CNN model called DenseNet-121 was used to classify the CXR images into Pleural Effusion or Non-Pleural Effusion. In the second step, as the objective was to enforce the network focus only in the relevant areas, saliency maps were trained and used as input to fine-tune the already trained CNN architecture.

Although this method achieved better values of similarity between the medical images of the same condition and better image retrieval than the state-of-the-art approaches, it did not allow radiologists to query cases based on specific lesions and the possibility of multiple lesions was disregarded as it only considered two classes.

In 2022 Guan et al. released a study [11] where a framework applied to a CXR dataset was described and addressed the problem of reduced accuracy of medical image retrieval systems due to the difficult separability of classes and the constant omission of medical lesions. A medical image hashing retrieval method that conjugated interpretability and feature fusion was employed, with the objective of improving the system's precision and the similarity ranking of the retrieved images. A global network is created using global image learning that produces saliency maps that function as attention mechanisms to get local discriminate regions and further extract the local features. The resulting feature fusion operations and a jointed loss function are high-quality hash codes that are used for the calculation of similarity and for the retrieval of the closest images. An example of the top-five retrieved images, when some queries are tested, is presented in figure 3.6.

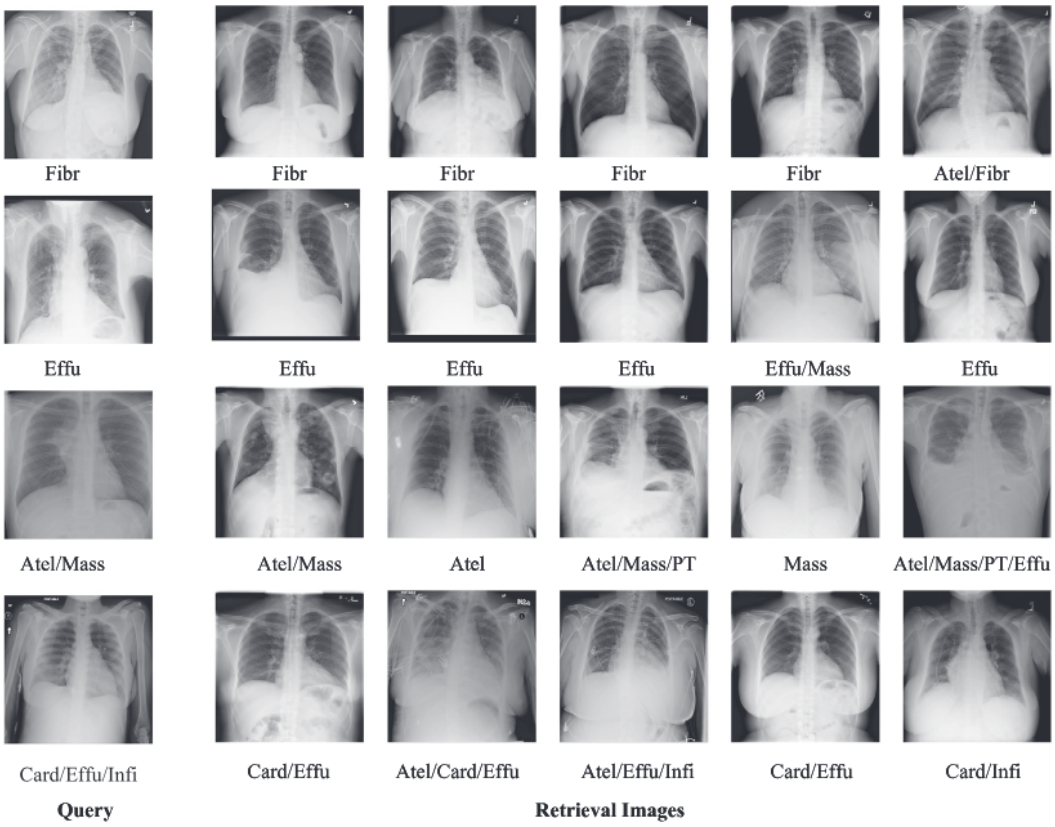


Figure 3.6: Example of the top-five retrieved images and when query images are tested. Below each image, it is presented the correspondent disease label [11].

This system can achieve better accuracy in medical image retrieval, however, as this framework detects the small regions with lesions instead of the whole region of interest, it is difficult to detect large lesion regions. Another limitation is that the size of the database used was not large, so the efficiency and accuracy of the retrieval method may not be the best.

The method proposed by Pedrosa et al. allowed to detect efficiently the areas of the lesions, by extracting the relevant features and using them to retrieve similar cases from a database of CXR images [12], as shown in figure 3.7.

This method proposes a Lesion-based CXR Image Retrieval (LXIR) framework, shown in figure 3.8, that uses an architecture called YOLOv5x as a CXR pathology object detection network trained on previously annotated CXRs. The framework identifies abnormal regions that require feature representation and, by extracting the features of each predicted object at the corresponding location, performs the feature representation itself.

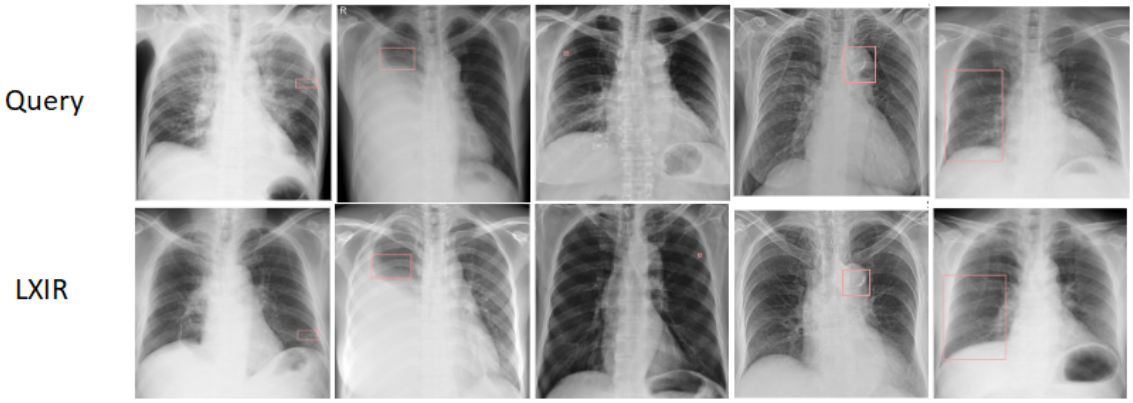


Figure 3.7: Example of a query and the retrieved images when this method is employed [12].

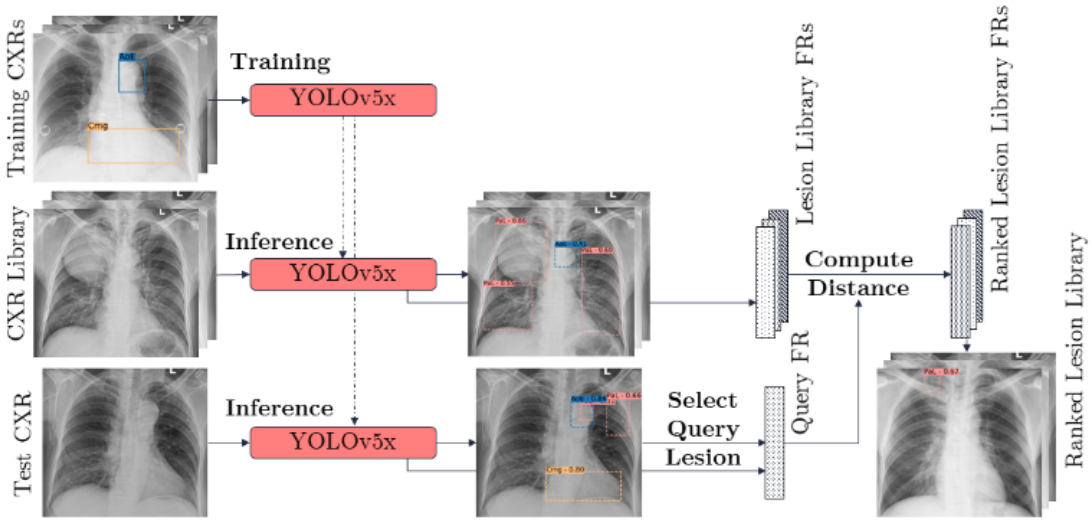


Figure 3.8: Overview of LXIR workflow [12].

However, this framework has a limited amount of labelled data, not allowing the representation of more specific features of all pathology classes, especially the less representative ones. If, in the lesion retrieval process, a query for a rare lesion that is not found in the training/validation set is performed, a relevant retrieval might not be found.

Chapter 4

Content-Based Image Retrieval in CXR

The integration of deep learning techniques with large-scale public databases has been extensively employed to develop methods for detecting abnormal or pathological regions. Despite their promising results, these approaches are not frequently implemented in clinical settings due to a lack of explainability regarding the decision-making process.

In this chapter, it is provided an in-depth explanation of the content-based image retrieval (CBIR) system utilised in this study. The first section focuses on describing the public databases utilised, while the subsequent section delves into the methodology employed.

The CBIR system presented herein is referred to as Lesion-Based CXR Image Retrieval (LXIR) and is built upon the architectural framework known as YOLOv5x.

4.1 Datasets

Two distinct databases comprising CXR images were utilised. These databases exhibit divergent characteristics, and to provide clarity regarding their dissimilarities and comprehend their distinctive features, a detailed description of each database is presented in the subsequent subsections.

4.1.1 VinDr-CXR

The VinDr-CXR database is a collection of 18 000 posteroanterior (PA) CXR images with no information about the identity of the patients.

The images contain the classification of typical thoracic diseases, performed by radiologists, and the localisation of the found lesions. The annotations include the location of each finding via a bounding box, each one is identifying a pathology class from 14 possible labels shown in the table 4.1. The training set of this database is composed of 15 000 images independently labelled by 3 radiologists and the test set is constituted of 3 000 images and each one presents annotations resulting from the agreement among 5 radiologists [50].

Table 4.1: The 14 pathology classes of VinDr-CXR.

Class Number	Class Abbreviation	VinDr-CXR Classes
0	AoE	Aortic Enlargement
1	Atl	Atelectasis
2	Clc	Calcification
3	Cmg	Cardiomegaly
4	Cns	Consolidation
5	ILD	Interstitial Lung Disease
6	Inf	Infiltration
7	LOp	Lung Opacity
8	Nod	Nodule/Mass
9	OtL	Other Lesion
10	PlE	Pleural Effusion
11	PlT	Pleural Thickening
12	Pnm	Pneumothorax
13	PuF	Pulmonary Fibrosis

Some classes have a low representation in the database, and in some cases, the radiologists' opinion about the class of the lesion identified as an abnormal region is controversial. Thus, Pedrosa et al. [13] proposed to group the 14 pathology classes into 6 pathology classes in order to reduce the radiologists' variability and subjectivity. These 6 groups represent the 14 lesions as shown in the table 4.2.

Table 4.2: The new 6 pathology classes grouping 14 pathology classes of VinDr-CXR.

Class Number	Class Abbreviation	6 Classes	14 Classes
0	AoE	Aortic Enlargement	Aortic Enlargement
3	Cmg	Cardiomegaly	Cardiomegaly
7	PaL	Parenchymal Lesion	Atelectasis Calcification Consolidation Interstitial Lung Disease Infiltration Lung Opacity Nodule/Mass Pulmonary Fibrosis
9	OtL	Other Lesion	Other Lesion
10	PlL	Pleural Lesion	Pleural Effusion Pleural Thickness
12	Pnm	Pneumothorax	Pneumothorax

By decreasing the pathology classes from 14 to 6 to reduce the radiologists' diagnosis variability, the critical regions' localisation is enhanced. This modification promotes better explainability to the human reader and makes this system more suitable to be applied in clinical practice [13].

4.1.2 ChestX-ray14

The ChestX-ray14 database was collected from 30850 patients. It comprises 112 120 frontal-view CXR images with image labels of 14 possible diseases with no indication of the pathological region through bounding boxes [51]. Table 4.3 presents the number of images associated with each pathology label.

Table 4.3: Number of images of each pathology.

Item #	X-Ray14
Report	112 120
Atelectasis	11 535
Cardiomegaly	2 772
Effusion	13 307
Infiltration	19 871
Mass	5 746
Nodule	6 323
Pneumonia	1 353
Pneumothorax	5 298
Consolidation	4 667
Edema	2 303
Emphysema	2 516
Fibrosis	1 686
Pleural Thickness	3 385
Hernia	227
No Findings	60 412

4.1.3 Correspondences Between Categories of Radiological Findings

As the labels in the VinDr-CXR and the ChestX-ray14 do not perfectly match, a table of correspondences (table 4.4) was constructed in order to facilitate the manipulation of the two databases.

Table 4.4: Establishment of correspondences between categories of VinDr-CXR and the ChestX-ray14 databases [13].

VinDr-CXR	ChestX-ray14
Aortic Enlargement	-
Cardiomegaly	Cardiomegaly
Atelectasis	Atelectasis
Calcification	-
ILD	-
Infiltration	Infiltration
Pulmonary Fibrosis	Pulmonary Fibrosis
Lung Opacity	-
Consolidation	Pneumonia Consolidation
Nodule/Mass	Nodule Mass
Pleural Thickening	Pleural Thickening
Pleural Effusion	Pleural Effusion
Pneumothorax	Pneumothorax
Other Lesion	Hernia Emphysema

4.2 Lesion-Based CXR Image Retrieval

It is common for radiologists to apply direct comparisons between similar findings in images to support their decisions. LXIR is a system based on an object detection framework called YOLOv5x [52], which is a version of You Only Look Once (YOLO), that was used for the detection of abnormalities in the CXR images.

4.2.1 Training Phase

The YOLOv5x model was trained using annotated data from the VinDr-CXR dataset. The dataset was partitioned into train, validation, and test sets, following a ratio of 60%, 20%, and 20%, respectively. Random selection was employed to allocate data within each set while maintaining the percentage distribution of different pathologies as closely identical as possible. Patient identities were not considered in the dataset division since such information was not available. Each CXR image, even if presenting multiple pathologies, was assigned to only one set.

For every CXR image, YOLOv5x extracted the feature representation of each predicted lesion. This information was utilised for creating a collection of CXR representations afterwards used in the retrieval phase.

The YOLOv5x architecture consists of a CSPDarkNet backbone, incorporating a Cross-Stage Partial Network (CSPNet) [53] within DarkNet [54], followed by a Path Aggregation Network

(PANet) [55]. This sub-architecture effectively reduces the number of parameters, initially 87,7 million, to mitigate the issues of repeated gradients and enhance gradient flow refinement.

Figure 4.1 illustrates the YOLOv5x architecture, including the BottleNeckCSP module featuring the CSPNet bottleneck. Convolutional layers with 1×1 and 3×3 kernels are utilised, which can be repeated up to four times. The schematic representation showcases a Spatial Pyramid Pooling (SPP) layer and a spatial upsampling operation with a factor of 2. Additionally, the framework incorporates input concatenation across channels.

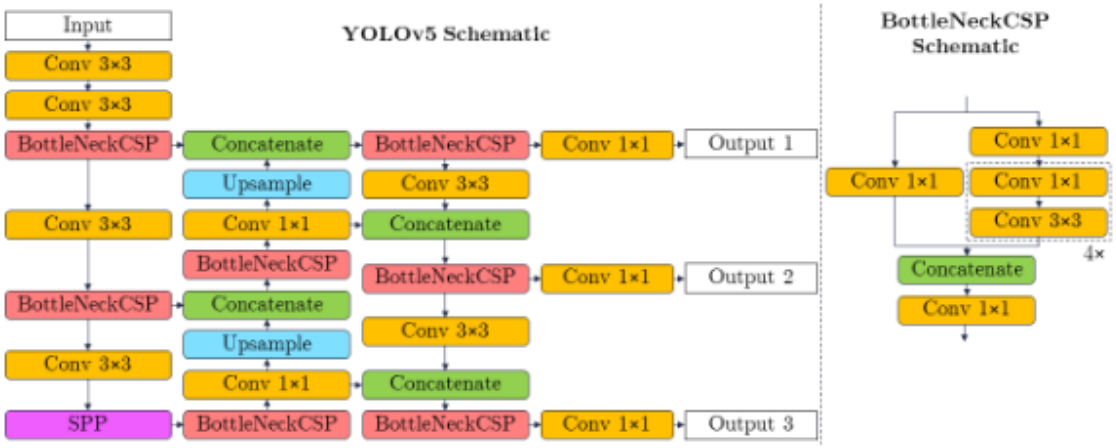


Figure 4.1: The YOLOv5x architecture and the BottleNeckCSP module.

This model divides each image into an MxM grid, where each grid position predicts a bounding box, generating a unique Feature Representation (FR). Each FR captures the higher-order semantic description of the predicted lesion's characteristics. It is important to note that a CXR image can contain multiple independent lesions.

YOLOv5x incorporates automatic optimisation of bounding box anchors, which are predefined bounding boxes with a specific height and width. The architecture also applies mosaic augmentation during training to improve object recognition independently of the background.

YOLOv5x enables multi-scale predictions at three distinct sizes, making it suitable for detecting abnormalities in CXR images and allowing the model to predict objects of various sizes, ranging from entire lung opacities to calcifications and small nodules. Consequently, the model has three possible output branches, as depicted in Figure 4.1, that generate the bounding box predictions.

The architecture was initialised with weights pre-trained on the COCO database [56] and trained using stochastic gradient descent optimisation with an initial learning rate of 0.01 for a duration of 150 epochs.

The proposed framework involves the creation of lesion FR classes based on predicted bounding boxes from the VinDr-CXR database's train and validation sets, where lesions with confidence scores exceeding 0.1 are considered. Non-maximum suppression (NMS) is employed to suppress overlapping bounding boxes, discarding the box with the lowest confidence score if the Intersection over Union (IoU) between two boxes is 0.6 or higher.

For grid locations where predicted lesions are present, an abnormal region detection framework performs FR on these lesions. Each lesion FR contains a fixed number of features depending on the output branch, such as 640, 1280, or 2560 features. Normalisation is applied to the FR vectors, ensuring a zero mean and a standard deviation of 1 to standardise the feature values and eliminate any potential bias or scale differences among the features.

4.2.2 Retrieval Phase

Following the training phase, the YOLOv5x model is utilised to predict lesions in the non-annotated ChestX-ray14 database. Each prediction is assigned a corresponding class.

The query set comprises the test set of the VinDr-CXR database. Lesion retrieval is conducted within a search collection consisting of the ChestX-ray14 database, as well as the train and validation sets of the VinDr-CXR database.

To index the features, the L_2 norm (Euclidean distance) is computed between the FR of a query lesion and the FRs of lesions from the search collection belonging to the same class. Only FRs with matching feature sizes (640, 1280, or 2560) are compared, disregarding lesions with dissimilar sizes or different classes.

The computed distances are ranked, and the lesion FRs with lower distances from the search collection are retrieved, indicating higher similarity to the query lesion.

It is important to note that this unsupervised retrieval framework does not guarantee the veracity of the retrieved lesions. To address this limitation, a suggestion is made to implement a supervised retrieval approach, where lesion predictions are compared with radiologist annotations from the VinDr-CXR database. In the supervised retrieval phase, the search collection is confined to the annotated train and validation sets of the VinDr-CXR database.

Supervised learning can enhance the reliability of lesion analysis in independent learning scenarios. However, in instances where the number of annotated CXRs is limited compared to publicly available databases and for rare pathologies in query lesions, unsupervised learning has demonstrated its superiority.

Chapter 5

Experiments

The sections below describe the experiments that were performed in this project as well as the obtained results and the conclusions drawn. These experiences aim to comprehend the behaviour of the LXIR framework and to test the performance of the system when the retrieval method suffers some changes.

5.1 Generalisation and Robustness of the Model

The LXIR framework initially employed the VinDr-CXR database. For the purposes of training, validation, and testing, this database is smaller in size when compared to larger medical databases such as the ChestX-ray14 database.

The utilisation of a small database in a CBIR system can cause several drawbacks. Low data diversity within the database may result in an incomplete representation of lesions and their classes, which may lead to biased model training. Consequently, the CBIR system may exhibit reduced robustness and heightened sensitivity to minor variations and perturbations in the input data. Moreover, small databases are more prone to noise and outliers, which can detrimentally affect the training process and subsequent retrieval performance of the CBIR model. Furthermore, the feature representation of lesions may prove inadequate due to the limited number of examples characterising each class within the small database. Therefore, the model may fail to capture the underlying distribution of features adequately. Additionally, the evaluation of the CBIR system may not reflect its true performance, as the small size of the validation and test datasets may limit its representativeness and realism. In the context of CBIR systems, the utilisation of small databases imposes constraints on the search process for identifying analogous lesions to the given query. This limitation results in an inaccurate assessment of the CBIR system's performance.

In this experiment, the model is tested in a bigger and more diverse database, the ChestX-ray14 database in order to get a more comprehensive and reliable acknowledgement of the performance of the model, its ability to generalisation, its robustness and scalability.

The focus in this experiment was exclusively directed towards the Parenchymal Lesion (PaL) class when considering queries and search collection images. The decision to concentrate solely

on the PaL class stems from its high representativeness within the database employed. This class comprises a total of 8 subclasses, which ensures a comprehensive coverage of different lesion types, thereby facilitating a targeted analysis of the model's performance and its ability to retrieve lesions accurately within this specific context. By narrowing the investigation to the PaL class, potential biases arising from a class imbalance in the database were effectively mitigated. Class imbalance refers to a scenario where the number of instances varies significantly across different classes, potentially leading to uneven representation. Given that class imbalance is a common phenomenon encountered in medical image databases, the decision to focus on the PaL class contributes to a more balanced representation of subclasses, thereby reducing the potential impact of skewed class distributions.

This approach ensures a more reliable and robust evaluation of the model's generalisation and retrieval performance within the specific context of parenchymal lesions.

The query set used in this experiment was derived from the test set of the VinDr-CXR database. In this analysis, the occurrences were recorded to determine the frequency with which the top three retrieved results originated from the search collection composed of either the train and validation sets of the VinDr-CXR database or the train and validation sets of the ChestX-ray14 database when the distance metric applied was the L_2 norm.

Specifically, the number of times that the first, second, and third retrievals belonged to the two search collections was tallied. This approach allowed for a comparative evaluation of the retrieval performance and the respective contributions of the VinDr-CXR and ChestX-ray14 databases in producing the top-ranked results.

In order to complement this experiment and to assess the impact of database size and diversity on the representativeness of lesion retrieval in the CBIR system, an investigation about whether a larger and more diverse search collection improved the similarity and visual correspondence between query lesions and retrieved lesions was initiated.

To accomplish the aim of this analysis, the query set was compared against the two distinct search collections of CXR images. For each query, two retrieval scenarios were generated, one for each search collection, accompanied by the corresponding distances from the query.

Additionally, this complementary study aimed to investigate whether query lesions that present a larger distance to the retrieved lesions in the VinDr-CXR search collection could be better represented. The objective was to explore if lesions that are not visually similar or have a greater dissimilarity in feature space could potentially benefit from a larger and more diverse search collection, such as the ChestX-ray14 database. By comparing the retrieval scenarios and considering the distances between the query lesions and the retrieved lesions, the experiment aspired to determine if a more extensive search collection could provide closer similar retrieved lesions to rare query lesions, thus enhancing the overall retrieval performance.

To conduct this investigation, one of the used procedures was to analyse the examples where the retrieval distances between the retrieved lesions using the VinDr-CXR or the ChestX-ray14 were the largest. This first approach also called *The Largest Difference Approach*, aimed to maximise the contrast between the two search collections by identifying retrieval scenarios where the

ChestX-ray14 search collection had lower distance values than the VinDr-CXR search collection.

By evaluating the retrieval distances, the experiment aspired to determine the extent to which the ChestX-ray14 search collection outperformed the VinDr-CXR search collection in terms of distance values and visual similarity. Theoretically, when first comparing the query to the VinDr-CXR search collection and then comparing it to the ChestX-ray14 search collection, a larger decrease in distance values indicates that the ChestX-ray14 search collection yielded a retrieved lesion that was visually more similar to the query lesion compared to the VinDr-CXR search collection retrieved lesion. By focusing on maximising the difference in distance values, this procedure aimed to demonstrate the potential benefits of utilising a larger and more diverse search collection in achieving better visual similarity and retrieval performance.

In addition to assessing the difference in distance values between the two search collections, in a second approach, the largest retrieval distances in the VinDr-CXR search collection were identified and compared to the distances obtained for the same query in the ChestX-ray14 search collection. By examining the largest retrieval distances in the VinDr-CXR search collection, the second procedure, also named *The Largest Distance Approach*, aimed to identify cases where the model struggled to find relevant and visually similar lesions. These instances represent challenging scenarios for the CBIR system, particularly for rare lesions that may have limited representation in the training data.

The comparison of the distances for the same query in the ChestX-ray14 search collection provided insights into the potential improvement in retrieval performance when utilising a larger and more diverse database.

Yet, the visual similarity was evaluated by an untrained person, the author of the thesis, which may indicate that the evaluation may not be optimal or reflective of professional judgement.

5.1.1 Results and Discussion

Given that the FR of the lesions in this study can have one of three possible sizes, namely 640, 1280, or 2560, each table presents information for all the sizes. This approach enables a comprehensive and organised presentation of the FR data, facilitating a detailed analysis and comparison of the results based on the specific FR size employed.

Firstly, Table 5.1 provides an overview of the dataset sizes for the train and validation sets derived from the VinDr-CXR database, constituting one of the search collections. Additionally, this table displays the dataset sizes for the train and validation sets sourced from the ChestX-ray14 database, comprising the other search collection. Furthermore, to facilitate comprehensive comparison and analysis, the quantity of data encompassed within the query/test set of the VinDr-CXR database is also included.

Table 5.1: Dataset size analysis of the Train and Validation Sets in the VinDr-CXR Database, and Query/Test Set for all Lesion Feature Representations sizes.

FR Size	VinDr-CXR	# Data	ChestX-ray14	# Data
640	Train Set	35916	Train Set	241528
	Validation Set	9012	Validation Set	85042
	Total	44928	Total	326570
	Test Set	751	Test Set	-
1280	Train Set	53541	Train Set	494181
	Validation Set	15234	Validation Set	165867
	Total	68775	Total	660048
	Test Set	870	Test Set	-
2560	Train Set	31890	Train Set	316678
	Validation Set	9557	Validation Set	104637
	Total	41447	Total	421315
	Test Set	855	Test Set	-

Table 5.2 focuses on the experimental results pertaining to the precise number of features used for lesion representation. It offers a consolidated summary of the outcomes achieved in the experiment, allowing for a detailed evaluation and comparison of the findings.

Table 5.2: Practical results of experiment 1 using feature representations with different sizes.

FR Size	Database	1^{st} Retrieval		2^{nd} Retrieval		3^{rd} Retrieval	
640	VinDr-CXR	209	(28%)	187	(25%)	197	(26%)
	ChestX-ray14	**542**	(**72%**)	**564**	(**75%**)	**554**	(**74%**)
1280	VinDr-CXR	161	(19%)	137	(16%)	152	(17%)
	ChestX-ray14	**709**	(**81%**)	**733**	(**84%**)	**718**	(**83%**)
2560	VinDr-CXR	177	(21%)	145	(17%)	142	(17%)
	ChestX-ray14	**678**	(**79%**)	**710**	(**83%**)	**713**	(**83%**)

Through the documentation of the retrieval distribution across the two search collections, it is evident that the system demonstrates a higher number of first, second, and third retrievals within the search collection comprised of the train and validation sets from the ChestX-ray14 database.

These findings provide evidence that the model has the capability to generalise its learned knowledge from the training data to unseen data. By successfully retrieving lesions from the ChestX-ray14 database, which contains different images and potentially different types of abnormalities, the model exhibits the ability to apply its learned features and patterns to diverse datasets. The fact that the model can effectively retrieve lesions from a distinct database indicates its capacity to adapt to different data domains and handle variations in the CXR images, like another kind of image acquisition, patient demographics or labelling protocols.

Given the larger number of lesions present in the ChestX-ray14 database, it is expected that the CBIR system would yield a higher count of retrievals from this search collection. The increased lesion diversity and quantity in ChestX-ray14 naturally leads to a higher likelihood of matching

query lesions with relevant instances within this search collection. Consequently, the retrieval results tend to be more abundant in comparison to the VinDr-CXR search collection, which contains a comparatively smaller number of lesions.

Also, the higher quantity of retrievals in the ChestX-ray14 search collection indicates that the model is not overly biased towards the VinDr-CXR data used for training, which reveals that the model's learned representations are more generalised and not restricted to specific features or patterns present only in the training set. By retrieving lesions from an alternative database, the model demonstrates a reduced bias and broader applicability in real-world scenarios.

Overall, the model can be considered more robust due to its ability to effectively retrieve lesions from the ChestX-ray14 search collection, which is a larger and distinct database from the VinDr-CXR database used for training.

In figures 5.1 and 5.2, two illustrative examples are provided to demonstrate the retrieval process in the CBIR system using a FR size of 640 features. The CBIR system is employed to find similar lesions in two distinct search collections: one constituted of the train and validation sets from the VinDr-CXR database, and the other composed of the train and validation sets from the ChestX-ray14 database.

In figure 5.1, *The Largest Difference Approach* was applied to identify the retrieval with the largest distance difference between the query lesion and the corresponding retrieved lesion from each search collection. The figure presents the first retrieved lesions from both search collections that exhibit the maximum distance difference. Figure 5.2, on the other hand, showcases an example of *The Largest Distance Approach*. Here, the figure illustrates the second-largest retrieval distance to a query lesion from the VinDr-CXR search collection and the respective retrieved lesion from the ChestX-ray14 search collection. This is because figure 5.1 already displayed the retrieval with the largest distance in the VinDr-CXR search collection.

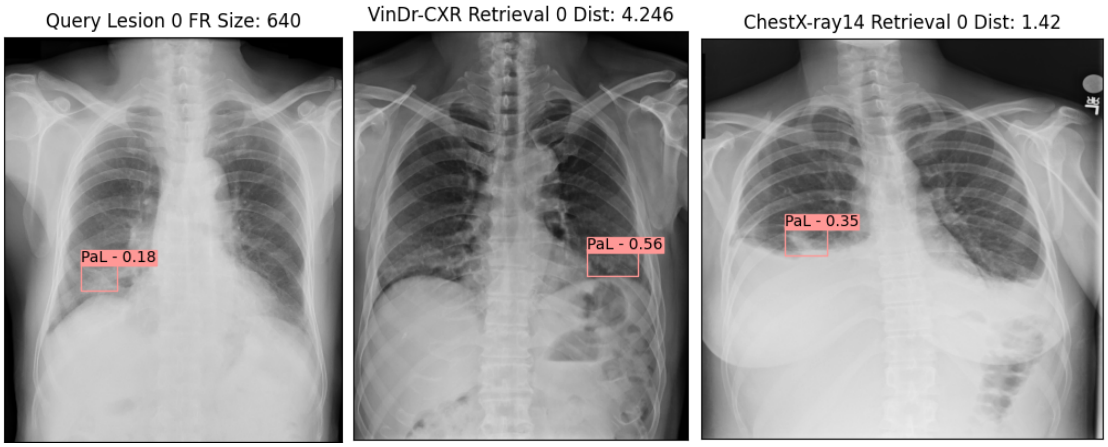


Figure 5.1: Retrieved lesions from different search collections when a query lesion with a FR size of 640 features is tested. In this case, it was applied *The Largest Difference Approach*.

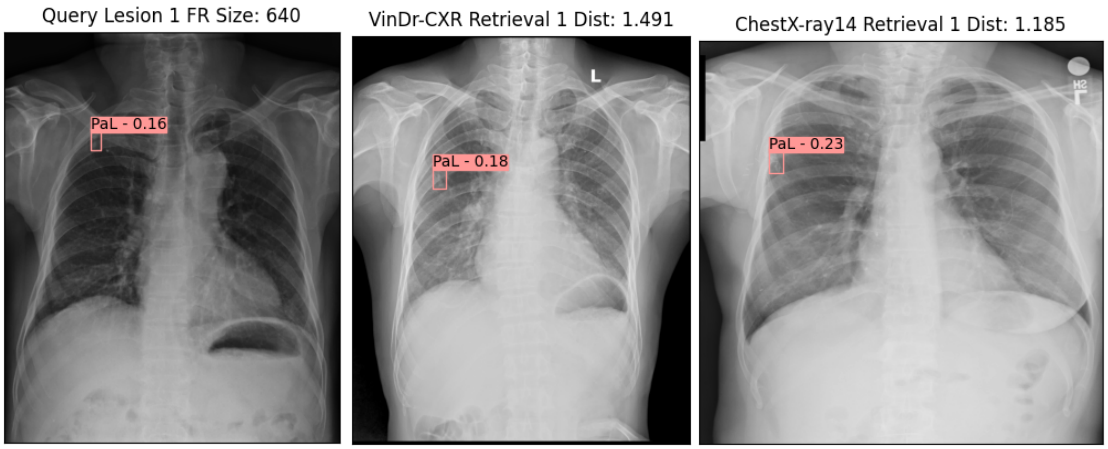


Figure 5.2: Retrieved lesions from different search collections when a query lesion with a FR size of 640 features is tested. In this case, *The Largest Distance Approach* was used.

In the same way, figure 5.3 and figure 5.4 show two examples of a query lesion and the first retrieved lesion belonging to the search collection constituted by train and validation sets of the VinDr-CXR database, and first retrieved lesion belonging to the other search collection constituted by train and validation sets of the ChestX-ray14 database, when the FR of the lesions is composed of 1280 features.

In figure 5.3 *The Largest Difference Approach* was the one applied. Figure 5.4 shows an example of when *The Largest Distance Approach* is applied. The figure reveals one of the largest VinDr-CXR retrieval distances to a query and the respective retrieved lesion from the ChestX-ray14 search collection.

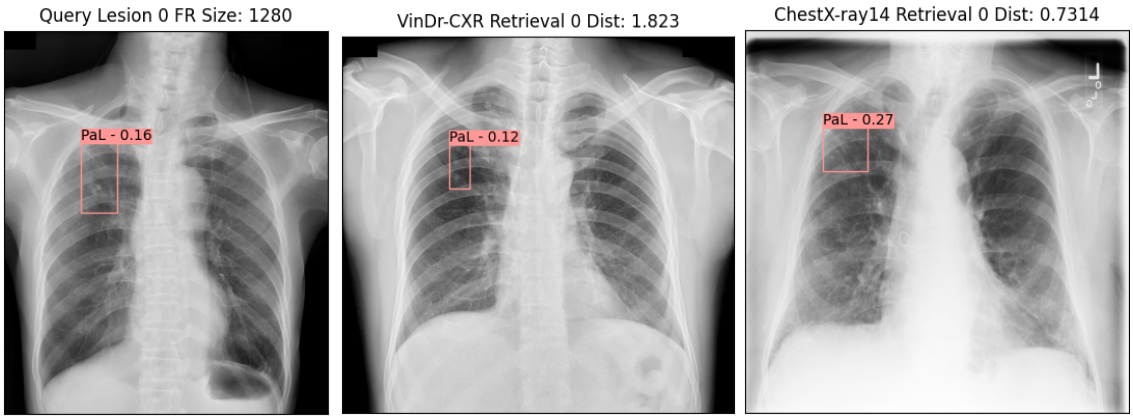


Figure 5.3: Retrieved lesions from different search collections when a query lesion with a FR size of 1280 features is tested. In this case, it was applied *The Largest Difference Approach*.

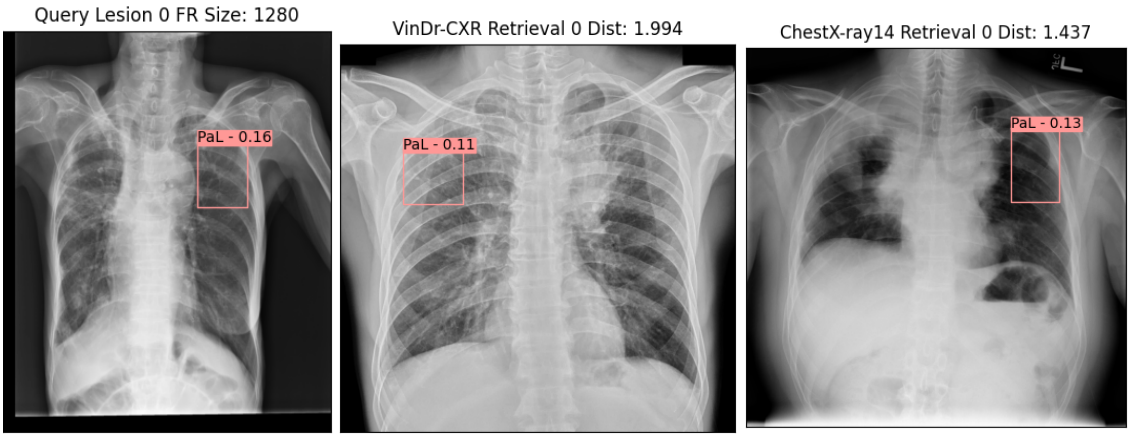


Figure 5.4: Retrieved lesions from different search collections when a query lesion with a FR size of 1280 features is tested. In this case, *The Largest Distance Approach* was used.

Lastly, two retrieval examples of when the FR of the lesions is composed of 2560 features are presented. Figure 5.5 and figure 5.6 possess, each one of them, a query lesion and the respective first retrieved lesions, one belonging to the search collection constituted by the train and validation set of the VinDr-CXR database and the other belonging to the search collection constituted by the train and validation set of the ChestX-ray14 database. In figure 5.5 *The Largest Difference Approach* was the one applied. Figure 5.6 shows an example of when *The Largest Distance Approach* is applied. The figure reveals the second largest VinDr-CXR retrieval distance to a query and the respective retrieved lesion from the ChestX-ray14 search collection as figure 5.5 already shows the retrieval with the retrieved lesion from the VinDr-CXR search collection with the largest distance to the query.

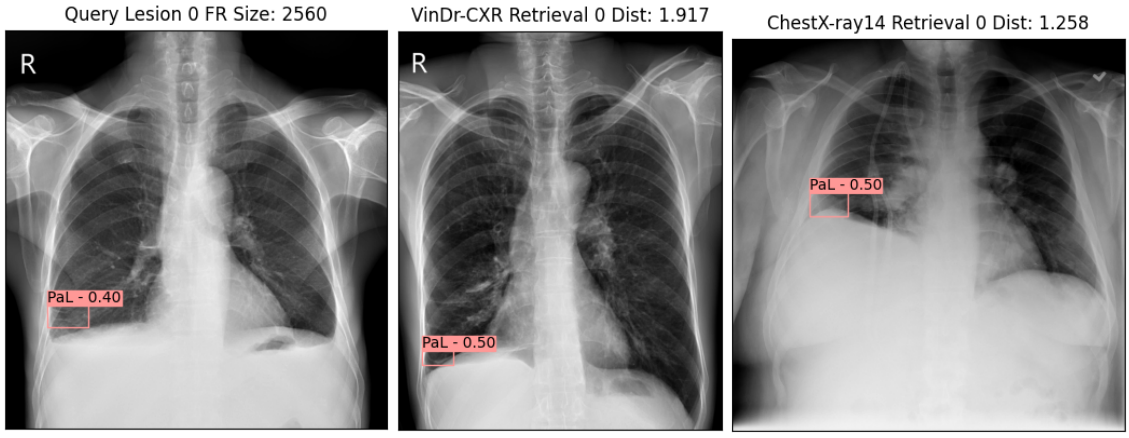


Figure 5.5: Retrieved lesions from different search collections when a query lesion with a FR size of 2560 features is tested. In this case, it was applied *The Largest Difference Approach*.

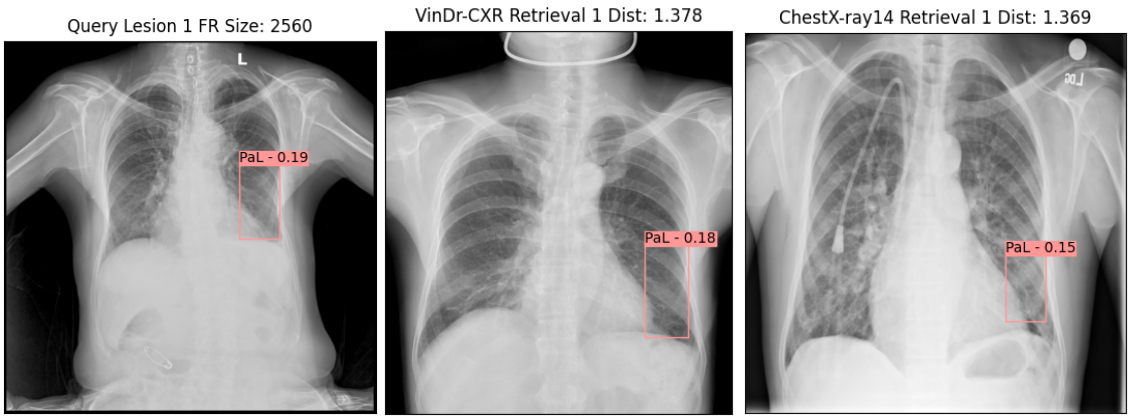


Figure 5.6: Retrieved lesions from different search collections when a query lesion with a FR size of 2560 features is tested. In this case, *The Largest Distance Approach* was used.

Upon observing the examples, it becomes evident that in certain cases, the difference in distance values between the search collections is significantly high and the visual similarity improves when the retrieval system uses the ChestX-ray14 search collection, as for example, the case presented in figure 5.1. However, it is also noticeable that in some cases, like figure 5.5, even though the distance to the query decreased, the retrieved lesion from the ChestX-ray14 search collection does not appear to be more similar to the query compared to the retrieved lesion from the VinDr-CXR search collection.

Also, by expanding the size of the search collection, the model found lesions spatially closer to the location of the query lesion in other CXR images which indicates an improvement in retrieval accuracy and localisation precision. This finding supports the notion that a more extensive and diverse search collection contributes significantly to improving the system's generalisation and retrieval performance, making it more effective in real-world medical image analysis and diagnosis.

5.2 Dissimilarity between Classes

The main objective of this experiment is to assess the effectiveness of the utilised distance metric in separating lesion classes. Since lesions from different classes are expected to exhibit distinct appearances, lesions from other classes shouldn't be retrieved when a query lesion from one class is tested, as their distances to the query lesion are likely to be greater. By eliminating lesions from other classes during the retrieval stage, the computational workload is expected to decrease. The advantages of this approach include the potential reduction in computational requirements due to the exclusion of irrelevant lesions during retrieval.

By focusing solely on queries from the PaL class and incorporating data from all classes in the train and validation sets of the VinDr-CXR database into the search collection, a comprehensive evaluation of the distance metric's capability to discriminate between lesion classes can be conducted.

The effectiveness of the distance metric in class separation relies heavily on the quality of the features and the appropriateness of the metric itself. If the features fail to adequately capture the variations across different lesion classes, the performance of the distance metric may be compromised. The distance metric that was used was the Euclidean distance.

Ideally, the objective of this experiment is to identify a discriminative threshold or boundary that effectively separates the search collection composed only of the PaL class from the other search collections composed of the other classes. This threshold or boundary would enable accurate retrieval of queries from the PaL class of the test set in the VinDr-CXR database when compared against diverse search collections composed of distinct classes from the train and validation sets of the same database.

To facilitate this analysis, histograms were generated for each specific feature representation (FR) size utilised in the experiment, namely 640, 1280, and 2560. These histograms provide a graphical representation of the distribution patterns exhibited by the relevant retrieval distances. Furthermore, they serve as a valuable tool for assessing and contrasting the performance of query retrieval across different FR sizes within the experimental framework.

The focus was also on exploring how the change in FR size affected the distribution of distances between queries and the retrieved lesions.

It is noteworthy that the normalisation procedure applied during this experiment aimed to standardise the distances across all classes, utilising the mean and standard deviation values of class 7, the PaL class, on all lesions FR of all search collections.

Furthermore, in order to understand the variation of the inter-class distance values, it is shown in each histogram, the Kernel Density Estimation (KDE) curve for each class.

5.2.1 Results and Discussion

The train and validation sets of class 3, denoted as the Cmg class, did not contain any lesions or instances suitable for representation by the chosen FR size of 640 features. As a consequence of this absence, the histogram presented in Figure 5.7 focused on the remaining five search collections.

Upon observation of Figure 5.7, it becomes evident that the distances between the queries of the PaL class and the Pnm class (class 12) search collection were notably higher, exhibiting distance values around 30. Despite applying normalisation techniques, the distances for queries of the PaL class and the Pnm class remained significantly greater, indicating substantial differences in feature representations and visual similarities between these classes.

In order to facilitate the observation of the distance distribution in the other search collections, figure 5.8 was created, where the Pnm class (class 12) was excluded.

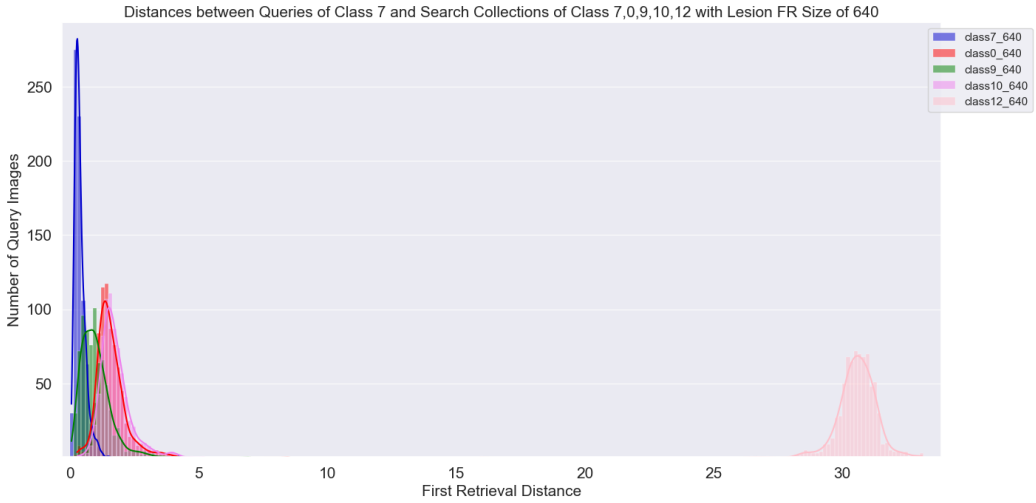


Figure 5.7: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances when the FR of the lesions contains 640 features. The step between the bins is 0.15. See the example presented in figure A.1 of appendix A.

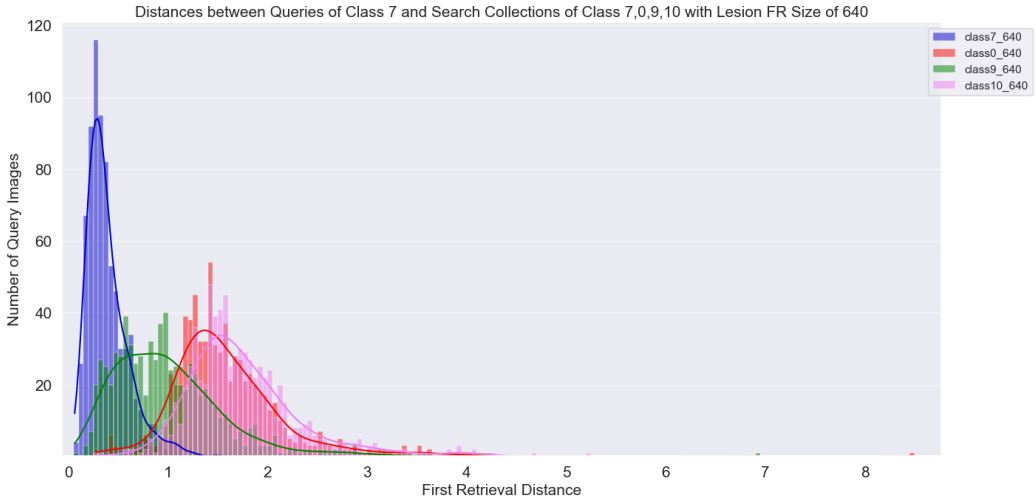


Figure 5.8: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances when the FR of the lesions contains 640 features. The step between the bins is 0.05.

However, upon examining the histograms presented in Figure 5.9 and Figure 5.10, it becomes apparent that the phenomenon observed in the Pnm class in the FR size of 640 features did not persist as the FR size increased to 1280 and 2560 features. The larger FR sizes allowed for a more detailed and informative representation of the lesions. Consequently, the model could better capture similarities and visual correspondences between queries of the PaL class and the Pnm class search collection, reducing the disparity in distances observed in the FR size of 640 features.

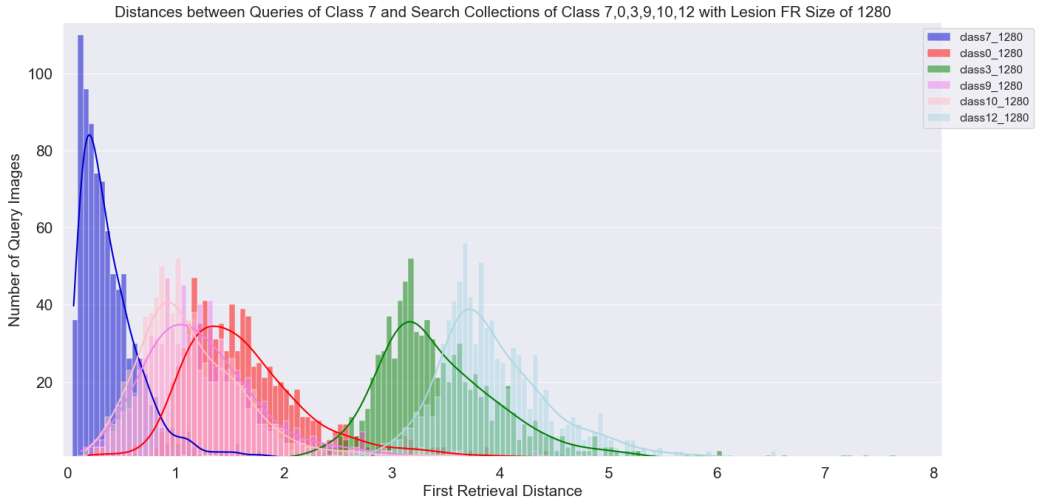


Figure 5.9: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances when the FR of the lesions contains 1280 features. The step between the bins is 0.05. See the example presented in figure A.2 of appendix A.

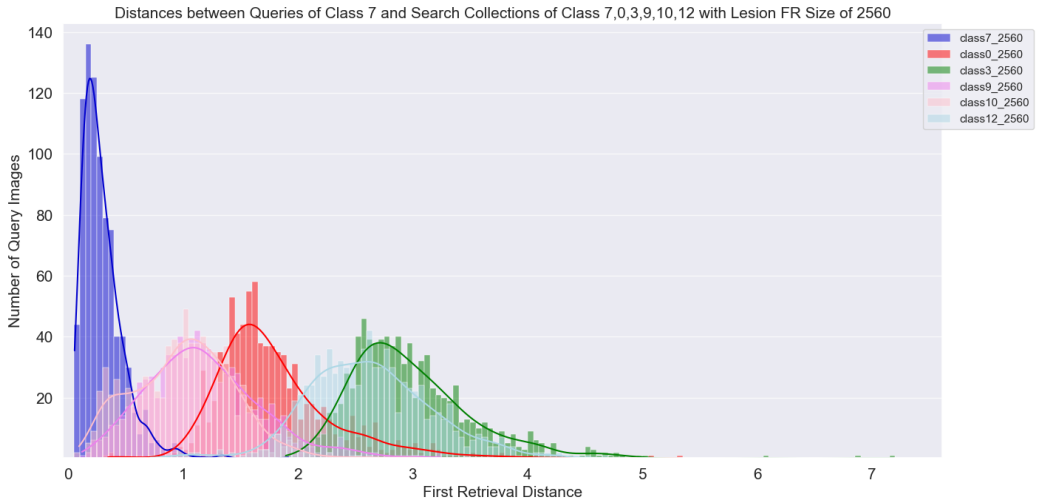


Figure 5.10: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances when the FR of the lesions contains 2560 features. The step between the bins is 0.05. See the example presented in figure A.3 of appendix A.

When comparing figure 5.9 with figure 5.10, it can be observed that the order of the first four retrieved search collections to appear was consistent between the two histograms, suggesting that the ranking of the search collections remained relatively stable when increasing the FR size. However in the histogram with the FR size of 2560, figure 5.10, the second class to appear, the PlL class (class 10), exhibited a higher quantity of lower distances near the peak of the PaL KDE curve. This characteristic suggests that there might be a significant overlap between the PlL class and the

PaL class in the FR size of 2560, which could lead to challenges in effectively distinguishing between these two classes based on the distances from the retrieved lesions to the query and leading to worse discrimination of the PlL class. This outcome emphasises the impact of feature sizes on retrieval performance in the CBIR system.

By analysing all three histograms together, excluding class 12, shown in figure 5.11, it is possible to observe that across all FR sizes, the search collection formed by PaL data consistently appears as the closest retrieved class, indicating that the L_2 norm performs well in correctly identifying PaL lesions.

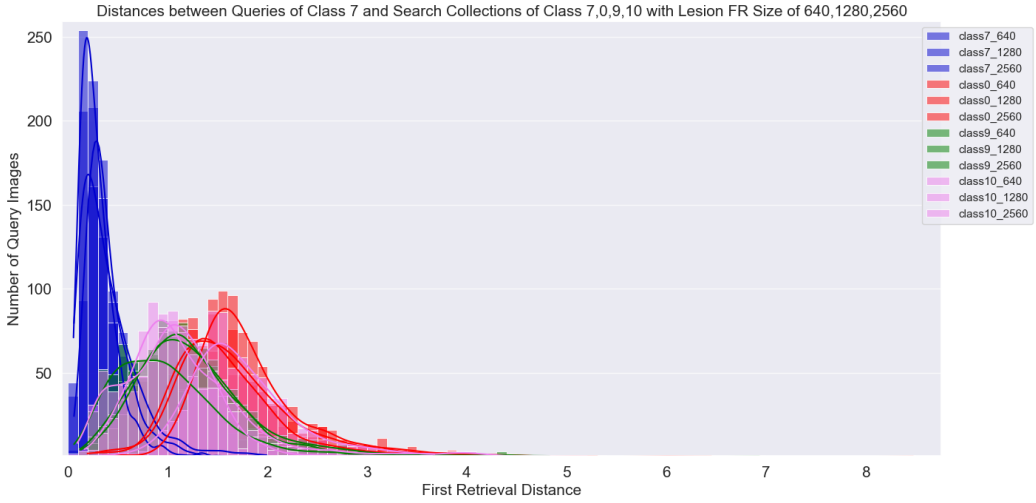


Figure 5.11: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances for all the lesion FR sizes. The step between the bins is 0.10.

The second closest class to be retrieved is from the histogram with FR size of 640, the OtL class, class 9. This suggests that with this specific FR size, the L_2 norm may not be as effective in distinguishing the PaL class from other classes. If the defined distance threshold is too close to the KDE peak of the retrievals from the PaL class, it might lead to a lower threshold value, making it challenging to separate PaL retrieved lesions from other classes.

The determination of an optimal distance threshold plays a crucial role in fine-tuning the retrieval performance of the CBIR system. Setting a higher threshold can result in a larger overlap between the retrieved classes, potentially leading to difficulties in effectively distinguishing between them. Also, adopting a lower threshold may discriminate against a greater number of retrieved lesions from the PaL class and can inadvertently exclude relevant retrieved lesions from this class that exhibit visual similarity to the query. Hence, the selection of an appropriate threshold value assumes critical importance in enhancing the CBIR system's capability to accurately discern and discriminate between different lesion classes, ultimately contributing to the improvement of retrieval accuracy

Based on the histograms, it appears that the Euclidean distance metric performs better in separating PaL retrieved lesions from PlL retrieved lesions when the FR size is 1280 as the distance

between the two classes is more distinct compared to the FR size of 640 and the FR size of 2560.

By visually inspecting each histogram individually it is possible to define a threshold that discriminates the other search collections that do not comprehend the class of the query. As a way of defining a threshold, by observing the distribution of the distances associated with the first two search collection classes to be retrieved, a distance threshold can be drawn. So, for each lesion FR size: 640, 1280, and 2560 the threshold identified was 0.2, 0.45 and 0.2 respectively.

Further investigation and fine-tuning of the distance metric and threshold settings can help refine the model's performance and achieve more accurate and reliable lesion retrievals. However in order to evaluate if the suggested thresholds are a good tool to discriminate the retrievals that do not represent similar lesions to the query, these retrievals should be analysed by specialists in the area.

5.3 Dissimilarity between Subclasses of the PaL Class

In the previous experiment, the primary objective was to conduct a comprehensive analysis of the CBIR system's overall capacity to differentiate between lesions from different classes during image retrieval. In contrast, the current experiment takes a more specific and in-depth approach by assessing the CBIR system's discriminative performance within lesions originating from the same class. Both experiments play an important role in elucidating the inherent strengths and limitations of the CBIR system, thereby offering valuable insights into enhancing retrieval accuracy and discriminative prowess. The combination of both experiments enables researchers to obtain a holistic comprehension of the CBIR system's behaviour across varying levels of granularity in lesion classification, thereby promoting knowledgeable decisions for system optimisation and improvement. In this experiment, the distance metric applied was also the Euclidean distance.

Through this experiment, it is possible to determine the dissimilarity between subclasses within the PaL class by comparing the minimum distance when a query lesion is matched against a search collection consisting of lesions from the same subclass versus a search collection comprising lesions from different subclasses. Table 4.2 displays the subclasses of the PaL class, encompassing the following categories: Atl, Clc, Cns, ILD, Inf, LOp, Nod, and PuF.

The analysis of dissimilarity between subclasses within the PaL class provides insights into the performance of the supervised CBIR system. By comparing the minimum distance between query lesions and search collections, it is possible to assess the impact of subclass variations on the retrieval results. This examination allows the evaluation of whether lesions from the same subclass exhibit greater similarity and thus yield better retrieval performance compared to lesions from different subclasses.

By explicitly focusing on subclasses within the PaL class, this experiment provides a more granular understanding of the dissimilarity between lesions, enabling more accurate retrieval and subclass-specific analysis. Utilising annotated data from expert radiologists ensures the reliability and accuracy of the CBIR system, as the training and evaluation process is guided by expert knowledge.

A supervised CBIR system is employed in this experiment leveraging annotated data exclusively provided by radiologists. The matching criterion utilised to evaluate the similarity between predicted bounding boxes and radiologists' bounding boxes is based on an IoU threshold of 0.4.

However, the reliance on radiologists' annotations introduces the possibility of subjective variations in lesion classification and bounding box annotations, which may affect the accuracy of the matching criterion and subsequent retrieval results.

Thus, this experiment aims to investigate the dissimilarity between subclasses within the PaL class in a supervised CBIR system. By comparing the minimum distance between query lesions and search collections, this analysis sheds light on the impact of subclass variations on retrieval performance. Although potential biases and limited generalisation should be considered, this approach provides valuable insights into subclass-level dissimilarity.

The generated histograms illustrate the fluctuations observed in distances between queries and two distinct search collections. The first search collection comprises lesions belonging to the same subclass as the queries, while the second search collection encompasses lesions from all other subclasses within the PaL class. Notably, these search collections are composed of data extracted from both the train and validation sets of the VinDr-CXR database. The queries of each subclass belong to the test set of the VinDr-CXR database.

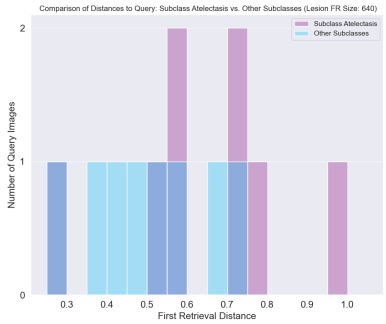
To gain a more direct understanding of the density of cases for each distance value, KDE curves were applied to the subclass distributions with a higher data quantity.

5.3.1 Results and Discussion

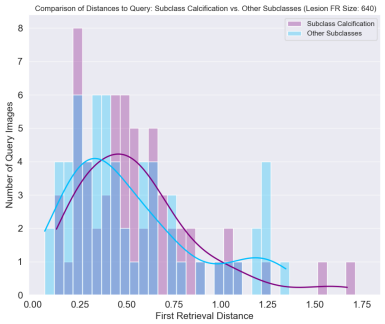
Figure 5.12, figure 5.13 and figure 5.14 present the histograms for each subclass when the lesions FR possesses a specific size, respectively, 640, 1280 and 2560 features.

Upon careful observation of the figures presented, it becomes evident that the number of queries for each subclass is relatively limited, which may potentially impact the statistical significance of the obtained results. Notably, a discernible trend emerges, indicating a substantial mixture of the two search collections. The KDE curve peaks and the presence of dark blue bins (indicative of the combination of search collection colours) in each histogram demonstrate a substantial overlap between the two search collections.

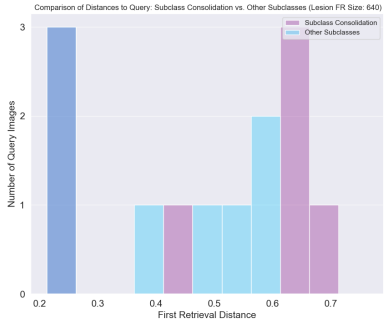
Furthermore, a general pattern is discernible, wherein the retrieved lesions that closely match the query lesions tend to belong to different subclasses rather than originating from the same subclass. This may occur because of the quantity of data that represents each subclass. Even the lesions presented in each subclass may not always be equal, so the method may identify closer characteristics in lesions from the other subclasses.



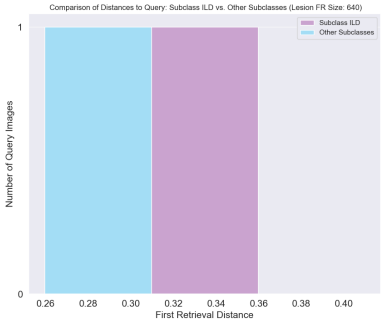
(a) Atelectasis



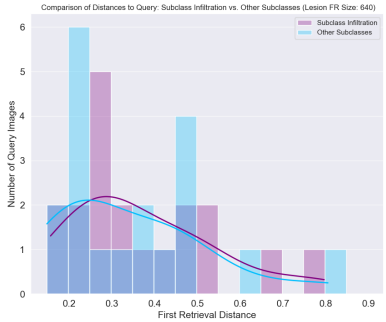
(b) Calcification



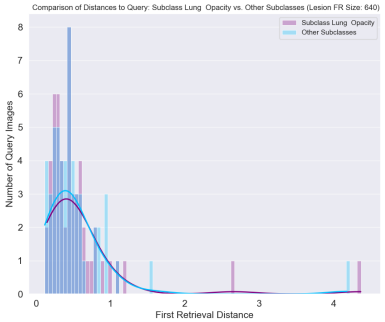
(c) Consolidation



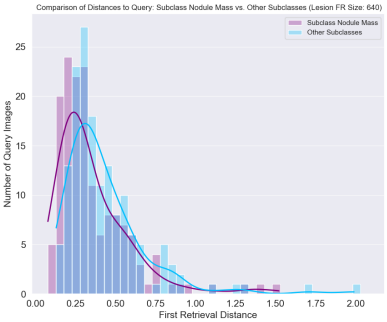
(d) ILD



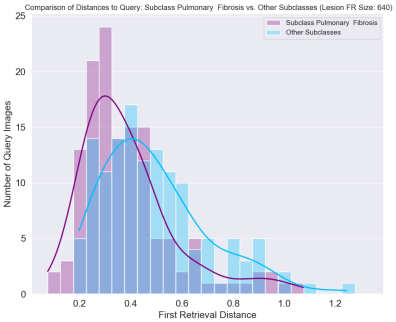
(e) Infiltration



(f) Lung Opacity

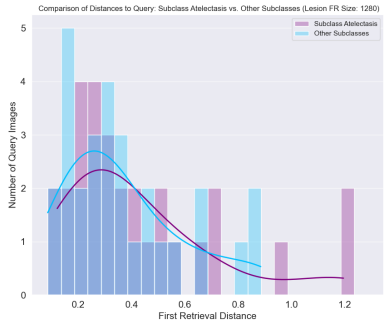


(g) Nodule/Mass

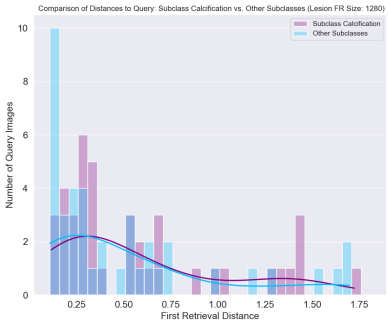


(h) Pulmonary Fibrosis

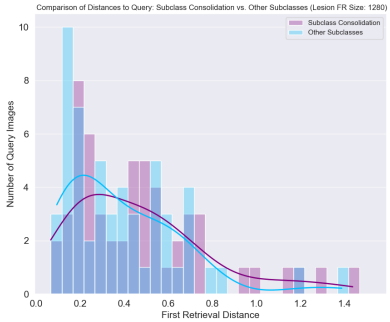
Figure 5.12: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances for each subclass of the PaL Class when the lesions FR size is 640. Appendix B shows a retrieval example for each subclass, from figure B.1 to B.8.



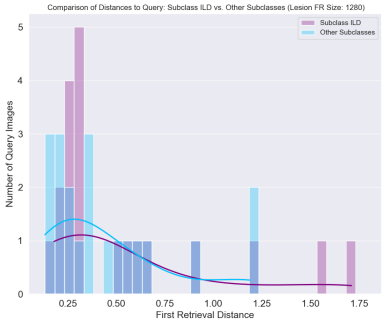
(a) Atelectasis



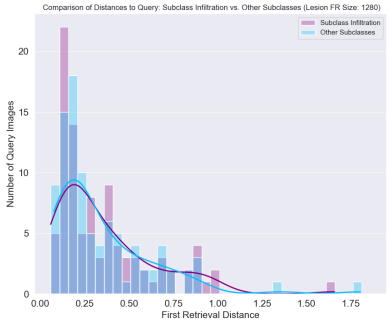
(b) Calcification



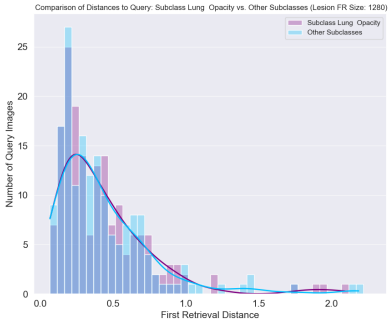
(c) Consolidation



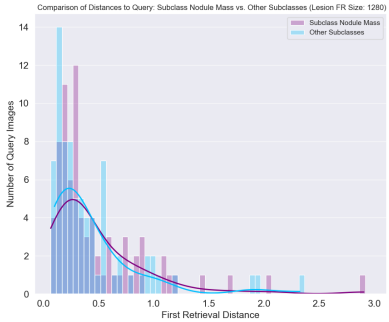
(d) ILD



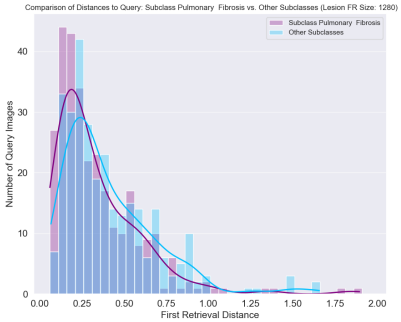
(e) Infiltration



(f) Lung Opacity

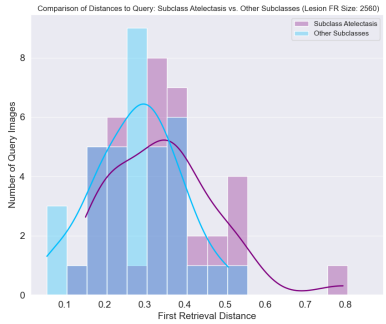


(g) Nodule/Mass

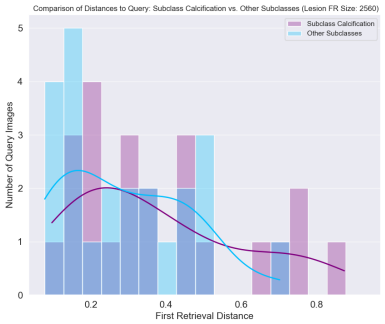


(h) Pulmonary Fibrosis

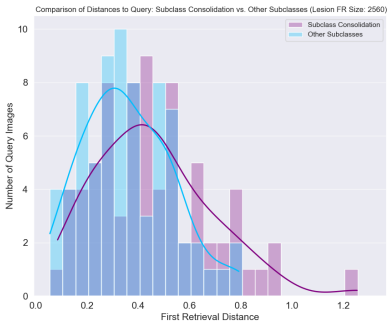
Figure 5.13: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances for each subclass of the PaL Class when the lesions FR size is 1280. Appendix B shows a retrieval example for each subclass, from figure B.9 to B.16.



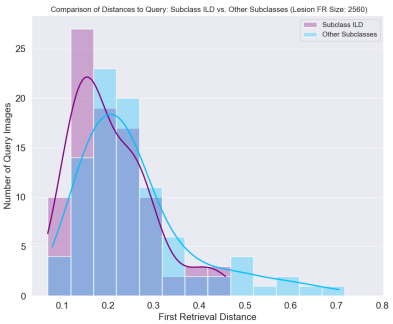
(a) Atelectasis



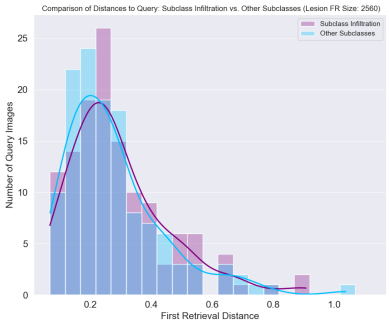
(b) Calcification



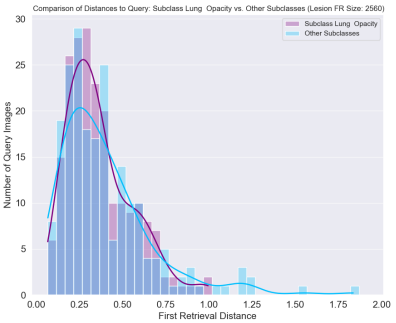
(c) Consolidation



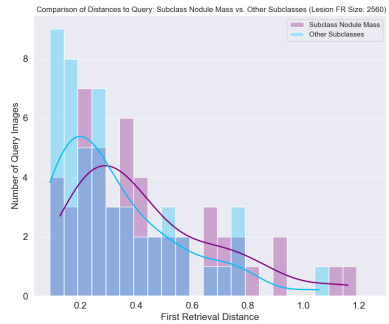
(d) ILD



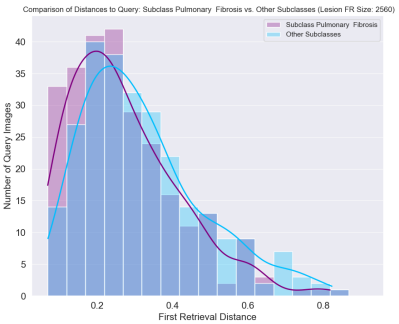
(e) Infiltration



(f) Lung Opacity



(g) Nodule/Mass



(h) Pulmonary Fibrosis

Figure 5.14: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances for each subclass of the PaL Class when the lesions FR size is 2560. Appendix B shows a retrieval example for each subclass, from figure B.17 to B.24.

It is important to note that these findings may also be influenced by the relatively low number of queries per subclass. Consequently, caution must be exercised in interpreting the results and drawing definitive conclusions. The presence of overlapping search collections and the difficulty in distinguishing between closely related subclasses underscore the need for further investigation. Also, efforts to augment the dataset with more annotated data for each subclass and refine the feature representation may contribute to enhancing the system's performance and addressing the challenges observed in accurately retrieving lesions from the same subclass.

Figure 5.15 provides a comprehensive summary of the experiment's findings, emphasising the recurring trend of substantial overlap among the retrieved lesions from different search collections. The graph highlights the challenges faced by the CBIR system in accurately distinguishing between subclasses of lesions within the PaL class.

One crucial aspect to bear in mind is the relatively low number of queries available for each subclass. The limited number of queries may impact the statistical significance of the observed results, potentially introducing variability and influencing the overall performance assessment of the CBIR system.

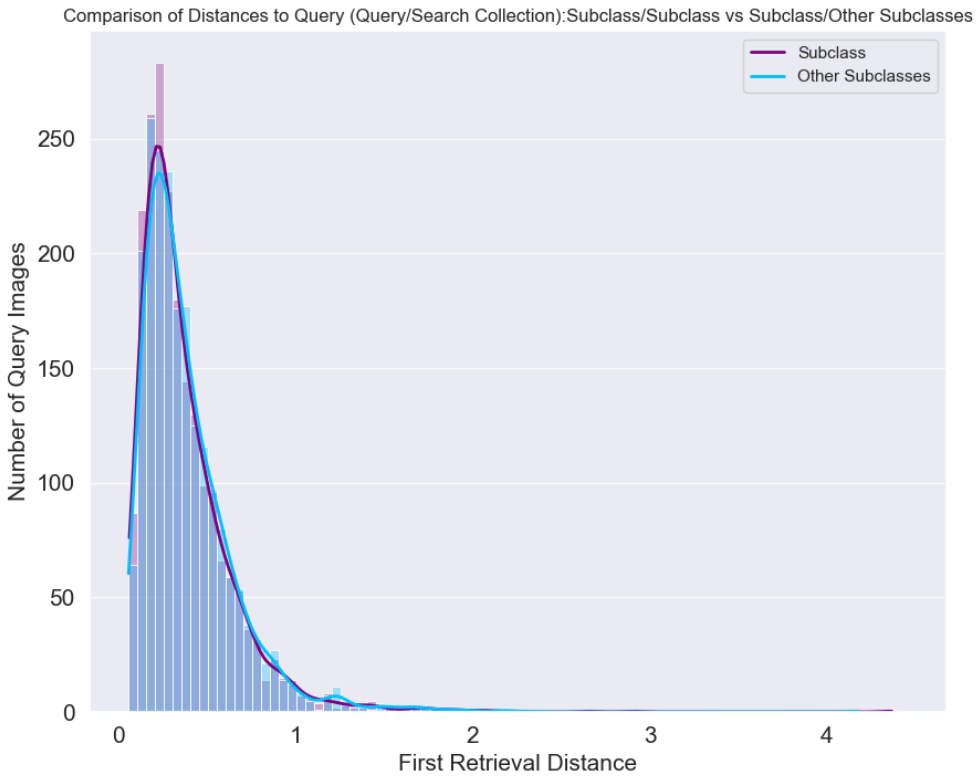


Figure 5.15: Graphical representation of the distribution patterns exhibited by the relevant retrieval distances.

Indeed, the observed results in the experiment may be attributed to the confusion or ambiguity in lesion annotations provided by radiologists. The presence of lesions with multiple annotated classes, such as a calcified nodule annotated with both "Nodule/Mass" and "Calcification" may introduce challenges and complexities in accurately determining the true class of each lesion.

The overlapping annotations may lead to confusion in the distributions of lesions within different subclasses, potentially causing the CBIR system to struggle in precisely differentiating between subclasses. This confusion could result in the substantial overlap observed in the retrieved lesions from different search collections, as illustrated in Figure 5.15. The presence of multiple annotations for a single lesion may introduce uncertainty and impede the CBIR system's ability to assign a precise class label, impacting the overall retrieval performance and subclass discrimination. Addressing the issue of overlapping annotations and resolving the confusion between classes in lesion annotations is crucial to improve the reliability and accuracy of the CBIR system. Despite these reviews, the experiment provides valuable insights into the challenges and potential areas of improvement for subclass-specific medical image retrieval.

Chapter 6

Conclusions and Future Work

CBIR is a vast family of methods that optimise the process of searching for similar images in massive databases. As medical images demand a more detailed and fine-grained analysis, in large-scale image repositories of this domain, the employment of traditional CBIR methods may not be the most suitable. New opportunities to innovate traditional medical retrieval systems have emerged as the size of medical image databases increased and new ramifications in this area started to make use of automated systems to support the experts' decisions.

In large-scale datasets, the creation of very sparse spaces, the reduction of the dimension of the feature vectors and the improvement of the adopted strategies for similarity search or data indexing are topics that may be explored for efficient medical image retrieval. Semi-supervised and unsupervised methods are preferred when image annotations are difficult to obtain. Non-linear methods may yield better retrieval performance for challenging image differentiation, although training such hashing functions may be more time-consuming than linear ones. Efficiency measures are essential in medical retrieval systems as they serve as indicators of laborious tasks, such as feature indexing.

The previous studies in CXR image retrieval described in section 3.4 presented some limitations, including the size of the databases used and the lack of explanations for pathological locations. Additionally, some queries had a low number of relevant retrievals due to limited data representation of rare lesions and classification methods that did not allow their inclusion in specific classes.

This work aimed to address these limitations and conduct a deeper analysis of the CBIR system applied to a CXR image database, the LXIR system, through the realization of experiments. By augmenting the representation of pathologies using another database, a more rigorous evaluation of the method's performance was achieved. The results of the conducted experiments yield valuable insights.

In the first experiment, the CBIR system based on YOLOv5x architecture demonstrated promising generalisation and robustness by effectively retrieving lesions from two distinct databases, VinDr-CXR and ChestX-ray14. This shows its potential to be applied to diverse databases and

real-world medical scenarios. Also, the experiment highlighted the importance of search collection diversity in improving retrieval performance. The CBIR system performed competently when a larger and more diverse search collection (ChestX-ray14) was used compared to a more restricted one (VinDr-CXR). The smaller retrieved distances from the ChestX-ray14 database's retrieved lesions to the queries suggest that the ChestX-ray14 search collection contributed significantly to improving the retrieval accuracy and visual similarity.

In a second experiment, the dissimilarity between the different classes was evaluated and it was possible to observe that depending on the lesions FR size, the threshold setting and the proximity between the different search collections varied. Also, the CBIR system showed to be able to separate the PaL search collection from the other search collections.

In the third experiment, the dissimilarity between the subclasses of the PaL class was evaluated. Due to the scarcity of data representing each subclass and the variation of annotations in each lesion by the radiologists, drawing definitive conclusions proved to be challenging. The overlap of the distributions of search collection retrieved lesions further complicated the analysis. For better evaluation of this experiment, it is required a larger amount of annotated data from the PaL subclasses.

To enhance the CBIR system's performance, future work should focus on acquiring more annotated data and balancing and increasing the quantity of class and subclass data. Furthermore, by incorporating outlier detection methods into the CBIR system, the retrieval accuracy can be improved, the system reliability can increase, and it can provide more valuable insights for medical professionals. As medical image databases continue to grow in size and complexity, addressing outliers and refining the retrieval process will be essential for advancing the application of CBIR in medical diagnosis, research, and patient care.

References

[1] J. Anthony Seibert. One hundred years of medical diagnostic imaging technology. *Health Physics*, 69:695–720, 11 1995. URL: https://europepmc.org/article/med/7558862, doi:10.1097/00004032-199511000-00006.

[2] Terri L. Fauber. *Radiographic imaging & exposure*. Elsevier, fifth edition, 2017.

[3] R.A. Ehrlich and J.A. Daly. *Patient Care in Radiography: With an Introduction to Medical Imaging*. Elsevier Science Health Science Division, 2008. URL: https://books.google.pt/books?id=KzWtPAAACAAJ.

[4] Sheila Dixon. Diagnostic imaging dataset annual statistical release 2021/22. NHS England, december 15, 2022. URL: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2022/12/Annual-Statistical-Release-2021-22-PDF-1.3-MB.pdf.

[5] Eric J. Stern Jannette Collins. *Chest Radiology: The Essentials*. LWW, 2015.

[6] Tor Ercleve. Normal chest x-ray, Nov 2020. URL: https://litfl.com/normal-chest-x-ray/.

[7] Zhongyu Li, Xiaofan Zhang, Henning Müller, and Shaoting Zhang. Large-scale retrieval for medical image analytics: A comprehensive review. *Medical Image Analysis*, 43:66–84, 2018. URL: https://www.sciencedirect.com/science/article/pii/S136184151730138X, doi:https://doi.org/10.1016/j.media.2017.09.007.

[8] Jiansheng Fang, Huazhu Fu, and Jiang Liu. Deep triplet hashing network for case-based medical image retrieval. 1 2021. doi:10.1016/j.media.2021.101981.

[9] Nandinee Fariah Haq, Mehdi Moradi, and Z. Jane Wang. A deep community based approach for large scale content based x-ray image retrieval. *Medical Image Analysis*, 68:101847, 2 2021. doi:10.1016/j.media.2020.101847.

[10] Wilson Silva, Alexander Poellinger, Jaime S. Cardoso, and Mauricio Reyes. Interpretability-guided content-based medical image retrieval. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 305–314, 2020.

[11] Anna Guan, Li Liu, Xiaodong Fu, and Lijun Liu. Precision medical image hash retrieval by interpretability and feature fusion. *Computer Methods and Programs in Biomedicine*, 222:106945, 7 2022. doi:10.1016/j.cmpb.2022.106945.

[12] João Pedrosa, Pedro Sousa, Joana Silva, Ana Maria Mendonça, and Aurélio Campilho. Lesion-based chest radiography image retrieval for explainability in pathology detection. pages 81–94, 2022.

[13] João Pedrosa, Pedro Sousa, Ana Maria Mendonça, and Aurélio Campilho. Location-aware chest radiography abnormality classification with object detection yolov5.

[14] Arthur C Christie. *A Manual of X-Ray Technic*. J. B. Lippincott Company, 1913.

[15] Montague Cohen and Nigel G. Trott. Radiology, physical science, and the emergence of medical physics. *Medical Physics*, 22:1889–1897, 11 1995. URL: https://aapm.onlinelibrary.wiley.com/doi/10.1118/1.597444, doi:10.1118/1.597444.

[16] E. Ammann and W. Kutschera. X-ray tubes–continuous innovative technology. *The British journal of radiology*, 70 Spec No, 1997. URL: https://pubmed.ncbi.nlm.nih.gov/9534712/, doi:10.1259/BJR.1997.0002.

[17] F. E. Zink. X-ray tubes. *Radiographics : a review publication of the Radiological Society of North America, Inc*, 17:1259–1268, 1997. URL: https://pubmed.ncbi.nlm.nih.gov/9308113/, doi:10.1148/RADIOGRAPHICS.17.5.9308113.

[18] Zhenwei Zhang and Ervin Sejdić. Radiological images and machine learning: Trends, perspectives, and prospects. *Computers in Biology and Medicine*, 108:354–370, 5 2019. doi:10.1016/j.compbiomed.2019.02.017.

[19] Muhammad Kashif, Gulistan Raja, and Furqan Shaukat. An efficient content-based image retrieval system for the diagnosis of lung diseases. *Journal of Digital Imaging*, 33:971–987, 8 2020. doi:10.1007/s10278-020-00338-w.

[20] Grigory Antipov, Sid-Ahmed Berrani, Natacha Ruchaud, and Jean-Luc Dugelay. Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1263–1266, 2015.

[21] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157 vol.2, 1999. doi:10.1109/ICCV.1999.790410.

[22] Sivic and Zisserman. Video google: a text retrieval approach to object matching in videos. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 1470–1477 vol.2, 2003. doi:10.1109/ICCV.2003.1238663.

[23] Guohui Wei, Hui Cao, He Ma, Shouliang Qi, Wei Qian, and Zhiqing Ma. Content-based image retrieval for lung nodule classification using texture features and learned distance metric. *Journal of Medical Systems*, 42:13, 1 2018. doi:10.1007/s10916-017-0874-5.

[24] Tizita Nesibu Shewaye and Alhayat Ali Mekonnen. Benign-malignant lung nodule classification with geometric and appearance histogram features, 2016. arXiv:1605.08350.

[25] Amal A. Farag, Asem Ali, Salwa Elshazly, and Aly A. Farag. Feature fusion for lung nodule classification. *International Journal of Computer Assisted Radiology and Surgery*, 12:1809–1818, 10 2017. doi:10.1007/s11548-017-1626-1.

[26] Amal Farag, Asem Ali, James Graham, Shireen Elhabian, Aly Farag, and Robert Falk. Feature-based lung nodule classification. pages 79–88, 2010. doi:10.1007/978-3-642-17277-9_9.

[27] Xinqi Wang, Keming Mao, Lizhe Wang, Peiyi Yang, Duo Lu, and Ping He. An appraisal of lung nodules automatic classification algorithms for ct images. *Sensors (Switzerland)*, 19, 1 2019. doi:10.3390/s19010194.

[28] Yoshua Bengio. Learning deep architectures for ai. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009. URL: http://dx.doi.org/10.1561/2200000006, doi:10.1561/2200000006.

[29] Qi Dou, Hao Chen, Lequan Yu, Jing Qin, and Pheng-Ann Heng. Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection. *IEEE Transactions on Biomedical Engineering*, 64:1558–1567, 7 2017. doi:10.1109/TBME.2016.2613502.

[30] Xinglong Liu, Fei Hou, Hong Qin, and Aimin Hao. Multi-view multi-scale cnns for lung nodule type classification from ct images. *Pattern Recognition*, 77:262–275, 5 2018. doi:10.1016/j.patcog.2017.12.022.

[31] Keming Mao, Renjie Tang, Xinqi Wang, Weiyi Zhang, and Haoxiang Wu. Feature representation using deep autoencoder for lung nodule image classification. *Complexity*, 2018:1–11, 2018. doi:10.1155/2018/3078374.

[32] Guorong Wu, Minjeong Kim, Qian Wang, Yaozong Gao, Shu Liao, and Dinggang Shen. Unsupervised deep feature learning for deformable registration of mr brain images. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 649–656, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. doi:10.1007/978-3-642-40763-5_80.

[33] P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. In *Parallel distributed processing: Explorations in the microstructure of cognition*, pages 194–281. MIT Press, Cambridge, MA, 1986. URL: https://www.bibsonomy.org/bibtex/2007c1ca25e600146c7b6eb20907fddf5/jabreftest.

[34] Yu Cao, Shawn Steffey, Jianbiao He, Degui Xiao, Cui Tao, Ping Chen, and Henning Müller. Medical image retrieval: A multimodal approach. *Cancer Informatics*, 13s3:CIN.S14053, 1 2014. doi:10.4137/CIN.S14053.

[35] Fuhui Long, Hongjiang Zhang, and David Dagan Feng. *Fundamentals of Content-Based Image Retrieval*. 2003. doi:10.1007/978-3-662-05300-3_1.

[36] Akinori Higaki, Naoto Kawaguchi, Tsukasa Kurokawa, Hikaru Okabe, Takuro Kazatani, Shinsuke Kido, Tetsuya Aono, Kensho Matsuda, Yuta Tanaka, Saki Hosokawa, Tetsuya Kosaki, Go Kawamura, Tatsuya Shigematsu, Yoshitaka Kawada, Go Hiasa, Tadakatsu Yamada, and Hideki Okayama. Content-based image retrieval for the diagnosis of myocardial perfusion imaging using a deep convolutional autoencoder. *Journal of Nuclear Cardiology*, 2022. doi:10.1007/s12350-022-03030-4.

[37] Akinori Higaki, Naoto Kawaguchi, Tsukasa Kurokawa, Hikaru Okabe, Takuro Kazatani, Shinsuke Kido, Tetsuya Aono, Kensho Matsuda, Yuta Tanaka, Saki Hosokawa, Tetsuya

Kosaki, Go Kawamura, Tatsuya Shigematsu, Yoshitaka Kawada, Go Hiasa, Tadakatsu Yamada, and Hideki Okayama. Content-based image retrieval for the diagnosis of myocardial perfusion imaging using a deep convolutional autoencoder. *Journal of Nuclear Cardiology*, 30:540–549, 4 2023. doi:10.1007/s12350-022-03030-4.

[38] R. T. Akash Guna and O. K. Sikha. *Content-Based Image Retrieval Using Deep Features and Hamming Distance*. 2023. doi:10.1007/978-3-031-20541-5_7.

[39] Computational intelligence based secure three-party cbir scheme for medical data for cloud-assisted healthcare applications. *Multimedia Tools and Applications*, 81, 2022. doi:10.1007/s11042-020-10483-7.

[40] D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2161–2168, 2006. doi:10.1109/CVPR.2006.264.

[41] Menglin Jiang, Shaoting Zhang, Hongsheng Li, and Dimitris N. Metaxas. Computer-aided diagnosis of mammographic masses using scalable image retrieval. *IEEE Transactions on Biomedical Engineering*, 62:783–792, 2 2015. doi:10.1109/TBME.2014.2365494.

[42] Jun Wang, Wei Liu, Sanjiv Kumar, and Shih-Fu Chang. Learning to hash for indexing big data—a survey. *Proceedings of the IEEE*, 104(1):34–57, 2015.

[43] Lei Zhu, Jialie Shen, Liang Xie, and Zhiyong Cheng. Unsupervised visual hashing with semantic assistant for content-based image retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 29:472–486, 2 2017. doi:10.1109/TKDE.2016.2562624.

[44] Amal Farag, Shireen Elhabian, James Graham, Aly Farag, and Robert Falk. Toward precise pulmonary nodule descriptors for nodule type classification. pages 626–633, 2010. doi:10.1007/978-3-642-15711-0_78.

[45] Sandra Jardim, João António, Carlos Mora, and Artur Almeida. A novel trademark image retrieval system based on multi-feature extraction and deep networks. *Journal of imaging*, 8, 9 2022. URL: https://pubmed.ncbi.nlm.nih.gov/36135404/, doi:10.3390/jimaging8090238.

[46] Wei Chen, Yu Liu, Weiping Wang, Erwin M. Bakker, Theodoros Georgiou, Paul Fieguth, Li Liu, and Michael S. Lew. Deep learning for instance retrieval: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20, 2022. doi:10.1109/TPAMI.2022.3218591.

[47] Hypothesis transfer learning based on structural model similarity. *Neural Computing and Applications*, 31:3417–3430, 8 2019. URL: https://link.springer.com/article/10.1007/s00521-017-3281-4, doi:10.1007/S00521-017-3281-4/TABLES/6.

[48] José Ramos, Thessa T.J.P. Kockelkorn, Isabel Ramos, Rui Ramos, Jan Grutters, Max A. Viergever, Bram Van Ginneken, and Aurélio Campilho. Content-based image retrieval by metric learning from radiology reports: Application to interstitial lung diseases. *IEEE Journal of Biomedical and Health Informatics*, 20:281–292, 1 2016. doi:10.1109/JBHI.2014.2375491.

[49] Yong Zhang, Weihua Ou, Jiacheng Zhang, and Jiaxin Deng. Category supervised cross-modal hashing retrieval for chest x-ray and radiology reports. *Computers & Electrical Engineering*, 98:107673, 3 2022. doi:10.1016/j.compeleceng.2021.107673.

[50] Ha Q. Nguyen, Khanh Lam, Linh T. Le, Hieu H. Pham, Dat Q. Tran, Dung B. Nguyen, Dung D. Le, Chi M. Pham, Hang T. T. Tong, Diep H. Dinh, Cuong D. Do, Luu T. Doan, Cuong N. Nguyen, Binh T. Nguyen, Que V. Nguyen, Au D. Hoang, Hien N. Phan, Anh T. Nguyen, Phuong H. Ho, Dat T. Ngo, Nghia T. Nguyen, Nhan T. Nguyen, Minh Dao, and Van Vu. Vindr-cxr: An open dataset of chest x-rays with radiologist's annotations. *Scientific Data*, 9:429, 7 2022. doi:10.1038/s41597-022-01498-w.

[51] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *arXiv:1705.02315*, 6 2017.

[52] G. Jocher, A. Stoken, J. Borovec, A. Chaurasia, L. Changyu, V. Laughing, A. Hogan, J. Hajek, L. Diaconu, and Y. Kwon. Ultralytics/yolov5, 2021. URL: https://github.com/ultralytics/yolov5.

[53] Chien-Yao Wang, Hong-Yuan Mark Liao, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh, and I-Hau Yeh. Cspnet: A new backbone that can enhance learning capability of cnn. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1571–1580, 2020. doi:10.1109/CVPRW50498.2020.00203.

[54] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[55] Kaixin Wang, Jun Hao Liew, Yingtian Zou, Daquan Zhou, and Jiashi Feng. Panet: Few-shot image semantic segmentation with prototype alignment. In *proceedings of the IEEE/CVF international conference on computer vision*, pages 9197–9206, 2019.

[56] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.

Appendices

Appendix A

Dissimilarity Between Classes: Examples

To gain a deeper understanding of the results depicted in histograms 5.7, 5.9 and 5.10, illustrative examples of retrieved lesions for queries belonging to class 7 are provided, considering the three different lesions FR sizes, 640,1280 and 2560. Specifically, figures A.1, A.2 and A.3 show an example of the retrieved lesions for a query from class 7 for each lesion FR size.

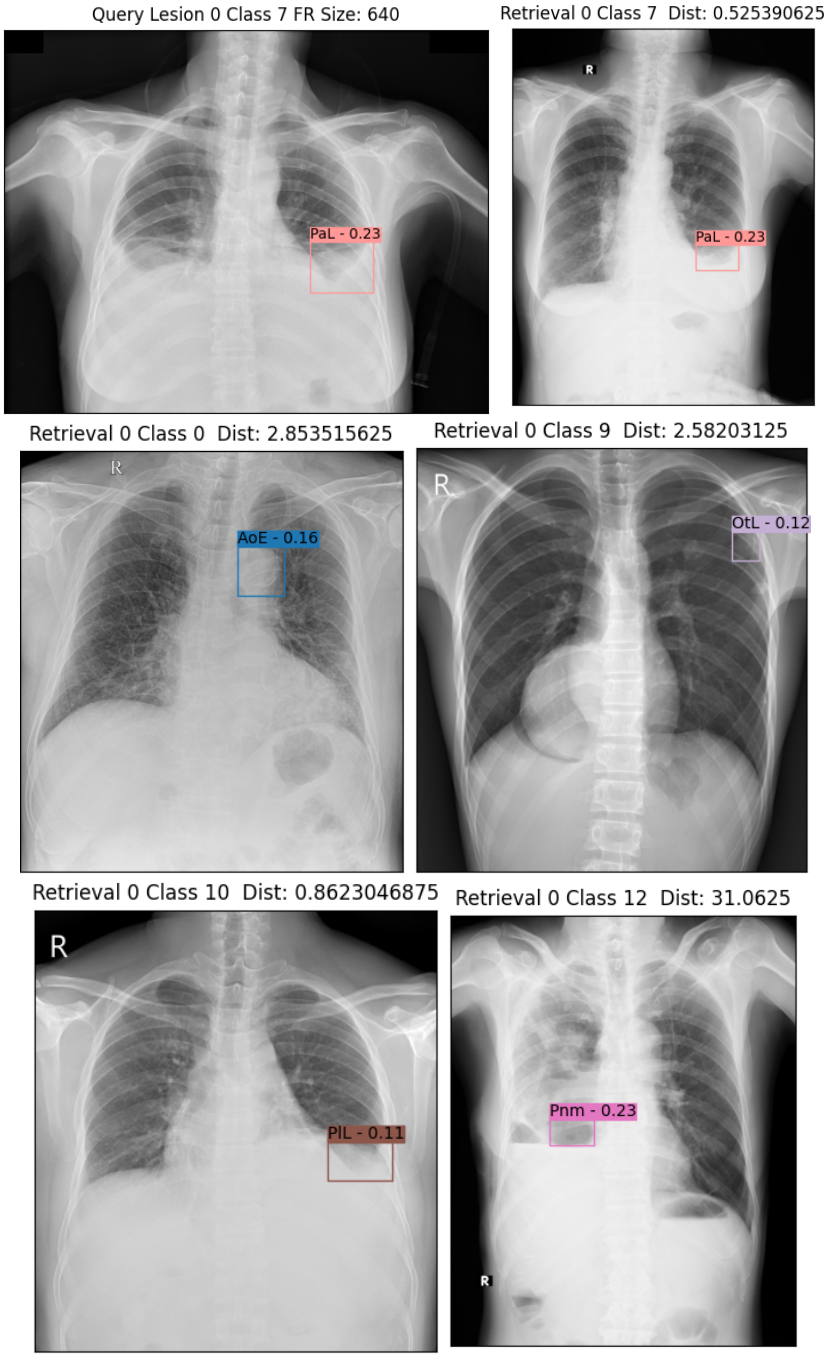


Figure A.1: In this figure is presented a query from class 7 and the retrieved lesions from class 7,0,9,10,12 search collections when the lesions FR size is 640.

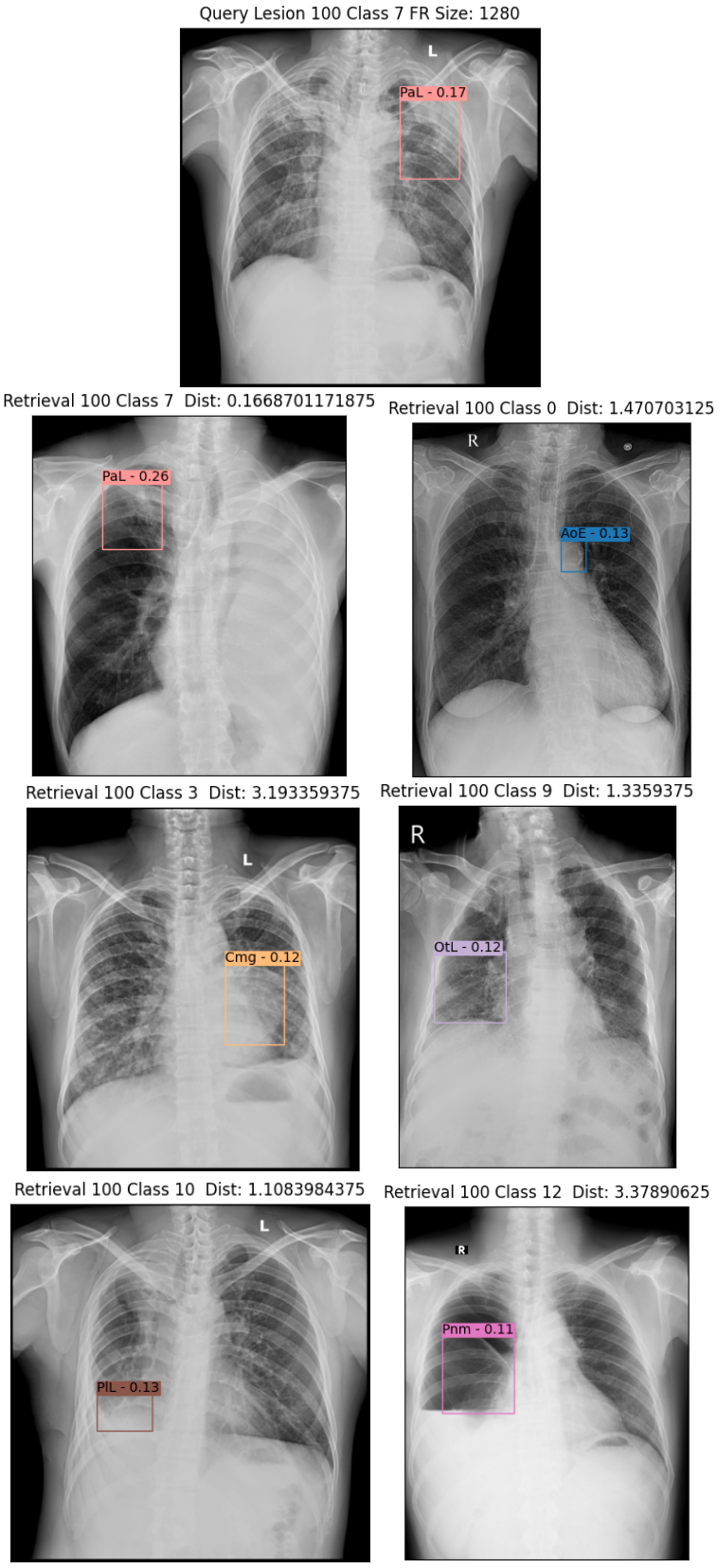


Figure A.2: In this figure is presented a query from class 7 and the retrieved lesions from class 7,0,3,9,10,12 search collections when the lesions FR size is 1280.

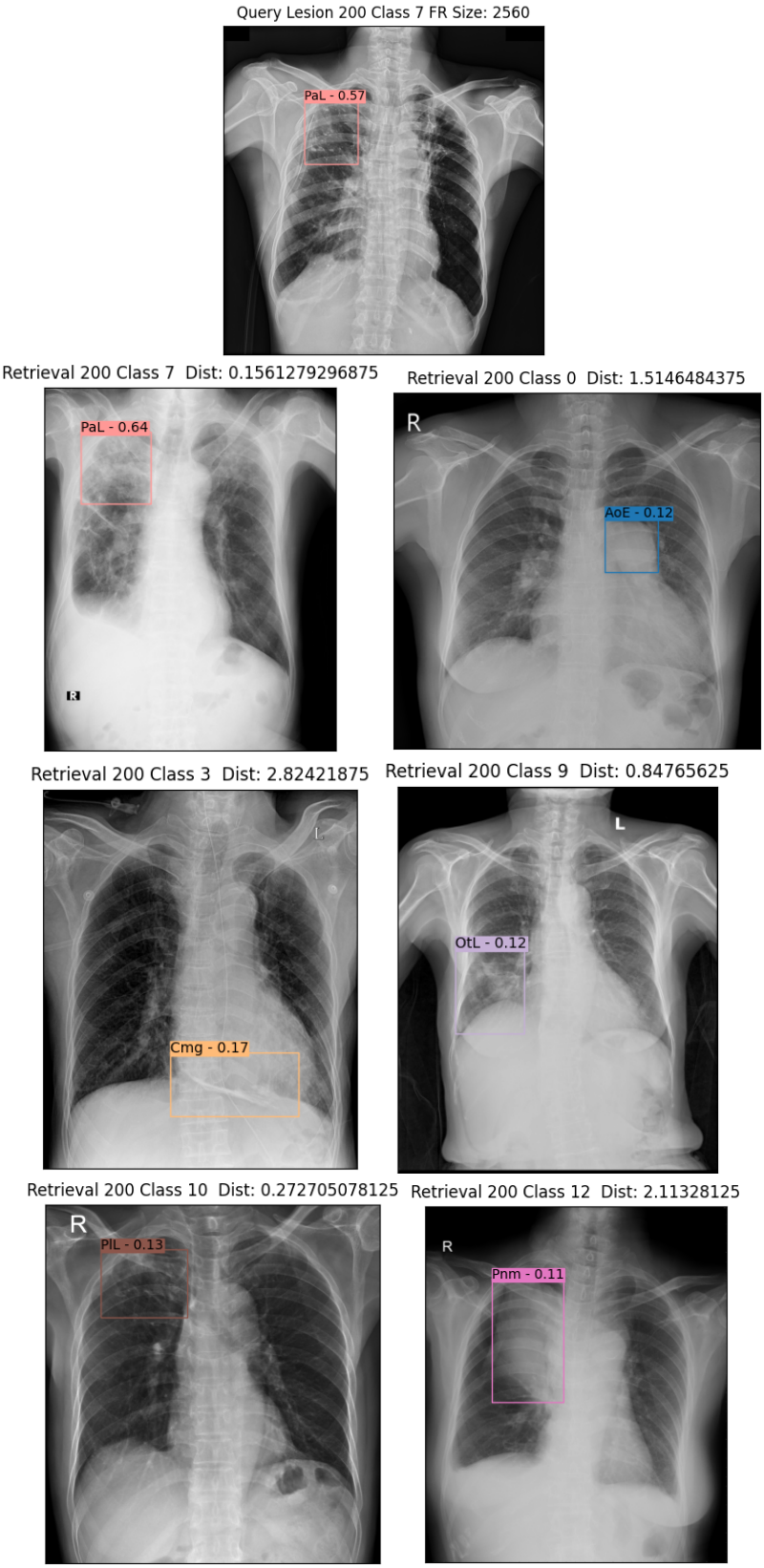


Figure A.3: In this figure is presented a query from class 7 and the retrieved lesions from class 7,0,3,9,10,12 search collections when the lesions FR size is 2560.

Appendix B

Examples of Dissimilarity Between Subclasses of the PaL Class

In order to understand better the results demonstrated in figure 5.12, it is presented here an example of the retrieved lesions for a query from each subclass of the PaL class for lesions FR size of 640.

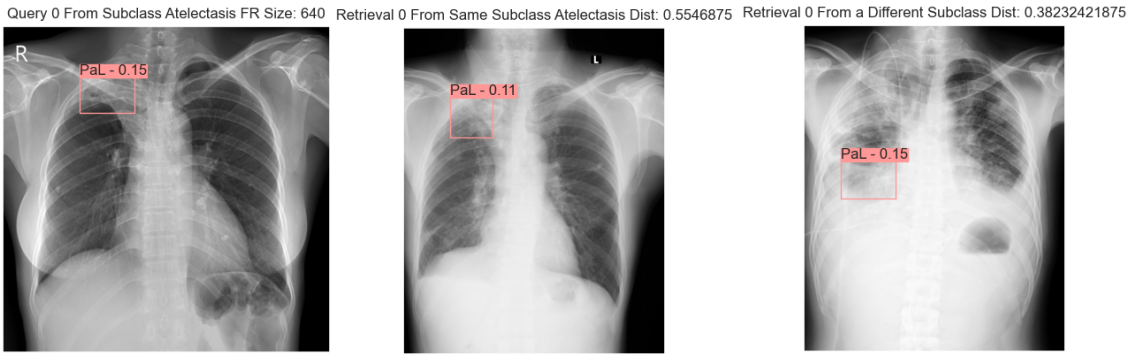


Figure B.1: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Atelectasis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

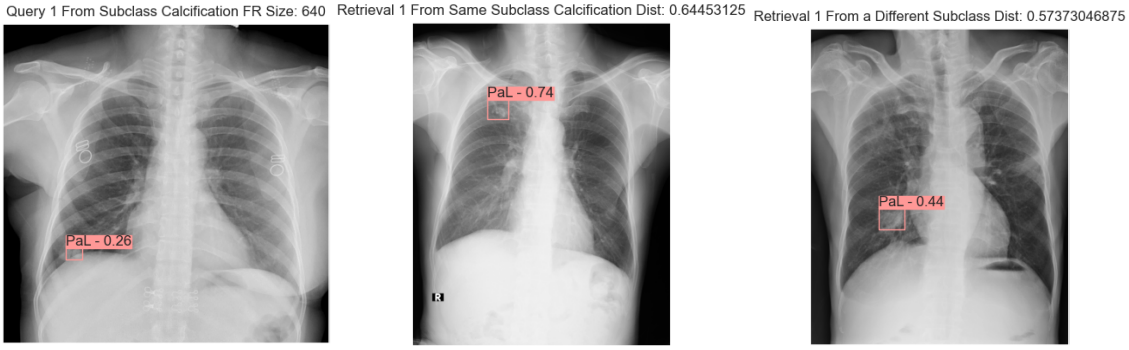


Figure B.2: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Calcification is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

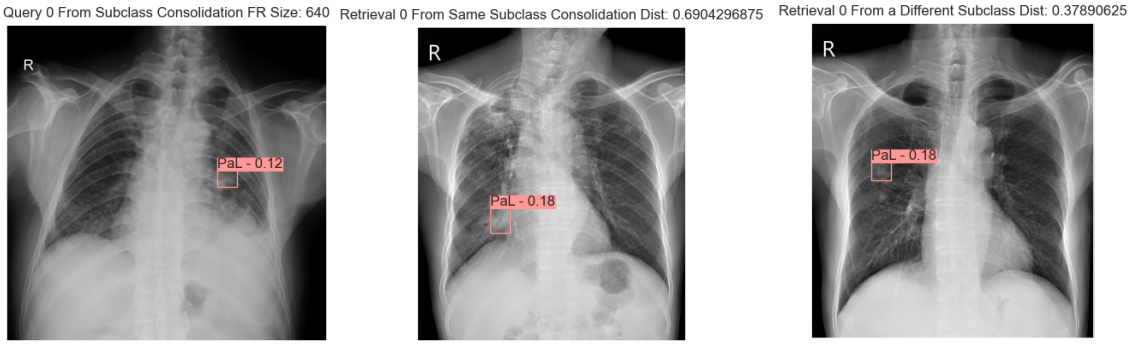


Figure B.3: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Consolidation is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

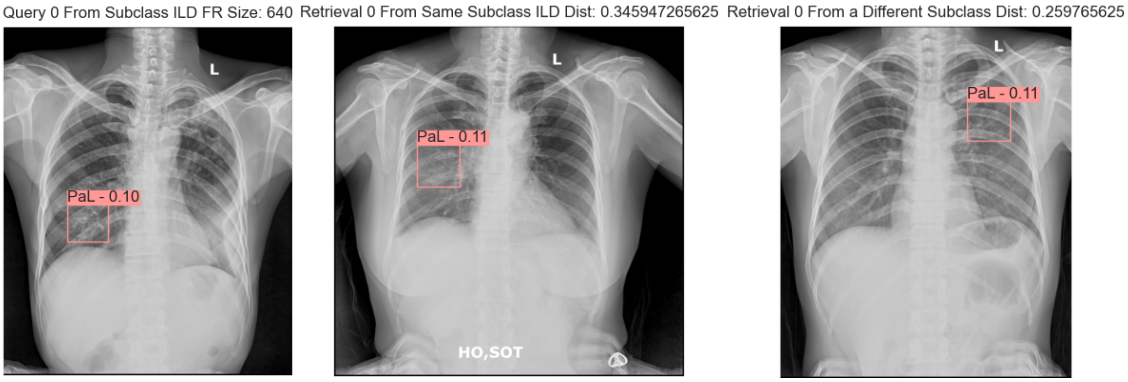


Figure B.4: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass ILD is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

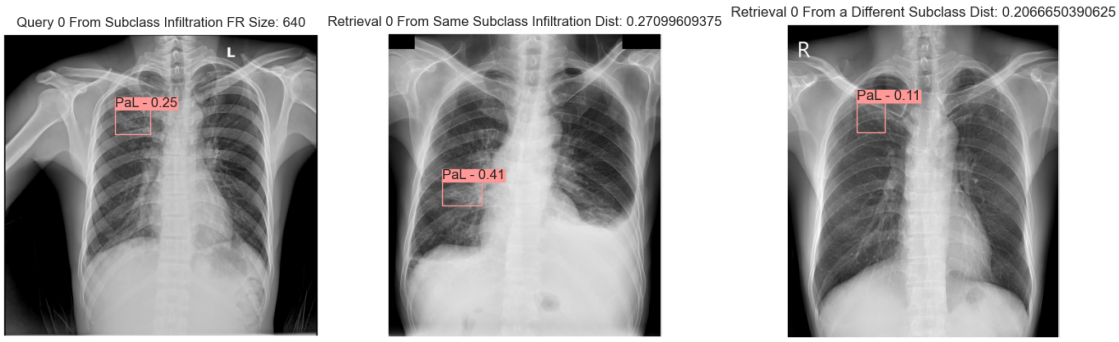


Figure B.5: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Infiltration is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

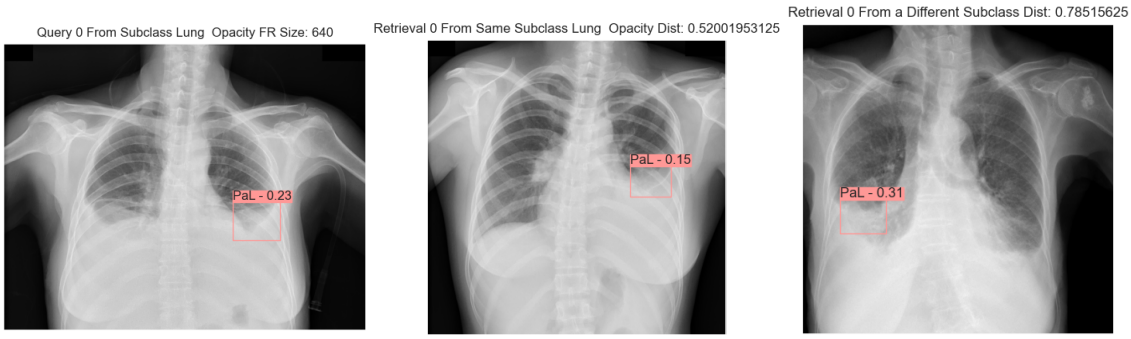


Figure B.6: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Lung Opacity is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

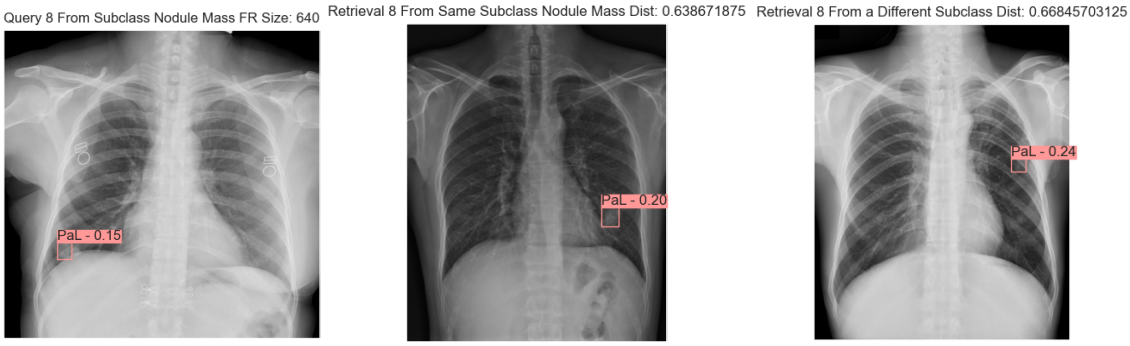


Figure B.7: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Nodule/Mass is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

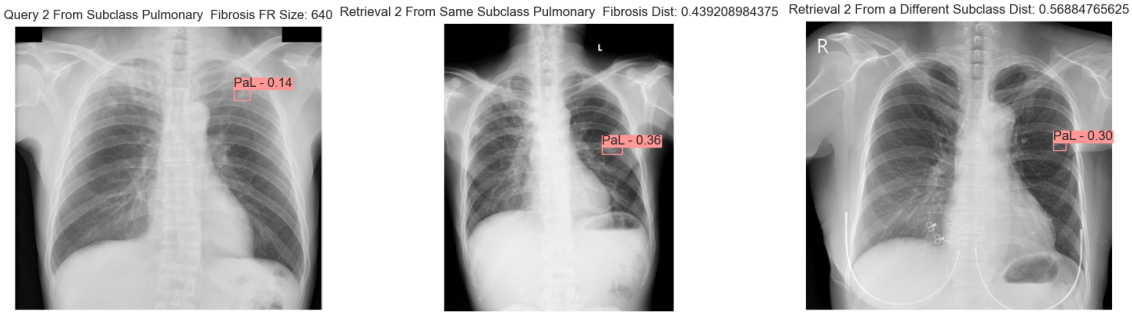


Figure B.8: Retrieved lesions from two search collections when a query lesion with a FR size of 640 features from the subclass Pulmonary Fibrosis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

In order to understand better the results exhibited in figure 5.13, it is presented here an example of the retrieved lesions for a query from each subclass of the PaL class for lesions FR size of 1280.

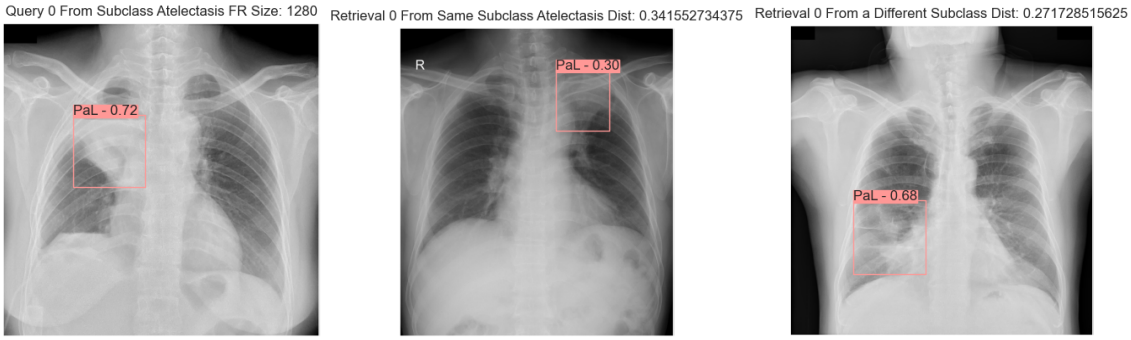


Figure B.9: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Atelectasis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

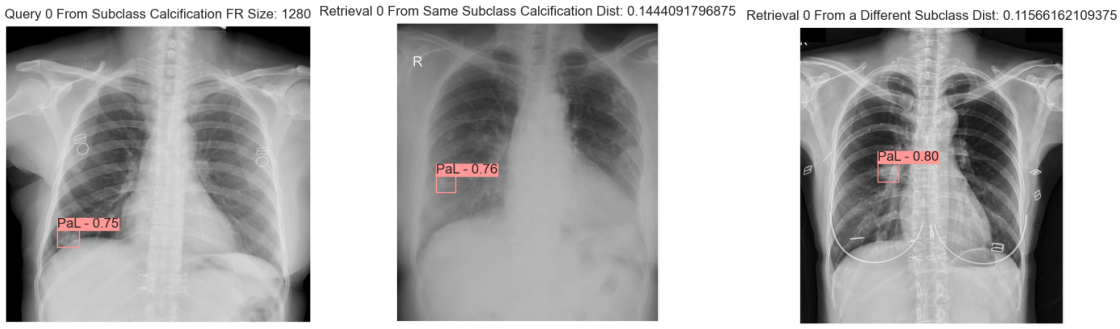


Figure B.10: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Calcification is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

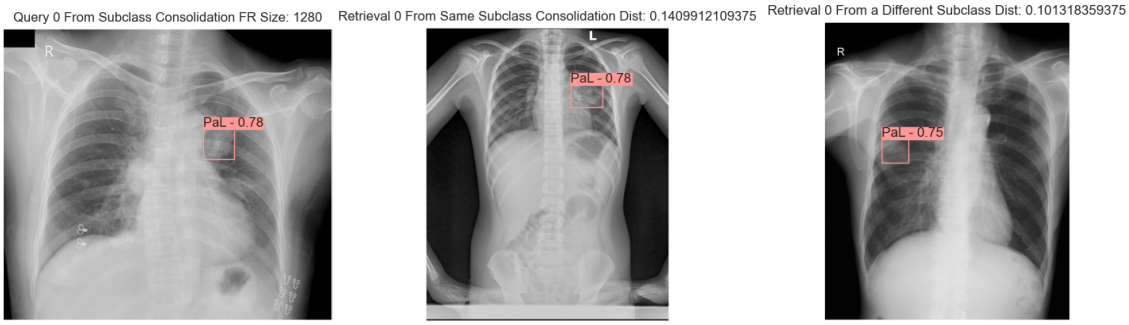


Figure B.11: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Consolidation is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

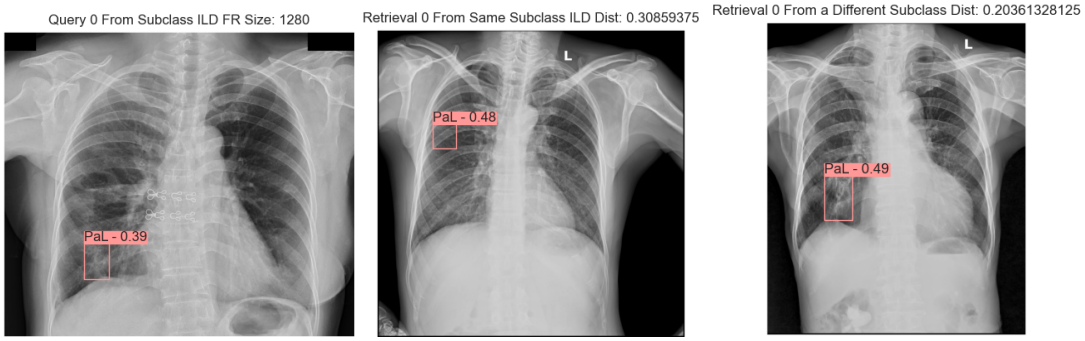


Figure B.12: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass ILD is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

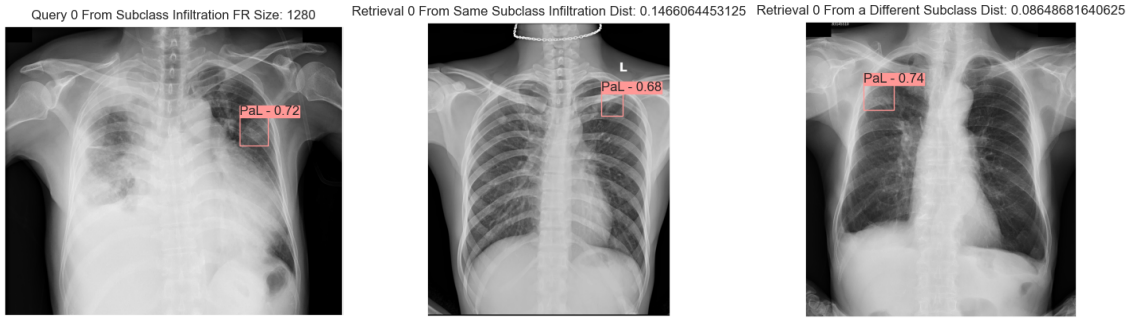


Figure B.13: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Infiltration is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

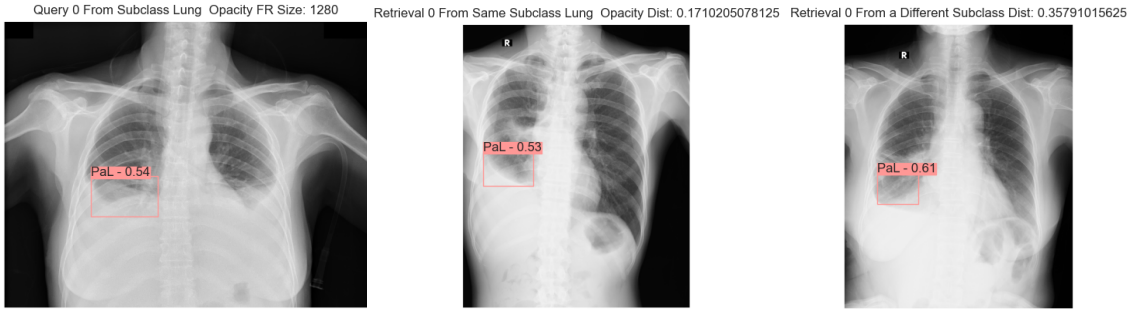


Figure B.14: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Lung Opacity is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

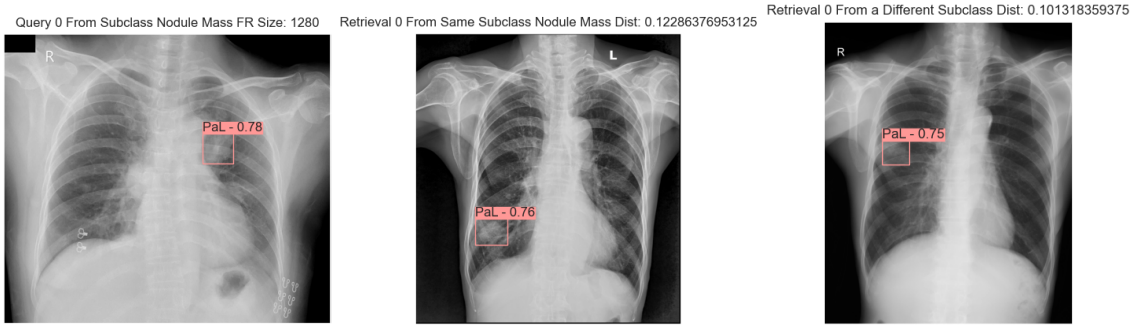


Figure B.15: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Nodule/Mass is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

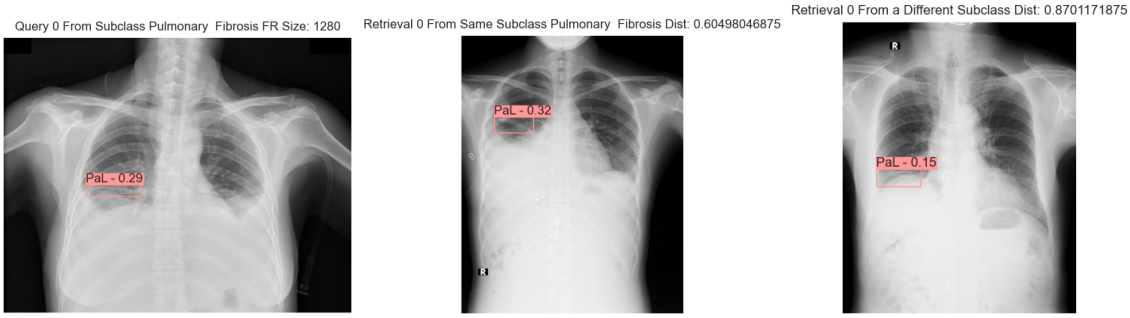


Figure B.16: Retrieved lesions from two search collections when a query lesion with a FR size of 1280 features from the subclass Pulmonary Fibrosis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

In order to understand better the results displayed in figure 5.14, it is presented here an example of the retrieved lesions for a query from each subclass of the PaL class for lesions FR size of 2560.

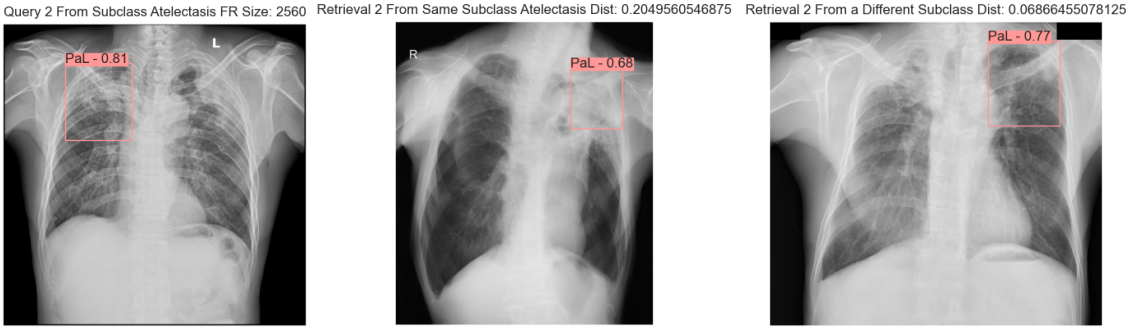


Figure B.17: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Atelectasis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

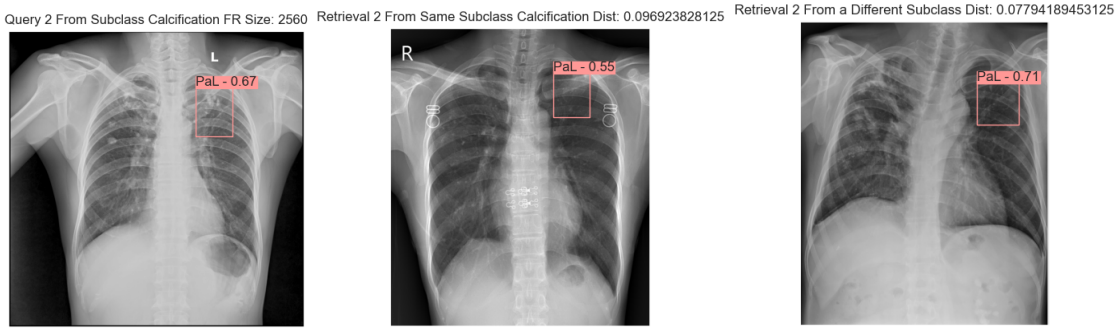


Figure B.18: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Calcification is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

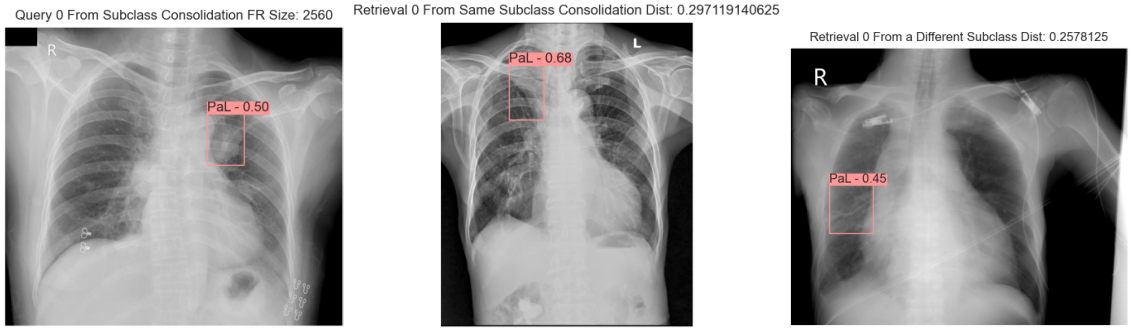


Figure B.19: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Consolidation is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

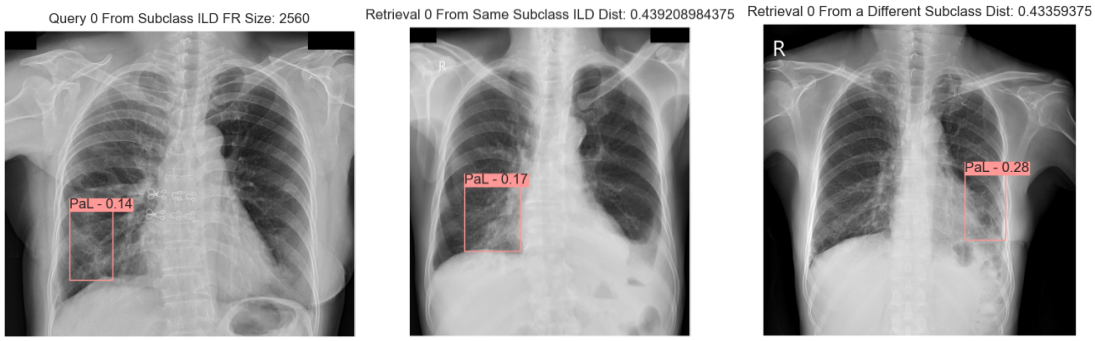


Figure B.20: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass ILD is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

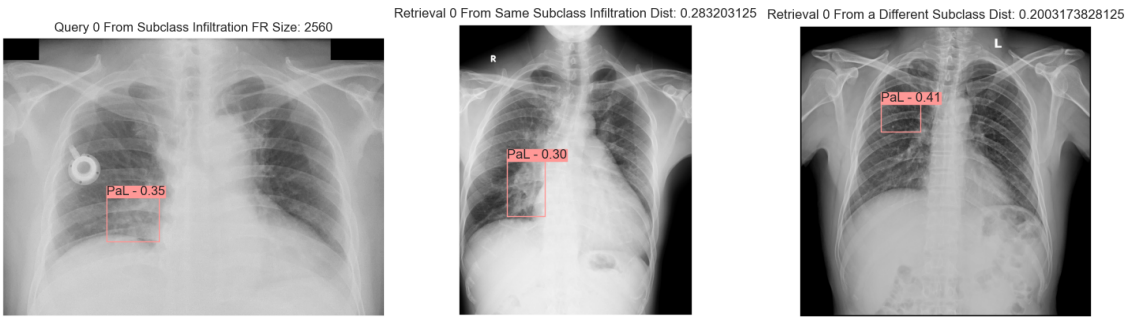


Figure B.21: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Infiltration is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

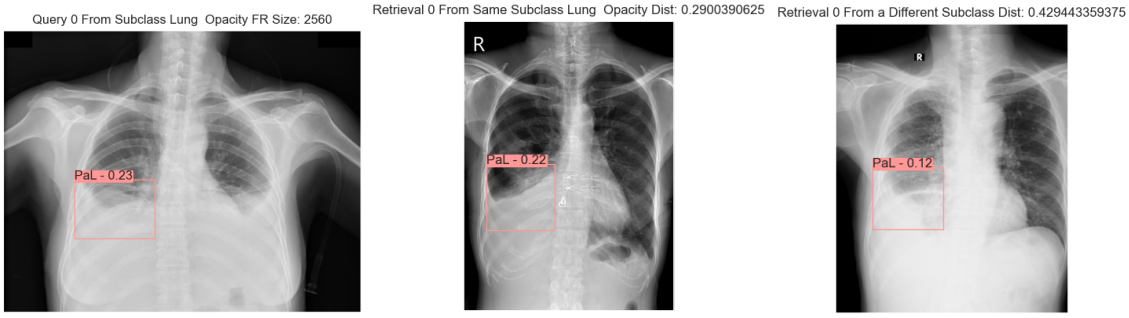


Figure B.22: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Lung Opacity is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

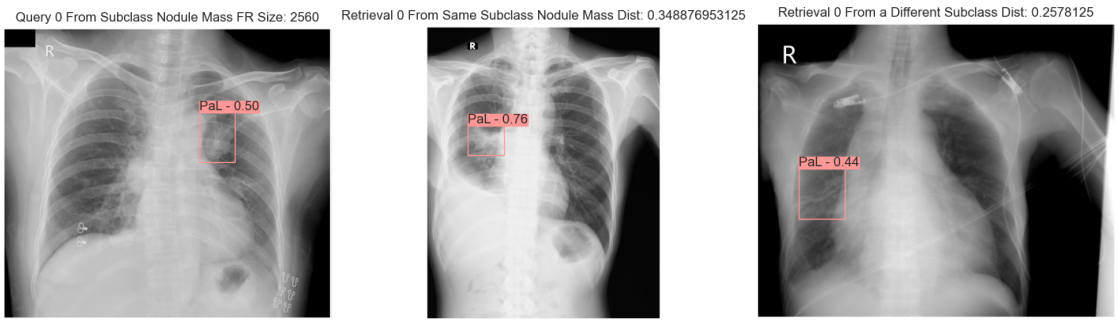


Figure B.23: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Nodule/Mass is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.

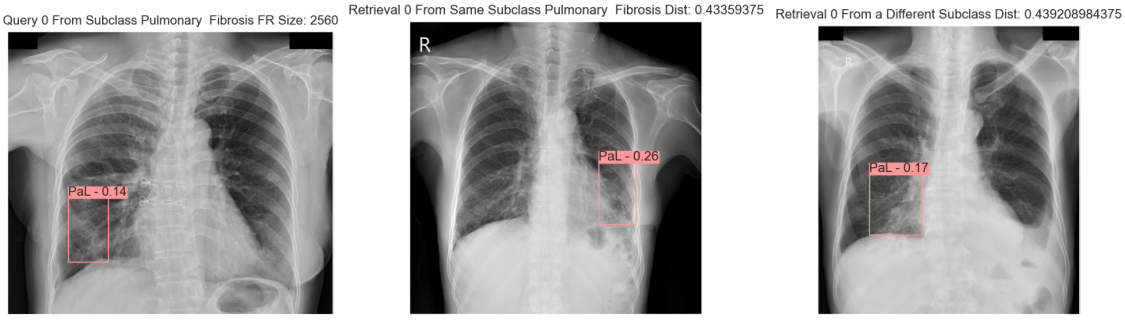


Figure B.24: Retrieved lesions from two search collections when a query lesion with a FR size of 2560 features from the subclass Pulmonary Fibrosis is tested. One of the search collections is constituted of data from the same subclass of the query and the other one is constituted of data from all of the other subclasses of the PaL class.