FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Autoencoder-based Image Recommendation for Lung Cancer Characterization

Guilherme Carlos Salles

U. PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Master in Engineering and Data Science

Supervisor: João Vinagre

Co-Supervisor: Tânia Pereira

July 28, 2023

Autoencoder-based Image Recommendation for Lung Cancer Characterization

Guilherme Carlos Salles

Master in Engineering and Data Science

July 28, 2023

Abstract

This study examines the application of Recommender Systems in the healthcare domain, specifically focusing on their potential to assist physicians and clinicians in characterizing lung cancer. The research provides a comprehensive overview of fundamental techniques and state-of-the-art approaches used in Health Recommender Systems. Furthermore, it explores the utilization of an Image Retrieval System as the core engine for a Health Recommender System dedicated to lung cancer characterization.

The objective is to leverage medical annotations and image similarity search to assist healthcare professionals in diagnosing and characterizing lung cancer based on the retrieval of similar cases. The study investigated mainly two employed models, a medical annotation-based approach and Convolutional Autoencoder. Additionally, a primary investigation is performed on Supervised Autoencoder.

The results presented the capabilities and limitations of the employed models. The medical annotation-based approach demonstrated promising potential in retrieving similar cases based on medical annotations. By leveraging the annotated information, it assists in identifying cases with similar medical characteristics, thereby facilitating the retrieval of relevant past cases. However, one of the main limitation is that it heavily relies on the availability of annotated cases in order to perform an effective retrieval. On the other hand, the Convolutional Autoencoder model focuses on capturing low-level visual features to retrieve visually similar lung cancer images with dependency of medical annotations. While it presented promising results in terms of visual similarity, it encountered challenges in effectively incorporating the necessary medical-specific characteristics for accurate lung cancer characterization. Further exploration was conducted on the Supervised Autoencoder model to address the challenges encountered in the previous model, particularly the presence of a semantic gap. However, the results of this exploration did not yield the expected outcomes.

Overall, this study highlights the potential of Recommender Systems in the healthcare domain, particularly in supporting lung cancer characterization. It emphasizes the importance of leveraging medical annotations and image retrieval techniques. The findings contribute to a better understanding of the capabilities and limitations of the employed models, paving the way for further research and development in the field of Health Recommender Systems.

Index Terms — Medical Image, Computed Tomography , Chest CT, Lung Cancer, Health Recommender System, Content-based, Content-based Image Retrieval, Convolutional Autoencoder.

Acknowledgment

To my parents and brother for their support and encouragement over the years.
To Joana, my great partner and girlfriend, for all her patience, encouragement and support during this journey.
To my friends and colleagues who trough knowledge sharing, insightful discussions, and collaborative mindset have contributed significantly to my learning experience.
To the professors at FEUP who, beyond the great lessons, were always open and willing to support.
Finally, thank my supervisors João Vinagre, Tânia Pereira and INESC TEC group for the guidance, expertise, and continuous support throughout this research. Their invaluable insights, constructive feedback, and dedication have shaped this project and enhanced its quality.

Guilherme Salles

“ Do the best you can until you know better.
Then when you know better, do better.”

Maya Angelou

Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
AE	Autoencoder
AP	Average Precision
BCE	Binary Cross Entropy
CAE	Convolutional Autoencoder
CB	Content-based
CF	Collaborative filtering
CT	Computed Tomography
CAE	Convolutional Autoencoder
CBIR	Content-based Image Retrieval
CNN	Convolutional Neural Network
DiL	Doctor-in-Loop
EDA	Exploratory Data Analysis
FN	False-Negative
FP	False Positive
IDRI	Image Database Resource Initiative
HRS	Health Recommender System
KB	Knowledge-based
K-NN	K-Nearest Neighbours
LIDC	Lung Image Database Consortium
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
NCC	Normalized Cross-Correlation
NRI	Number of Relevant Images
PET	Positron Emission Tomography
ROI	Region Of Interest
RS	Recommender System
SGD	Stochastic Gradient Descent
TN	True-Negative
TP	True-Positive

Chapter 1

Introduction

Lung cancer remains one of the most prevalent and deadliest forms of cancer worldwide, posing a significant public health challenge [1]. Early detection and accurate characterization of lung cancer are crucial for improving patient outcomes, as timely interventions can lead to more effective treatment strategies. With the advancements in artificial intelligence (AI) and machine learning, there is a growing interest in developing AI systems for lung cancer characterization to assist healthcare professionals in diagnosis, prognosis, and personalized treatment decisions.

The field of lung cancer research has witnessed significant advancements in recent years, particularly in the areas of molecular profiling and personalized medicine. The identification of specific genetic mutations and biomarkers has led to the development of targeted therapies that can selectively inhibit tumor growth and improve treatment outcomes. Additionally, immunotherapies have shown promising results in enhancing the immune system's ability to recognize and eliminate cancer cells [2].

AI systems, particularly deep learning models, have shown remarkable capabilities in analyzing medical images and extracting meaningful information for disease detection and classification. In the case of lung cancer, medical imaging modalities such as computed tomography (CT) scans and positron emission tomography (PET) scans provide detailed anatomical and functional information about lung nodules and lesions [3]. However, the interpretation of these images is a complex task that requires expertise and experience. AI systems can aid in this process by automatically analyzing and characterizing lung cancer features, facilitating more accurate and efficient diagnoses.

1.1 Motivation

In recent years, the healthcare industry has witnessed significant advancements in technology and data availability, revolutionizing the way diseases are diagnosed, treated, and managed. New development of AI systems have emerged as promising tools in the field of healthcare. The AI-based solutions aim to support diagnostic and improve the clinical workflow and tasks, and increase accuracy of complex activities. Generally, lung cancer mortality depends upon the precise and accurate early detection, and examination of pulmonary nodules [4].CT images have proven to be a rapid and non-invasive way to characterize lung cancer.

The evaluation and characterization of lung cancer nodules based on CT images are essential tasks for clinicians in their daily activities. However, manually assessing patient similarities while researching similar cases and treatment protocols can be challenging and time-consuming. Additionally, the process of annotating nodules is typically performed by radiologists, introducing the possibility of human errors and inconsistencies.

Machine learning models are typically designed with the assumption that the data follows an identical distribution. However, in many real-life scenarios, including the clinical environment, this assumption can be easily violated. Healthcare data is often scarce and unbalanced, making it challenging to develop accurate and reliable machine learning models.

In the field of healthcare, machine learning applications have predominantly focused on tasks such as nodule detection and classification for lung cancer diagnosis. While these tasks are critical for early detection and treatment planning, there is a need for more comprehensive systems that can support clinicians in the decision-making process based on medical images.

The main motivation of this study is to develop a system to focus on providing clinicians with a tool that can present them with past similar cases based on CT images, supporting their decision-making process.

1.2 Objectives

This dissertation aims to develop an AI system that recommends a set of relative (past) cases to guide the decision-making process and diagnostic for the clinician. In order to achieve this goal the sub-objetives are set:

- Understand and review the state-of-art on the topic;
- Perform a study for the State-of-the-art approaches;
- Propose and assess a search engine for Image Recommender system architecture for lung cancer characterization.

1.3 Main contributions

The contribution of this work is the following:

- Identification of Image Retrieval System strategies that can support on HRS development.
- Development of a Image Retrieval engine for a Health Recommender System (HRS).

1.4 Strutucture

This document is structured into five chapters. The first chapter introduces the dissertation by providing the context and the motivation for the work proposed, as well as the objectives and expected contributions. Chapter 2 presents some concepts and contextualises the problem, including a description of cancer, focusing on lung cancer. Chapter 3 presents main techniques involving recomendation systems and content-based image retrieval, enfazise the applications on Health sector. It also introduces the overview of literature review on RS and CBIR and describes the state-of-the-art applications related and main challenges. Chapter 4 presents the initial system design architecture and propose the search image model to be used on the core engine for HRS. On the Chapter 5 is presented the results and discussion about the architecture developed. Finally, Chapter 6 concludes the dissertation by summarizing the key findings and contributions of the research. It also outlines potential avenues for future work and highlights areas of improvement or further investigation.

Chapter 2

Background

This chapter introduces some concepts and contextualization of the problem. The first section offers an overview of cancer. The second section discusses the specific characteristics of lung cancer. In the third and fourth sections the focus shifts to the diagnosis of lung cancer and introduces the concept and details about the biopsy process are introduced.

2.1 Cancer

Cancer continues to be a major global health issue, accounting for a significant number of deaths worldwide. In 2020 alone, approximately 10 million individuals lost their lives to this devastating disease [1]. Among the various types of cancer, breast cancer stands out as one of the most diagnosed cancer globally as it illustrated on Figure 2.1, affecting a significant portion of the population. On the other hand, lung cancer takes the lead as cause of cancer-related deaths, emphasizing the urgent need for effective prevention, early detection, and treatment strategies to combat this highly lethal form of cancer. Understanding the incidence and mortality rates of different cancer types can help guide public health initiatives and research efforts aimed at reducing the burden of cancer on a global scale.

The cancer diagnoses continue to increase worldwide, impacting tremendously on multiple fronts, including physical, emotional, and financial aspects for individuals, families, communities, and healthcare systems [1]. Unfortunately, in many low and middle income countries, access to timely and quality cancer diagnosis and treatment remains limited, enhancing the already severe impact of the disease. In countries with robust healthcare systems, advancements in early detection, high-quality treatment, and survivorship care have led to improved survival rates for various types of cancers [1].

To further illustrate the magnitude of the problem, Figure 2.2 depicts the projected growth of new cancer cases worldwide from 2020 to 2040. These estimates demonstrate the alarming upward trend in cancer incidence, underscoring the urgent need for effective prevention, early detection, and comprehensive treatment strategies to mitigate the escalating impact of cancer on a global scale.

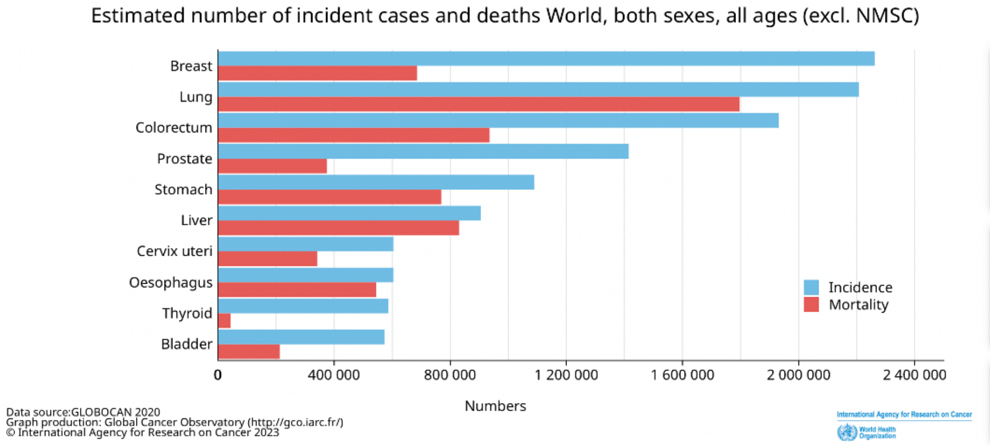

Figure 2.1: Estimated number of cancer cases and deaths worldwide by type of cancer [1].

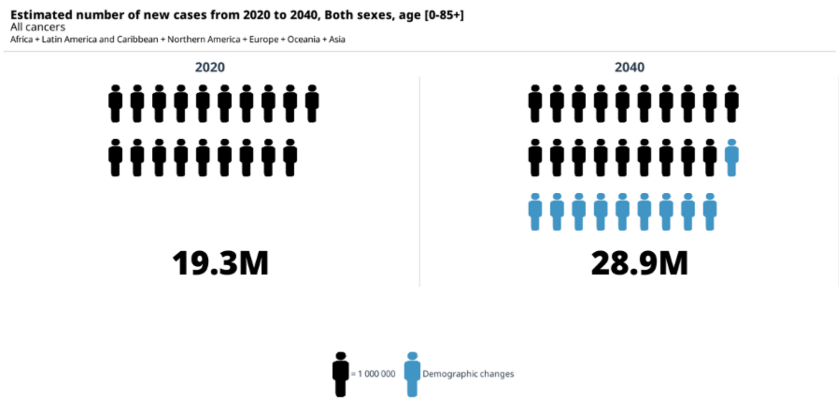

Figure 2.2: Estimated growth of new cases of cancer worldwide [5].

Early cancer diagnoses improve the cancer outcomes as the treatment is more likely to be effective, with a higher rate of survival, less morbidity and a less expensive treatment. Early diagnose is, therefore, an important public health strategy. By emerging effective approaches to identify cancer in initial stages, lives can be spared as well as the personal, societal and economic costs of cancer care can decrease.

According to World Health Organization [5], there are two different approaches that encourage early detection:

- Screening – identification of individuals with abnormalities indicative of a certain cancer or pre-cancer who still have not developed any symptoms and refer them quickly for a diagnosis and treatment
- Early diagnosis – identification of symptomatic cancer cases at the earliest possible stage;

2.2 Lung Cancer

In the context of lung cancer, the situation is particularly alarming, with only 15% of patients surviving beyond five years after diagnosis, primarily due to the fact that around 70% of cases are diagnosed at an advanced stage [6]. Despite efforts to reduce smoking rates in Western countries, the prevalence of smoking remains significant, with approximately 17 to 28% of adults currently being smokers. Moreover, the initiation of smoking among young individuals continues to pose a significant challenge [7]. As a result, lung cancer and other tobacco-related diseases are anticipated to remain significant global health concerns for many years to come [6] [8]. Efforts to improve early detection, develop effective treatment strategies, and implement preventative measures is essential for combating this disease.

Early detection of lung cancer is critical for improving patient outcomes, as the disease is often diagnosed at advanced stages when treatment options are limited. Symptoms may not manifest until the disease has progressed, further emphasizing the need for effective screening and diagnostic tools. Imaging techniques such as chest X-rays, CT scans, and PET scans play a crucial role in detecting and staging lung cancer[3].

2.3 Diagnosis

CT imaging has emerged as the most effective technique for detecting lung cancer, particularly in its early stages. Compared to traditional radiography techniques, CT scans provide more detailed information about the localization and size of nodules, leading to improved accuracy in diagnosis. In fact, studies have demonstrated a significant reduction in mortality rates, up to 20%, with the implementation of low-dose CT screening programs. Additionally, the use of CT imaging has resulted in a higher rate of positive screening tests, enabling prompt intervention and treatment for patients. These findings underscore the crucial role of CT imaging in lung cancer detection and emphasize its superiority over conventional radiographic methods [4].

The classification of lung nodules as malignant or benign heavily relies on the careful examination of 2D CT slices, which in turn requires a comprehensive analysis of 3D lung voxel data. CT scans provide a wealth of information regarding the characteristics of nodules, but as the number of images increases, accurately assessing them becomes an increasingly challenging task for radiologists. Furthermore, the detection of lung cancer nodules can be significantly influenced by human error, which introduces the possibility of misdiagnosis or missed detections. This highlights the need for advanced computational methods and decision-support systems to aid radiologists in achieving more accurate and consistent nodule classification [4].

2.4 Biopsy

Biopsy is a commonly performed medical procedure that entails extracting a small tissue sample from the body for microscopic examination. Its primary purpose is to identify and characterize abnormal cells or tissues. Biopsies serve as valuable diagnostic tools, aiding in the determination of specific medical conditions. Moreover, they facilitate the assessment of disease severity and grading, which are crucial factors in treatment decision-making. Biopsy results are instrumental in selecting appropriate treatment strategies, predicting treatment response, and providing valuable prognostic information. The information obtained from biopsies greatly contributes to enhancing patient care and improving overall patient outcomes [9].

Biopsies, although widely used for diagnostic purposes, can sometimes yield inconclusive results or may not provide a representative sample due to the heterogeneity of tumors. In such cases, additional tests or repeated biopsies may be necessary to confirm the diagnosis. However, this approach is often impractical due to logistical and financial constraints. Moreover, biopsies can be invasive and painful for patients, particularly when the tumor is located in a challenging or inaccessible area. In certain instances, obtaining tissue samples may even be impossible. These limitations highlight the need for alternative or complementary methods to overcome the drawbacks associated with traditional biopsy procedures [10].

2.5 Summary

In summary this chapter introduces contextualization on the problem. It provided background information on the nature of cancer, its prevalence worldwide, and its impact on individuals and society. Followed by a emphasis on Lung cancer and their methods of diagnosis. On the last section presented the purpose and importance of biopsies in the diagnostic process for lung cancer.Thus, it set the stage for further exploration of the research problem and the development of innovative approaches to improve and support the diagnosis and treatment of lung cancer.

Chapter 3

Literature Review

This chapter is structured into three main sections. The first section provides an overview of the literature pertaining to Recommender Systems (RS), with a specific focus on their application in the healthcare sector. It introduces the fundamental concepts and techniques on Recommendation Systems, including collaborative filtering, content-based filtering, and hybrid approaches. This section aims to establish a solid theoretical foundation for understanding the subsequent discussions on RS in healthcare. The second section reviews the most recent and relevant studies in the field of RS and Image Retrieval Systems. This section aims to explore the intersection between these two areas and identify potential synergies that can be leveraged for the development of effective recommendation models for healthcare. By examining the latest advancements and approaches in RS and Image Retrieval Systems, this section provides valuable insights into the state-of-the-art techniques and methodologies that can be applied to healthcare recommender systems. The last section of this chapter introduces the main challenges and limitations associated with the development and implementation of Healthy Recommender Systems.

3.1 Recommender Systems

Recommender Systems are a subclass of information filtering system that use different algorithms to analyze large amounts of data in order to provide personalized recommendations to users based on their past behavior and preferences. The main models for Recommender Systems use two types of data: (i) user-item interactions, such as ratings or purchasing patterns, and (ii) attribute data about users and objects[11].

3.1.1 Collaborative filtering

One of the most popular techniques used in RS is collaborative filtering (CF), which takes in consideration the behavior and preferences of similar users in order to generate recommendations for a specific user. The basic idea of these technique is that if users shared the same interests in the past they will, more likely have similar preferences in the future[12]. For instance, if they read the same books, or bougth the same items. The Fig. 3.1 represents an illustration of the collaborative filtering.

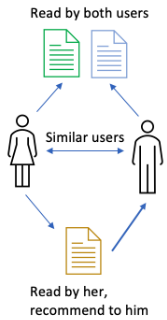

Figure 3.1: Representation of collaborative filtering.

The Matrix factorization is a commonly used approach within collaborative filtering, which seeks to approximate the user-item interaction matrix as the product of two low-rank matrices.

The main challenge in designing collaborative filtering methods is that the underlying rating matrices are sparse [11]. For instance, the RS of a streaming platform could be impacted if most users would have viewed only a small fraction of the large universe of available movies. Additionally, collaborative filtering can also suffer from the cold-start problem, where the system struggles to make recommendations for new users or items that have little or no prior information.

3.1.2 Content-based

In content-based recommender systems (CB),the recommendations are based on characteristics or descriptive attributes of items that a user has previously interacted with [11]. The main idea behind content-based filtering is that if a certain user enjoys item A, the user is likely to enjoy item B if item B has similar characteristics to item A. The Figure 3.2 exemplifies the content-based filtering.

On these systems the first step consists in representing each item in the dataset as a feature vector that describes its characteristics or attributes. This representation can be achieved through various methods such as text analysis, image processing or audio analysis. The second step is to measure the similarity between the items based on their feature vectors. This similarity measurement can be achieved using various distance metrics such as Cosine similarity, Euclidean distance or Pearson correlation. The final step is to generate recommendations based on the similarities between the items and the user's preferences. This can be achieved by sorting the items based on their similarity to the items that the user has liked in the past.

The main advantage of content-based filtering is its ability to generate recommendations based on the characteristics of items, which allows a more personalized recommendations. It is also effective in situations where there is a limited amount of information regarding the preferences and behaviors of users, such as the cold-start problem. However, one of the limitations of content-based filtering is that it relies on the representation of items, which can be subject to biases and

Figure 3.2: Simplified representation of content-based.

errors. Additionally, content-based filtering may not be able to capture the diversity of users' preferences and may generate repetitive recommendations.

3.1.3 Knowledge-based

Knowledge-based Recommendation is another technique that incorporates knowledge by logic inferences. This type of filtering uses explicit knowledge about an item, user preferences and other recommendation criteria as illustrated on Figure 3.3. The definition of rules or requirements for items are explicitly, for example "the food should not contain cheese since I am allergic to dairy products" [13].

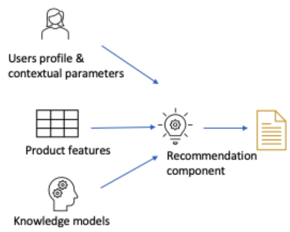

Figure 3.3: Simplified representation of knowledge-based.

3.1.4 Hybrid

Hybrid methods combine the strengths of multiple techniques, for instance it can combine the collaborative and content-based filtering. By using both the user-item interactions and the item attributes to generate recommendations, it can tackle the cold-start problem of collaborative filtering by using the available information about the items for prediction[13].

Overall, the state-of-the-art in recommendation systems is constantly evolving, with new techniques and approaches being proposed and evaluated regularly. The best approach for a specific application will depend on the characteristics of the data and the goals of the system.

3.1.5 Health Recommender Systems

Despite the extensive use of recommender systems in the e-commerce and leisure domains, their application in healthcare is still on the beginning. These systems may be used to create tailored health recommendations, thus reducing the cost of healthcare and fostering a healthier lifestyle in the population. These systems can be used to recommend healthy lifestyle choices, such as exercise and nutrition plans, or to suggest personalized treatment options for patients. By leveraging data from electronic health records, wearable devices, and other sources, these systems can help individuals make more informed decisions about their health and wellness [14].

One area of research in this field is the use of Recommender Systems for personalized nutrition. Studies have shown that Recommender Systems can be used to provide personalized dietary recommendations based on individual characteristics such as age, gender and health conditions [15] . These systems could be be effective in promoting healthy eating habits and reducing the risk of chronic diseases. Another area of research is the use of Recommender Systems for physical activity, where they are used to provide personalized recommendations for physical activity based on individual characteristics such as fitness level and preferences. These systems have been showing effectiveness in increasing physical activity levels and improving overall health [16].

In the field of lung diseases, there have been studies using Recommender Systems for personalized treatment recommendations. These systems analyze patient data such as medical history and genetic information to provide personalized treatment options and support less experienced physicians [17]. For Cancer, when the HRS effectively designed and implemented, could greatly contribute to the management of cancer-related screenings, diagnoses, treatments, operations, and rehabilitation programs. By leveraging HRS, individuals and healthcare professionals can gain access to reliable and relevant information, enabling informed decision-making throughout the various stages of cancer care. These systems play an essential role in guiding patients towards appropriate treatment options and providing them with health recommendations [18]. The development and implementation of well-designed recommender systems in the field of cancer care hold significant potential for improving patient outcomes and facilitating efficient information management. By harnessing the power of these systems, healthcare professionals can enhance the delivery of personalized care and ensure that patients and their families are equipped with the information they need to make informed decisions regarding their health and well-being [18].

Overall, the related work in the field of Recommender Systems for healthcare shows that these systems have the potential to promote healthy lifestyles and prevent chronic diseases. However, it is important to note that the majority of these studies have predominantly focused on specific areas such as nutrition and physical activity. There is a significant opportunity for further investigation and development of Recommender Systems tailored specifically for cancer detection and personalized treatment recommendations.

3.2 Image Retrieval Systems

In the modern world, there is a massive need for retrieving digital images of large proportions and multiple ranges in the most variable fields. The image retrieval system is a computer system that allows browsing, searching and retrieving images from a database of digital images or text annotations [19].

The number and volume of medical images has been growing intensely and image retrieval can be extremely important in the medical field for therapeutic diagnosis and early detection of diseases as well as in the medical research and education areas, bringing valuable insights through image analysis [20]. In order to perform a search in a substantial volume of image databases, efficient tools for image searching, browsing and retrieval are mandatory.

In the context of image retrieval systems, there are different approaches to retrieve relevant images. In query-by-text, users describe the desired image using textual descriptions, providing specific details or criteria. Alternatively, in query-by-example, users can provide an example image that closely resembles the desired image. This example image serves as a reference for the system to retrieve visually similar images from the database. Each of these query modalities offers a unique way for users to convey their interests and preferences, facilitating effective image retrieval in different scenarios [20].

There are two common approaches for image retrieval systems, text-based and content-based, that will be detailed in the next subsection.

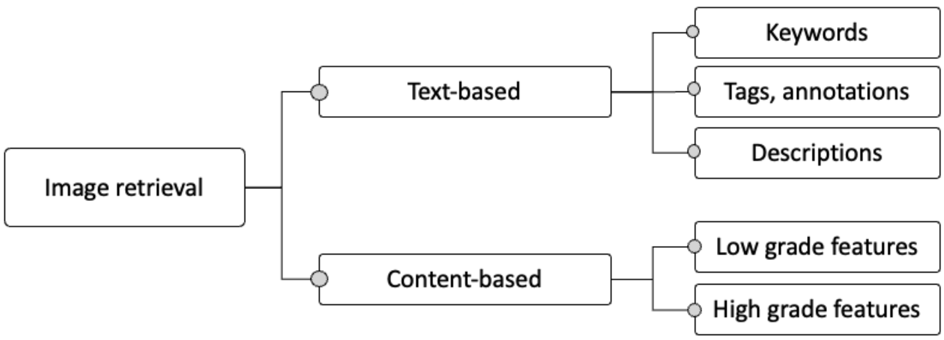

Figure 3.4: Simplified taxonomy of image retrieval systems[20].

3.2.1 Text-based Image Retrieval

In the field of image retrieval, text-based approaches have long been employed to retrieve images relevant to a given input query. In this traditional approach, the input query is given as text, and the system matches this text with the textual descriptions associated with each image in the database. Various techniques have been developed to facilitate this matching process, including bag-of-words models, natural language processing algorithms, and Boolean retrieval methods. These text-based approaches leverage the annotated textual descriptions to effectively retrieve images that align with the user's query, enabling efficient and accurate image retrieval based on textual information [19].

This method has the advantages of being fast and reliable when the images are well annotated as computation overheads associated to the image content processing and analyzing can be reduced by only analyzing the medical annotations and by using annotations as its features [20].

On the other hand, text-based image retrieval methods are not able to perform a search on non-annotated image databases and require an annotation process which leads to a substantial amount of human labor, making it time-consuming and expensive, but also this method can present some subjectivity inherent to the human perception [21].

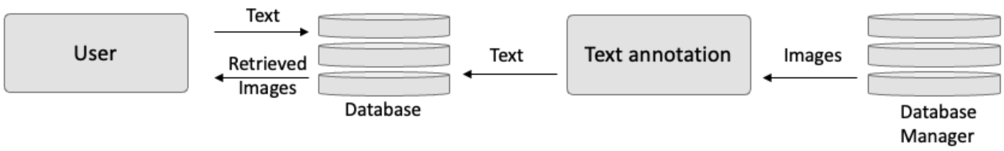

Figure 3.5: Text-based image retrieval systems[20].

3.2.2 Content-based Image Retrieval

In the content-based approach to image retrieval, the input query is provided as images with the objective of obtaining similar images. This approach addresses the challenge of searching for specific digital images in large databases by analyzing the actual content of the images, including their texture, shape, and color, and comparing them with the query image [19].

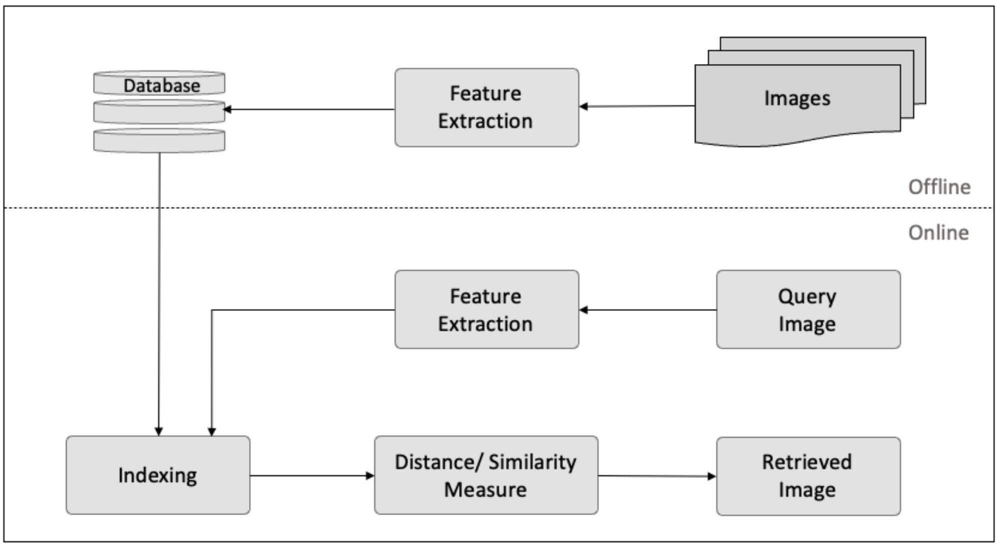

Figure 3.6: Content-based image retrieval systems[20].

Content-based Image Retrieval (CBIR) is a fully automated method. However, CBIR faces a challenge known as the "semantic gap," which refers to the disparity between low-level image features and high-level concepts or perception contained within the images. This gap often leads to the retrieval of irrelevant images [22].

To bridge this semantic gap, several studies have focused on developing methods to convert high-level concepts in images into features. Depending on the feature extraction methods, features are generally categorized into global features and local features.

Researchers have explored various techniques to extract and utilize these features effectively in CBIR systems. The goal is to enhance the system's ability to retrieve images that align with the high-level concepts or perception that users have in mind when formulating their queries. Global features, including color, texture, shape, and spatial information, provide a representation of the entire image. They offer the advantage of faster feature extraction and similarity computations, making them efficient for tasks such as object classification and detection. However, global features fall short in distinguishing between the background and the object within the image, making them less suitable for retrieval in complex scenes or object recognition. Complex scenes may involve different image parts with varying levels of importance or relevance [22].

In contrast, local features are well-suited for image retrieval, matching tasks, and recognition. Object recognition involves identifying and labeling objects within an image, while object detection focuses on determining the presence and location of objects belonging to predefined classes in an image. Object detection encompasses the task of object classification as a subset [22].

Local features are defined as key points or specific parts of an image, such as corners, blobs, or edges. They possess robustness to scale, rotation, translation, changes in backgrounds, clutter, and partial occlusions. These features allow for more precise and detailed analysis, facilitating accurate matching and recognition tasks[22].

The distinction between content-based and text-based retrieval systems lies in the level of human interaction involved. Text-based systems rely on human interpretation using high-level features or concepts, such as keywords or text descriptors, to measure image similarity. In contrast, content-based systems formerly utilize low-level features automatically extracted through computer vision techniques, such as color, texture, shape, and spatial layout. It is important to note that there is generally no direct link between these high-level concepts and the low-level features [23] .

3.3 Related work

This section aims to provide information about relevant studies and related work to RS and content-based image retrieval. It provides details about the research method and most recent information about other Recommender Systems and Content-Based Image Retrieval.

3.3.1 Method

This section was conducted in conformance with the main steps required for systematic reviews according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

The literature search was conducted using the Scorpus, ScienceDirect, Elsevier and Google Scholar databases. This literature review is focused initially on Health Recommender Systems and later on, due similarity, also extended to content-based image retrieval systems. The inclusion criteria were:

- Studies focusing on HRS and CBIR in general;
- Studies focusing on HRS and CBIR for a specified disease;
- Conference Proceedings, Journals and paper published after 2014;

In addition, a backward search was performed by examining the bibliographies of the reviewed papers discussed in the introduction section and the reference list of included studies to identify any additional studies.

3.3.2 State of the art

A comprehensive literature review of the current state-of-the-art in recommendation systems can be found in [24] [14] [25] [26] [18] [27] [13] [28] . These papers cover various aspects of Recommendation Systems, including algorithmic techniques, evaluation metrics, domain applications, main challenges and scalability issues.

As presented on Table 3.1, the scope of the existing literature has an extensive range. For example, Stark et al. [24] focused on HRSs for medicine recommendation. De Croon et al. [26] researched the approach of RS on different domains such Lifestyle, nutrition and general health. Yue et al. [14] presented similar categories including two more fields, the decision-making for patients and physicians and related to disease-related prediction.

Related to main techniques on RS, De Croon et al. [26] reported that most of the papers based on which the study was conducted, stated the use of hybrid approach, representing 44% of the papers, followed by a knowledge-based with 22% and 10% with the content-based. Although collaborative filtering is a popular technique on RS, it was not used frequently in the HRS domain. In total, 4 review studies were focused on HRS applications for the patient. On the other hand, Calero Valdez et al. [27] reinforced the new paradigma of Doctor-in-Loop (DiL) approach.

> “It pictures the doctor not only as a consumer of digital information, but also as a someone who can interactively manipulate algorithms and tools. The doctor as a final authority inside the loop of an expert system can make sure that expert knowledge is integrated in the decision making process, by finding patterns and supplying tacit knowledge, while the recommender system can integrate patient data as well as treatment results and possible side-effects related to previous decisions.” [27]

Calero Valdez et al.[27] also supported the idea that procedures must be incorporated on the design of HRS. In order to clarify the tasks definitions on HRS, the author proposed guidance questions and additional procedures to enrich the contextual picture of the usage scenario. The Figure 3.7 illustrates the proposed framework and guidance question elaborated by the author.

Author	Objetive	Contribution
Stark et al. (2019)[24]	Provides an overview of the existing Recommender Systems with more focus on medicine recommendation engines.	It is demonstrated the existing solutions for the healthcare providers in order to improve the medicine selection process and select an appropriate medication for the patients.Mostly systems are Ontology rule-based model and it is proposed graph database model the relationships between patients.
Yue et al. (2021)[14]	It presented a comprehensive review of typical recommendation techniques and their applications in the field of healthcare.	Five applications proposed on health RS: Dietary recommendations, lifestyle recommendation, training recommendation, decision-making for patients and physicians, and disease-related prediction.
Wiesner and Pfeifer (2014)[25]	It provides a motivation, a strategy and importancy of individually tailored health information in personal health records and the benefits for patients	It presents a strategy for integrated a Personal Health Record Systems in a HRS and the main challenges and requirements.
De Croon et al. (2021)[26]	It conducts a systematic literature review and synthesized the results. It proposes techniques, evaluation designs, and deployment of recommended items to the users on HRS	It present the systematic review of the state-of-art in HBS. It reports the hybrid recommendation algorithm as the majority strategy applied on HRS cases It also recommend five guidelines that can serve as a reference frame for future HRS studies.
Aalipour and Ghazisaeedi (2017)[18]	It introduces the main techniques used in RS, the main challeges and some applications in the field of health	It explores the Recommender systems as suitable tool for the information management of cancer-related screenings, diagnoses, treatments, operations, and rehabilitation programs. Access to treatment and health recommend
Calero Valdez et al. (2016)[27]	It presents the techniques used in RS, evaluation methods, a framework and main consideration on the development of HRS	It proposed a framework to be used on the development of HBS and medical applications and presents the concept of Doctor is inside the loop of an expert system to interactively manipulate algorithms and tools .
Tran et al. (2021)[13]	It provides a systematic overview of existing research on healthcare recommender systems and insights into recommendation scenarios and recommendation approaches	It provides insights on RS scenarios of food recommendation, drug recommendation, health status prediction, physical activity recommendation. Also, provide details about the application of the main recommendation techniques CF, CB, KB, HyR, and context-based.
Pincay et al. (2019)[28]	It conducts a systematic literature review and presents the results. It highligh the main techniques used and similarity analysis to build the recommendation engine	It mention initiatives of machine learning and fuzzy logic techniques as a way of achieving higher levels of accuracy and performance of the RS

Table 3.1: An overview of the existing literature review on health recommender system.

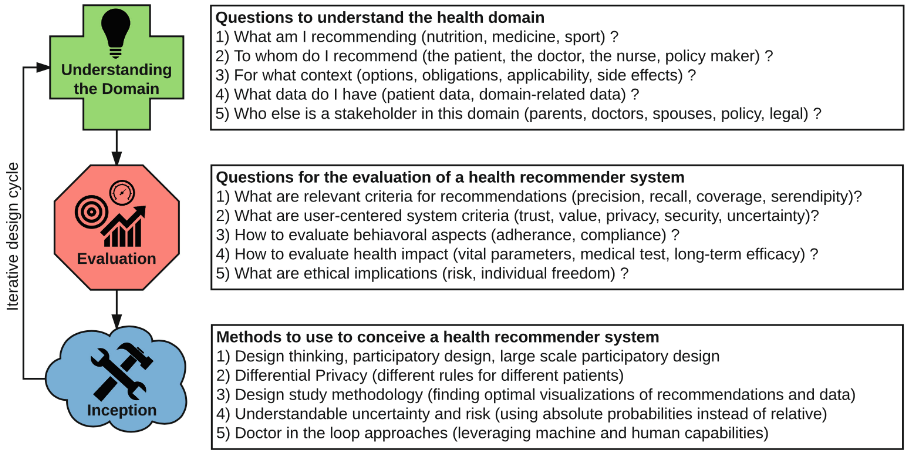

Figure 3.7: Framework with three iterative step proposed by Calero Valdez et al.[27].

De Croon et al.[26] also performed some recommendations regarding the best practices for future research on HRS. The frame for report is presented in Figure 3.8. It is requested the information about the target user, what item is recommended and how, information about database, outline the algorithms and techniques used as well as the methods of evaluation.

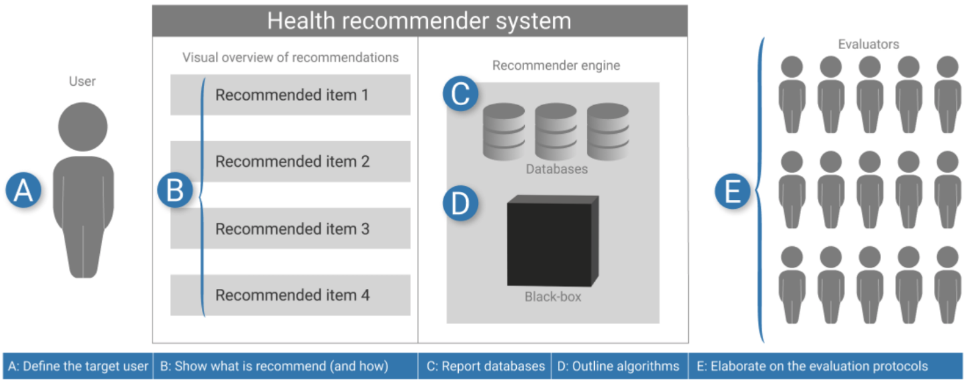

Figure 3.8: A reference guidelines for HRS studies sugested by De Croon et al.[26].

Pincay et al.[28] stated the relevance of using fuzzy logic and machine learning techniques in order to a achieve higher levels of accuracy on RS. Zhang et al.[29] also mentioned those methods as well as other AI techniques that might support the RS, as for example the usage of CNN's or auto-encoder for processing images.

Furthermore, Zhang et al.[30] proposed an a HRS called iDoctor offering users personalized doctor recommendations. The system integrated three approachs: sentiment analysis, topic modeling and hybrid matrix factorization.The sentiment analysis algorithm calculates the emotional offset from user reviews. The modeling module extracts user's preferences and doctor's features (e.g., specialty, fee range, and prescribing habits) from user reviews. The information is then used in the hybrid matrix factorization in order to predict the ratings.

Regarding the evaluation, the most common evaluation method applied in the aforementioned recommendation approaches is offline evaluation, estimating the prediction quality of a recommendation approach using existing datasets. This approach compares recommendations made by a recommender system with a predetermined collection of actual user feedback, or "ground truth," using accuracy metrics [13].

3.3.2.1 State of the art CBIR

In recent years, there has been a notable shift in content-based image retrieval systems towards the utilization of machine learning algorithms. This paradigm shift aims to develop models capable of effectively handling new input data and providing accurate predictions, thereby enhancing the overall image search process [22].

Traditionally, CBIR systems relied on handcrafted features and similarity measures to retrieve images based on their content. However, these approaches often faced challenges in adapting to new and diverse image datasets. As a result, the search accuracy and robustness of these systems were limited. To overcome these limitations, researchers have increasingly turned to machine learning algorithms, which can learn and adapt from data. By training models on large-scale image datasets, these algorithms can capture complex patterns and relationships within the data. This enables them to generalize demonstrating efficiency with new images and improve the accuracy of image retrieval [22].

Siradjddin et al. [31] contributed with their work on a Convolutional Neural Network (CNN)-based Autoencoder for feature extraction. The encoder layer has used the feature learning capability of CNN to extract important representations of images and reduce their dimensions. The decoder layer has reconstructed the representations to make the output of the Autoencoder closer to the input data. The extracted features from the encoder layer are used for content-based image retrieval by calculating the similarity distance between the query image's features and the database.

Moreover, Siradjddin et al. [31] have highlighted the influence of the number of features on the quality of the reconstructed images. Additionally, they emphasized that the number of layers within the encoder and decoder play an essential role in feature learning and extraction. Their experimental results demonstrated promising outcomes in terms of retrieving relevant images.

In a recent study, Agrawal et al. [32] proposed a similar model for content-based image retrieval system for medical images, specifically focusing on the retrieval of lung X-ray images for early detection and classification of lung diseases, including COVID-19.

Unlike previous methods that relied on Autoencoder for feature extraction, Agrawal et al. [32] utilized pre-trained deep neural models like VGG19 and ResNet50 to extract features from medical images and classify them into classes of diseases. Once the classifier was trained, the embedding vectors obtained from the feature extraction stage were utilized for image retrieval.

Besides, the feature extraction on CBIR, Hameed et al. [22] underlined the importance of similarity measurement on the the performance of image retrieval systems. The similarity measurement determines which images are considered most relevant to the query image and should be returned from the dataset, which means that the similarity measure determines the accuracy of the CBIR indirectly and has an effect on the computational complexity of the system.

3.3.3 Limitations and challenges

The literature review was focused on Heath Recommendation Systems and extended to Image Retrieval Systems. Topics involving Machine learning, Convolutional Neural Networks and Auto-encoders have a strong interface with the field of study and were partially evaluated.

The studies reported a few main challenges involving HRS, such as cold start and data sparsity, which are challenges well known in recommender system.

Zhang et al. [29] included data privacy as a major concern in HRS, as the vast amount of sensitive and personal information collected and processed by these systems can easily be misused if not properly protected. Ensuring data privacy and protecting user privacy while still providing accurate recommendations is a delicate balance that must be achieved in order to build successful and trustworthy recommendation systems. The increase of transparency of algorithms and the interpretability of the HRS can be seen as a challenge as reliable explanation about the recomentations can improve the confidence level of end-users on the System.

In the field of CBIR, Hameed et al. [22] stated that a significant challenge is the semantic gap that exists between the high-level meaning conveyed by an image and the low-level visual features that CBIR algorithms typically operates on. This gap arises due to the inherent complexity of interpreting and representing the rich semantics of images solely based on their visual characteristics. To address this challenge, extensive research efforts have been dedicated to bridge the semantic gap in CBIR. Researchers have explored various strategies and techniques to enhance the retrieval performance by incorporating novel and diverse features, as well as by leveraging the fusion of multiple feature representations.

3.4 Summary

In summary, RS play an essential role in providing personalized recommendations to users in various domains. Despite the significant progress made in recent years, there are still many challenges and limitations to be addressed, including scalability, privacy, and interpretability. This Chapter presents the progress and evolution of recommendation systems on healthcare. Besides the literature review presented, it also introduces the main challenges and limitations that HRS migth face. On the side of CBIR, it presents the new researches and architectures based on machine learning as an innovative approach as well as the semantic gap and main challenge for CBIR.

Chapter 4

Methods and Experimental design

This chapter is divided in two main parts. The first section presents the requirements for the system design, the steps involved, the definition and the characterization of the system. The second section provides details about potential models for search core algorithms and future implementation.

4.1 System Design Requirements

Initially, in order to design the HRS, a prior knowledge regarding the domain is mandatory. To support the system specifications, the framework proposed by Calero Valdez et al. [27] was used as reference.

4.1.1 Understanding the Domain

1. What items are being recommended?
   As the goal is to support on lung cancer characterization, the item of recommendation is a CT of Lung from past cases.

2. Who is the target user for the recommendation?
   Typically the RS is designed for an end-user, which could be the patient. But health recommender systems may extend their audience to health professionals. On this application, the target user are doctors, clinicians and radiologists

3. For what context?
   Initially, the context will be the semantic and structural similarity of an input image with the images from the database to retrieval the ones with the most similarities according to established factors. For personalising the recommendations, demographics information could be include as a variant factor. Information such age, ethnicity, gender, a flag if patient is a smoker or other could support the personalising tasks and provide different and relevant results for a specif input image.

4.1.2 Dataset

For the initial exploration the Lung Image Database Consortium(LIDC) and Image Database Resource Initiative (IDRI) dataset were selected. The LIDC was a result of the collaboration of seven academic centers and eight medical imaging companies [33]. This dataset contains information regarding the thoracic CT images data along with annotations of those images by experienced radiologists.

The LIDC/IDRI database contains 1018 thoracic CT scans and its associated XML-based annotations. It has been created to foster research and development of computer methods for lung nodule detection, classification, and quantitative assessment. In this database there are information about 2669 lesions marked as a nodule higher or equal than 3 mm by at least one of four radiologists and 928 lesions marked as such by all four radiologists. Each radiologist's annotations for these lesions include nodule outlines and subjective nodule characteristic ratings [33]. The images are similar to the one presented on Fig. 4.1

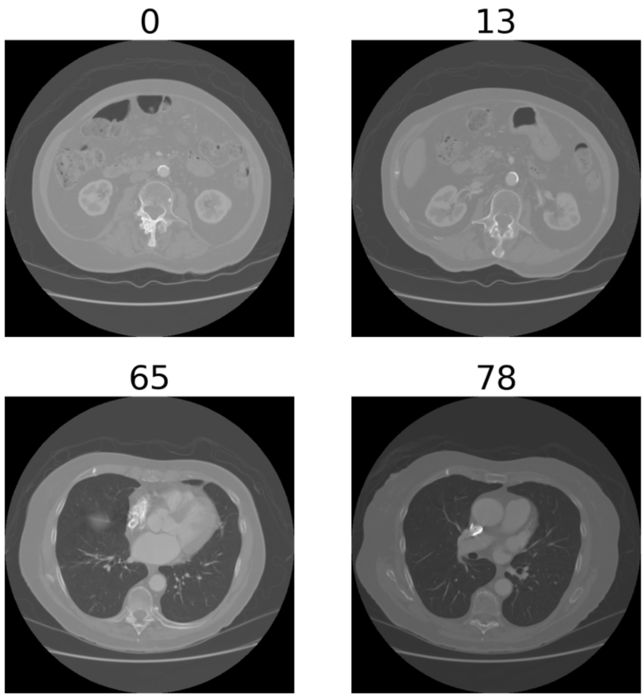

Figure 4.1: Slices of thoracic CT from a specific patient of LIDC.

The fields available on the annotations are presented on Table 4.1 proposed by Opulencia et al. [34] :

Besides the information available about the LIDC, an Exploratory Data Analysis (EDA) was performed in order to gain deeper insights into the main characteristics of the dataset, it generated a series of charts based on the annotations provided by the radiologists. These charts provide valuable visualizations that allow us to analyze and understand the key features and patterns present

Characteristic	**Description**	**Rating**
Calcification	Calcification appearance in the nodule—the smaller the nodule, the more likely it must contain calcium in order to be visualized. Benignity is highly associated with central, non-central, laminated, and popcorn calcification	Popcorn, Laminated, Solid, Non-central, Central, Absent
Internal Structure	Expected internal composition of the nodule	Soft tissue, Fluid, Fat, Air
Lobulation	Whether a lobular shape is apparent from the margin or not—lobulated margin is an indication of benignity	Marked, - - - None
Malignancy	Likelihood of malignancy of the nodule—malignancy is associated with large nodule size while small nodules are more likely to be benign. Most malignant nodules are non calcified and have speculated margins	Highly unlikely, Moderately unlikely, Indeterminate, Moderately suspicious, Highly suspicious
Margin	How well defined the margins of the nodule are	Poorly defined, - - - Sharp
Sphericity	Dimensional shape of the nodule in terms of its roundness.	Linear, Ovoid, - Round
Spiculation	Degree to which the nodule exhibits spicules, spike-like structures, along its border—spiculated margin is an indication of malignancy	Marked, - - - None
Subtlety	Difficulty in detection—refers to the contrast between the lung and its surroundings	Extremely subtle, Moderately subtle, Fairly subtle, Moderately obvious, Obvious
Texture	Internal density of the nodule—texture plays an important role when attempting to segment a nodule, since part-solid and nonsolid texture can increase the difficulty of defining the nodule boundary	Nonsolid, - Part-solid/mixed, - Solid

Table 4.1: An overview of LIDC Nodule Characteristics and definitions proposed on [34].

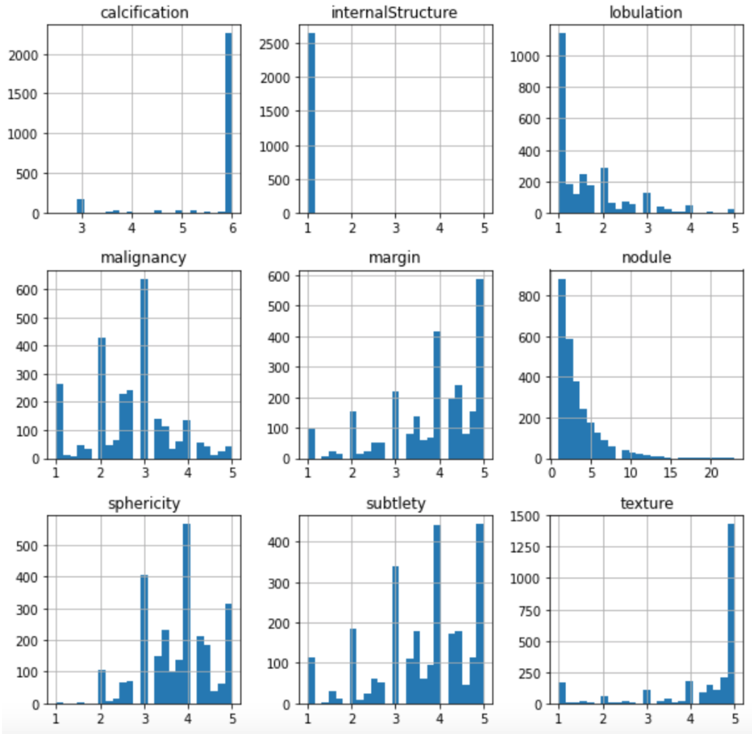

Figure 4.2: Histogram of field from CT annotations.

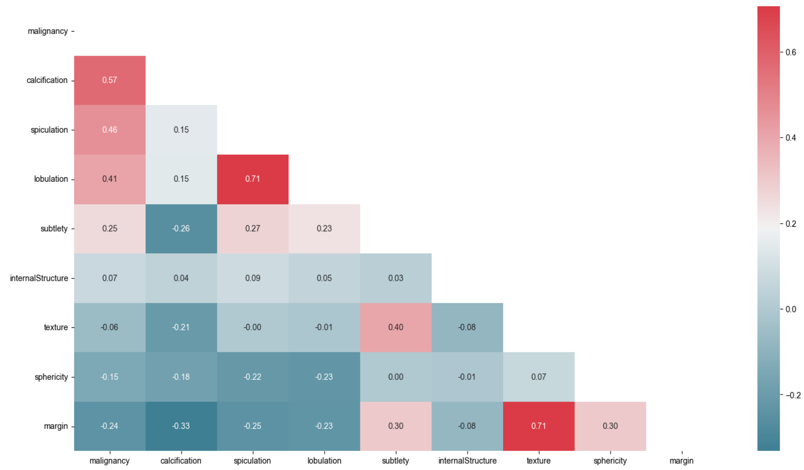

Figure 4.3: Correlation matrix of the Medical annotations.

in the dataset. The Figure 4.2 presents the histograms depicting the distribution of these key characteristics. These visualizations provide a comprehensive overview of the frequency and range of values associated with each variable. Furthermore, to explore the interrelationship between variables, various data manipulation techniques were employed. This analysis is illustrated in Figure 4.3, which highlights the correlations and dependencies among the variables under investigation. These visual representations serve as valuable tools for comprehending the relationships within the dataset.

4.1.3 Data Transformation

In this section, it is discussed the data transformation process applied on LIDC images in order to verify and compare images by region of interest (ROI) of the object of study, in this cases the lung nodules. The goal was to analyze and compare specific regions of interest within different images to determine their level of similarity. This is a essential process in various applications, including image retrieval, object recognition, and content-based image analysis.

To facilitate the comparison of specific regions within images, a cropping technique was applied to isolate the desired portions. Cropping involves selecting a rectangular or an irregular shaped region of interest (ROI) within an image and extracting only that region for further analysis. The selection of ROIs was guided by the annotations made by radiologists.

A Python script was implemented to leverage the LIDC framework for image analysis. The script successfully accessed the LIDC dataset and extracted the coordinates of the nodules annotated by radiologists. Subsequently, for each patient, the script cropped the images around the nodules, but only for the images where nodules were detected. For a same nodule that appears in multiple CT, it was selected coronal slice the one with more area on the annotation by the radiologist. To ensure consistency and compatibility across the dataset, the original images were resized from their original size of 512x512 to a standardized shape of 128x128. Furthermore, the images were converted to the Portable Network Graphics file format (.png). This process resulted in the generation of a new dataset consisting of images of the cropped region with nodule in the center.

4.2 Methodology

4.2.1 Initial system architecture

In this subsection, it is presented the propose architecture of the recommender model and provide an overview of system's design characteristics. The proposed architecture, illustrated in Figure 4.4, has been developed based on the requirements outlined in the previous section.

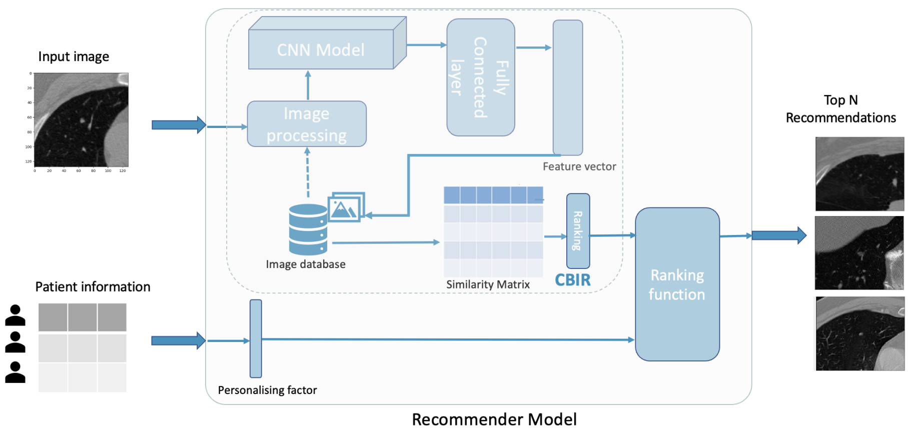

Figure 4.4: Proposed architecture for recommender model.

The design of the HRS incorporated a fundamental mechanism that involves taking an input image and leveraging a Convolutional Neural Network to extract characteristics and features from the image. This process is analogous to content-based image retrieval techniques, where the visual content of the image is thoroughly analyzed to identify and retrieve similar cases. Additionally, an alternative approach that can be explored within the HRS framework is the utilization of text-based retrieval methods. By incorporating text-based information, such as medical annotations or textual descriptions, the system can enhance the retrieval process by considering both visual and textual similarities between images.

Next, the HRS can generate a ranking of retrieved images from database based on their similarity to the input. In addition to the layer of image retrieval, a personalization factor can be incorporated to further enhance the relevance of the retrieved images. This personalization factor could take into account additional characteristics such as gender, age, smoking history, presence of multiple nodules, and other relevant patient-specific information. By considering these personalized factors, the system can recommend images that not only share visual similarity with the input image but also align with the specific characteristics and context of the patient being evaluated.

Once the architecture of the HRS is pre-established, the next step involves developing the content-based image retrieval component that can be integrated into the system for recommending past cases.

On the next sections, an Image Retrieval system is proposed and tested as a core engine for the recommendation system. This study is divided into three sub-sections, each focusing on a specific model development as a potential core component of the recommendation engine. For each sub-section, a brief discussion is provided .These analyses serve as a guide for improving different aspects of the solution in the subsequent iterations.

4.2.2 Model 1 - Text-based on medical annotations

In this first model, the lung nodule annotations performed by radiologists were explored. The objective is to implement a model that leverages the medical annotations provided to enable accurate and targeted retrieval of relevant medical cases based on the similarity of the features annotations. The Figure 4.5 illustrates the initial proposed model.

The features annotations available on LIDC dataset and utilized on this approach were: Subtlety, Internal Structure, Calcification, Sphericity, Margin, Lobulation, Texture and Malignancy. The annotations were transformed into a structured representation, specifically a dataframe, where each row corresponds to a patient's annotated nodule provided by a radiologist and the columns represent the various annotated features. Prior to the computation of similarity, these features were normalized within the range of zero to one during a pre-processing step.

4.2.2.1 Similarity/Dissimilarity

In the context of image retrieval systems, the performance is significantly influenced by the method used to measure similarity.

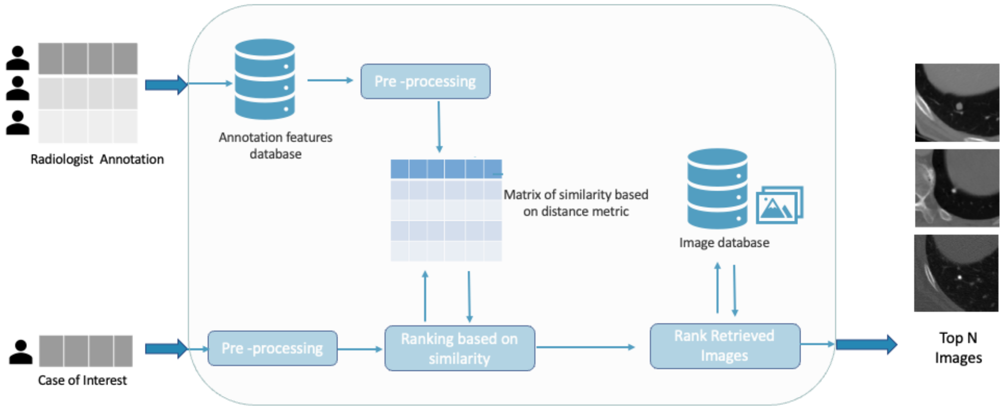

Figure 4.5: Model based on medical annotation.

The similarity measurement determines which images should be retrieved from the dataset [22]. There are two main categories of similarity measures: distance measures and similarity metrics. Distance measures quantify the dissimilarity between two annotations by calculating the distance between their feature vectors. In this study, three commonly used measures will be highlighted: Manhattan distance, Euclidean distance, and Cosine similarity. These measures provide different ways to assess the similarity between the feature vectors and offer distinct perspectives on the relationships among images.

For Manhattan distance (also known as L1 distance) and the Euclidean distance (also known as L2 distance) was utilized the Minkowski family of distances. Both are included in the general formula as follows:

$$Minkowski(x,y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

The feature vectors for the calculation distance are represented by X (x_1,x_2,x_3, ...,x_n) and Y (y_1,y_2,y_3, ...,y_n). When parameter p=1 it is equivalent to the Manhattan distance and when p=2 is equivalent to the Euclidean distance. The Minkowski distance encompasses both of these distances as well as other variations when p takes different values. For these distance measures, the smaller the distance between two images, the more similar they are considered to be in terms of their features.

Cosine similarity calculates the cosine of the angle between two vectors, providing a measure of their similarity. A larger value indicates a higher degree of similarity between the vectors. [35].

$$CosineSimilarity(X,Y) = \frac{X \cdot Y}{\|X\|.\|Y\|}$$

4.2.2.2 Ranking results

The rank of similar annotations is obtained by utilizing a similarity search algorithm called K-Nearest Neighbor (KNN). This algorithm calculates the distance between two features based on a similarity metric and returns the k most similar features. In this study, the retrieval process was implemented and tested using all the distance metrics mentioned earlier.

The results of the retrieval process provided a ranking of similarity based on the medical annotations. For example, a nodule annotation retrieved with smaller euclidean distance are ranked higher, indicating a higher level of similarity to the query image. This ranking enables the presentation of the most relevant and similar cases at the top of the recommendation list. To enhance the relevance of the retrieved similar images, a threshold technique could be applied. By setting a similarity threshold, only images that reached the threshold are considered as similar or relevant for retrieval system.

Below, on Figure 4.6 it is presented an example of the query annotation and the annotations results. On the Figure 4.7, demonstrate the respective images from the previous annotation query.

Query annotation:

	patient_id	subtlety	internalStructure	calcification	sphericity	margin	lobulation	spiculation	texture	malignancy	pid_nod
30	LIDC-IDRI-0033	0.687500	0.0	1.0	0.687500	0.562500	0.125000	0.062500	0.750000	0.5	LIDC-IDRI-0033-1

Results annotation:

	patient_id	subtlety	internalStructure	calcification	sphericity	margin	lobulation	spiculation	texture	malignancy	pid_nod
233	LIDC-IDRI-0251	0.687500	0.0	1.0	0.750000	0.562500	0.062500	0.125000	0.812500	0.5	LIDC-IDRI-0251-1
796	LIDC-IDRI-0916	0.625000	0.0	1.0	0.687500	0.625000	0.125000	0.125000	0.875000	0.5	LIDC-IDRI-0916-1
393	LIDC-IDRI-0440	0.625000	0.0	1.0	0.687500	0.500000	0.187500	0.187500	0.750000	0.5	LIDC-IDRI-0440-1
742	LIDC-IDRI-0848	0.666667	0.0	1.0	0.666667	0.416667	0.250000	0.083333	0.750000	0.5	LIDC-IDRI-0848-1
387	LIDC-IDRI-0434	0.625000	0.0	1.0	0.750000	0.500000	0.250000	0.187500	0.750000	0.5	LIDC-IDRI-0434-1
521	LIDC-IDRI-0591	0.833333	0.0	1.0	0.583333	0.500000	0.166667	0.083333	0.833333	0.5	LIDC-IDRI-0591-1

Figure 4.6: Results annotation from Query patient.

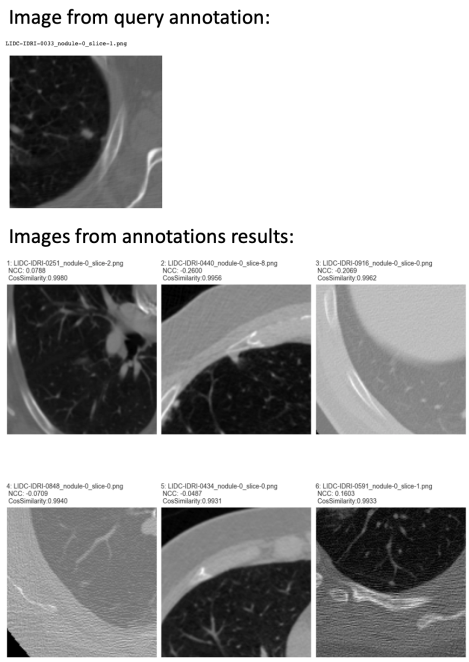

Figure 4.7: Image results for query annotation.

This approach performs an interesting retrieval specially because it is based on the similarity of radiologist's annotations, leading to retrieval similar annotation cases. An enhanced approach for this model could utilize the potential uses of weight euclidean distance, where it is possible to attribute distinct weight according to the importance of features considered by the user. A more detailed analysis of the results of this model is discuss on the next chapter.

Additionally, it's important to note that the success of this approach relies on the quality and consistency of the radiologist annotations. Furthermore, it is necessary to consider the limitations and variability in radiologist interpretations when inferring similar results. A dependency of the existing annotation is also a limitation on this model, since it is only possible to perform a search when the annotation are available, which in some cases are very time consuming or might have not a clear consensus on the annotation. Presenting this limitation, a new approach is explored focusing on content image analysis of the nodule patients for similar cases retrieval.

4.2.3 Model 2 - Convolutional Autoencoder architecture

This section presents a methodology employed focusing on the utilization of Autoencoder as a core engine component of an image recommender system. The proposed architecture builds upon a previous research reported by Siradjuddin et al. [31] and aims to leverage the potential of Autoencoder in extracting essential structural features from the dataset images.

The Figure 4.8 illustrates the proposed model, wherein a Convolutional Autoencoder architecture is employed. This architecture is responsible for extracting and compressing the key structural features of the dataset images into a latent space representation, which subsequently allows the reconstruction of the original images. The encoder component of the Autoencoder serves as a feature extractor, with the objective of processing the images and generating their respective latent representations. This latent space retains the intrinsic features of the images and serves as a mechanism for indexing and comparing the dataset images, thereby facilitating efficient image retrieval from the CT image database.

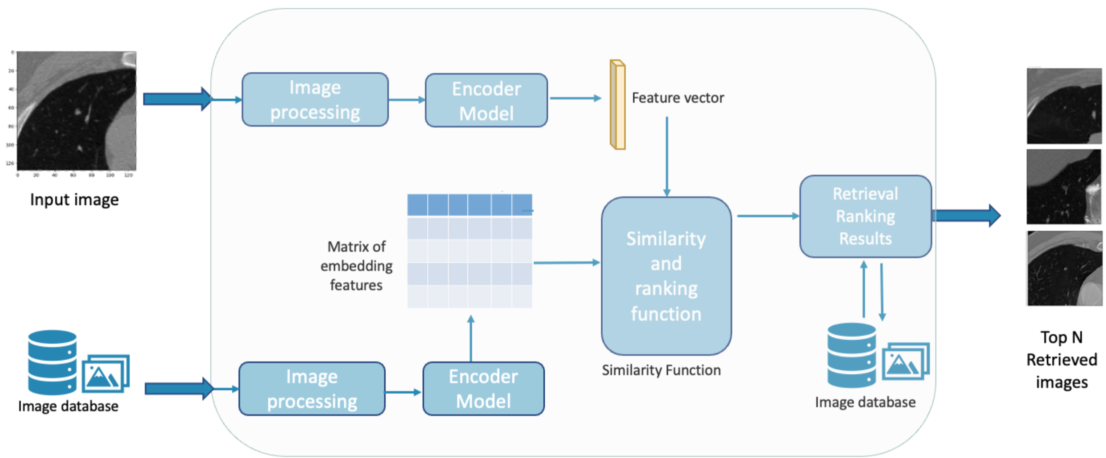

Figure 4.8: Convolutional Autoencoder model.

The Convolutional Autoencoder employed in this study undergoes a self-training process on the pre-processed images of LIDC dataset. The model is designed to learn the ability to compress the input images into a lower-dimensional latent space representation and capturing the essential features of the images. The encoder component of the Autoencoder assumes the responsibility of mapping the high-dimensional input image to a lower-dimensional latent space representation. This latent space representation is situated within a bottleneck layer, reducing the dimensions of the data. On the other hand, the decoder component is responsible to reconstruct the input image based on the information encoded in the latent space [36]. Consequently, the latent vector serves as a compact representation of the original high-dimensional input. The architecture of the Convolutional Autoencoder is outlined on Figure 4.9 , where it is illustrated the components and their role in the compression and reconstruction process.

After encoding the images, the similarity between two images can be determined by comparing their respective latent space representations. Various distance metrics, such as Euclidean distance or cosine similarity, can be employed for this purpose. By calculating the distance or similarity score between the latent representations of two images, we can quantitatively measure their level of similarity.

To retrieve similar images on visual and semantic aspect for a given query image, a nearest neighbor search is conducted in the latent space. The query image is encoded using the trained Autoencoder, generating the corresponding latent representation. Subsequently, the similarity between the query image's latent representation and the latent representations of all other images in the dataset is computed. The images with the closest latent representations to the query image are considered the most similar.

For Euclidean distance, the minimum distance value signifies the highest degree of similarity. The retrieved nodules are presented to the user in descending order based on the similarity measure. This ranking facilitates the presentation of the most relevant and visually similar images first. The results visualizations may include displaying the query image alongside its top-ranked similar images, enabling users to visually compare and assess the level of similarity between the query image and the retrieved images, similar to the retrieval from previous model, as illustrated on the Figure 4.7.

4.2.3.1 Architecture and Training

The proposed architecture for the Convolutional Autoencoder (CAE) in this study is illustrated in Figure 4.9. The encoding phase of the CAE comprises a series of convolutional layers, followed by the application of the non-linear activation function ReLU, and downsampling operations such as max pooling with a 2x2 kernel size as illustrated on 4.2. Throughout the encoding phase, the number of channels in the convolutional layers typically increases while the spatial dimensions of the feature maps gradually decrease. This gradual compression of information in the image allows for the learning of low-dimensional representations of nodule images. In the initial architecture, the size of the layers was configured to generate a latent vector of size 4096 (16 x 16 $\times$ 16) which serves as the compressed representation of the input image thought convolutional layers. The initial experiments were performed using the parameters presented on 4.2. Besides this initial setup layer, different configuration for layers were also tested.

Table 4.2: Parameters for the initial architecture of Convolutional Autoencoder.

Layer	**L1**	**L2**	**L3**	**L4**	**L5**	**L6**	**L7**	**L8**	**L9**
Type	conv	pool	conv	pool	conv	pool	deconv	deconv	deconv
Channel	16	-	32	-	32	-	32	32	16
Kernel size	3x3	2x2	3x3	2x2	3x3	2x2	2x2	2x2	2x2
Stride	1	2	1	2	1	2	2	2	
Padding	1x1	-	1x1	-	1x1	-	-	-	-
dilation	-	1	-	1	-	1	-	-	-
Activation	ReLU	-	ReLU	-	ReLU	-	ReLU	ReLU	ReLU

The decoding phase of the CAE aims to reconstruct the image from the latent vector. It involves gradually increasing the spatial dimensions of the feature maps while decreasing the number of channels, enabling the decoder to properly reconstruct the original image. Similar to the encoder, non-linear activation functions such as ReLU may be applied after the deconvolutional layers to introduce non-linearity into the reconstruction process. Through this encoding-decoding process, the CAE learns to capture and preserve the essential features of the input images in a compressed latent space representation.

The output of the decoder in the Autoencoder model possesses the same dimensions as the input image. By training the model with an undercomplete representation, it is imposed a constraint that encourages the Autoencoder to learn the most significant and essential features present in the training data. This constraint is achieved by imposing limitations on the activity of the hidden representations within the model [37]. When the architecture is appropriately adjusted by incorporating regularization terms, the Autoencoder might acquires additional qualities beyond its ability to replicate the input image as the output. These additional qualities include the extraction and encoding of the most salient features from the input data, which can enhance its capacity for feature representation.

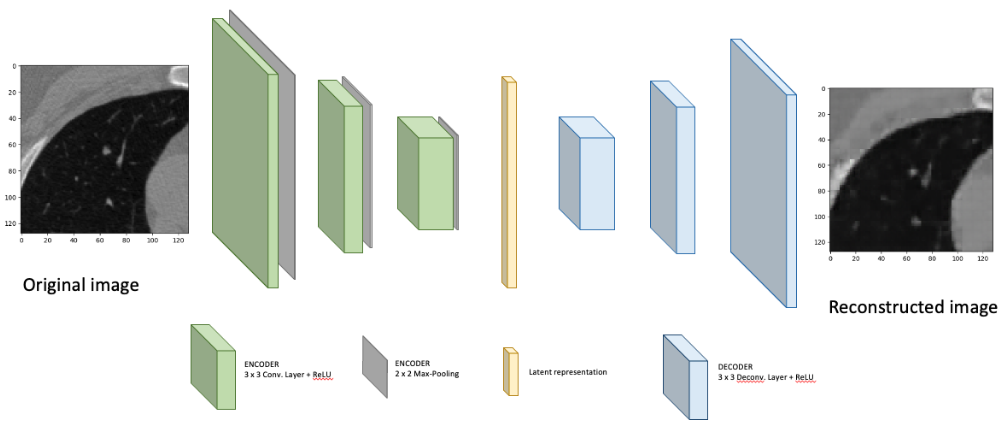

Figure 4.9: Convolutional Autoencoder architecture.

The objective of training is to minimize the reconstruction loss. The loss function used for the Autoencoder model was the Mean Squared Error (MSE), which measures the average squared difference between the input data and the reconstructed output. Mathematically, the MSE loss is defined as:

$$MSELoss = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{4.1}$$

where N represents the number of data samples, x_i denotes the original input data, and $\hat{x}_i$ represents the corresponding reconstructed output. The MSE loss quantifies the dissimilarity between the input and output, with a lower value indicating a better reconstruction performance. Minimizing the MSE loss during the training process enables the Autoencoder to learn the most representative latent features and effectively reconstruct the input data.

The training of the Convolutional Autoencoder was conducted utilizing two optimization algorithms: stochastic gradient descent (SGD) and the AdamW optimizer. SGD is widely recognized as the go-to algorithm for effective training of Convolutional Neural Networks (CNNs). It excels in learning discriminative linear classifiers when applied to convex loss functions, rendering it a popular choice in the field. On the other hand, AdamW, an improved variant of the Adam optimizer, was employed. Introduced by Loshchilov and Hutter in their work titled 'Decoupled Weight Decay Regularization', AdamW rectifies the Weight Decay Regularization issue in Adam, leading to models with superior generalization capabilities [38]. As a result, AdamW can effectively compete with SGD while significantly speeding the training process.

The entire set of nodules was divided into a training set (80%) and a testing set (20%), and the split was employed at nodule-level and patient-level, i.e slices of the patient's nodule were only used for training or only for testing, to prevent data leakage.

Table 4.3: Hyperparameters for Convolutional Autoencoder.

Hyper-parameters	**Values**
Batch size	16, 32, 64
Learning Rate	1e-2, 1e-3, 1e-4
Loss Function	MSE, L1
Optimizer	AdamW, SGD

The hyper-parameters selected for this experiment were illustated on Table 4.3 The best results were achieved when using batches size of 32,learning rate of 1e-4 with Loss function MSE and AdamW as optimizer.

During the traning, it was also evaluated the impact of changing the latent representation size, specifically exploring the range of 512, 1024, 2048, and 4096 as the output dimensions of the Convolutional Layers. The focus was mainly on testing three-layer configurations with different setups for the first and second output layers, such as 32,64, 16,64, and 16,32 . By altering the size of the last output layer, it is possible to control the dimensionality of the latent representation.

The Figure 4.10 illustrates the impact of the latent layers in the reconstruction of an image using the Convolutional output for the first and second layers with a configuration of (32, 64, x), where x is variable to adjust the size of last layer, manipulating the latent representation size. It demonstrates how varying the last output layer allows changes in the size of the latent representation.

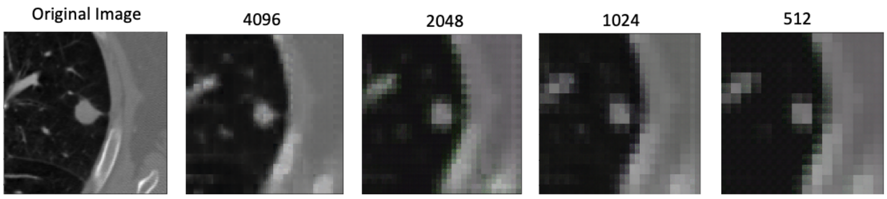

Figure 4.10: Reconstructed images for different latent representation sizes.

4.2.3.2 Retrieval and Relevant items

For retrieval of similar images, the algorithm used was the K-Nearest Neighbours (KNN).The KNN is a well-established technique for measuring similarity between data points based on their feature vectors. In this study, the KNN is employed to build a model that identifies the K most similar images to a given query image. By selecting the K nearest neighbors, it can retrieve images that share similar visual characteristics with the query image[39].

The proposed algorithm uses the information provided by the encoder's latent representation to perform the similarity search. The K-NN combined with the encoder will be able to:

1. Generate a Matrix of latent representation of all the images of the dataset.
2. Process the query image as an input and calculate the latent representation of the query image.
3. Measure the distance between the query latent representation and the latent representation of the images in the dataset. In order to calculate the similarity, the Euclidean distance will be applied.
4. Retrieve K nearest images based on the distance measured in the previous step.

Initially, a range of k values from 1 to 6 will be experimented with. One of the noteworthy features of this approach is that when the nearest neighbors are used, the retrieved images are ordered based on their similarity. This allows the system to rank the retrieved images, enabling the possibility of providing recommendations.

Regarding the evaluation of the recommendation system, the accuracy of the retrieved images needs to be assessed. However, this measurement cannot be directly obtained due to the unique characteristics of the system and the absence of labeled classes in the dataset. In such cases, alternative evaluation metrics and methodologies need to be employed to gauge the effectiveness of the recommendation system.

The results of this model will be analysed and discussed in detail in the next chapter. However, as a summary, one of the main challenges encountered during the evaluation was the semantic gap between the features learned by an Autoencoder and manual medical annotation. It revealed limitations on assessing features such as malignancy, spheracity, subtlety, margin as well as others that required high expertise on medical domain.

The decrease of the semantic gap is an ongoing research topic, and the effectiveness of these techniques can vary depending on the specific dataset and task. It is recommended to experiment with different approaches and evaluate their impact on reducing the semantic gap in each particular use case. In this study, it was decided to proceed with an implementation of a supervised learning architecture based also on the annotations to attempt to mitigate this gap.

4.2.4 Model 3 - Hybrid Autoencoder architecture

This section introduces a new architecture proposal intended to address the limitation of the previous model, specifically targeting the semantic gap issue identified. The proposed architecture builds upon the principles of Supervised Autoencoder, drawing inspiration from prior research reported about personalized classification model using similarity learning via Supervised Autoencoder by Jo and Jun [40].

In the previous experiment, although the retrieved images exhibited a general similarity in terms of their overall structure, a deeper analysis revealed that the medical annotations associated with those images were not sufficiently similar to the query image. This phenomenon highlights the presence of a semantic gap, which refers to the disparity between the high-level medical semantic concepts contained on the annotations and the low-level visual features from retrieved image [41].

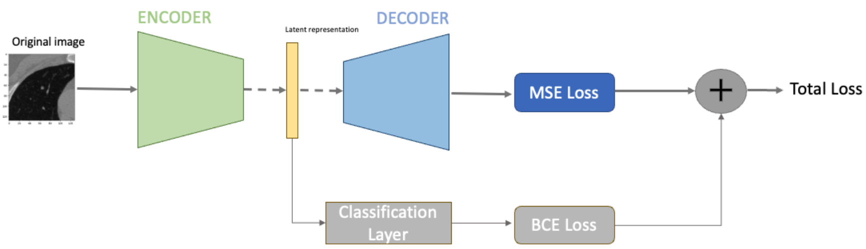

Figure 4.11: Supervised Autoencoder architecture.

The primary objective of this new architecture is to enforce the latent vector to preserve the desired features of the input images and minimize the semantic gap by incorporating a supervised layer. To achieve this, a new mechanism was implemented to indirectly manipulate the latent space, aiming to retain the desired information within it.

The proposed solution performs an adjustment on the loss function. The proposed solution included the implementation of a loss function that aimed to learn image features while aligning them with manual medical annotations. To achieve this, it was included a new classification layer on the network and it was performed a combination of reconstruction loss and classification loss, as it is demonstrated on Fig. 4.11.

The purpose is that the reconstruction loss measures the difference between the original image and its reconstructed image, ensuring that the Autoencoder learns to preserve important features. On the other hand, classification loss aims to reduce the discrepancy between the learned features and the desired feature indicated by the annotations. On the initial settings was chosen the feature Malignancy to be included on the model. For the target class, a rank of malignancy 1 and 2 it was labeled as benign.For rank of malignancy 4 and 5 as malignant, 3 is neglected for this initial evaluation and further scenarios could be explored [42] .

In terms of retrieval images, the algorithm was designed to work similar to previous model. It generates a Matrix from process of transforming images into latent representations, calculates the distance between the query image and dataset images, and retrieving the K nearest images based on their similarity to the query.

The design of this new system is centered around the preservation of the original image output, while disregarding the output of the classifier. The primary objective is to retrieve the most similar images, thereby reducing the existing semantic gap. By focusing solely on image similarity, the system aims to prioritize the retrieval of visually and structurally similar images, aligning with the intended purpose of the image retrieval task.

However, it is worth noting that in certain scenarios, the incorporation of a classifier can provide additional benefits. For medical professionals, the classifier's output can serve as valuable support in the interpretation and explanation of the retrieved images. In these specific cases, the classifier's output can be utilized to provide relevant insights and explanations to the medical practitioners. This integration of classification and retrieval serves to enhance the interpretability and comprehensibility of the image retrieval system, providing valuable support and explanations to the medical practitioners.

4.3 Evaluation Metrics

In the context of content-based image retrieval for medical image similarity, several evaluation metrics are commonly used to assess the performance of retrieval algorithms. These metrics provide insights into the effectiveness and efficiency of the retrieval system.

Precision (P) is the ratio of the number of relevant images within the first k results to the total number of images that are retrieved and is expressed as follow, (1) where TP refers to the relevant images retrieved and FP refers to the false positive, i.e., the images misclassified as relevant images. Precision quantifies the system's ability to retrieve only relevant images, disregarding any irrelevant ones.

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

Precision@K is a metric that focuses on evaluating the precision of the top K retrieved images. It measures the proportion of relevant images among the top K retrieved results. Precision@K is particularly useful when users are only interested in examining a limited number of top-ranked images.

$$Precision@k = \frac{Number\ of\ relevant\ items\ in\ top\ k}{k}$$

Average Precision (AP) for a single query for k items is obtained by taking the mean over the precision values at each relevant image, providing a comprehensive evaluation of the system's performance. It considers the precision at various levels and provides a summarized measure of the system's retrieval quality:

$$AveragePrecision = \frac{1}{n}\sum_{k=1}^{n} P(k) \cdot rel(k)$$

Where:

n: Total number of retrieved instances.

P(k): Precision at the k-th position in the ranked list when the k-th instance is retrieved.

rel(k): An indicator function that is 1 if the k-th instance is relevant, and 0 otherwise.

Normalized Cross-Correlation (NCC) is a valuable metric used in image analysis and has found applications in the medical field. The NCC, along with its variations, is employed to estimate the similarity between images [43]. This metric considers both spatial and intensity information present in the images. It provides a similarity score ranging from -1 to 1, where a value of 1 indicates perfect similarity and -1 indicates perfect dissimilarity. In the context of this study, a range of 0.5 to 1 can be considered as indicative of a moderate to strong correlation between images and utilizes the formula below:

$$NCC(I_1, I_2) = \frac{\sum_{i,j}(I_1(i,j) - \bar{I}_1)(I_2(i,j) - \bar{I}_2)}{\sqrt{\sum_{i,j}(I_1(i,j) - \bar{I}_1)^2 \cdot \sum_{i,j}(I_2(i,j) - \bar{I}_2)^2}}$$

4.4 Hypothetical relevant images

One of the limitations encountered in content-based image retrieval for the LIDC dataset was a lack comprehensive ground-truth that accurately represents relevant images for a given query. This limitation makes it challenging to evaluate and analyze the effectiveness of the proposed models solely based on ground-truth annotations.

By acknowledging the need for expert medical validation, the proposed strategy aims to bridge the gap between computational retrieval methods and the expertise of medical specialists. While the models can retrieve images based on their visual similarity to a query image, the final interpretation and validation of the results heavily rely on the assessment of a medical specialist.

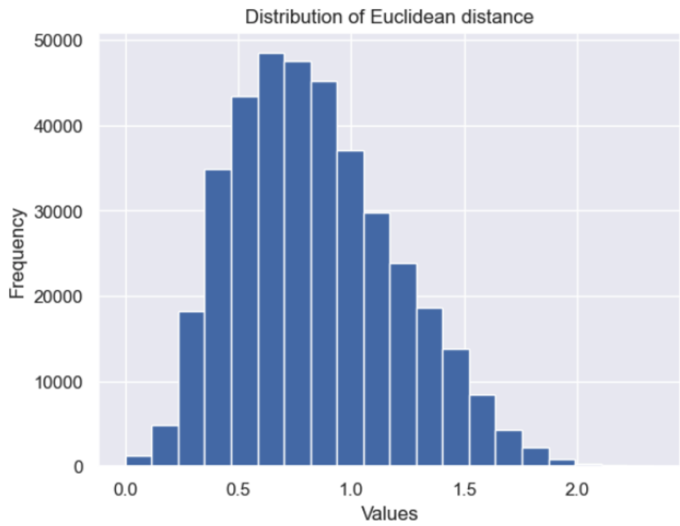

Figure 4.12: Distribution of the Euclidean distance for the top 6 retrieval.

To enable a meaningful comparison between the proposed models and deeper analysis, a hypothetical scenario is introduced. This scenario establishes a similarity threshold based on the Euclidean distance between the query image's annotation and all the images in the dataset. A distance of medical annotations was calculated between the query image and the first 6 retrieval.It was performed this calculation for all images on the dataset and the Figure 4.12 represents this distribution. For the Euclidean distance, a lower distance value signifies a higher level of similarity between two entities. It was established the value that retrieves the 20% lowest values for the distribution.Thus, the euclidean distance on the query below this value of 0.517 would be considered hypothetically relevant for the retrieval analysis and model comparison.

Ultimately, this strategy aims to provide a foundation for evaluating and comparing the performance of the models by considering the hypothetical relevance of retrieved images, thereby facilitating a more in-depth analysis and interpretation of the results.

4.5 Summary

In summary, on this Chapter is presented an initial architecture for the Health Recommender System, along with the development of three models: medical annotation-based, Convolutional Autoencoder, and Supervised Autoencoder. Acknowledging the limitations encountered during the study, various modifications were incorporated into the design to address the main challenges associated with image retrieval, with a specific focus on text-based retrieval and content-based image retrieval (CBIR) for the characterization of lung cancer cases.

Chapter 5

Experimental Results and Discussion

This chapter presents the results obtained for the proposed models for the core engine of the recommender system presented in the methods section.

5.1 Results and discussion of Text-based on medical annotations

In this section, it is presented the results of the study regarding the first proposed model, which aimed to retrieve similar images based on the similarity of medical annotation. The evaluation of the system involved several experiments and analyses with the objective of assessing its performance and effectiveness in retrieving relevant images.

The retrieval starts when a patient id and a nodule is selected. Then the system performs a search to find similar cases through the calculation of a metric distance between the feature vectors and return them by a ranking, according to the relevancy.

On this scenario three metrics were tested: Euclidean distance, Cosine similarity and the Manhattan distance.

For example, taking in consideration the retrieval from the image below Fig. 5.1 :

LIDC-IDRI-0033_nodule-0_slice-1.png

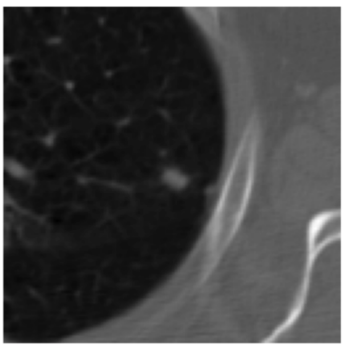

Figure 5.1: Example of a query image.

For this single instance, the retrieval results demonstrate that the first image was the same for all the three calculation methods. Additionally, out of the six retrieved images in each of the

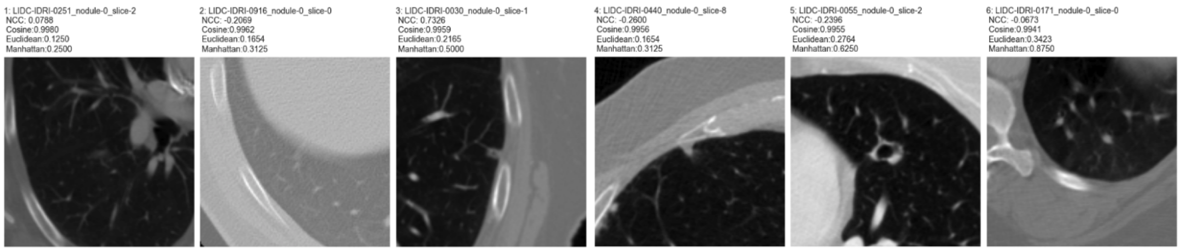

Figure 5.2: Retrieval images from Cosine similarity calculation.

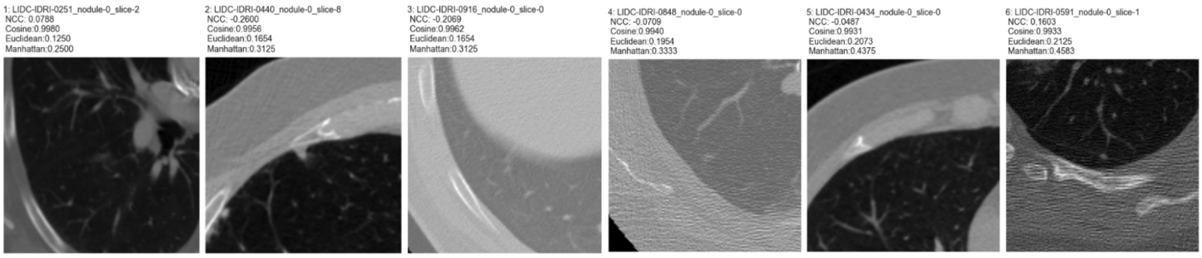

Figure 5.3: Retrieval images from Euclidean distance calculation.

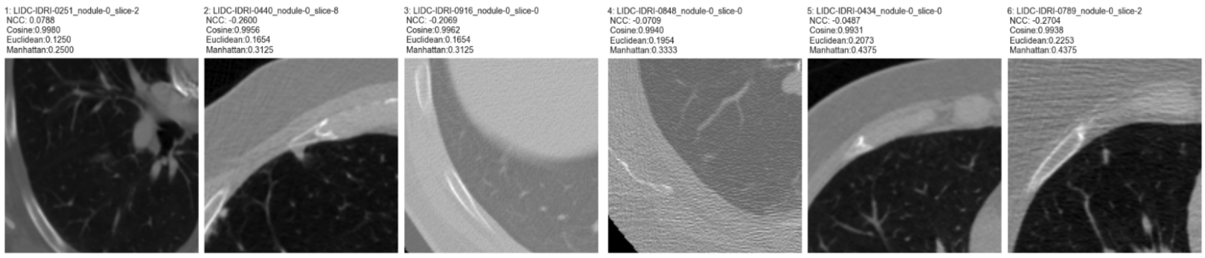

Figure 5.4: Retrieval images from Manhattan distance calculation.

calculation methods, three images were the same in all methods. Furthermore, the Euclidean and Manhattan distance presented similarity on the images retrieved and their order, with only the last image retrieved being different. The first image retrieval, the same for all methods, was able to match five of the nine features with zero difference, as it is illustrated on Table 5.5. In general,the three method presented some similarities on images results but, when analysing the results for the medical annotation similarity they presented a strong similarity (Table 5.5).

	patient_id	subtlety	internalStructure	calcification	sphericity	margin	lobulation	spiculation	texture	malignancy	pid_nod
30	LIDC-IDRI-0033	0.6875	0.0	1.0	0.687500	0.5625	0.125000	0.062500	0.750000	0.500000	LIDC-IDRI-0033-1
233	LIDC-IDRI-0251	0.6875	0.0	1.0	0.750000	0.5625	0.062500	0.125000	0.812500	0.500000	LIDC-IDRI-0251-1
796	LIDC-IDRI-0916	0.6250	0.0	1.0	0.687500	0.6250	0.125000	0.125000	0.875000	0.500000	LIDC-IDRI-0916-1
28	LIDC-IDRI-0030	0.6250	0.0	1.0	0.625000	0.5000	0.125000	0.000000	0.625000	0.375000	LIDC-IDRI-0030-1
393	LIDC-IDRI-0440	0.6250	0.0	1.0	0.687500	0.5000	0.187500	0.187500	0.750000	0.500000	LIDC-IDRI-0440-1
52	LIDC-IDRI-0055	0.7500	0.0	1.0	0.833333	0.7500	0.166667	0.083333	0.833333	0.583333	LIDC-IDRI-0055-1
164	LIDC-IDRI-0171	0.8125	0.0	1.0	0.812500	0.6875	0.187500	0.187500	0.937500	0.625000	LIDC-IDRI-0171-1

Figure 5.5: Medical annotation from retrieved images.

5.1.1 Overall evaluation

It was performed an overall analysis to evaluate the performance on retrieving other images of the dataset. To evaluate the overall retrieval based on semantic annotation similarity, it is important to define a ground truth or a set of relevant images for each query to perform an evaluation. With lack of this information, the analysis of the system is limited. For an accurate analysis it is necessary to involve medical experts to support the validation.

In the hypothetical scenario, to establish a reference performance for the proposed model, it was considered as relevant retrieval an annotation with a value below the define threshold (0.517), as discussed in the evaluation metrics section. The average distance for the annotation retrieval is 0.232 $\pm$ 0.121, calculated over a total of 4310 samples for k=6. This corresponds to an hypothetical precision of 97.7% based on medical annotations, which will serve as baseline for further comparisons with other models.

In terms of evaluating the structural image similarity, it was compared the six images retrieved from each patient and nodule using the three different algorithms. To measure the similarity between the images, it was employed the Normalized Cross-Correlation (NCC) method. The NCC operates within a range of -1.0 to 1.0, where a range of 0.5 to 1.0 was adopted as an indicator of a moderate to strong correlation between the image structures [44]. The average scores obtained from the retrieval of all images are presented in Table 5.1.

Table 5.1: Normalized Cross-Correlation from images.

Distance calculation	Average Precision based NCC score
Cosine	0.056
Euclidean	0.053
Manhattan	0.057

Overall, the three methods presented a similar average of NCC score between themselves. However, the average values presented were low, indicating poor structural similarity between the query image and retrieved images.

5.1.2 Model 1 with feature of structural enhancement

In the first model, the implementation takes into account the medical annotations and their respective images. To improve the retrieval based on structural similarity, an additional step on the retrieval algorithm was added that not only orders the images by their lower distance from the annotation but also considers the NCC score of each image. By incorporating this step, the retrieved images were different, emphasizing their structural composition, which in turn affects the Euclidean distance. The Figure 5.6 illustrates the variation and the structural changes of the retrieved images. This approach allows setting specific requirements in the retrieval process, making the results more aligned with the user's preferences or requirements.

The impact on performance can vary depending on the retrieval restrictions imposed. In the case of Figure 5.6, the requirement was to retrieve images with an NCC score of at least 0.4.

As a result, the percentage of images meeting this criteria increased from approximately 5.5% to 48.3%, while the average Euclidean distance increase to 0.353 ± 0.167 and the hypothetical precision based on annotations decreased from 97.7% to 85.5%. It is important to note that these are hypothetical scenarios used for model comparison purposes only. Expert medical evaluation is necessary to validate and assess the results obtained from the model.

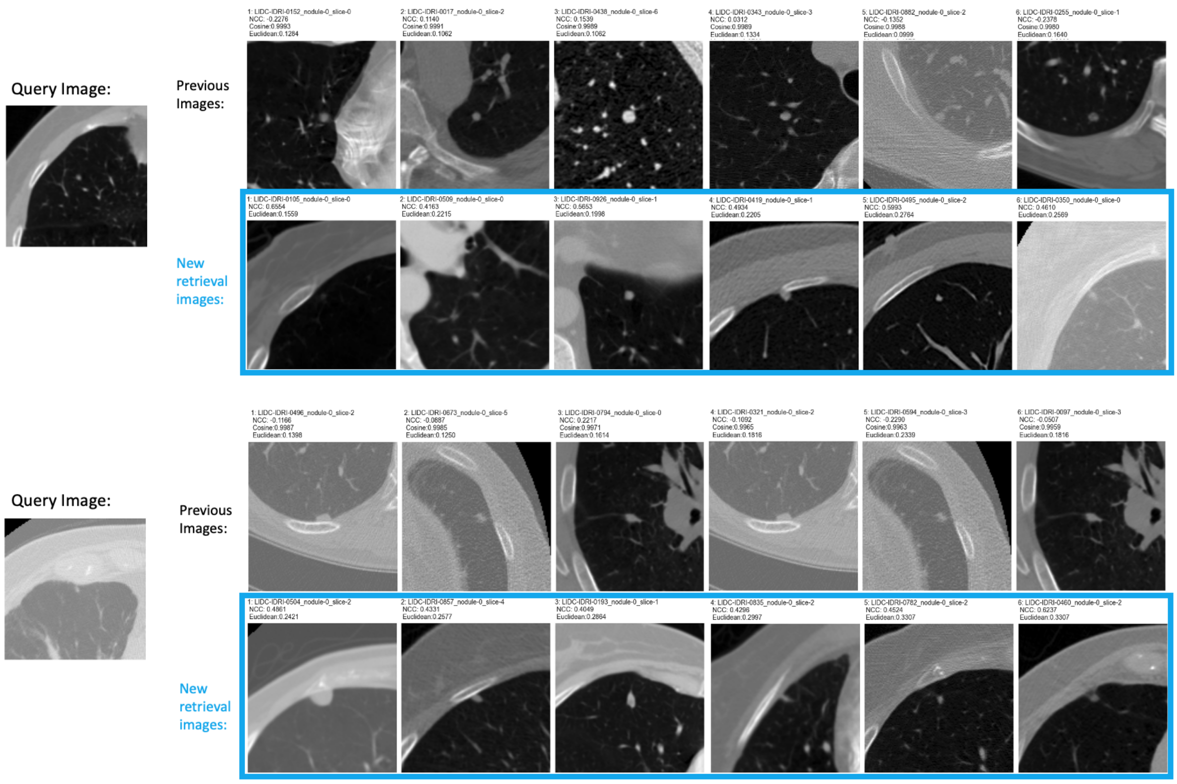

Figure 5.6: Retrieval images before and after the application of a new calculation layer based on NCC on the model.

More experiments were conducted with different images for the model 1 and besides the capability to retrieval images with similar medical annotation, an extensive study involving specialist is required to validate the results in terms of accuracy, precision and relevance of the model 1. Especially, because the similarity computed by NCC can be different in terms of medical structural similarity.

The obtained results indicated a weak to moderate capability of the model to retrieve images with similar structural composition of the query. However, as it accesses directly the medical annotations, it is able to retrieve similar cases based on the medical annotation.

Considering the specific requirements of the medical domain, alternative approaches can be explored. For instance, if certain features have different degrees of relevance compared to others, a distance calculation based on the Weighted Euclidean distance could be employed. This approach would enable the retrieval system to prioritize and emphasize the search based on one or more specific features. Additionally, in order to address the aspect of structural similarity in images, the inclusion of convolutional neural networks (CNNs) becomes a viable option. By incorporating

CNNs into the model architecture, it becomes possible to leverage their capacity to learn and extract relevant image features, without the need of medical annotations to perform the retrievals.

5.2 Results and discussion of Convolutional Autoencoder

In this section, the results of the second proposed model are presented, which aimed to retrieve similar images based on the similarity of the encoder's latent representation of the image.

The Convolutional Autoencoder (CAE) was trained in various scenarios, hyper-parameters, and latent representation sizes, as described in the previous section. As it is shown in the Figure 5.7, the training phase did not suffered from any visible problem that could be spotted in its loss values, as the final loss of the reconstruction dropped and stabilized with batch size 32, using loss function MSE and optimizer AdamW. A mechanism of early stopping was implemented to prevent the model from overfit. This mechanism stops the trainning when it has passed 10 Epochs without a reconstruction loss improvement on validation set. For Learning rate 1e-2, it can be seen this strategy of stopping the model close to Epoch 50, as illustrated the Figure 5.8, while with the learning rate of 1e-3 it was stopped on the Epoch 250 as presented in the Figure 5.7

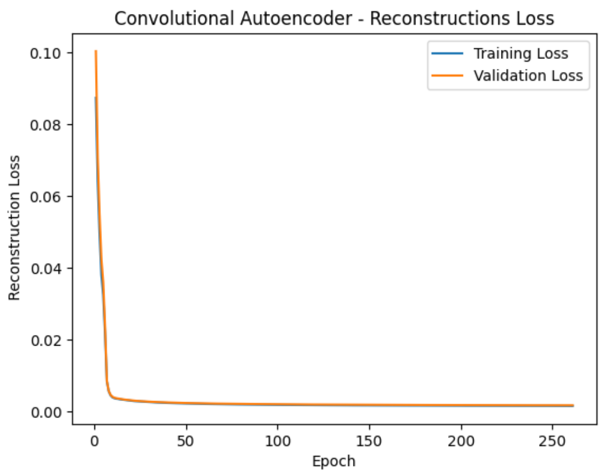

Figure 5.7: Training Loss for Learning Rate 1e-3.

As shown in Fig. 5.9, the reconstruction of the input images was conducted accurately. After the dimension reduction of the latent space, the major structural features of the images were preserved. This observation guarantees that the latent vector effectively retains the major image-reconstructed features. However, it is essential to note that some details and areas of the generated output images lost some definition due to the latent vector being eight times smaller than the original image.

An evaluation of image feature extraction process based on recommending the nearest neighbor was performed. The Figure 5.10 illustrates the query image and the top 6 retrieved images results. In addition, it also exhibits their distance from the query image, the Cosine similarity, the

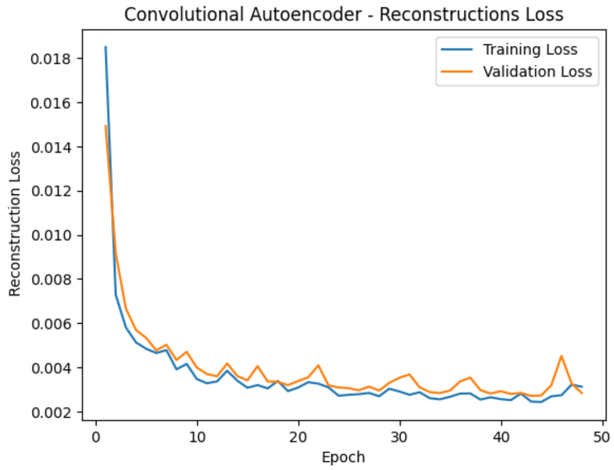

Figure 5.8: Training Loss for Learning Rate 1e-2.

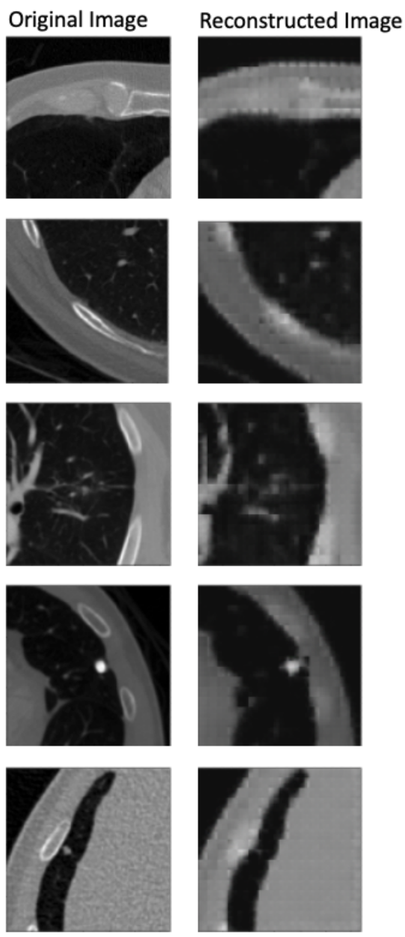

Figure 5.9: Reconstruction images from the model.

Euclidean distance based on Medical annotation, and NCC score of each image plotted at the top of the image.

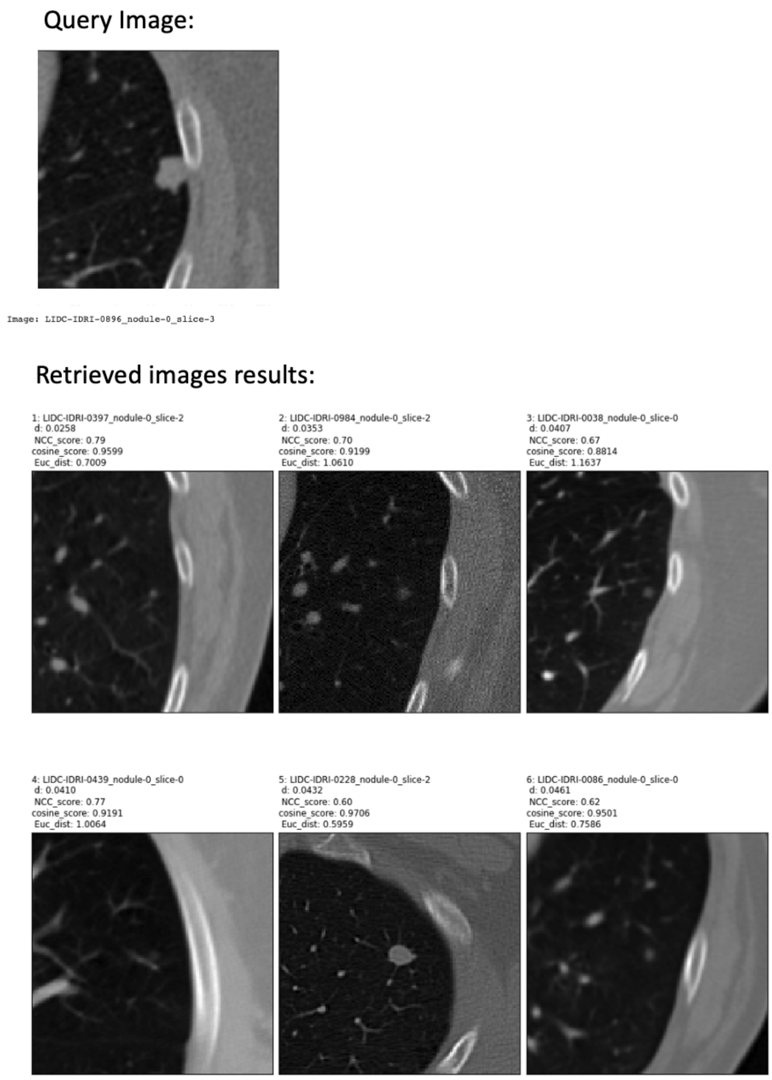

Figure 5.10: Query results from the model 2.

By analyzing the network results, the importance of using an Autoencoder as a feature extractor for the recommendation was demonstrated. As shown in Figure 5.10, the composition of the retrieved images presents similarity to the query. In objective terms, the NCC score was higher than 0.5, which means a structural image similarity and not necessary medical similarity. This is one of the most valuable features of the architecture, as it enables specialists to compare visually similar images but otherwise distinct cases.

5.2.1 Overall performance

Besides the retrieval presented on Figure 5.10, to evaluate the consistency of retrieval results, multiple queries were performed with slight variations in order to assess the system's robustness and reliability. The Figure 5.11 presents the retrieval of other 7 cases chosen randomly, with 6 retrieval images as results.

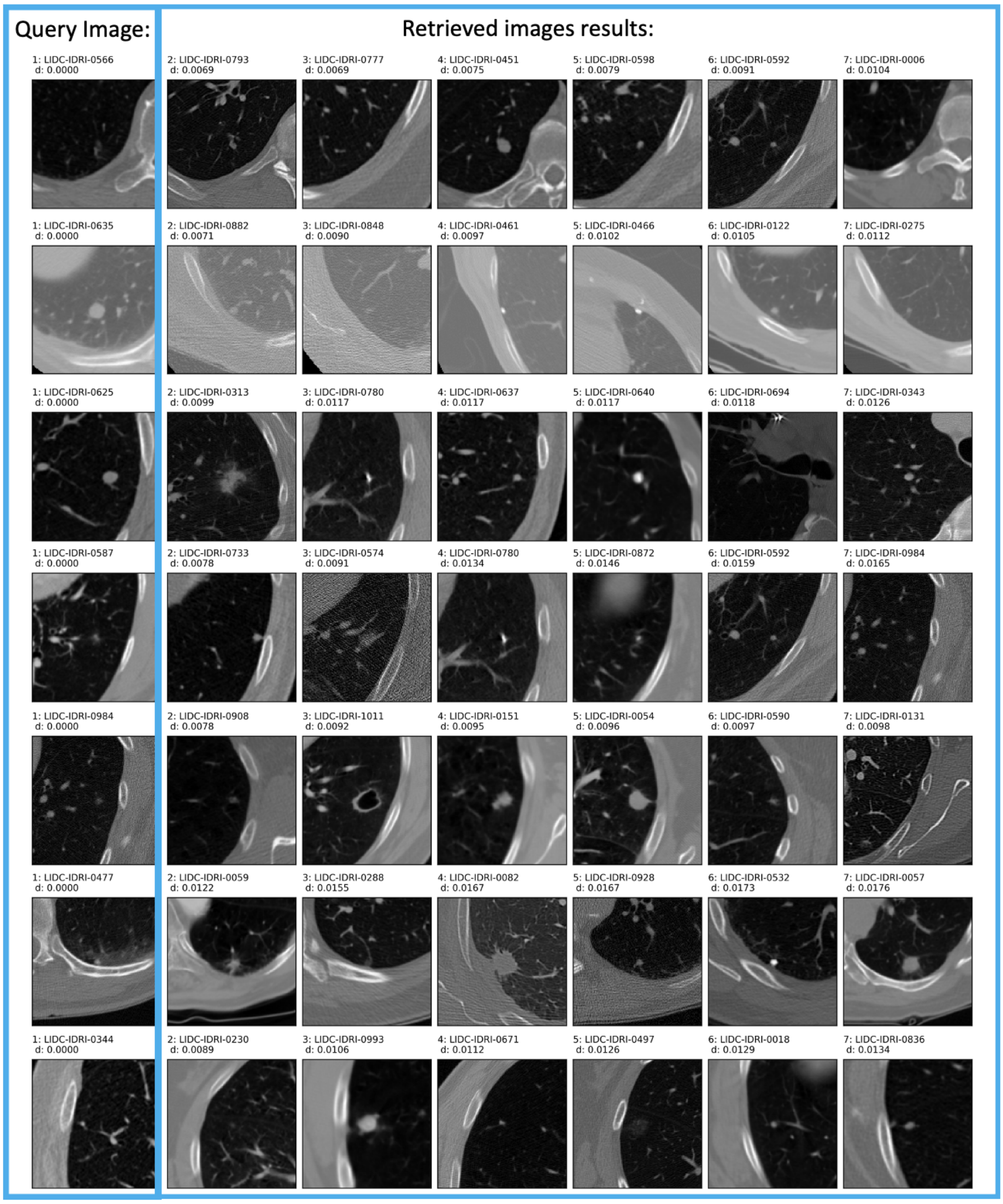

Figure 5.11: Query results from the model 2 from multiples images.

Considering the inclusion of NCC in the retrieval process, a comprehensive analysis was conducted using the entire validation dataset. The average precision was calculated for each image retrieval, ranking from 1 to 6 the retrieved images. Table 5.2 provides an overview of the results, specifically focusing on the case where k=1, indicating relevance based on structural similarity. In contrast to the first model, which achieved a score of approximately 5.5% in terms of structural similarity, the Autoencoder model reached a 73,8% of precision, indicating a strong capability to retrieve images based on structural composition.The NCC precision showed a decrease as the number of retrieved images increased, indicating that fewer relevant images were identified in terms of structural similarity as more images were requested. In the context of the medical image annotations, the average Euclidean distance increase to 0.861 ± 0.361, it became evident that the semantic meaning of these annotations was increasingly diverging.

Table 5.2: Hypothetical precision based on NCC Score.

Precision@k	Precision based on NCC
k=1	0.738
k=2	0.668
k=3	0.613
k=4	0.575
k=5	0.542
k=6	0.521

In addition, there are some important factors that must be taken into account when analyzing the results of image recommendation. Specifically, it becomes necessary to evaluate the similarity between the retrieved images and the medical annotations associated with each patient's nodule.The Spearman correlation was used to investigate the relationship between the Euclidean distance derived from the medical annotations and the distance of the image retrieval in order to assess this similarity. As shown in Figure 5.12, the results reveal a Spearman's rho coefficient of 0.08, indicating a very weak linear association between these measures.

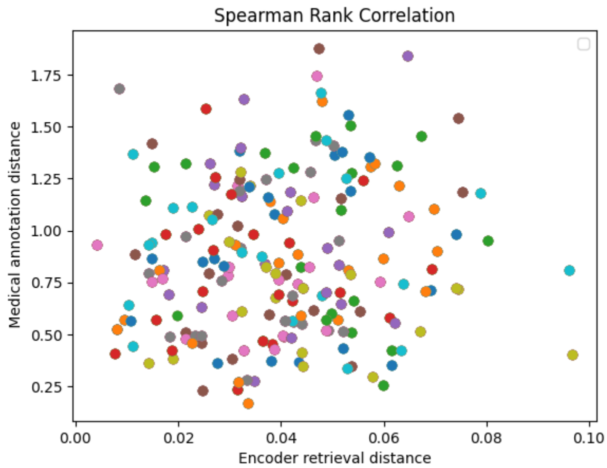

Figure 5.12: Spearman correlation from the model 2.

As previously discussed, a hypothetical evaluation scenario was considered, where images with an Euclidean distance below a predefined similarity threshold were considered relevant for the purpose of comparing the models. The objective of this analysis was to assess the hypothetical precision of the system based on medical annotations and to compare the performance of the different models.

In contrast to the previous model, the precision based on the Euclidean distance of the medical annotations exhibited low values, as illustrated in Table 5.3. These results indicated a limited capability of the models to effectively learn and capture the more medical-specific characteristics from the images. Furthermore, various experiments were conducted involving different latent representations, as shown in Table 5.4, in an attempt to reduce the latent space and encourage the models to learn beyond low-level characteristics. Unfortunately, these attempts did not yield the desired outcomes, suggesting the need for further exploration and refinement of the models' learning capabilities in order to enhance their performance in capturing the more nuanced medical-specific characteristics.

Table 5.3: Hypothetical precision based on Euclidean distance of the medical annotations.

Precision@k	**Hypothetical precision**
k=1	0.294
k=2	0.292
k=3	0.285
k=4	0.274
k=5	0.279
k=6	0.270

Table 5.4: Effect of several latent representation size on the Encoder with same number of layers.

Latent representation	**Reconstruction Loss**	**NCC Score**	**Hypothetical precision**
(16,16,16)	0.00325	0.747	0.228
(8,16,16)	0.00461	0.803	0.285
(4,16,16)	0.00509	0.789	0.280
(2,16,16)	0.00592	0.799	0.285

The results obtained from the Autoencoder evaluation presented a promising performance in retrieving lung cancer images based on their visual composition, indicating its ability to capture low-level visual features for similarity-based retrieval. However, when it was evaluated the medical annotations of the query image with the medical annotations of the retrieval images it was identify a low level of similarity. This model faces a semantic gap, which means it lacks the incorporation of medical-specific characteristics necessary to accurately characterize lung cancer. As a result, the retrieved images from this model exhibit a high degree of structural similarity but a low level of semantic similarity. Besides this, a medical expert would be required to perform a deep analysis and validate the results from the proposal model.

5.3 Results and discussion of Supervised Autoencoder

In this section, the results of the third proposed model are presented, which aimed to retrieve similar images based on the similarity of the encoder's latent representation of the image, including a supervised learning layer to address the challenge of reducing the semantic gap.

In the training phase of the new model, it was necessary to find appropriate hyperparameters to optimize its performance. Thus, the Supervised Autoencoder was trained in various scenarios, hyper-parameters and latent representation sizes.

After the incorporation of a classifier loss into the training process, it became essential to carefully balance the weights assigned to each loss term. The weight assigned to each loss determines the relative importance or contribution of that loss term in the overall training objective. In the initial experiments, a weight of 0.5 was assigned to the classifier loss, implying equal importance of both the classifier loss and the reconstruction loss terms. Later on, during the experiments adjusts were made on the iterative refinement process.

During the experiments conducted on the proposed architecture, a significant challenge arose in stabilizing and finding a balance between the two losses incorporated in the model. The objective was to ensure that the model effectively and simultaneously performed well on both the classification task and the image reconstruction task.

As observed in Figure 5.13, the training loss curves exhibited fluctuations and instability during the training process, reaching the early stopping several times. This behavior indicated the difficulty in optimizing the model to achieve optimal performance on both objectives simultaneously. The fluctuations in the loss curves implied that the model was struggling to strike a balance between the two competing objectives.

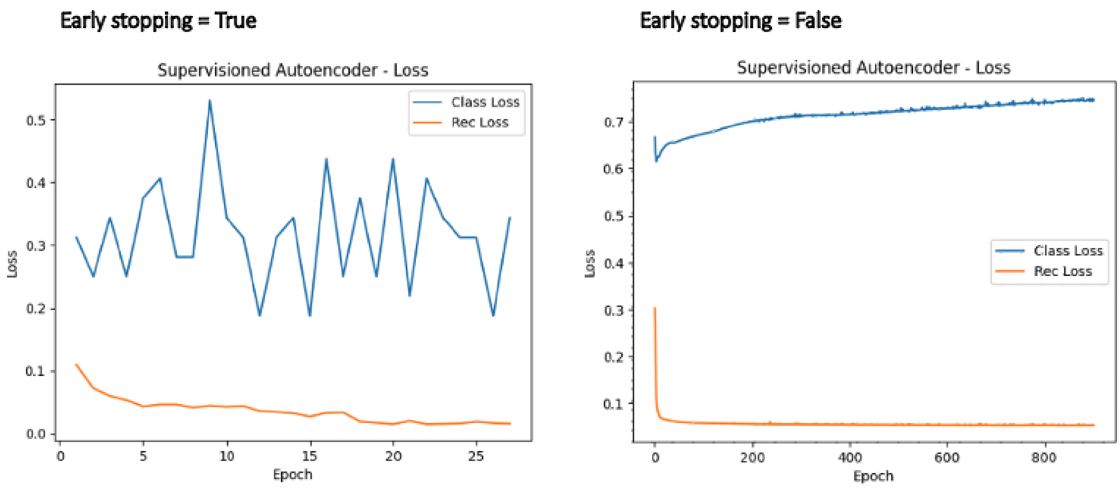

Figure 5.13: Model trained with early stop parameter True and False.

Based on the classification loss instability of the previous training section, a two-step training process was implemented to provide the model with a better initial understanding of the input data.

In the first step, the model underwent a pre-training phase that prioritized the reconstruction loss. The main objective was to train the model to accurately reconstruct the input images. By focusing on the reconstruction task initially, the model was able to develop a robust understanding of the underlying structure and patterns present in the data. At this stage, it was expected that the model had already acquired underlying features of the data through the pre-training phase.

Once the pre-training phase was completed, the model was transitioned to the second step of training, which involved fine-tuning on the classification task as it is illustrated on Figure 5.14. The focus shifted towards training the model to perform well on the specific classification task between benign or malign classes.

The main goal of incorporating both reconstruction and classification losses in the training process was for the model being able to learn both the low-level features necessary for accurate

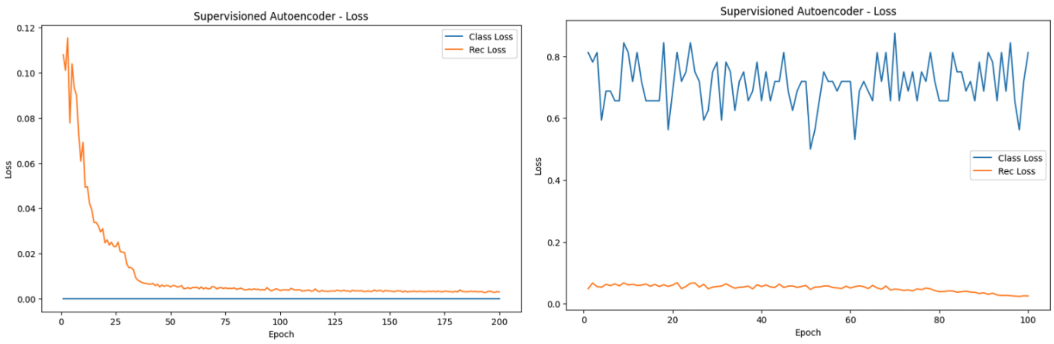

Figure 5.14: Model trained two-steps approach.

reconstruction and the higher-level features relevant to classification. This two-step training approach aimed to strike a balance between capturing fine details in the input images and extracting discriminate features for effective classification. However, by analyzing the query image results, the Supervised Autoencoder as a tool of feature extractor presented lower capability to retrieve images with correlated composition similarity, as shown in Figure 5.15.

In this instance, it was also considered an hypothetical scenario where the images with Euclidean distance below a particular threshold were considered as relevant images for the study to evaluate if the semantic gap was reduced. Based on the table 5.5, the Euclidean distance from the annotation compared with the previous model did not reduced.

Table 5.5: Hypothetical precision based on Euclidean distance - Model 3.

Precision@k	**Hypothetical precision**
k=1	0.262
k=2	0.254
k=3	0.258
k=4	0.254
k=5	0.249
k=6	0.242

Despite exploring various MLP architectures with different numbers of hidden layers and trying different combinations of learning rates, optimizers (SGD, Adam, and AdamW), and batch sizes (16, 32, and 64), the achieved results did not meet the initial expectations. The selected approach, consisting of pre-training on the reconstruction loss followed by fine-tuning with a learning rate of 1e-5 and incorporating a classification loss with a weight of 0.2, did not effectively reduce the gap between the embedding representations.

The observed reduction in retrieval images with similarity composition suggests that the pre-training phase focused primarily on the reconstruction aspect rather than on learning discriminative features relevant to the classification task. This mismatch between the objective of the pre-training and the desired outcome of reducing the gap layer may have contributed to the suboptimal results.

Further investigation and experimentation are required to identify the underlying factors contributing to the limited effectiveness of the embedding representation. It may be necessary to ex-

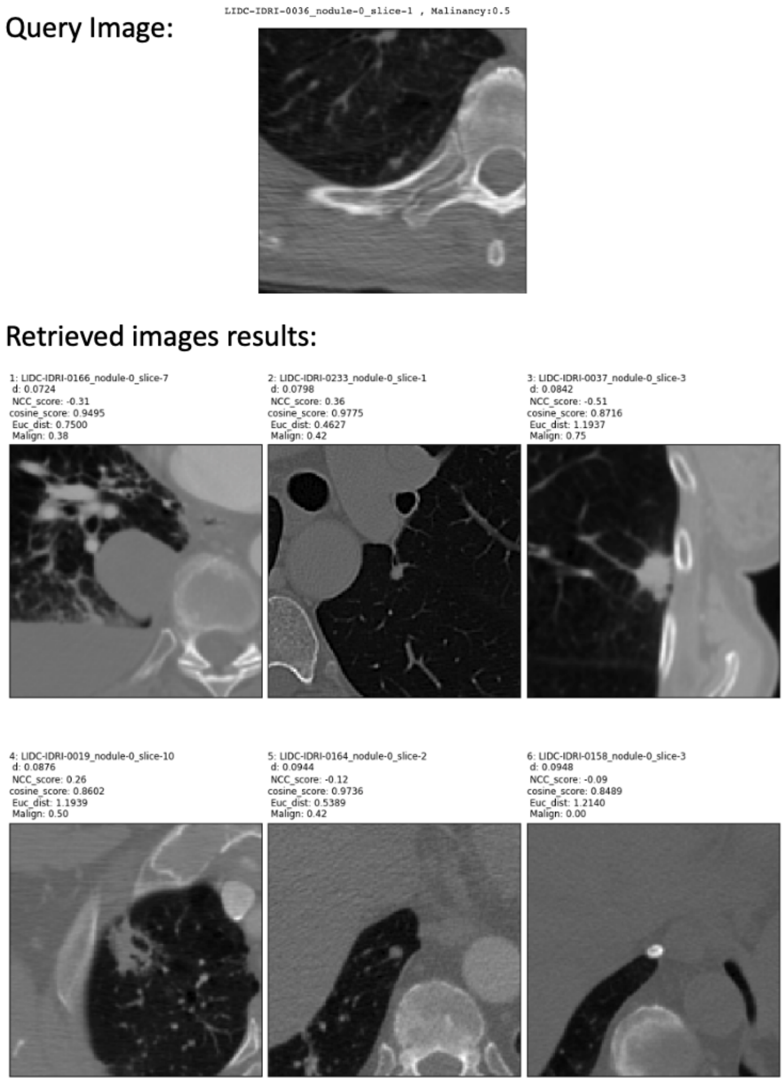

Figure 5.15: Query results from the model 3.

plore alternative pre-training strategies or modify the fine-tuning process to encourage the model to learn more discriminative features. Additionally, considering other architectural variations or incorporating domain-specific knowledge into the model design could potentially yield improved results.

5.4 Evaluation of Retrieval Time

This section examines the impact of CPU time on the performance of similarity search algorithms, specifically focusing on the k-nearest neighbors (kNN) algorithm, that was the algorithm used to find the most similar items to a given query item based on a similarity metric.

Retrieval time is an essential factor in image retrieval systems, as it directly impacts the user experience and system efficiency. The retrieval time refers to the duration it takes for the system to process a query and retrieve relevant images from a large database. Efficient retrieval time is essential to ensure real-time or near-real-time performance.

One important factor that affects the performance of kNN is the size of the search (number of images) or size of the vectors being considered, in this instance the size of latent representation. As the number of vectors increases, the computational complexity of finding the nearest neighbors also increases. Consequently, the CPU time required for the search process may increase proportionally.

To evaluate the impact of CPU time on the search similarity, experiments were conducted by varying the size of the latent space in 256 , 512, 1024 , 2048 and 4096 and the search space in 1k, 10k, 100k. It was measured the corrresponding time taken on the interation of latent vector and search loop using different configurations and the results were compared. In order to evaluate the CPU time performance on a larger dataset, it was generated synthetic data to overcome the size limitations presented on LIDC dataset.

Table 5.6: Retrieval time in ms per loop interaction for 6 images on different sizes of dataset.

Latent representation	**1k images**	**10k images**	**100k images**
(1,16,16)	453 µs ± 12.5 µs	1.75 ms ± 59.7 µs	19.5 ms ± 931 µs
(2,16,16)	572 µs ± 33.1 µs	3.43 ms ± 231 µs	39.7 ms ± 2.3 ms
(4,16,16)	796 µs ± 53.2 µs	7.17 ms ± 187 µs	70.7 ms ± 1.12 ms
(8,16,16)	1.26 ms ± 78.1 µs	14.2 ms ± 273 µs	142 ms ± 3.08 ms
(16,16,16)	5.35 ms ± 1.01 ms	49.9 ms ± 2.03 ms	472 ms ± 4.63 ms

The kNN algorithm was executed to find the top 6 nearest neighbors using the Colab Pro environment with 25.5Gb of memory RAM available. The results demonstrated a noticeable increase in CPU time as the latent space grew larger. This increase can be attributed to the higher computational effort required to search through a larger space and calculate the similarity between vectors.

As the search space increases, the computational effort required for similarity search algorithms also increases. This is evident in the processing times observed in Table 5.6, where the latent representation size of 4096 exhibits a significant increase in processing time as the number of images in the dataset grows.

For instance, when considering a dataset of 1K images and latent representation of 4096 (16,16,16), the processing time for the similarity search is recorded as 5.35 ms. However, when the dataset size is increased to 100K images, the processing time rises to 474 ms, representing a substantial increase of approximately 88 times in the processing time.

In the context of this study, the retrieval times posed no significant issues due to the limited number of slides used. Nevertheless, it is essential to recognize that concerns may arise with respect to integration more images and scaling the system as it progress further. The impact of CPU time on search similarity is an important consideration in practical applications. In scenarios where real-time or near-real-time performance is crucial, such as interactive systems or time-sensitive applications, it becomes necessary to optimize the algorithm or employ alternative approaches to mitigate the computational burden. This may involve using dimension reduction techniques, approximate nearest neighbor search methods or distributed computing frameworks to distribute the workload and reduce the overall CPU time.

Understanding the relationship between CPU time and vector size provides insights regarding the scalability and the efficiency of similarity search algorithms. By considering the impact of CPU time, researchers and practitioners can make informed decisions when selecting and optimizing algorithms for large-scale similarity search tasks, ensuring that the system can handle increasing workloads without compromising search accuracy or response times.

5.5 Summary

In this Chapter it was presented the experimental results obtained for the three types of image retrieval models proposed: text-based on the medical annotations, a content-based using Convolutional Autoencoder and Supervised Autoencoder.

The annotation-based model demonstrated a promising performance in retrieving similar cases based on medical annotations, effectively leveraging the annotated information to identify cases with similar characteristics. However, regarding the structural aspects, it presented a NCC Score of 5,5%, indicating low physical correlation between the retrieval images and demonstrated limitations in capturing the structural aspects of the images, as it initially relied on textual annotations. Furthermore, new features were implemented on the model to facilitate the search, being able to enhance the match on structural aspects. Although it retrieved images with similar semantic meaning, a medical validation is necessary to evaluate the relevance of the retrieved cases. On this model the main challenge relies heavily on dependency of annotated cases, which can be time-consuming and resource-intensive to obtain.

The Convolutional Autoencoder model showed promising performance in retrieving lung cancer images based on their visual composition, indicating its ability to capture low-level visual features for similarity-based retrieval. With a NCC score of 73%, the percentage of the retrieved images presented a strong correlation with the query image. However, when it was evaluated the medical annotations of the query image with the medical annotations of the retrieval images it was identified a low level of similarity. This model faces a semantic gap, which means it lacks the incorporation of medical-specific characteristics necessary for accurately characterizing lung cancer. As a result, the retrieved images from this model exhibit a high degree of structural similarity but a low level of semantic similarity. To validate the relevance and accuracy of the retrieved images, it is again crucial to involve medical experts who can provide their expertise and insights in evaluating the retrieved results. This step is essential to ensure that the retrieved images are not only visually similar but also medically relevant and informative.

In contrast, the Supervised Autoencoder model was proposed as a solution to address the semantic gap in lung cancer characterization. However, the results obtained from this model were ineffective. It struggled to retrieve images with sufficient similarity and failed to bridge the semantic gap adequately. To improve the performance of the Supervised Autoencoder model, further research is needed. Exploring alternative approaches, refining the model architecture, and incorporating additional medical-specific characteristics could help to overcome these limitations and enhance its effectiveness in accurately characterizing lung cancer.

Furthermore, it was evaluated the computational cost of the size of latent representation and the size of the search space and their impact on the retrieval time in retrieval system. It was highlighted that for system implementation it is necessary to evaluate the trade-off between computational cost, search space size, and retrieval time to support the efficiency and scalability of these systems.

Chapter 6

Conclusion and future work

In this study, it was investigated the utilization of a Content-Based Image Retrieval system as the core engine for a health recommender system focused on lung cancer characterization. The aim was to leverage medical annotations and image similarity search to assist healthcare professionals in diagnosing and characterizing lung cancer based on retrieval of similar cases. It was explored and highlighted the capabilities and limitations of the employed models, namely medical annotation-based, Convolutional Autoencoder and Supervised Autoencoder, in addressing the challenges associated with the semantic gap between visual features and medical annotations.

The primary results indicated that the text-based model showed promising performance in retrieving similar cases based on medical annotations. The model effectively leveraged the annotated information to identify cases with similar characteristics, leading to the retrieval of relevant past cases. However, when it comes to image similarity, the model exhibited a lower correlation with the structural aspects of the images as the model was primarily relied on the textual annotations. Additionally, a required medical validation is necessary to evaluate the relevance of the retrieval cases. One of the main challenges encountered is that this model heavily relies on the availability of annotated cases in order to perform effective retrieval. This poses a limitation, as obtaining comprehensive and accurate annotations for the past cases can be time-consuming and resource-intensive.

The results of the Autoencoder model exhibited promising performance in retrieving lung cancer images with a strong correlation to their image composition. This capability indicates that the Autoencoder effectively captured and leveraged low-level visual features to retrieve visually similar images. However, a notable limitation of the Autoencoder-based approach was the presence of a semantic gap when compared to the medical annotations associated with the lung cancer images. The model struggled to bridge this gap, as it focused primarily on visual similarity and failed to incorporate the necessary medical-specific characteristics required for accurate lung cancer characterization.

On the other hand, the Supervised Autoencoder, designed with the attempt to address the semantic gap presented an ineffective results. It exhibited limitations in retrieving images with sufficient similarity and failed to effectively bridge the semantic gap. The Supervised Autoencoder presented new challenges and complexities associated with training and architecture of image retrieval system in this particular scenario of lung cancer characterization.

Further research and exploration of different types of architectures are essential to overcome the technical challenges encountered in this study. One promising avenue is the utilization of pre-trained models, fine-tuning techniques and restrict the problem for a multi-class classification. Fine-tuning these models on the targeted dataset can enhance their effectiveness in capturing relevant patterns and structures related to lung cancer characterization.

Additionally, exploring the development of an embedded model that incorporates the analysis of slices from a 3D perspective could be a interesting approach for enhancing the capabilities of the recommendation system.The use of 3D slices allows for a more comprehensive representation of the lung cancer cases, capturing additional spatial information and contextual cues that may contribute to more accurate retrieval and characterization

Moreover, collaboration with medical experts and domain-specific guidance is essential for developing effective health recommender systems. Close collaboration will ensure that the system aligns with the requirements and expectations of healthcare professionals while it supports the improving of system's performance and enhancing its clinical relevance.

References

[1] T. I. A. for Research on Cancer (IARC), "Global cancer observatory. available at https://gco.iarc.fr/. accessed last time at 09 january 2023," 2020.

[2] J. Koury, M. Lucero, C. Cato, L. Chang, J. Geiger, D. Henry, J. Hernandez, F. Hung, P. Kaur, G. Teskey, and A. Tran, "Immunotherapies: Exploiting the Immune System for Cancer Treatment," *Journal of Immunology Research*, vol. 2018, pp. 1–16, 2018.

[3] "Understanding your chest ct scan pet scan for lung cancer detection. available at https://www.docpanel.com/blog/post/understanding-your-chest-ct-scan-pet-scan-lung-cancer-detection. accessed last time at 25 february 2023."

[4] S. K. Thakur, D. P. Singh, and J. Choudhary, "Lung cancer identification: a review on detection and classification," *Cancer and Metastasis Reviews*, vol. 39, pp. 989–998, Sept. 2020.

[5] "World health organization. promoting cancer early diagnosis. available at https://www.who.int/activities/promoting-cancer-early-diagnosis. accessed last time at 10 february 2023."

[6] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018: Cancer Statistics, 2018," *CA: A Cancer Journal for Clinicians*, vol. 68, pp. 7–30, Jan. 2018.

[7] "World health organization. tobacco. available at https://www.who.int/news-room/fact-sheets/detail/tobacco. accessed last time at 10 february 2023."

[8] H. J. de Koning, C. M. van der Aalst, P. A. de Jong, E. T. Scholten, K. Nackaerts, M. A. Heuvelmans, J.-W. J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W. P. Mali, F. A. Mohamed Hoesein, P. M. van Ooijen, J. G. Aerts, M. A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegenthart, J. E. Walter, K. ten Haaf, H. J. Groen, and M. Oudkerk, "Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial," *New England Journal of Medicine*, vol. 382, pp. 503–513, Feb. 2020.

[9] "Lung biopsy. available at https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lung-biopsy. accessed last time at 18 february 2023."

[10] "Lung biopsy techniques and clinical significance. available at https://www.ncbi.nlm.nih.gov/books/nbk563153/. accessed last time at 25 february 2023."

[11] C. C. Aggarwal, *Recommender Systems*. Cham: Springer International Publishing, 2016.

[12] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich, *Recommender Systems: An Introduction*. New York, NY, USA: Cambridge University Press, 1st ed., 2010.

[13] T. N. T. Tran, A. Felfernig, C. Trattner, and A. Holzinger, "Recommender systems in the healthcare domain: state-of-the-art and research issues," *Journal of Intelligent Information Systems*, vol. 57, pp. 171–201, Aug. 2021.

[14] W. Yue, Z. Wang, J. Zhang, and X. Liu, "An Overview of Recommendation Techniques and Their Applications in Healthcare," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, pp. 701–717, Apr. 2021.

[15] T. N. Trang Tran, M. Atas, A. Felfernig, and M. Stettinger, "An overview of recommender systems in the healthy food domain," *Journal of Intelligent Information Systems*, vol. 50, pp. 501–526, June 2018.

[16] U. Bhimavarapu, M. Sreedevi, N. Chintalapudi, and G. Battineni, "Physical Activity Recommendation System Based on Deep Learning to Prevent Respiratory Diseases," *Computers*, vol. 11, p. 150, Oct. 2022.

[17] L. Verboven, T. Calders, S. Callens, J. Black, G. Maartens, K. E. Dooley, S. Potgieter, R. M. Warren, K. Laukens, and A. Van Rie, "A treatment recommender clinical decision support system for personalized medicine: method development and proof-of-concept for drug resistant tuberculosis," *BMC Medical Informatics and Decision Making*, vol. 22, p. 56, Dec. 2022.

[18] E. Aalipour and M. Ghazisaeedi, "Recommender system introduction for requests of cancer world," *International Journal Of Community Medicine And Public Health*, vol. 4, p. 275, Jan. 2017.

[19] P. M. Ashok Kumar, T. Subha Mastan Rao, L. Arun Raj, and E. Pugazhendi, "An Efficient Text-Based Image Retrieval Using Natural Language Processing (NLP) Techniques," in *Intelligent System Design* (S. C. Satapathy, V. Bhateja, B. Janakiramaiah, and Y.-W. Chen, eds.), vol. 1171, pp. 505–519, Singapore: Springer Singapore, 2021. Series Title: Advances in Intelligent Systems and Computing.

[20] L. R. Nair, K. Subramaniam, and G. K. D. Prasannavenkatesan, "A Review on Multiple Approaches to Medical Image Retrieval System," in *Intelligent Computing in Engineering* (V. K. Solanki, M. K. Hoang, Z. J. Lu, and P. K. Pattnaik, eds.), vol. 1125, pp. 501–509, Singapore: Springer Singapore, 2020. Series Title: Advances in Intelligent Systems and Computing.

[21] P. m. T.karthikeyan and S. n. aprabhu, "A Survey on Text and Content Based Image Retrieval System for Image Mining," *International Journal of Engineering Research*, vol. 3, no. 3, 2014.

[22] I. M. Hameed, S. H. Abdulhussain, and B. M. Mahmmod, "Content-based image retrieval: A review of recent trends," *Cogent Engineering*, vol. 8, p. 1927469, Jan. 2021.

[23] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, pp. 262–282, Jan. 2007.

[24] B. Stark, C. Knahl, M. Aydin, and K. Elish, "A Literature Review on Medicine Recommender Systems," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 8, 2019.

[25] M. Wiesner and D. Pfeifer, "Health Recommender Systems: Concepts, Requirements, Technical Basics and Challenges," *International Journal of Environmental Research and Public Health*, vol. 11, pp. 2580–2607, Mar. 2014.

[26] R. De Croon, L. Van Houdt, N. N. Htun, G. Štiglic, V. Vanden Abeele, and K. Verbert, "Health Recommender Systems: Systematic Review," *Journal of Medical Internet Research*, vol. 23, p. e18035, June 2021.

[27] A. Calero Valdez, M. Ziefle, K. Verbert, A. Felfernig, and A. Holzinger, "Recommender Systems for Health Informatics: State-of-the-Art and Future Perspectives," in *Machine Learning for Health Informatics* (A. Holzinger, ed.), vol. 9605, pp. 391–414, Cham: Springer International Publishing, 2016. Series Title: Lecture Notes in Computer Science.

[28] J. Pincay, L. Teran, and E. Portmann, "Health Recommender Systems: A State-of-the-Art Review," in *2019 Sixth International Conference on eDemocracy & eGovernment (ICEDEG)*, (Quito, Ecuador), pp. 47–55, IEEE, Apr. 2019.

[29] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex & Intelligent Systems*, vol. 7, pp. 439–457, Feb. 2021.

[30] Q. Zhang, G. Zhang, J. Lu, and D. Wu, "A Framework of Hybrid Recommender System for Personalized Clinical Prescription," in *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, (Taipei, Taiwan), pp. 189–195, IEEE, Nov. 2015.

[31] I. A. Siradjuddin, W. A. Wardana, and M. K. Sophan, "Feature Extraction using Self-Supervised Convolutional Autoencoder for Content based Image Retrieval," in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, (Semarang, Indonesia), pp. 1–5, IEEE, Oct. 2019.

[32] S. Agrawal, A. Chowdhary, S. Agarwala, V. Mayya, and S. Kamath S., "Content-based medical image retrieval system for lung diseases using deep CNNs," *International Journal of Information Technology*, vol. 14, pp. 3619–3627, Dec. 2022.

[33] S. G. Armato and McLennan, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans: The LIDC/IDRI thoracic CT database of lung nodules," *Medical Physics*, vol. 38, pp. 915–931, Jan. 2011.

[34] P. Opulencia, D. S. Channin, D. S. Raicu, and J. D. Furst, "Mapping LIDC, RadLex™, and Lung Nodule Image Features," *Journal of Digital Imaging*, vol. 24, pp. 256–270, Apr. 2011.

[35] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, "Comparing Dissimilarity Measures for Content-Based Image Retrieval," in *Information Retrieval Technology* (H. Li, T. Liu, W.-Y. Ma, T. Sakai, K.-F. Wong, and G. Zhou, eds.), vol. 4993, pp. 44–50, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Series Title: Lecture Notes in Computer Science.

[36] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. `http://www.deeplearningbook.org`.

[37] A. Ng, "Sparse autoencode - lecture notes cs294a stanford university." Available at `https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf` Accessed last time at 22 April 2023.

[38] I. Loshchilov and F. Hutter, “Decoupled Weight Decay Regularization,” Jan. 2019. arXiv:1711.05101 [cs, math].

[39] J.-M. Ho, S.-Y. Lin, C.-W. Fann, Y.-C. Wang, and R.-I. Chang, “A novel content based image retrieval system using K-means with feature extraction,” in *2012 International Conference on Systems and Informatics (ICSAI2012)*, (Yantai, China), pp. 785–790, IEEE, May 2012.

[40] H. Jo and C.-H. Jun, “A personalized classification model using similarity learning via supervised autoencoder,” *Applied Soft Computing*, vol. 131, p. 109773, Dec. 2022.

[41] B. Barz and J. Denzler, “Content-based Image Retrieval and the Semantic Gap in the Deep Learning Era,” Nov. 2020. arXiv:2011.06490 [cs].

[42] A. K. Dhara, S. Mukhopadhyay, A. Dutta, M. Garg, and N. Khandelwal, “Content-Based Image Retrieval System for Pulmonary Nodules: Assisting Radiologists in Self-Learning and Diagnosis of Lung Cancer,” *Journal of Digital Imaging*, vol. 30, pp. 63–77, Feb. 2017.

[43] S. Liu, B. Yang, Y. Wang, J. Tian, L. Yin, and W. Zheng, “2D/3D Multimode Medical Image Registration Based on Normalized Cross-Correlation,” *Applied Sciences*, vol. 12, p. 2828, Mar. 2022.

[44] M. Mukaka, “Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research,” 2012.