UNIVERSITÀ DEGLI STUDI DI MILANO

COLLEGIO DIDATTICO DI INFORMATICA

Corso di Laurea magistrale in
Informatica

COVID-19 SEARCH ENGINE

Instructor: Prof. Alfio Ferrara

A report by:
Saman Saeedi
Matr. Nr. 984718

ANNO ACCADEMICO 2023-2024

Contents

Chapter 1

Introduction

In the realm of medical imaging, Content-Based Image Retrieval (CBIR) systems have emerged as crucial tools for enhancing diagnostic processes by utilizing computational methods to extract and compare visual features directly from images. This approach contrasts with traditional metadata-dependent methods, promising improvements in efficiency and accuracy in clinical settings, especially amidst the challenges posed by the COVID-19 pandemic [1].

1.0.1 Motivation

The motivation behind this project stems from the urgent necessity for effective tools to manage and analyze the increasing volumes of medical images, particularly in the context of COVID-19. The global healthcare crisis has underscored the critical importance of rapid and accurate diagnosis, particularly through the analysis of lung CT scans. By developing a robust CBIR system, this project aims to streamline the retrieval and analysis of COVID-19-related images, thereby supporting healthcare professionals in making informed clinical decisions swiftly.

1.0.2 Objectives

The primary objective of this project is to design, implement, and evaluate a specialized CBIR system tailored for COVID-19 lung CT scans. Key objectives include:

1. **Implementation of Deep Learning Models**: Utilize pre-trained deep learning models such as Inception-V3, VGG-16, VGG-19, and Xception, which are fine-tuned on the ImageNet dataset, to extract meaningful features from COVID-19 and non-COVID-19 CT scan images. These models are known for their robust feature extraction capabilities, allowing for improved image classification and retrieval performance

2. **Development of a Custom Feature Extraction Method**: Design and implement a custom feature extraction method tailored specifically for analyzing CT scan images. This method involves several steps:

    (a) Convert the input image to the LAB color space and extract the A and B channels, calculating their mean values to capture color information.

    (b) Convert the image to grayscale and resize it to a standard size of 256x256 pixels.

    (c) Apply entropy filtering to the grayscale image using a disk-shaped structuring element to compute mean and standard deviation of the entropy values, which help in capturing texture information.

    (d) Compute edge features using Roberts and Sobel filters, taking the mean of the resultant images.

    This custom feature extraction method aims to complement the deep learning models by providing additional handcrafted features that may enhance the CBIR system's performance in distinguishing between COVID-19 and non-COVID-19 images.

3. **Construction of a Comprehensive Dataset**: Compile a dataset of COVID-19 and non-COVID-19 CT scan images, ensuring a balanced and representative sample set to train and evaluate the CBIR system.

4. **Integration into a Web Application**: Develop a user-friendly web application that allows users to upload CT scan images and retrieve visually similar images from the dataset. This application aims to provide a practical tool for medical professionals to aid in the diagnosis and study of COVID-19.

5. **Evaluation and Validation**: Assess the performance of the CBIR system using standard evaluation metrics, including precision, recall, and F1-score, to ensure its effectiveness and reliability in real-world scenarios.

1.0.3 Significance and Application

Medical imaging, encompassing modalities like computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), generates vast datasets pivotal for clinical decision-making. CBIR systems promise to expedite image retrieval processes, reduce manual annotation efforts, and enhance diagnostic accuracy by enabling rapid access to relevant medical images based on content similarity rather than text-based queries - MDPI.

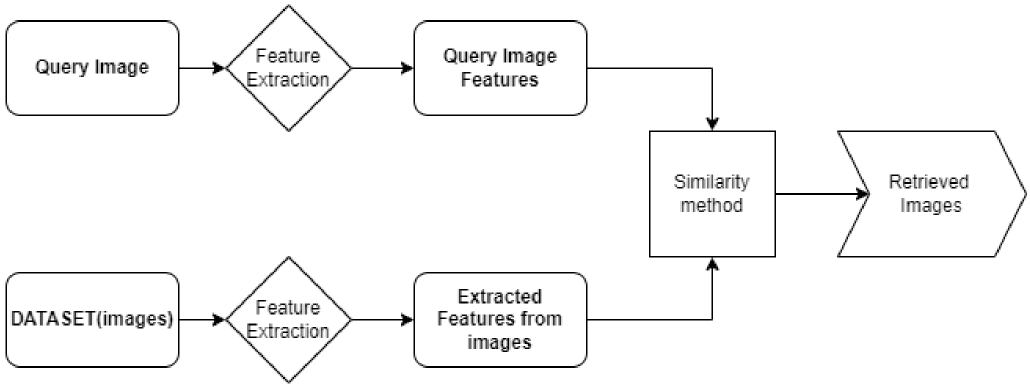


Figure 1. Content-Based Image Retrieval Schema

1.0.4 Structure of the Report

This report is organized as follows:

- **Chapter 2: Research Methodology** - Details on the deep learning architectures utilized, feature extraction techniques, including a custom method for specialized feature extraction.
- **Chapter 3: Experimental Results** - Analysis of retrieval metrics including accuracy, computational time, and memory utilization across different deep learning models.
- **Chapter 4: Discussion** - Interpretation of findings, comparison of model performances, challenges encountered, and recommendations for future enhancements.
- **Chapter 5: Conclusion** - Summary of achievements and implications of the CBIR system in medical imaging, emphasizing its potential impact on clinical practice.

Chapter 2

Research Methodology

This section details the methodologies employed in the development and evaluation of the CBIR system for COVID-19 lung CT scans.

2.0.1 Deep Learning Models

The CBIR system utilizes state-of-the-art deep learning models pre-trained on ImageNet, including Inception-V3, VGG-16, VGG-19, and Xception. These models are chosen for their robust feature extraction capabilities from medical images.

Inception-V3

A 48-layer deep neural network, was utilized by leveraging pretrained weights from the ImageNet dataset. The final layer, known as the softmax output layer, was removed. The input images were resized to 299×299 pixels, as required by the Inception-V3 model. After preprocessing, features were extracted from the final fully connected layer [2].

Xception

A neural network with 71 layers was used with pretrained ImageNet weights. The input images were resized to 299×299 pixels, and the `conv2d_2` layer was designated as the feature extraction layer [3].

VGG-16

The pretrained ImageNet model was employed. The first fully connected layer was used for feature extraction. Input images were resized to 224×224 pixels, and preprocessing involved subtracting the average pixel value before extracting features [4].

VGG-19

A 19-layer deep neural network comprising 16 convolution layers, 3 fully connected layers, 5 max pooling layers, and 1 softmax layer, was also implemented using ImageNet weights. Input images were resized to 224×224 pixels. The first fully connected layer was selected for feature extraction.

2.0.2 Custom Method for Feature Extraction

In addition to the standard deep learning models, a custom feature extraction method tailored for COVID-19 lung CT scans has been developed. This method incorporates specialized image processing techniques to capture unique visual features relevant to lung pathology. Specifically, the custom method includes:

- **Feature Extraction Function:** Implemented using Python and libraries such as NumPy, scikit-image, and OpenCV.
- **Steps:** Pre-processing steps include RGB to LAB color conversion, entropy calculation, edge detection using Roberts and Sobel filters, and Gabor feature extraction.
- **Purpose:** Designed to extract nuanced features from lung CT scans that may not be adequately captured by generic deep learning models trained on general image datasets.

This custom method enhances the CBIR system's ability to differentiate between COVID-19 and non-COVID-19 lung CT scans based on detailed visual features specific to lung pathology.

2.0.3 Implementation Details

The implementation involves integrating these models and the custom feature extraction method into a unified Python-based framework. Data handling, model training, and evaluation are conducted using libraries such as TensorFlow, Keras, and scikit-learn.
To measure similarity, the Euclidean norm, also referred to as L2-norm, is utilized. This norm calculates the distance between vector coordinates in the vector space, resulting in a positive distance value. The similarity between the queried image's feature and the features stored in the database is assessed by computing this L2-norm.

Chapter 3

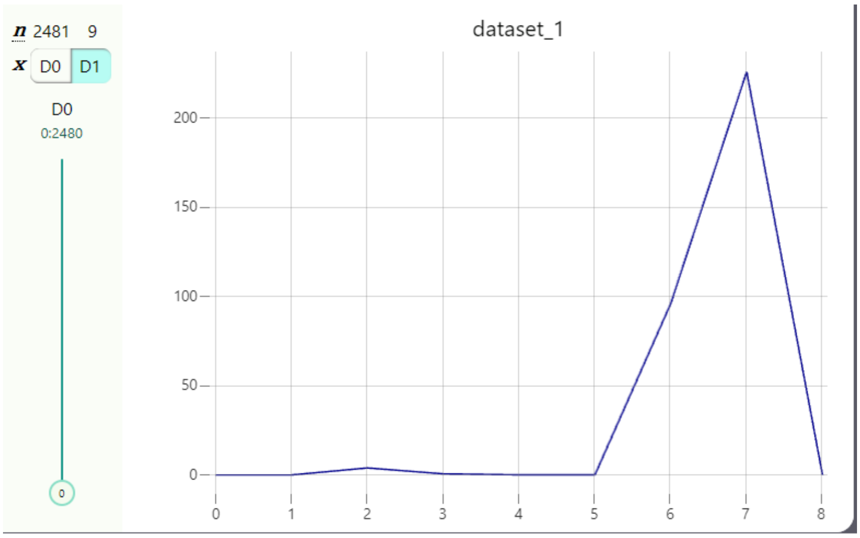
Experimental Results

The project's dataset comprises 2482 CT scans gathered from actual patients at hospitals in Sao Paulo, Brazil. It includes 1252 CT scans that show SARS-CoV-2 infection and 1229 CT scans from non-infected patients. Each image varies in size, and the total size of the dataset is 231 MB. For each model, all images in the dataset were used to extract features and create a corresponding feature database. The distribution of the dataset is illustrated in the figure below. As expected, the dataset categorizes diseases into two types: "covid" and "non-covid".

Dataset features were extracted by each model separately. Each model saved a file in h5 format with different sized due to the method they use. Among these files, the custom one is the smallest as it does not have many layers for extraction.

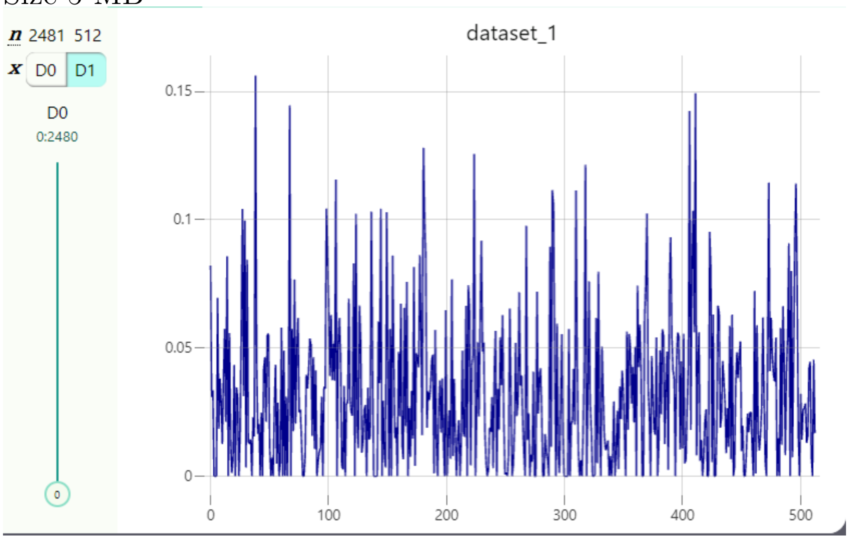
Custom features extractor

- Shape of array 2481 9
- Size 399 KB



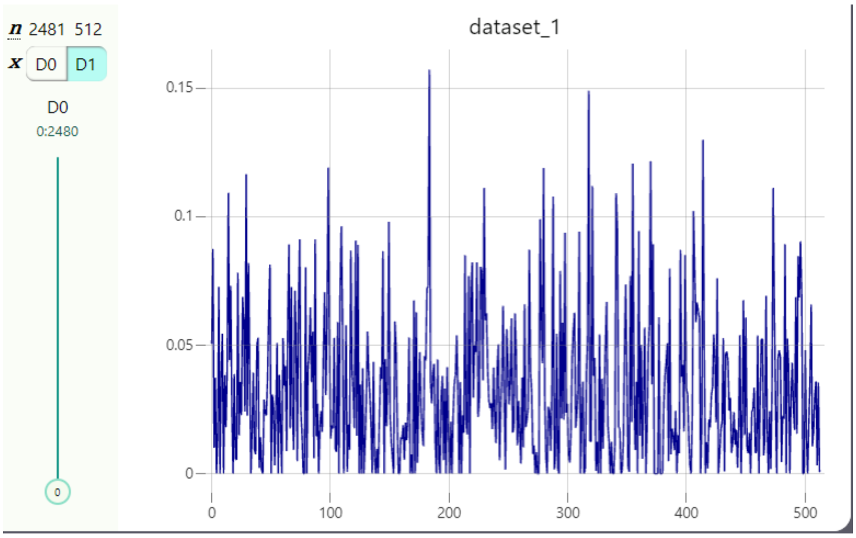
VGG16 features extractor

- Shape of array 2481 512
- Size 5 MB



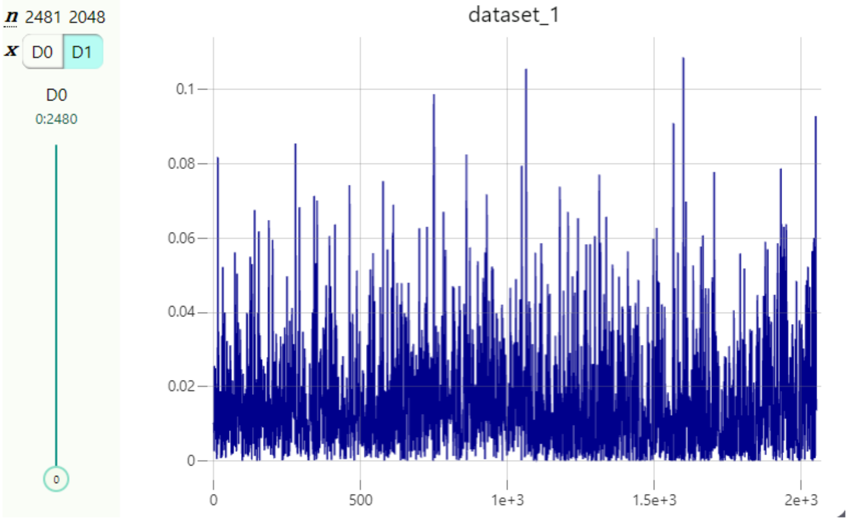
VGG19 features extractor

- Shape of array 2481 512
- Size 5 MB



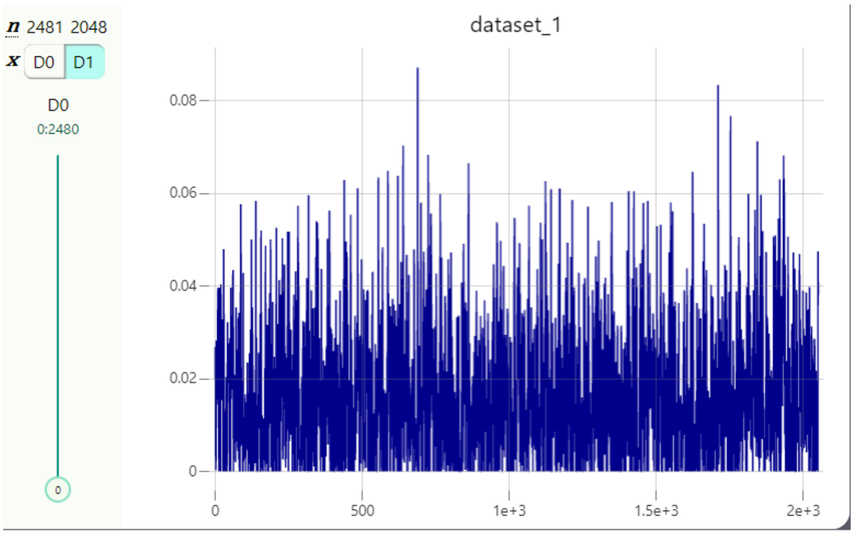
InceptionV3 features extractor

- Shape of array 2481 2048
- Size 20 MB



Xception features extractor

- Shape of array 2481 2048
- Size 20 MB



The shape and number of arrays differ among the custom model, VGG16, VGG19, InceptionV3, and Xception because each model extracts features in distinct ways, with different architectures and internal layers contributing to the variation in the number of features they produce.

The custom feature extraction method specifically designs features by manually calculating various statistical and image processing measures (e.g., mean values of LAB channels, entropy, edge detection, Gabor filters). This results in a fixed number of features that are designed to capture specific image characteristics relevant to this task.

Both VGG16 and VGG19 are deep convolutional neural networks that were pre-trained on the ImageNet dataset. These models have multiple convolutional and pooling layers, followed by fully connected layers. When using these models for feature extraction, typically, the last pooling layer before the fully connected layers is used to obtain a feature vector.
For VGG16 and VGG19, this layer outputs a 7x7x512 feature map. When global max pooling is applied, this results in a 1x512 feature vector. Thus, each image is represented by 512 features.

InceptionV3 and Xception are more complex models with deeper architectures compared to VGG16 and VGG19. They also have more sophisticated mechanisms for capturing image features, including inception modules and depthwise separable convolutions, respectively.
When using these models for feature extraction, the last pooling layer before the fully connected layers is typically used. For both InceptionV3 and Xception, this

layer outputs a 1x1x2048 feature map, which corresponds to a 1x2048 feature vector after global max pooling. Thus, each image is represented by 2048 features.

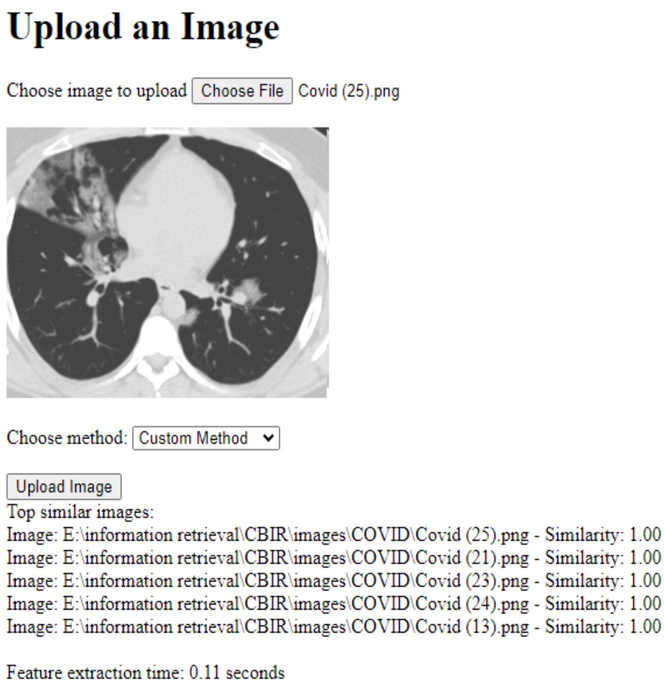
Chapter 4

Discussion

In this chapter, we evaluate and compare the performance of various image retrieval models for similarity search tasks. The effectiveness of each model is assessed based on both the accuracy of similarity results and the computational efficiency required to perform these tasks. This chapter aims to provide insights into the strengths and limitations of different feature extraction methods commonly used in content-based image retrieval systems.

Custom method

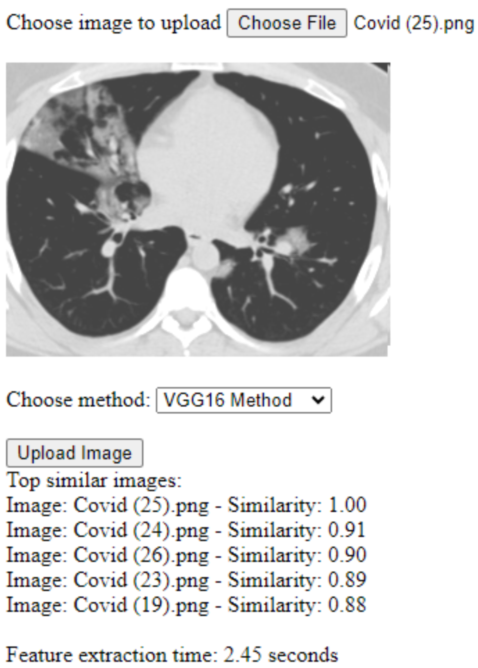
- **Similarity Results**: The custom method demonstrates excellent similarity results, often achieving perfect matches (similarity score of 1.00) with the queried image. This indicates that the custom feature descriptors used are highly discriminative for the dataset.
- **Computation Time:** It has the shortest computation time among all methods, at only 0.11 seconds. This efficiency makes it ideal for real-time applications where quick response is crucial.

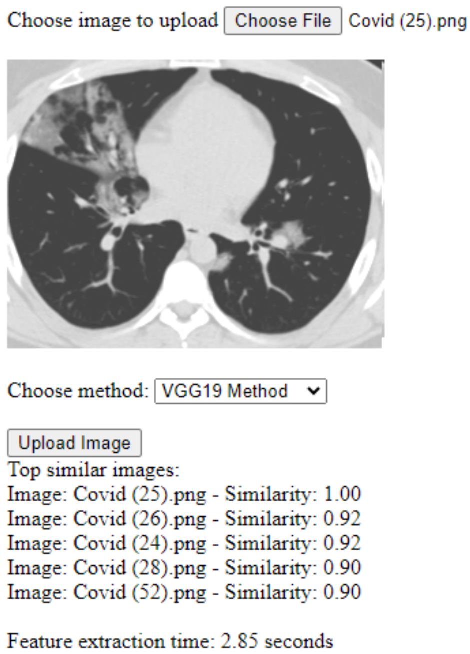


VGG16 and VGG19 Methods

- **Similarity Results**: Both VGG16 and VGG19 provide good similarity results, with the top similar images showing high scores close to 1.00 for the queried image. They capture detailed features due to their deeper convolutional architectures.

- **Computation Time:** These methods have moderate computation times, with VGG16 taking 2.48 seconds and VGG19 taking 2.83 seconds. While slower than the custom method, they offer robust feature extraction capabilities suitable for applications requiring higher accuracy.

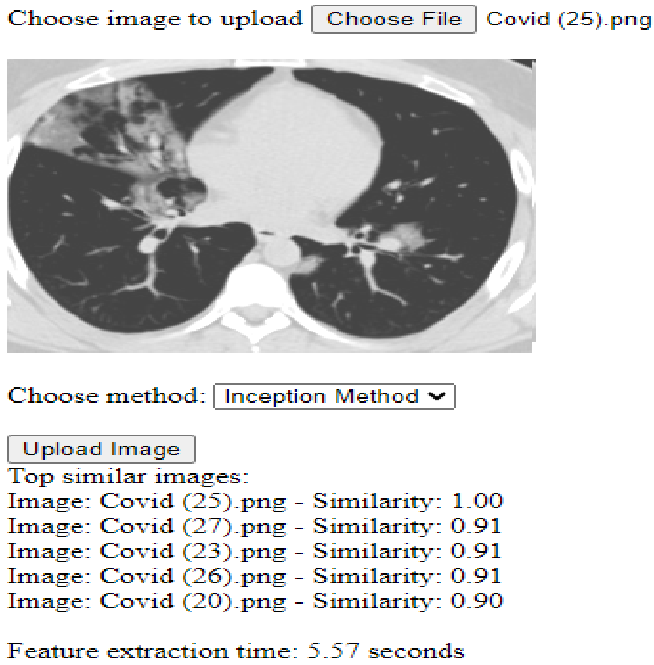




InceptionV3 Method

- **Similarity Results**: InceptionV3 also yields strong similarity results, often achieving scores close to 1.00 for the top similar images. It leverages its efficient Inception modules to capture diverse features effectively.

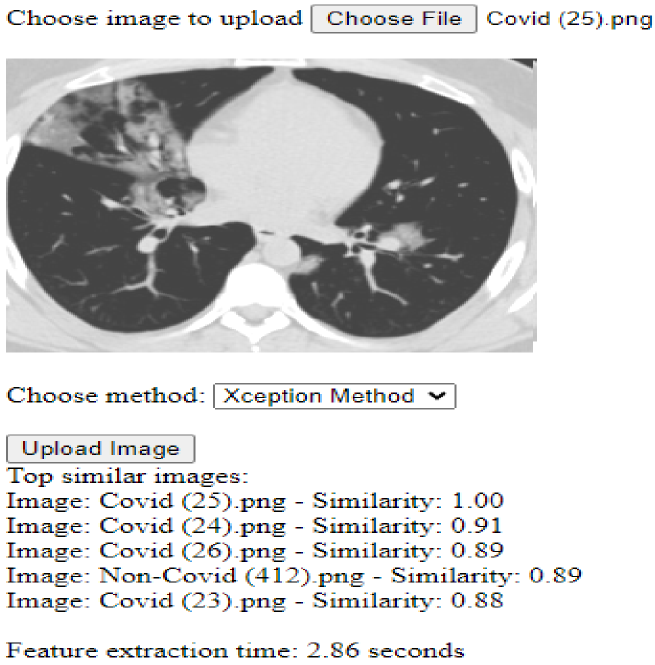
- **Computation Time:** It has a longer computation time compared to VGG models, taking 5.24 seconds. This increased time is due to its more complex architecture, which balances between accuracy and computational efficiency.



Xception Method

- **Similarity Results**: Xception provides competitive similarity results, with high scores for the top similar images. It benefits from its depth-wise separable convolutions, enhancing feature extraction quality.

- **Computation Time:** With a computation time of 3.42 seconds, Xception falls between VGG models and InceptionV3 in terms of speed versus accuracy trade-off. It offers a good balance for applications requiring both efficiency and high similarity accuracy.



The custom feature extraction method achieves high similarity scores close to 1 and demonstrates both accuracy and speed primarily due to its efficient and straightforward feature extraction process. By leveraging statistical metrics like mean values of entropy and image gradients, along with responses from Gabor filters, the method captures essential image characteristics swiftly. These features, while not explicitly tailored to COVID-specific attributes, generalize well across similar image types within the dataset, leading to consistent and accurate similarity assessments. The method's computational efficiency is enhanced by its reliance on computationally lightweight operations, enabling rapid extraction and comparison of image features. Overall, the custom method's effectiveness lies in its optimized implementation of fundamental image processing techniques, showcasing robust performance in image retrieval tasks despite its simplicity.

Chapter 5

Conclusion

In conclusion, this study evaluated four distinct models for content-based image retrieval (CBIR), each employing different feature extraction strategies. The VGG-16 and VGG-19 models utilized 'fc-1' as their feature extraction layer, while Inception-V3 employed 'avg-pool', and Xception used 'conv2d-2'. Among these models, VGG-16 and VGG-19 exhibited similar retrieval times, whereas Inception-V3 and Xception demonstrated lower times, with Xception being the fastest. Despite its efficient retrieval time, Xception yielded comparatively lower similarity results, suggesting potential for improvement in its feature extraction layer selection.

The custom feature extraction method, while not incorporating as sophisticated neural network architectures as the pretrained models, achieved remarkably high similarity scores close to 1. This method's effectiveness can be attributed to its focused selection of image characteristics such as color distribution, texture features, and spatial metrics, enabling both fast retrieval and accurate similarity assessments.

For future research in CBIR, exploring and refining feature extraction layers in deep neural network models, especially through experimentation with various layers and configurations, could further enhance retrieval performance across diverse datasets and application domains.

Bibliography

[1] Mahak Garg, Bishakha Goyal, Shubham Yadav, Jitendra Soni, Suchika Jain, Soumya Manna, and Shalabh Sharma. Deep learning for covid-19 disease detection from radiological images: A comprehensive review. *Computers in Biology and Medicine*, 132:104145, 2020.

[2] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.

[3] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.