CONTENT BASED MEDICAL IMAGE RETRIEVAL: A DEEP LEARNING APPROACH

A Thesis Presented

by

Munir Ali Seid

to

The Faculty of Informatics

of

St. Mary's University

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

in

Computer Science

February, 2023

ACCEPTANCE

CONTENT BASED MEDICAL IMAGE RETRIEVAL: A DEEP LEARNING APPROACH

By

Munir Ali Seid

Accepted by the Faculty of Informatics, St. Mary's University, In Partial Fulfillment of the Requirements for the degree of Master of Science in Computer Science

Thesis Examination Committee:

_____

Internal Examiner
Alembante Mulu (PhD)

DR. MESFIN ABEBE 20/02/2023

External Examiner
Mesfin Abebe (PhD)

_____

Dean, Faculty of Informatics
Alembante Mulu (PhD)

February, 2023

DECLARATION

I, the undersigned, declare that this thesis work is my original work, has not been presented for a degree in this or any other universities, and all sources of materials used for the thesis work have been duly acknowledged.

Munir Ali Seid
Full Name of Student

_____
Signature

Addis Ababa

Ethiopia

This thesis has been submitted for examination with my approval as advisor.

Million Meshesha (PhD)
Full Name of Advisor

Signature

Addis Ababa

Ethiopia

February, 2023

Acknowledgment

Firstly, I would like to thank the almighty God for his divine guidance and constant support throughout this study.

Next, I'd want to express my gratitude to Dr. Million Meshesha, without whom none of this work would have been possible. I owe a special appreciation to all members of the zewdutu memorial hospital radiologist, especially wendwesen and his staff, for their assistance in the collection of the x-ray images.

It is also my pleasure to thank my wonderful mother Hayat Beyan, sisters Husna and kewser, brothers Mubarek, Sebri, and Muhammad deserve a sincere thank you because without them, I wouldn't be who I am today.

Last but not least, I'd want to convey my heartfelt appreciation to everyone who has helped and supported me during my thesis research, especially Miss Wagaye Asefa, miss aster shukur who offered me her valuable time to share ideas and moral and material support and encouraged me to finish it.

Munir Ali

DEDICATION

To: Ali Seid (**baba**) may God grant you paradise. I wanted you to wear my gown and see my achievement, my father.

TABLE OF CONTENTS

Acronyms

ALIPR: Automatic Linguistic Indexing of Pictures - Real Time

ASSERT: Automated System Security Evaluation and Remediation Tracking.

CBIR: Content Based Image Retrieval

CBMIR: Content Based Medical Image Retrieval

CD-R: Compact Disc Recordable

CNN: Convolutional Neural Network

DCNN: Deep convolutional neural networks

DICOM: Digital Imaging and Communications in Medicine

FC layer: Fully Connected

HRCT: High-Resolution Computed Tomography

IBM: International Business Machines Corporation

IMedline: Medical Literature Analysis and Retrieval System Online

IRMA: Information Resources Management Administration

NHANES II: National Health and Nutrition Examination Survey

PSO: Particle Swarm Optimization

QBE: Query by Image Example

QBIC: Query by Image Content

RedLex: Radiology Lexicon

ReLU: Rectified Linear Unit

RF: Random Forest

RESNET-50: Residential Energy Services Network

VGG16: Visual Geometry Group

X-ray: X-radiation

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Ethiopia is one of the countries where overall health service has been compromised by inadequate & poorly maintained infrastructure and scarcity of health professionals. Radiological service is a resource intensive unit in a hospital and most developing countries radiological service is expected to be poor or may not be available at all. However, there is no study conducted to assess the radiological service in Ethiopia. Content-based medical image retrieval systems are designed to retrieve images that are relevant, based on detailed analysis of latent image characteristics. A Content-based medical image retrieval system maintains high-level image visuals in the form of feature vectors, which the retrieval engine leverages for similarity based matching and ranking for a given query image.

In this study, a Content-based medical image retrieval system is proposed for the retrieval of medical images for enabling the early classification of different type of diseases based on X-ray images. The Content-based medical image retrieval system is built using the deep learning models for the retrieval and classification of disease specific features using transfer learning based like VGG16, VGG19 and ResNet50. The models have been trained on standard X-ray image datasets. The dataset contains 4194 X-ray images. From this, 80% of the images are used for training and the rest for testing the model. In this research work the distance of each query image measure by Euclidean distance, content based image retrieval based on medical database.

Experimental evaluation on the standard dataset revealed that the proposed approach achieved an accuracy of 96.74% for VGG16, accuracy of 96.46% for VGG19 and accuracy of 92.30% for ResNet50. Accordingly, VGG16 is proposed based on its performance. In this study there is no means to propose medicine for the disease, proper therapy for the disease, or there is no estimate the severity of the disease once it has been classified on the medical X-ray images, which are left as a way forward for further study.

Keywords: Medical Imaging, Content-based Image Retrieval, Deep Learning, Transfer Learning, Classification

CHAPTER ONE
INTRODUCTION

1.1. Background of the study

Imaging is a fundamental component of modern medicine and is used widely for diagnosis treatment, planning and assessing response to treatment [1]. Historical records show the use of images date back to paintings on walls of cave by early man. In the pre-Roman times images were seen mostly in the form of building plans and maps. The need and use of images grew with the ages, particularly with the advent of photography in the sixteenth century. In the twentieth century, introduction of computer and advances in science and technology gave birth to low cost and efficient digital storage devices and the worldwide web, which in turn became the catalyst for increasing acquisition of digital information in the form of images [2]. In this computer age virtually all spheres of human life including commerce, government, academics, hospitals, crime prevention, surveillance, engineering, architecture, journalism, fashion and graphic design and historical research are in need of, and use of images for efficient services [3].

A large collection of images is referred to as image database [4]. Image database is a system where image data are integrated and stored. Image data include the raw images and information extracted from images by automated or computer assisted image analysis [2]. In the early 1990s, the query by image content (QBIC) system of IBM was one of the first approaches to content-based image retrieval (CBIR), and the query by image example (QBE) paradigm has since been established. Representing images by means of numerical features, relevant images are identified by comparing the signature of an example with all signatures in a repository. Initially, CBIR was applied to images from the Internet or large volumes of photographs. The signatures were obtained by extracting features such as color, texture, and shape from images. Since color has been identified as the most relevant structure for CBIR, the semantic gap was recognized. It describes the differences between image similarity on the high level of human perception and the low level of a few numerical numbers describing a mean color [5].

Imaging of the body is often complicated by the fact that anatomic structures overlap each other [6]. Diagnostic accuracy of radiographs generally refers to how well an exam

can predict the presence (or absence) of a disease or condition [7]. The technologist plays a pivotal role in improving diagnostic accuracy by providing diagnostic images. This requires a technologist to be aware of the various positions and techniques required to isolate and provide a clearer view of a body part being imaged. In addition to better viewing an anatomic part, different projections also help anatomize an abnormality or localize a foreign body [6].

Images play a very important role in human communication and it is being used long ago. Pictures build the communication method more user-friendly and clear. CBIR is the way towards finding images from huge datasets or a library of computerized images. It is the retrieving of images which have identical content of color, textures, or shapes. Processing with too amount of content commonly required a bulk amount of memory, computational power &time. A method which is used to retrieve similar images depending upon a user query from large collection of images is called CBIR [8].

Content-based means the search will analyze the actual contents of the image. The term 'content' in this context refers to colors, shapes, intensity, textures, or any other information that can be derived from the image for the purpose of measuring visual similarity during searching for image retrieval [9].

Content based image retrieval (CBIR) is the process of retrieving images from a database or library of digital images according to the visual content of the images. In other words, it is the retrieval of images that have similar content of colors, textures or shapes. Images have always been an inevitable part of human communication and its roots millennia ago. Images make the communication process more interesting, illustrative, elaborate, understandable and transparent [10].

Image retrieval attempts to search through a database and find images that are perceptually similar to a query image. This work aims to develop an efficient visual-Content-based technique to search, browse and retrieve relevant images from large-scale of medical image collections Features play a vital role during the image retrieval. For efficient query processing extract the low-level features such as texture, intensity, shape and color in order to classify the query and retrieve the similar images from the huge scale image collection of database [11].

1.2. Motivation

The CBMIR system supports doctors in making critical decisions about a specific illness or injury. By retrieving similar images and case histories, doctors can reach a more intelligent conclusion about the patient's sickness stage and diagnosis by viewing the retrieved related images from database [15].

The old technique is text-based retrieval that is used in Ethiopian hospitals. This method used manually annotated text descriptions as well as typical database techniques for image management. They cannot search un-annotated images from databases. Furthermore, it requires more time or consuming more time for annotation the image in database, and the annotation is different for different people. But in CBMIR Images is retrieved using visual information such as color, shape, and texture approach [12].

Because of the technological time, in the medicine field at various hospitals in Ethiopia, the number of collections of computerized medical images has increased rapidly and frequently. Therefore, to manage such large medical databases, it needs development of effective medical image retrieval system is required. It is also important to note that the images are used by radiologists and doctors. These images are increasingly being used to transmit information about a patient's history and it help to doctors in making important decisions regarding a particular sickness or physical harm. Therefore, CBMIR is important in the health domain because it supports radiologists and doctors.

1.3. Statement of the Problem

Nowadays, because of the alarming growth of digitization of documents, there are a number of attempts to make them easily accessible by users. In general, there are two approaches in designing image retrieval system; description based and content based. The first method is text-based retrieval. In this method, manually annotated text descriptions and traditional database techniques to manage images are used. Although text-based methods are fast and reliable when images are well annotated, they cannot search in un-annotated image databases. Moreover, text-based image retrieval has the following additional drawbacks, it requires time consuming annotation procedures and the

annotation is subjective. In content-based medical image retrieval method, images are indexed and retrieved by visual content such as color, shape and texture [12].

There are tools developed to retrieve medical images based on their content, such as the IRMA system, NHANES II, and iMedline. The IRMA system splits the image retrieval process into seven consecutive steps, including categorization, registration, feature extraction, feature selection, indexing, identification, and retrieval. This approach permits queries on a heterogeneous image collection and helps identify images that are similar with respect to global features [13]. NHANES II system contains the Active Contour Segmentation (ACS) tool, which allows the users to create a template by marking points around the vertebra [12]. I Medline is a multimodal search engine [14]. That builds tools employing a combination of text and image features. It also improves the retrieval of semantically similar images from the literature and from image databases, with the goal of reducing the semantic gap that is a significant hindrance to the use of image retrieval for practical clinical purposes [12].

ALIPR System is enabling automatic photo tagging and visual search on the web, and to interpret imaging findings. However, much of radiological practice is currently not based on quantitative image analysis, but on “heuristics” to guide physicians through rules-of thumb. ASSERT is a CBIR system for the domain of HRCT images of the lung with emphysema-type diseases. Furthermore, the visual characteristics of the diseases vary widely across patients and based on the severity of the disease. In fact, the physicians decide on a diagnosis by visually comparing the case at hand with previously published cases in the medical literature and RedLex (Radiology Lexicon) enables numerous improvements in the clinical practice of radiology, from the ordering of imaging exams to the use of information in the resulting report are current practice in medical image retrieval [12].

There are researches done on medical image retrieval using Deep Convolutional Neural Network by Qayyum et al. [15], Content-based Image Retrieval Algorithm for Medical Image Databases is also attempted by Pilevar et al. [16], during which the researcher works with different feature extraction, Similarity measurement and classification techniques.

Shewatatek [17] attempted to develop a content-based image retrieval system using segmentation and color feature that helps to organize and retrieve digital images using their visual content from large databases. Hewan [18] designed and implemented content based search system for Ethiopian art content on the Web to allow users to access different art works of Ethiopian artists and makes searching information from the web very easy and quick. Search engines were originally described as automated programs that compile and update databases without human intervention to serve as a bridge between a user and the artworks. Lingamuthu [19] provided a content based image retrieval using color, gray, advanced texture, shape features and random forest classifier with particle swarm optimization.

With the widespread dissemination of picture archiving and communication systems in hospitals, the size of medical image collections is increasing rapidly. Therefore, to manage such large medical databases, development of effective medical image retrieval system is required. However, as to the researcher knowledge, there are no local research works done to design and develop a system for searching in medical images using their content.

Hence, the aim of this study is to apply deep learning approach for developing content based image retrieval from medical image collection.

1.4. Research Questions

In this study, an attempt is made to investigate and answer the following research questions:

RQ1. Which CNN framework is suitable for developing content-based medical image retrieval model?

RQ2. How CNN models works for relevant medical image retrieval?

1.5. Objective of the Study

1.5.1. General Objective

The general objective of this study is to develop content based medical image retrieval using deep learning approach.

1.5.2. Specific Objectives

- To review literature and select suitable methods and techniques for content based image retrieval
- To prepare medical image data set for the experiment.
- To experiment the feature extraction methods and select the one that is appropriate for medical image retrieval.
- To provide medical image retrieval interface.
- To evaluate the performance of the proposed content based medical images retrieval.

1.6. Methodology

1.6.1. Literature Review

Reviewing the literature on this issue to have a thorough understanding of CBMIR and pinpoint the problem. Analysis of comparable works will aid in gathering the essential information about a certain issue and determining the best technique and tools to meet the identified difficulty. The literature review technique focuses on the CBMIR employing various deep learning algorithms, which will provide us with superior performance to keep the study on track. As a result, books and scientific articles are subjected to more thorough scrutiny.

1.6.2. Data collection

the dataset required for designing the cbmir system was collected from public hospital, the oldest and known zewditu memorial hospital that are scanned using x-ray machines.

1.6.3. Models and implementation tools

In this thesis experimental research approach is used. Different CNN models like Vgg16, vgg19 resnet50 for feature extraction and classification. Euclidean distance for distance measurements. Python as a programming language with tensor-flow and keras libraries on an anaconda3, we have used Jupyter notebook to implement the coding part. Microsoft Visio, Open CV.

1.6.4. Evaluation methods

Evaluation methods the confusion used to determine the performance of the classification models for a given set of test data, and calculate the different parameters for the model, using such as accuracy, precision, recall and F-measure .

1.7. Scope and Limitation of the Study

The main goal of this study is to develop CBMIR system so as to enhance medical image searching. The focus of the study will be medical image retrieval from database by its visual content. To this end, the different techniques for feature extraction will be experimented. The study will attempt to cover retrieval of image from database using image query. And in addition, we will use CNN algorithm for classifying the image.

The X-ray data's are collected from the Zewditu Memorial Hospital. The collected medical images are with DICOM file format and the amount of the data is about 4194images. Accessing medical images was a difficult task. In most cases, the privacy and security regulations of most hospitals do not allow any information to be out from hospitals.

We have planned, but not done in this paper are to work with a big dataset and classes of different radiology like CT scan, MRI, Mammogram, Ultrasound but we have done only with the x-images. And second the content based image retrieval doesn't display detail patient information like name age gender and others.

1.8. Significance of the Study

The study has a great significance to Hospitals, such as Zewditu Memorial Hospital. Major benefits that would be obtained from the study are:

- It helps Hospitals to retrieve images easily by its visual from the database
- Ease data handling and the user can easily access.
- Time requires is less to find those entire related images.
- More than one related outcomes occur by only one search (If more than one equally likely image present in the database).
- To deliver the needed image on time in order to improve the quality and efficiency of care processes
- It facilitates researchers in getting better performance of image retrieval system using different algorithms into practice CBMIR. If researchers have established typical and effective medical image retrieval, they may spend more time to their study and research more.

1.9. Organization of the Thesis

This study will have five chapters. The initial chapter provides background of the study, motivation, statement of the problem, objectives of the study general and specific, methodology, scope and limitation of the study, significance of the study, and organization of the thesis.

The second chapter of this thesis reviews literatures and research gap is identified. and it involves the following main tasks: History of x-ray, x-ray definition, types of x-ray views, converting image file formats, overview of cbir, feature extraction, vgg16 and vgg19, resnet50, inception V3, classification retrieval using similarity matching, Euclidean distance and CNN layers.

Chapter three covers research design, data collection and preparation image preprocessing, feature extraction methods and implantation tools.

Chapter four: overview, The Proposed Architecture, feature extraction using proposed model, hyper parameter settings, per training CNN, training components of the proposed vgg16, 19 and resnet50 models, experimental results of vgg16, 19 and resnet50 models, selecting the best performing model, developing content based medical image retrieval, and evaluating CBMIR.

Chapter five will summarizes the investigation up on bringing the findings obtained, the conclusions drawn, and the future work recommendations made.

CHAPTER TWO
LITERATURE REVIEW

2.1. Overview

Medical image databases will play a crucial part in future diagnosis, allowing doctors to know and decide more about a patient by viewing the image that has been retrieved. In order to extract the key data from x-ray images, radiologists must efficiently arrange and store them. Medical image content representation and retrieval is playing an increasing role in a wide spectrum of applications within the clinical process. For the clinical decision-making it can be useful to refer x-ray images of the same modality or the same anatomic region for the identification of certain pathologies. In this context, Content Based Image Retrieval (CBIR) has emerged as a powerful tool to efficiently retrieve x-ray images visually similar to a query image during last several years. In CBIR, the main idea is to represent each image as a feature vector and to measure the similarity between images with distance between their corresponding feature vectors according to some metric. Finding the correct features to represent images, as well as the similarity metric depend on the image domain and the goal of the retrieval system [20].

Medical gray scale images can be represented as a set of low-level visual features such as color, and texture features. But several image retrieval systems focus only on feature extraction for searching relevant images. It has been shown that the combination of feature extraction methods can give better retrieval performance [21].

With the widespread dissemination of picture archiving and communication systems in hospitals, the size of medical image collections is increasing rapidly. Therefore, to manage such large medical databases, development of effective medical image retrieval system is required. Apart from this task of managing database, a specific CBMIR system helps the doctors in making critical decisions about a specific disease or injury. By retrieving similar images and case histories, the doctors could make a more informed decision about the patient's disease stage and diagnosis. [15].

2.2. Introducing X-Ray

The first application of X-ray techniques was made in Germany by Wilhelm Conrad Rontgen in November 1895 [22]. Shortly after the discovery of X-rays, another form of penetrating rays was discovered. In 1896, French scientist Henri Becquerel discovered natural radioactivity. Many scientists of the period were working with cathode rays, and other scientists were gathering evidence on the theory that the atom could be subdivided. Some of the new research showed that certain types of atoms disintegrate by themselves. It was Henri Becquerel who discovered this phenomenon while investigating the properties of fluorescent minerals. Becquerel was researching the principles of fluorescence, wherein certain minerals glow (fluoresce) when exposed to sunlight. He utilized photographic plates to record this fluorescence. Becquerel continued to test samples of uranium compounds and determined that the source of radiation was the element uranium. [23]. The development of modern science would clearly have been difficult without the benefits of Rontgen rays and subsequent technological improvements. Companies and institutions around the world are still investing in new technologies and process components, developing even more new X-ray tube patents. [22].

2.2.1. What is X-Ray?

X-rays are a form of electromagnetic radiation with wavelengths ranging from 0.01 to 10 nanometers [24]. In the setting of diagnostic radiology, X-rays have long enjoyed use in the imaging of body tissues and aid in the diagnosis of disease. Simply understood, the generation of X-rays occurs when electrons are accelerated under a potential difference and turned into electromagnetic radiation [25]. As with any electronic vacuum tube, there is a cathode, which emits electrons into the vacuum and an anode to collect the electrons, thus establishing a flow of electrical current, known as the beam, through the tube. A high-voltage power source, for example 30e150 kV, is connected across cathode and anode to accelerate the electrons. The X-ray spectrum depends on the anode material and the accelerating voltage. In many applications, the current flow is typically in the range 1 mA to 1A [22].

2.2.2. Types of X-Ray View

Common Types of x-ray view are Posterior Anterior (PA), Anterior - posterior (AP), and lateral views [6].

Posterior Anterior (PA) - is a view of looking at the patient from the front, face-to-face to obtain the image, the patient is asked to stand with their front against the film, PA (poster-anterior) view can be utilized when the patient is unable to achieve the position required for the AP view.

Anterior - posterior (AP) -is towards the back of the body image can also be taken with the patient sitting or supine on the bed. View when the patient is too unwell to tolerate standing or leaving the bed.

Lateral-is a view of lateral visual perspective from the side that obtains the side-view of the patient**.**

2.2.3. Converting image file format

Since we aim to work with the DICOM image, it needs to be converted to JPG file format because DICOM files are large in sizes; so needs to reduce file size to save memory space.

When DICOM file format is changed to jpg, it decreases their quality; so there is a need to convert them to gray scale images to enhance the contrast. Obviously, this makes searching better with respect to speed and performance.

2.3. Overview of Content Based image Retrieval

Content Based Image Retrieval (CBIR) is a technique which uses visual features to search user required images from large image database according to user's requests in the form of a query image [26]. Images are retrieved on the basis of similarity in features where features of the query specification are compared with features from the image database to determine which images match similarly with given features. Content-Based Image Retrieval follows the steps of feature extraction, Classification, and retrieval using similarity matching [27].

2.3.1. Feature Extraction

Feature extraction is the first process in CBIR that aims to convert human perception into a numerical description that can be manipulated by machines [28].It is the process where features such as shape, texture, color, etc. are used to describe the content of the image. The features further can be classified as low-level and high-level features. In this stage visual information is extracts from the image and saves them as features vectors in a features database. For each pixel, the image description is found in the form of feature value (or a set of value called a feature vector) by using the feature extraction. These feature vectors are used to compare the query with the other images and retrieval [29]. Feature extraction is a means of extracting compact but semantically valuable information from images. This information is used as a signature for the image. Similar images should have similar signatures [30]. In the proposed feature extraction we extract using CNN image feature extractors, such asVGG16, VGG19, RESNET-50 and select one of the best for the retrieval.

2.3.1.1. VGG16 and VGG19

VGG (Visual Geometry Group) is a convolution neural net (CNN) architecture and used to win ILSVR(ImageNet) competition in 2014. The major characteristic of this architecture is instead of having a large number of hyper parameters, they concentrated on simple 3×3 size kernels in convolutional layers and 2×2 size in max pooling layers. In the end, it has 2 FC (Fully Connected layers) trailed by a softmax for output. The most familiar VGG models are VGG16 and VGG19 which include 16 and 19 layers, respectively. The difference between VGG-16 and VGG-19 is that VGG-19 has one more layer in each of the three convolutional blocks [31].

The Visual Geometry Group network (VGG-16) can serve as a high accurate feature extractor. The input image to the VGG-16 network is of fixed size, i.e., 3×224×224. It is passed through a stack of various convolutional layers of different receptive fields. The stride rate for convolutional layers and pooling layers remains the same throughout the VGG-16 network which is 3×3 with stride 1 in convolutional layer and 2×2 with stride 2 in pooling layer. The first two convolutional layers have 64 and 128 filters, respectively. The rest of the convolutional layers include 256, 512 and 512 filters, respectively. Border pixels are padded before each convolutional operation, which can preserve the features

maps size same to the input. The VGG-16 is ended with three fully connected layers. The first two FC layers consist of 4096 neurons while the final FC layer compresses the features to 1000 dimensions. [32] .

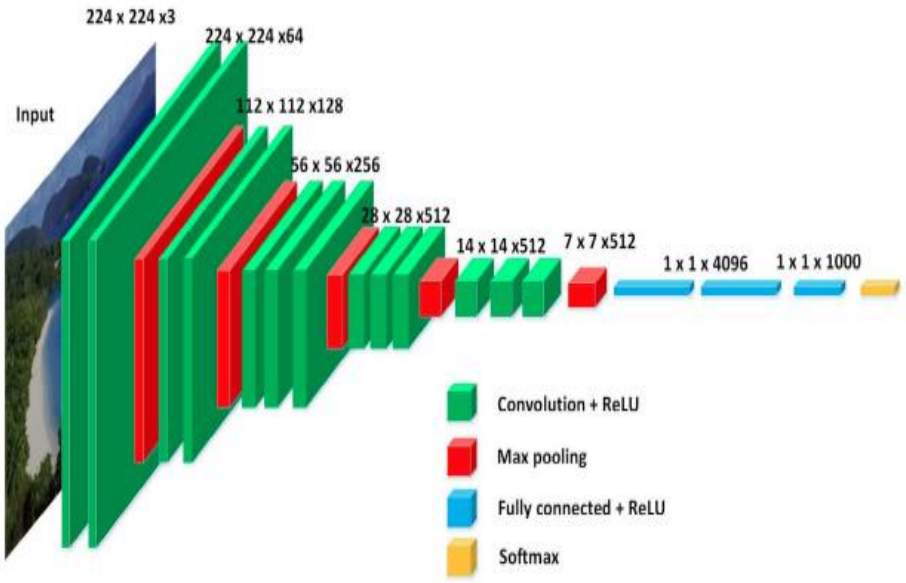


Figure 1. VGG-16 network architecture for feature extraction **[32]**

2.3.1.2. Resnet50

ResNet50 is the winner of ILSVRC 2015. The principal innovation is the introducing of the new architecture network-in-network using residual layers. The Renset50 consists of five steps each with a convolution and Identity block and each convolution block has 3 convolution layers and each identity block also has 3 convolution layers. Resnet50 has 50 residual networks and accepts images of 224 × 224 pixels [31].

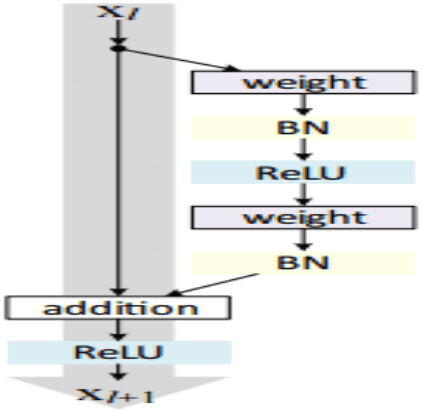


Figure 2. ResNet50 architecture for feature extraction **[33]**

2.3.1.3. Inception V3

The "Inception" micro-architecture was first introduced in 2014. The goal of the inception module is to act as a "multi-level feature extractor" by computing 1×1, 3×3, and 5×5 convolutions within the same module of the network. The output of these filters are then stacked along the channel dimension and before being fed into the next layer in the network [34]. The original incarnation of this architecture was called GoogLeNet, but subsequent manifestations have simply been called Inception vN where N refers to the version number put out by Google [33].

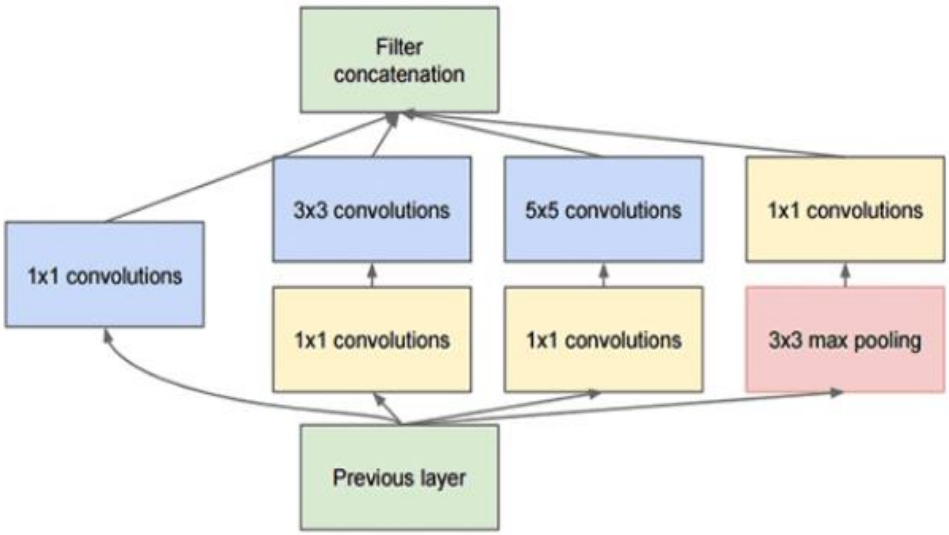


Figure 3.Architecture of Inception V3 for feature extraction **[33]**

2.3.2. Classification

During the classification phase, a deep convolutional neural network is trained for classifying medical images by following supervised learning approach. For this purpose, the medical images are divided into various classes based on body part or organ information, more details on the dataset. Images have been used for analysis hence the task is to classify each image into a class, which ultimately formulates into a multiclass image classification problem.

Typically, image classification algorithms have two modules i.e., feature extraction and classification module. DCNN learns both hierarchy of deep convolutional features and classifier from the training image data in an end-to-end learning framework. Deep learning algorithm learns low-level, mid-level and abstract features directly from the images as opposed to making domain specific assumptions, which is the case for handcrafted features. Hence, it can identify the class of a query image more effectively and therefore the learned features can be used for image retrieval task. Inspired by this property, a DCNN model is trained and optimized for multiclass classification problem in context of image retrieval, image classification has often been used as a pre-processing step for reducing the response time to query image in large databases and improving accuracy. [35].

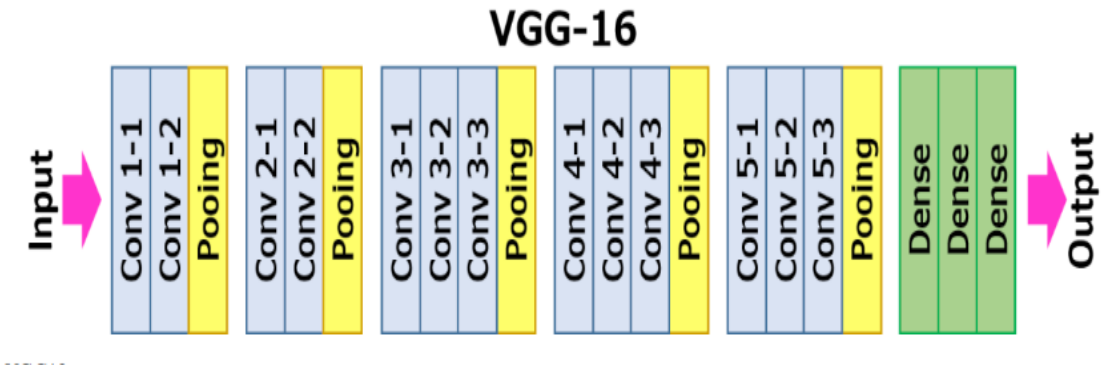


Figure 4.Image classification-based framework using VGG16 framework **[36]**

2.3.3. Retrieval using similarity Matching

The similarity functions seek to calculate the content difference between two images based on their features. One of the images is given as search parameter and another is stored in the database and had their features previously extracted. There are four major classes of similarity measures: color similarity, texture similarity, shape similarity, and object and relationship similarity [27]. The information about each image is stored in its feature vectors for computation process and these feature vectors are matched with the feature vectors of query image (the image to be search in the image database whether the same image is present or not or how many are similar kind images are exist or not) which helps in measuring the similarity. This step involves the matching of the above stated features to yield a result that is visually similar with the use of similarity measure method

called as Distance method. Most common distance measurement is Euclidean distance [37].

2.3.3.1. Euclidean Distance

This distance metric is most commonly used for similarity measurement in image retrieval because of its efficiency and effectiveness. It measures the distance between two vectors of images by calculating the square root of the sum of the squared absolute differences and it can be calculated [38]. Euclidean distance measures the similarity between the two different feature vectors. The formula for Euclidean distance is shown below. Q and D feature vectors of the Query image and database image [39].

$$\text{ED} = \sqrt{\sum_{i=0}^{n} [Qi - Di]^2}$$

2.4. Cnn (Convolutional Neural Network)

In recent years, using features get through deep CNNs has achieved impressive results in generic image classification, object recognition, detection, retrieval, and other related tasks. But in the medical field, there is not much attention on exploring deep neural networks CBMIR task, partially because the amount of labeled medical images is typically limited [40]. The CNN architecture includes several building blocks, such as convolution layers, pooling layers, and fully connected layers. A typical architecture consists of repetitions of a stack of several convolution layers and a pooling layer, followed by one or more fully connected layers. The step where input data are transformed into output through these layers is called forward propagation although convolution and pooling operations described in this section [41] . There are two main parts to a CNN architecture one convolution tool that separates and identifies the various features of the image for analysis in a process called as Feature Extraction and the second part fully connected layer that utilizes the output from the convolution process and predicts the class of the image based on the features extracted in previous stages [42].

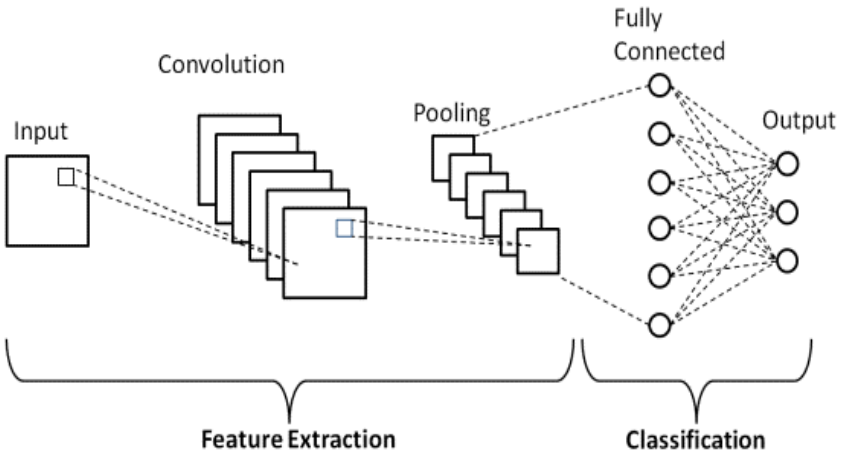


Figure 5. CNN structure for feature extraction and classification **[42]**

2.4.1. Cnn Layers

There are three types of layers that make up the CNN which are the convolutional layers, pooling layers, and fully-connected (FC) layers. When these layers are stacked, CNN architecture will be formed. In addition to these three layers, there are two more important parameters which are the dropout layer and the activation function which are defined below [42].

2.4.1.1. Convolutional Layer

This layer is the first layer that is used to extract the various features from the input images. In this layer, the mathematical operation of convolution is performed between the input image and a filter of a particular size MxM. By sliding the filter over the input image, the dot product is taken between the filter and the parts of the input image with respect to the size of the filter (MxM).

The output is termed as the Feature map which gives us information about the image such as the corners and edges. Later, this feature map is fed to other layers to learn several other features of the input image [42].

2.4.1.2. Pooling Layer

In most cases, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrease the size of the convolved feature map to reduce the computational

costs [43]. This is performed by decreasing the connections between layers and independently operates on each feature map. Depending upon method used, there are several types of pooling operations.

In Max Pooling, the largest element is taken from feature map. Average Pooling calculates the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is computed in Sum Pooling. The Pooling Layer usually serves as a bridge between the Convolutional Layer and the Fully Connected Layer [42].

2.4.1.3. Fully Connected Layer

The Fully Connected (FC) layer consists of the weights and biases along with the neurons and is used to connect the neurons between two different layers. These layers are usually placed before the output layer and form the last few layers of a CNN Architecture. In this, the input image from the previous layers are flattened and fed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations usually take place. In this stage, the classification process begins to take place. [42].

2.4.1.4. Dropout

Usually, when all the features are connected to the FC layer, it can cause over fitting in the training dataset. Over fitting occurs when a particular model works so well on the training data causing a negative impact in the model's performance when used on a new data.To overcome this problem, a dropout layer is utilized wherein a few neurons are dropped from the neural network during training process resulting in reduced size of the model [42].

2.4.1.5. Activation Functions

Finally, one of the most important parameters of the CNN model is the activation function. They are used to learn and approximate any kind of continuous and complex relationship between variables of the network. In simple words, it decides which information of the model should fire in the forward direction and which ones should not at the end of the network. It adds non-linearity to the network. There are several commonly used activation functions such as the ReLU, Softmax, tanH and the Sigmoid functions [43]. Each of these functions has a specific usage.

For a binary classification CNN model, sigmoid and softmax functions are preferred and for a multi-class classification, the rectified linear unit (ReLU) function is the most widely used activation function in today's networks. There is an advantage of using the ReLU function compared to the other activation functions in that it does not activate all the neurons at once. If the input is negative, then it is converted to 0, and the neuron is not activated. If the input is positive, it returns the positive value of x and the neurons get activated. Consequently, only a few neurons are activated at a time, making the network sparse and very efficient. The ReLU function also served as a significant advancement in the field of deep learning by overcoming the vanishing gradient problem. The softmax function is ideally used in the output layer of the classifier where we are actually trying to get the probabilities to define the class of each input. As a result, it is easier for us to categorize data points and determine to which category they belong [42].

2.5. Related Works

CBMIR system helps the doctors in making critical decisions about a specific disease or injury. By retrieving similar images and case histories, the doctors could make a more informed decision about the patient's disease stage and diagnosis [15].

MortezaBabaie et al. [44] proposed the retrieval of similar x-ray images from big image data using radon barcodes with single projections. They used single Radon projections for CBIR in its both forms, namely real-valued (single projection Radon, SP-R) and binary (single projection Radon barcode, SP-RBC) implementations. First, they attempt to automate preprocessing steps. In the next step, they used Radon projections to reach top similarity for each projection separately. Also they used the exploitation method to find the most similar images in the pre-selected set.

Ganesan and Subashini [45] come up with a content based approach to medical X-ray image retrieval using texture features. The proposed system retrieves the nearest three similar query images based on the City block distance method. Image feature textures are extracted using Grey-Level Co-occurrence Matrix (GLCM) and five of the high level GLCM features like energy, entropy, contrast, homogeneity and correlation are fed to the BPNN for classification of X-Ray images. Classification is used to classify the X-ray images based on its features. Here back propagation neural network classifier is

employed to classify the X-Ray images into any of the six classes of images. The back propagation neural network is the most commonly used neural network in classification applications. The most popular classifier, Support Vector Machine (SVM) is then deployed to yield the results of the most relevant images out of the query image and for which the network is trained.

Shivamurthy et al. [46] designed a powerful computerized lung division framework for midsection X-beam pictures. DWT algorithm is applied to extract the feature of given query input image and classify the images as well as retrieve the chest images using Matlab tools. Bhattacharya coefficient algorithm is used to find the similarity score of all the images so as to determine the picture resemblance score for all images present in the database.

Pilevar [16] proposed content-based image retrieval algorithm for medical image databases. Firstly, the images are segmented after necessary edge detection, using a low pass filter. By editing, deleting or correcting the insignificant segments, the experts provide the desired segmentations, Different features and shapes. Next the images are classified into specific predefined anatomical pathology classes, and then the images are coded with character strings of a length of six characters which are called class-codes. Or, any extracted feature there exist a feature vector into which the images are mapped. All images are known by their related feature vector. Therefore, too many feature vectors have to be supported by the image database system, while different measurement strategies can be applied. There are different metric functions for determining the similarity degree of the images with each other. Euclidean distance is used to evaluate distances in n dimensional vector spaces. Ten of the most similar images to the query images are selected from database.

Puliet al. [47] developed an efficient content-based medical image retrieval system for clinical decision support in brain tumor diagnosis. There is main two part of this system one is to offline part to create a database and other is online part which we used to retrieve the similar image. In the off-line phase, MR images are automatically segmented using k-means clustering technique to extract brain tumor from MR image. Tumors can be well discriminated by their shape and texture characteristics. These features are fed

into SVM and ANN classifier and assign label to the image as cancerous or non-cancerous tumor. In second stage local feature such as Local Binary Pattern (LBP) is extracted are extracted from the brain tumor for discrimination between tumors within the class. Similarly, in the online phase, the class label of the query image is identified using ANN classifier based on rotation invariant global shape and texture features of tumor. Using this label, the similarity comparison is only to images with similar class labels in the database. This reduced search time from the database. Then, the features of the query image are compared with database using Euclidian distance and retrieved most similar K images. Classification or Recognition process is for decision making, like this query image fit in which class or looks like. It means, in the phase of classification characters are identified and assign labeling. Performance of the classification depends on good feature extraction and selection. Various classification techniques are available and they all are ultimately based on image processing and artificial intelligence. Support Vector Machine, Artificial Neural Network and Naive Bayes classifier are used.

Bhandi and Devi [48] proposed an image retrieval using features from pre-trained deep CNN. In this paper uses features extractor VGG16 CNN model extract every image from the given dataset. VGG16 model is trained on ImageNet VGG16 model consists of 5 convolution blocks and each convolution block contains two convolution layers (size 3X3) and one maxpooling layer (size 2X2). The final classification step of the model consists of fully connected (FC) layers. And implantation part is using python use cosine distances for calculating the image similarity and retrieve top N results based on fetch size. We repeat the image retrieval experiment for each query image by varying the number of images retrieved. Average precision rates for various fetch sizes and for different classes are measured.

Chaudhari et al. [49] attempted to develop content based image retrieval using color and shape features for segmenting the query image into 5 classes based on its brightness and calculates the Euclidean distance between the respective classes of query image and database image attributes. Both color and shape retrieval algorithms are implemented in MATLAB with the database of 570 images. All the images are stored in JPEG format. To evaluate the performance of the image retrieval algorithm we use the two most well-known parameters; precision and recall.

Lingamuthu et al. [19] Suggested a new approach for Content Based Image Retrieval (CBIR) Shape features and shape invariant features are computed by using Contour based shape feature extraction methods and image moment extraction methods. The extracted features had been selected and combined by using Particle Swarm Optimization (PSO). The Random Forest (RF) classifier is used for classifying the database images based on the training images. The performance of the proposed CGATSFRFOPSO method is compared to the CGATMDOPSO method. The parameters taken up for performance comparisons are accuracy, precision, recall, and execution time. The experimented values of the proposed model are tabulated in the tables and the performance is compared with the previous method.

As pointed out by Thomas et al. [50], IRMA splits the retrieval process into seven consecutive steps of processing. Each step represents a higher level of image abstraction and content understanding. First, the categorization step aims at determining for each image entry the imaging modality and its orientation as well as the examined body region and functional system. For that, a detailed hierarchical coding scheme was developed Automatic categorization is based on a reference database of images selected arbitrarily from clinical routine and classified by experienced radiologists. Second, registration in geometry (rotation, translation, scaling) and contrast generates a set of transformation parameters that is stored for the corresponding image in each of its likely categories and utilized at higher layers of abstraction. Registration is based on prototypes which are manually defined for each category, and further incorporate medical expert knowledge into the IRMA system.

Thomas et al. [50] Further noted that, the feature extraction step derives local image descriptions, i.e. a feature value (or a set of values) is obtained for each pixel. These can be category-free (e.g. resulting from edge detection or regional texture analysis) or category-specific, such as the application of an active shape model that explicitly uses a-priori knowledge derived from the respective category. Decoupling feature selection from feature extraction allows integrating both image category and querying context into the abstraction process. For instance, the same radiograph might be subject to fracture or cancer examination, resulting in a contour-based or texture-based combination of features (feature sets) such as the contour set or texture set, respectively. In order to avoid

exhaustive computation during query processing, these feature sets are pre-computed for each image in each likely category.

Indexing provides an abstraction of the previously generated and selected image features, resulting in a compact image description. According to the selected feature set, this is done via clustering of similar image parts into regions represented by their second area moment description as ellipses ("blobs"). In contrast to the Blob world approach, this is done at multiple resolutions yielding a multi-scale blob-representation of the image ("blob tree"). Note that hierarchical indexing enables the processing of regions of interest (ROIs), which are marked by the user when issuing a query.

The identification step provides linking of medical a-priori knowledge to certain blobs generated during the indexing step. It relies on the prototypes defined for each category, which are labeled locally by medical experts, and the corresponding parameters for geometry and contrast registration. Thus, identification is the fundamental basis to introduce high-level image understanding by analyzing regional or temporal relationships between the blobs.

In IRMA, the retrieval itself is processed either on the abstract blob level or referring to identified objects. Note that only the retrieval step requires online computations while all other steps can be performed automatically in batch mode at entry time of an image into the database. This, of course, requires offline computation of all paths generated by the categorization and the feature selection step.

Shewatatek [17]developed a content-based image retrieval system using segmentation and color feature. CBIR helps to organize and retrieve digital images using their visual content from large databases. Segmentation is done using the K-Means clustering algorithm, feature extraction by extracting the color feature and different similarity measurements such as Minkowski-Form Distance, Quadratic Form (QF) Distance, Mahalanobis Distance and Kullback-Leibler (KL) Divergence and Jeffrey-Divergence (JD). The metric used to measure the performance of the system is the one that is used by simplicity the performance of Simplicity is improved by the usage of classification and more number of features. Similarly, it is possible to improve the performance of the new system by using additional features and indexing.

Hewan [18]attempted to develop content based search system for Ethiopian art content on the web. The search system is designed and implemented to allow users to access different art works of Ethiopian artists and makes searching information from the Web very easy and quick. Search engines were originally described as automated programs that compile and update databases without human intervention to serve as a bridge between a user and the artworks. Using efficient and advanced compression techniques and domain-specific search engine simply by adding domain-specific keywords to the user's input query, large piece of artwork can be easily presented to users. By providing a reasonable digitized copy of the original artworks, we avoid directly handling some delicate artworks.

Review of related works shows the need for designing content based image retrieval to facilitate searching from medical image database. Accordingly most of the studies are conducted by foreign scholars. On the contrary, there are few local research conducted to design a generic content based image retrieval. Even such systems cannot be used for medical images, because of their file format and requirement of a system with high precision. Hence this study attempted to develop content based medical image retrieval by applying CNN for feature extraction and classification.

Table 1 Summary of related works

Author	Title	Methodology	Research Gap
MortezaBabaie et al. [44]	Retrieval of similar x-ray images from big image data using radon barcodes with single projections.	Single Radon projections, namely real-valued (single projection Radon, SP-R) and binary (single projection Radon barcode, SP-RBC) implementations, exploitation method to find the most similar images.	The methodology provided here short binary vectors, or barcodes. And more time consuming to retrieve the most similar images.

Ganesanand Subashini[45]	Content based approach to medical X-ray image retrieval using texture features.	City block distance method as distance measurement, Image feature textures are extracted using Grey-Level Co-occurrence Matrix (GLCM), Support Vector Machine (SVM) to classify.	The model proposed in this research uses an ML technique and uses only six classes of images and The system retrieves the nearest only three similar query images.
Shivamurthy et al. [46]	a powerful computerized lung division framework for midsection X-beam pictures	Bhattacharya coefficient algorithm as similarity measurement DWT algorithm, Matlab tools	In this paper images are retrieval only worked in chest radiological images.
Pulietal. [47]	an efficient content-based medical image retrieval system for clinical decision support in brain tumor diagnosis	g k-means clustering technique to extract brain tumor from MR image, ANN classifier, Euclidian distance and retrieved most similar K images	In this paper only done classification accuracy not done the cbir accuracy. And only 294 MRI images.
Chaudhari et al. [49]	to develop content based image retrieval using color and shape features	Euclidean distance used as distance measurement. Implementation tool MATLAB, for	The drawback of this effort is that due to the dataset's limitations,

		evaluations precision and recall.	

2.6. Research Gap

A separate study on the problem of CBIR was conducted, and they used different algorithms to feature extraction and classification to retrieve related images from databases and libraries. The majority of studies used retrieved images based on their contents, and most retrieval systems do as well. The mentioned paper employs machine learning classifiers such as BPNN and SVM, as well as other methods, and works for a specific or one part of the body such as the chest or hand, and they uses small datasets, and retrieves no more than three related images. As a result, it must be able to work with various x-ray images, which we have done in the proposed using deep learning algorithm models to retrieval. In addition to retrieving similar 5 related images. Accordingly most of the studies are conducted by foreign scholars. The related work reviews don't evaluate the performance of cbir but the classification. On the contrary, there are few local research conducted to design a generic content based image retrieval. Even such systems cannot be used for medical images. As a result, we have provided Content Based Medical Image Retrieval: A Deep Learning Approach.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1. Overview

Content based image retrieval becomes a wide area of research. The retrieval of medical x-ray image passes through a series of steps/procedures that would be applied to medical image. This chapter focuses on the description of methodology that is used in order to accomplish this thesis including methods to implement the model, data collection, data preparation, implementation and evaluation techniques which are used to evaluate the model. In this thesis experimental research approach is used. Different experiments are carried out by using different deep learning algorithms. In addition to this, there are a lot of experiments conducted. This research is designed to apply image processing technology for content based image retrieval. Methodology is used to define the step by step process followed in conducting the study. The CBMIR system assists doctors in making critical decisions on a certain sickness or harm. Doctors might make a better intelligent decision concerning the patient's illness stage and diagnosis by retrieving matching images and cases records. Content based image retrieval for x-ray images will do using CNN algorithm and similarity matching. CNN algorithm used to extract features and classification after the extraction. And then retrieve using by their similarity.

3.2. Research Design

This thesis is conducted with three main phases. The first phase includes identifying the domain of the problem that means understanding the problem by reviewing different kinds of literature. Then objectives of the thesis are formulated including the general and specific objectives. The second phase is about data preparation and design of the thesis. During data preparation data is collected from Zewditu Memorial Hospital that is scanned using X-ray machines, then labeled with medical experts, finally splitted into training, validation, and testing. After data preparation, design of the model is performed. The third phase is about implementation of the thesis, in this phase the designed model is implemented with appropriate tools and methods. The designed model is trained and tested with the appropriate data. During the training phase, the model is

constructed using different deep learning algorithms. After getting the optimal model during training and validation, this model is evaluated with test data to measure its performance.

3.3. Data Collection and Preparation

A proper dataset is crucial to train an accurate and efficient machine learning model. Therefore, data preparation is the first step in designing any machine learning model. This chapter presents the data collection and preparation of the x-ray medical image dataset that is used to design the model that retrieves different x-ray medical images from the data base.

The dataset required for designing the CBMIR system was collected from public hospital, the oldest and known Zewditu Memorial Hospital that are scanned using x-ray machines. Any experimental study is heavily reliant on the dataset. The Data for this research is a collection of x-ray DICOM images which is collected by employing x-ray scanner. The scanned images are stored in the form of DICOM (Digital Imaging and Communications in Medicine) file formats. This file format takes high File size and high resolution. for this research is a collection of x-ray images collected from the Zewditu Memorial Hospital using CD-R the image format is DICOM file formats. Totally we have 4194 images. All the images are checked for quality, labeled and certified by domain experts (x-ray medical experts) of Zewditu Memorial Hospital. Summary of x-ray medical different type and number of images of each X-ray image type are described in Table 1 below.

Table 2. Number of Samples per Class of the X-ray Medical Images Dataset

No.	Type of X-ray image	Number of original images
1	Abdomen	114
2	Chest	1773
3	C-spine	194
4	Foot	300
5	Hand	251
6	Hip Joint	31

7	Knee	387
8	L-spin	425
9	Pelvis	117
10	Shoulder Joint	97
11	Skull	272
12	T-fibula	112
13	Wrist	121
Total		**4194**

The following Figure 6 shows sample X-ray medical images of Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist that are collected from the Zewditu Memorial Hospital.

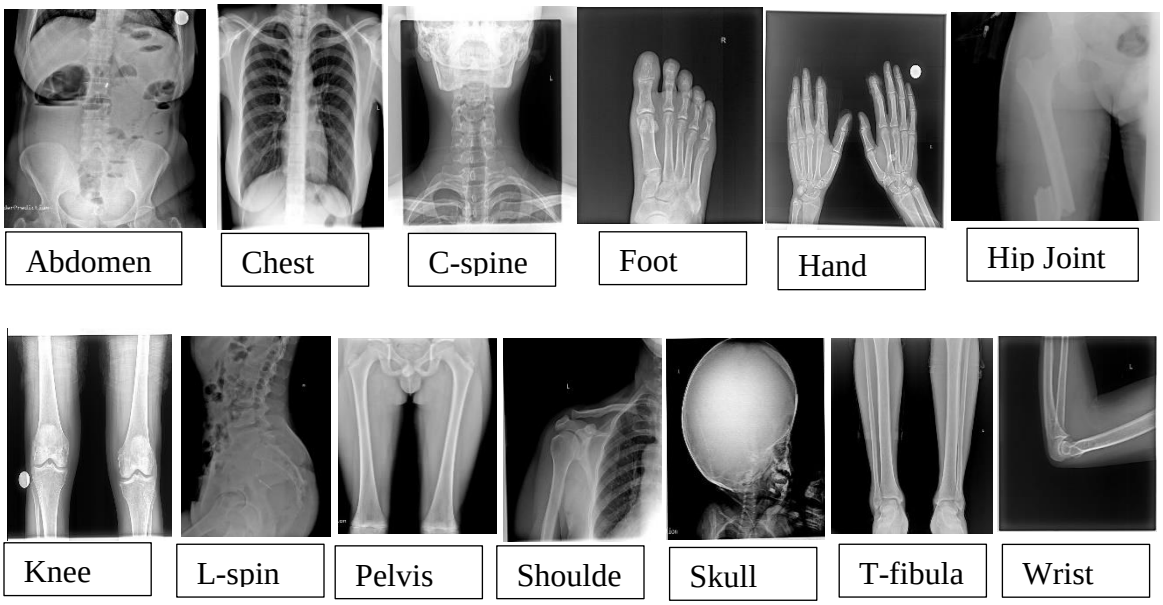


Figure 6.Sample images of the different parts of the body

3.4. Image Preprocessing

Preprocessing used to be the major important task when we want to work with image processing. It is a transformation of the raw data before feed into the neural network or deep learning algorithm. However, in CNN there is no need for explicit preprocessing on the dataset because the algorithm can take raw pixels of the image and learn the features

by itself. But the images contained in the prepared dataset have different sizes. Therefore, size normalization into 64 x 64 pixels is performed in the dataset in order to get similar size of all images for the CNN algorithm and to decrease the computational time of training because the model is trained on a standard PC with limited hardware resources such as processor, and memory.

3.5. Feature Extraction Methods

Feature extraction is a means of extracting compact but semantically valuable information from images. This information is used as a signature for the image. Similar images should have similar signatures. Deep learning methods, such as VGG is a convolutional neural network model pre-trained on ImageNet dataset. VGG is a convolutional neural network model for image recognition proposed by the Visual Geometry Group in the University of Oxford (COUNTRY).

There are two versions of the VGG model, the first one is the VGG16 and the second one is the VGG19 [30].The VGG16 has 16 weight layers and the VGG19 has 19 weight layers in the network. The model accepts 224 × 224 RGB images as an input and gives 1000 classes of ImageNet dataset (contains 14 million images belonging to 1000 classes) [45]. The input is passed through stacked convolution layers of the model with a 3 × 3 receptive field and followed by non-linearity which is ReLU. The model uses a stride of 1 and spatial padding 1 for all of the 3 × 3 convolutions. After every 3 consecutive convolution layers, there is a max-pooling layers. There is a total of 16 convolution layers in VGG19 architecture and 13 convolution layers in VGG16 architecture and 5 max-pooling layers in both. Finally, for the classification, there are 3 fully connected layers to which follow a stack of the conv layers. The first two layers have 4096 channel depth and the final layer has 1000 channel depth which is equal to the number of classes found in the ImageNet dataset with a SoftMax activation function.

3.6. Implementation Tools

Python programming language and Anaconda platform with Jupyter notebook are used for this research to display, edit, process, analyze, and recognize medical images. Because nowadays python becomes a common image processing tool and it is the state of

art higher level programming language for image processing, machine learning, and deep learning. It is a powerful language for image preprocessing and analysis. Software tools that we have used to implement the CNN algorithm are python as a programing language with TensorFlow and Keras libraries on anaconda environment. These tools fulfill all the consideration criteria's and they are used in python which is familiar to us. In order to identify the best tool for implementing the deep learning algorithm for medical picture classification, a review of various software tools and their libraries is carried out. During our in query, we discovered that there exist tools that are both general and tailored to deep learning algorithms. Before choosing the tools, we considered a few characteristics that would aid in the selection of the suitable software tools and libraries. The choice of programming language to implement the algorithm is the most important criterion

The tools must also be utilized in machines with limited resources. Other requirements include selecting tools with sufficient learning materials, such as literature, free video tutorials, and books, and the tools must be used in computers with low resources (like CPU only). Python as a programming language with Tensor-Flow and Keras libraries on an Anaconda environment were utilized to create the CNN algorithm. These tools meet all of the criteria for consideration and are written in Python, which we are familiar with.

1. **Anaconda3:** is used for the implementation of the model and it is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications, that aims to simplify package management and deployment. It contains different IDE is which are used to write the coding part such as Jupyter Notebook and Spyder. We have used Jupyter notebook to implement the coding part. It is easy and runs in a web browser.
2. **Keras**: is a model-level library, providing high-level building blocks for developing deep learning models. It does not handle itself low-level operations such as tensor manipulation and differentiation. Instead, it relies on a specialized, well-optimized tensor library to do so, serving as the "backend engine" of Keras. Rather than picking one single tensor library and making the implementation of Keras tied to that library, Keras handles the problem in a modular way, and several different backend engines can be plugged seamlessly into Keras. Currently, the three existing backend implementations are the TensorFlowbackend, the Theanobackend, and the

CNTK backend. In the future, it is likely that Keras will be extended to work with even more deep learning execution engines [51].

3. **Tensor Flow**, CNTK, and Theano are some of the main platforms for deep learning today. Theano is developed by the MILA lab at University de Montréal, while Tensor Flow is developed by Google, and CNTK is developed by Microsoft. Any piece of code that you write with Keras can be run with any of these back-ends without having to change anything to the code: you can seamlessly switch between the two during development, which often proves useful, for instance if one of these back-ends proves to be faster for a specific task. By default, I would recommend using the TensorFlow backend for most of your deep learning needs [51].
4. **Adobe Photoshop CS5:** is a popular image editing program that works in a similar way to Adobe Illustrator, Adobe in Design, Adobe Photoshop, and other Adobe Creative Suite programs. This lesson will teach you how to use Adobe Photoshop. You'll learn how to get started, navigate the interface, and alter photographs using fundamental editing skills [52].
5. **Microsoft Visio:** This tool is use for designing the research process and architecture of proposed system.
6. **Open CV**: (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision. The reason for using OpenCV is that it gives easy functionality to do different processes without going into implementations.

3.7. Evaluation Methods

The confusion matrix is the table used to describe the performance of the classifier or classification model on a set of test data for which the true values are known [53]. It represents the way in which the classification model gets confused in making predictions. Thus, a confusion matrix is a summary of prediction results on a classification problem. The confusion matrix is a matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known. The matrix itself can be easily understood, but the related terminologies may be confusing, since it shows the errors in the model performance in the form of a matrix, also known as an error matrix. Some of the metrics used for evaluating the

proposed approach with the help of the confusion matrix, we can calculate the different parameters for the model, such as accuracy, precision, recall and F1-score [54].

Classification Accuracy: It is one of the important parameters to determine the accuracy of the classification problems. It defines how often the model predicts the correct output. It can be calculated as the ratio of the number of correct predictions made by the classifier to all number of predictions made by the classifiers. The formula is given below:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (3.1)$$

Precision: It can be defined as the number of correct outputs provided by the model or out of all positive classes that have predicted correctly by the model, how many of them were actually true. It can be calculated using the below formula:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (3.2)$$

Recall: It is defined as the out of total positive classes, how our model predicted correctly. The recall must be as high as possible.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (3.3)$$

F-measure: If two models have low precision and high recall or vice versa, it is difficult to compare these models. So, for this purpose, we can use F-score. This score helps us to evaluate the recall and precision at the same time. The F-score is maximum if the recall is equal to the precision. It can be calculated using the below formula:

$$\text{F} - \text{measure} = \frac{2*\text{Recall}*\text{Precision}}{\text{Recall+Precision}} \quad (3.4).$$

CHAPTER FOUR
DESIGN, EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1. Overview

This chapter covers design and architecture of the proposed system. The many components of the suggested system architecture, as well as essential methods and algorithms, are also discussed. In generally, the proposed method goes through a number of steps, including feature extraction, classification and distance measurement. To retrieve similar x-ray image from the dataset we have to extract actually valuable information from images using different CNN feature extractor. Detailed implementation procedures, experimentation, analysis the result and how features are extracted and retrieved image are performed in the proposed model discussion are presented below.

4.2. The Proposed Architecture

The architecture of the proposed work is shown in figure 7. The major components of CBMIR are extraction, classification, and distance measurement that has been used in order to retrieve the similar images from the database. For retrieval process, Euclidean distance method has been used. The following processing steps have been applied in order to perform the retrieval of similar images from the database. The architecture design starts from input query x-ray image. Then for given input images, features are extracted using CNN algorithms. Based on the extracted features, image classification process has been performed using VGG16 tool. Finally, searching and retrieval process has been performed using one of the similarity measures such as Euclidean distance method.

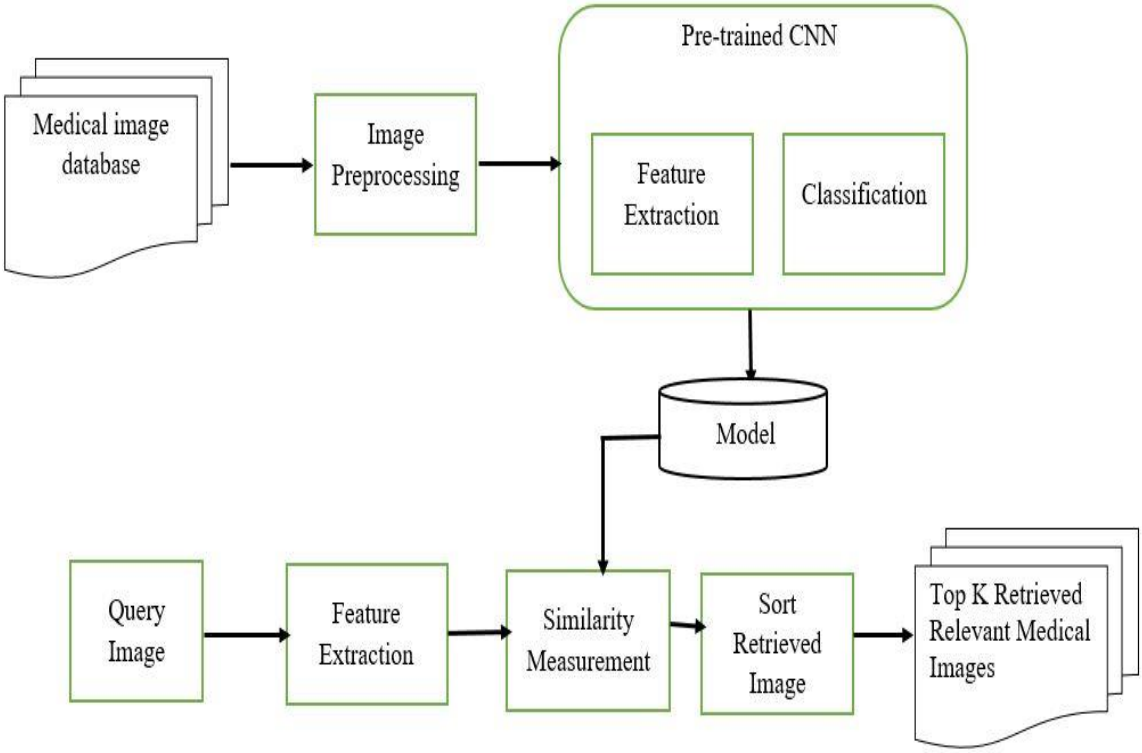


Figure 7. Proposed Architecture of Content Based Medical Image Retrieval

4.3. Feature Extraction using proposed Model

Feature extraction attempts to extract information from the input image to serve as an input into the conventional machine learning method. According to [55] feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input image will be transformed into a reduced representation set of features (also named features vector). If the features extracted are carefully chosen, it is expected that the features set will perform the desired task using the reduced representation instead of the full-size input. For an image, a feature can be defined as measures describing dataset properties and characteristics. These features play a fundamental role in classification. The features are necessary for differentiating one category from another. The method has to be used for describing the objects so that features of interest are highlighted.

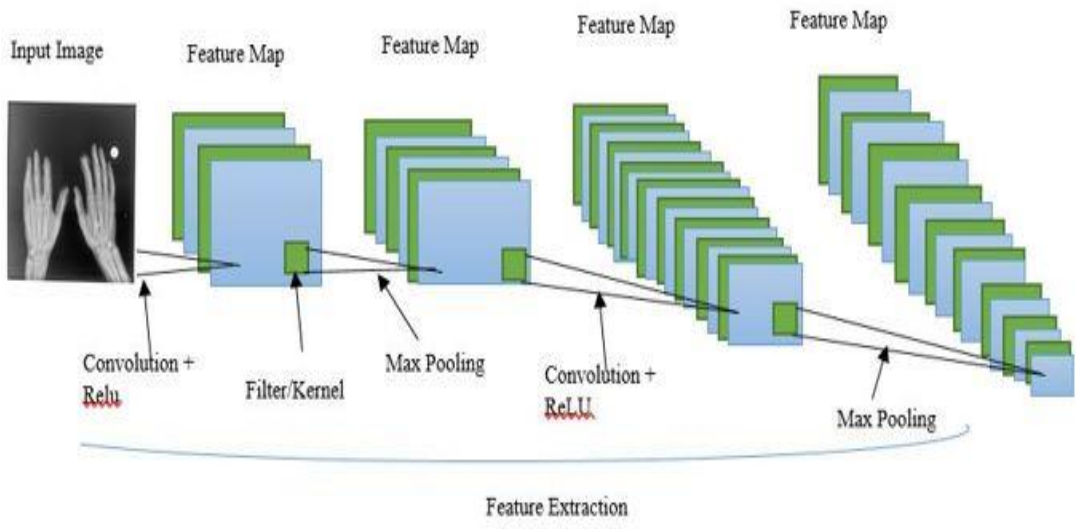


Figure 8. Feature Extraction of the proposed model

Features are extracted by convolution layers of CNN and feature extraction is the main purpose of this layer. These layers contain a series of filters or learnable kernels that aim to extract local features from the input image (see figure 8). To detect features, the filters in the convolution layer slide from left to right across the entry of the input image. During feature extraction, the convolution layer accepts pixel values from the input image and multiplies and sums them with the values of the filter (set of weights) to produce the feature map. The extracted feature of the input image is the feature map; it contains patterns that are used to distinguish the given images.

In this case, the features contain various color patterns from the given image. Then, to add nonlinearity to the network, each value of the feature map is passed through activation functions. Following nonlinearity, the feature map was fed back into the pooling layer to reduce the feature map's resolution and the network's computational complexity. The process of extracting useful features from an input image consists of several steps that are similar in nature, such as cascading convolution layers, adding nonlinearity, and pooling layers.

4.4. Hyper Parameter Settings

Hyper parameters are configurations that are defined before the training process begins and are external to the deep learning algorithm. There is no universal rule for selecting the optimum hyper parameters for a particular situation [56]. As a result, many experiments are carried out in order to select the hyper parameters. Hyper parameters for the model are described in the next sections.

- **Optimization algorithms:** the proposed model is trained using gradient scent optimization algorithm to minimize the error rate and back propagation of error algorithm is used to update the weights. Gradient descent is by far the most popular and the most widely used optimization algorithm in deep learning researches [57]at the same time, every state-of-the-art deep learning library contains implementations of gradient descent optimization algorithms such as Keras(used in this thesis). It updates the weight of the model and tunes parameters, therefore, minimize the loss function. To optimize the gradient descent, the Adaptive Moment Estimation (Adam) optimizer is used [58].Adam computes adaptive learning rate to each parameter and it uses squared gradients to scale learning rate, it also uses the moving average of the gradient.

- **Learning rate**: Because back propagation was used to train the proposed model, learning rate is used during weight update. It controls the amount of weight to updated during back propagation [56].The challenging part was to choose the proper learning rate during our experiment. In our experiment, we have seen that a learning rate with a value of too small takes longer to train than a value of larger. But when we give a smaller value the model is more optimal than a model with a learning rate of larger. The experiment was done by using learning rate le-3, 0.001 and 0.1. Then, learning rate le-3 is the optimal one for all of the experiments.

- **Loss function:** The choice of the loss function is directly related to activation functions that are used in the output layer (last fully connected layer) of the model and the type of problem we are trying to solve (whether regression or classification). In the proposed model, softmax is used as activation function in the last fully connected layer. The type of problem we are solving is a classification problem specifically categorical classification. We have used Categorical Cross-Entropy

(CCE) loss as a loss function for our model. Even though there is another loss function such as Binary Cross-Entropy (BCE), Mean Squared Error(MSE), but categorical cross-entropy is the recommended choice of loss function for multi class classification [58, 59]. It performs well for models that output of multiclass i.e.it measures the distance between the actual output and the desired output. The experiment was done by using CCE loss.

- **Activation function**: we have different activation functions. In our experiments for multiclass classification are conducted by using SoftMax activation functions. In the output layer of the model, the Softmax activation function is used because it is the best choice for a multiclass classification problem [59, 58].
- **Number of epochs**: is the number of iterations the entire dataset passes forward and backward through the model or the network. In our experiment, the model was trained by using different epochs starting from10to100. During the training, we have seen that when we use too small or too large epoch, the model gets a high gap between the training error and validation error. After many experiments, the model gets optimal with epoch thirty.
- **Batch size**: is the number of input data we pass into the network at once. It is too hard to give all the data to the computer in a single epoch so we need to divide the input into several smaller batches. It is preferred in model training to minimize the computational time of the machine. Batch size of 32 during model training is used in our experiment.

Table 3.Summary of hyper parameters used during model training

Parameter	**Epoch**	**Batch Size**	**Activation Function**	**Loss Function**	**Optimization Algorithm**	**Learning Rate**
Value	30	32	Softmax	CCE	Adam	Le-3

4.5. Pre-Trained Cnn

We used three pre-trained CNN models: VGG 16, VGG 19 and ResNet50 which are widely used pre-trained architectures in ImageNet since they are well fine-tuned. The VGG model is chosen because of its simplicity and the ResNet50 model is used because of its complicated features. Therefore, experiments are conducted in both a relatively simpler model and a complex one to get the classification accuracy of these models in our dataset. All the experiments are conducted in the same dataset.

4.5.1. Training Components Of The Proposed Vgg16 Model

The VGG model is characterized by its simplicity by using only 3 × 3 convolution layers which are staked on each other in increasing depth of the layer. In our experiment, down sampled RGB image of size 64× 64 is given as an input to the model and fine-tuned the model to give 13 classes of output in our dataset. The original VGG16 model has a total of 138million parameters which is very huge. We have trained the model by 17,883,981 parameters because the spatial dimension of the image in our model is smaller and we only trained some parts of the model. As we have discussed in previous sections, we have fine-tuned theVGG16 model by using only the conv base of the network. We have conducted several experiments in order to find the optimal pre- trained model by training different conv blocks of the model. The model is trained by using all the convbase of the network and changing only the fully connected layers, and the result shows high overfitting. Overfitting happens because the model weights are trained with millions of images which are different from our dataset and thousands of classes and we tried to train that model by using only 4194 original images. Hence, we need to update some of the weights of the network and increase the number of images by using data augmentation technique. Therefore, we have decided to freeze some of the layers (conv blocks) of the model. After several experiments, we noticed that freezing the first 3conv block is the optimal one compared to freezing the first2, and 4convblocks by using 4194 images which are generated by augmentation techniques.

Table 4. VGG16 model1: **"model"**

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 64, 64, 3)]	0
block1_conv1 (Conv2D)	(None, 64, 64, 64)	1792
block1_conv2 (Conv2D)	(None, 64, 64, 64)	36928
block1_pool (MaxPooling2D)	(None, 32, 32, 64)	0
block2_conv1 (Conv2D)	(None, 32, 32, 128)	73856
block2_conv2 (Conv2D)	(None, 32, 32, 128)	147584
block2_pool (MaxPooling2D)	(None, 16, 16, 128)	0
block3_conv1 (Conv2D)	(None, 16, 16, 256)	295168
block3_conv2 (Conv2D)	(None, 16, 16, 256)	590080
block3_conv3 (Conv2D)	(None, 16, 16, 256)	590080
block3_pool (MaxPooling2D)	(None, 8, 8, 256)	0
block4_conv1 (Conv2D)	(None, 8, 8, 512)	1180160
block4_conv2 (Conv2D)	(None, 8, 8, 512)	2359808
block4_conv3 (Conv2D)	(None, 8, 8, 512)	2359808
block4_pool (MaxPooling2D)	(None, 4, 4, 512)	0
block5_conv1 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv2 (Conv2D)	(None, 4, 4, 512)	2359808
block5_conv3 (Conv2D)	(None, 4, 4, 512)	2359808
block5_pool (MaxPooling2D)	(None, 2, 2, 512)	0
flatten (Flatten)	(None, 2048)	0
dense (Dense)	(None, 1024)	2098176
batch_normalization_v2 (Batc	(None, 1024)	4096
dropout (Dropout)	(None, 1024)	0

```
_________________________________________________________________
dense_1 (Dense)              (None, 1024)              1049600
_________________________________________________________________
batch_normalization_v2_1 (Ba (None, 1024)              4096
_________________________________________________________________
dropout_1 (Dropout)          (None, 1024)              0
_________________________________________________________________
dense_2 (Dense)              (None, 13)                13325
=================================================================
Total parameters: 17,883,981
Trainable parameters: 3,165,197
Non-trainable parameters: 14,718,784
_________________________________________________________________
```

The VGG16 consists of 13 sets of convolution and 3 dense layers. After the sets of steps are completed, the resulting values are given to the hidden layer and outputs provided. As shown in Table 3 the initial spatial size of the input volume is 64×64×3 and it gets changed after some convolution operations, the initial size of the filter F and Stride S are 3×3×3 and 2 respectively and these sizes are changed after some convolution and pooling operations, and there is same padding P in our network and the value of P is always same throughout the model. We have considered this setting of parameters to be valid during the process of resizing images contained in our dataset. If this arrangement is not considered, libraries that are used to implement the CNN model will throw exception or it will zero pad the rest of the area or it will crop the image to make it fit.

Input layer: the input layer of our CNN model accepts RGB images of size 64 × 64 × 3 with 13 different classes like Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, Lspin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist. This layer only passes the input to the first convolution layer without any computation. Therefore, there are no learnable features and the number of parameters in this layer is 0.

Convolutional layer: in the VGG16 model there are 4 convolutional layers. The first convolutional layer of the model filters the 64× 64×3 input image by using 64 kernels with a size of 3× 3×3 pixels. This layer has a depth of K=64, the output volume of this layer is 64×64× 64. The product of the output volume gives a total number of neurons in the layer (first conv layer) which is 1792. All the convolutional layers of the VGG16

model use ReLU non linearity as activation function ReLU is chosen because is faster than other non-linearity such as tanh to train deep CNNs with gradient descent.

Pooling layers: There are five max-pooling layers in the proposed model, such as second, fourth, seventh, tenth and thirteenth convolutional layers. The first max-pooling layer reduces the output of the second convolutional layer with a filter of size 5× 5. The second max- pooling layer takes as an input the output of the fourth convolutional layer and pools by using 5 X 5 filters size. The third max-pooling layer takes as an input the output of the seventh convolutional layer.

Fully Connected (FC) layer: in this VGG 16 model there are three fully connected layers including the output layer. The first and second fully connected layers have 1024 neurons; the second and final fully connected layer which is the output layer of the model has four neuron. The first FC layer accepts the output of the thirteen convolutional layers after converting the 3D volume of data in to a vector value (Flattening). This layer computes the class core and the number of neurons in the layer predefined during the development of the model. It is the same as ordinary NN and as the name implies, each neuron in this layer is connected to all the numbers in the previous layer.

Output layer: the output layer is the last (the third FC layer) of the model and it has 4 neurons with a softmax activation function. Because the model is designed to classify 13classes called Categorical classification. As we can see in the above table 3 the VGG 16 model have 17,883,981 parameters which are extremely small when we compare to the other deep learning architectures such as AlexNet which have 60 million parameters, VGG 138 million parameters, and GoogLeNet 4 million parameters. It is considered that deep learning models have a massive number of parameters; therefore, they need a huge computational power to train those models from scratch and they need a very large amount of data. But the proposed model is trained with a minimum amount of resources and data and it performs very well.

4.5.2. Training Components Of The Proposed Vgg19 Model

VGG -19, a type of CNN model is used in this thesis, which is 19 weight layers consisting of 16 convolutional layers with 3 fully connected layers and the same 5 pooling layers. The input is a 224 * 224 RGB image to VGG based convNet. The pre-

processing layer takes the RGB image with pixel values in the range 0–255 and subtracts the mean image values computed over the entire ImageNet training collection. After pre-processing, the input images are passed through layers of weight. The training images are processed through a stack of convolutional layers.

In this study RGB image of size 64× 64 is given as an input to the model and fine-tuned the model to give 13 classes of output in our dataset. The original VGG19 model has a total of 138 million parameters which is very huge. We have trained the model by 21,346,893 parameters because the spatial dimension of the image in our model is smaller and we only trained some parts of the model. As we have discussed in previous sections, we have fine-tuned the VGG19 model by using only the conv base of the network. We have conducted several experiments in order to find the optimal pre-trained model by training different conv blocks of the model. The model is trained by using all the conv base of the network and changing only the fully connected layers, and the result shows high overfitting. Overfitting happens because the model weights are trained with millions of images which are different from our dataset and thousands of classes and we tried to train that model by using only 4194 original images.

Table 5. VGG19 model：**"model_1"**

Layer (type)	Output Shape	Param #
input_2 (InputLayer)	[(None, 64, 64, 3)]	0
block1_conv1 (Conv2D)	(None, 64, 64, 64)	1792
block1_conv2 (Conv2D)	(None, 64, 64, 64)	36928
block1_pool (MaxPooling2D)	(None, 32, 32, 64)	0
block2_conv1 (Conv2D)	(None, 32, 32, 128)	73856
block2_conv2 (Conv2D)	(None, 32, 32, 128)	147584
block2_pool (MaxPooling2D)	(None, 16, 16, 128)	0
block3_conv1 (Conv2D)	(None, 16, 16, 256)	295168
block3_conv2 (Conv2D)	(None, 16, 16, 256)	590080
block3_conv3 (Conv2D)	(None, 16, 16, 256)	590080
block3_conv4 (Conv2D)	(None, 16, 16, 256)	590080
block3_pool (MaxPooling2D)	(None, 8, 8, 256)	0

```
_________________________________________________________________
block4_conv1 (Conv2D)        (None, 8, 8, 512)         1180160
_________________________________________________________________
block4_conv2 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block4_conv3 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block4_conv4 (Conv2D)        (None, 8, 8, 512)         2359808
_________________________________________________________________
block4_pool (MaxPooling2D)   (None, 4, 4, 512)         0
_________________________________________________________________
block5_conv1 (Conv2D)        (None, 4, 4, 512)         2359808
_________________________________________________________________
block5_conv2 (Conv2D)        (None, 4, 4, 512)         2359808
_________________________________________________________________
block5_conv3 (Conv2D)        (None, 4, 4, 512)         2359808
_________________________________________________________________
block5_conv4 (Conv2D)        (None, 4, 4, 512)         2359808
_________________________________________________________________
block5_pool (MaxPooling2D)   (None, 2, 2, 512)         0
_________________________________________________________________
flatten_1 (Flatten)          (None, 2048)              0
_________________________________________________________________
dense_3 (Dense)              (None, 512)               1049088
_________________________________________________________________
batch_normalization_v2_2 (Ba (None, 512)               2048
_________________________________________________________________
dropout_2 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_4 (Dense)              (None, 512)               262656
_________________________________________________________________
batch_normalization_v2_3 (Ba (None, 512)               2048
_________________________________________________________________
dropout_3 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_5 (Dense)              (None, 13)                6669
=================================================================
Total params: 21,346,893
Trainable params: 1,320,461
Non-trainable params: 20,026,432
_________________________________________________________________
```

Input layer: the input layer of our VGG19 model accepts RGB images of size 64 × 64 × 3 with thirteen different classes like Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist. This layer only passes the input to the first convolution layer without any computation. Therefore, there are no learnable features and the number of parameters in this layer is 0.

Convolutional layer: in the VGG19 model the reare13 convolutional layers. The first convolutional layer of the model filters the 64× 64×3 input image by using 64 kernels withasizeof3× 3×3pixels. This layer has a depth of K=64, the output volume of this layer

is 64×64×64. The product of the output volume gives a total number of neurons in the layer (first conv layer) which is 1792.

Pooling layer: The proposed model has five max-pooling layers such as the second,the fourth, seventh, tenth and thirteen convolutional layers. The first max-pooling layer reduces the output of the second convolutional layer with a filter of size 5 × 5. The second max- pooling layer takes as an input the output of the fourth convolutional layer and pools by using 5 x 5 filters size.

Fully Connected (FC) layer: in this VGG19 model there are two fully connected layers including the output layer. The first fully connected layers have 512 neurons; the second and final fully connected layer which is the output layer of the model has four neurons. The first FC layer accepts the output of the four convolutional layers after converting the 3D volume of data into a vector value (Flattening).This layer computes the class core and the number of neurons in the layer predefined during the development of the model. It is the same as ordinary NN and as the name implies, each neuron in this layer is connected to all the numbers in the previous layer.

Output layer: this layer is the last (the second FC layer) of the model and it has4 neuron with a softmax activation function. Because the model is designed to classify 13 classes called Categorical classification.

4.5.3. Training Components of the Proposed Resnet50 Model

While the Resnet50 architecture is based on the bottom model, there is one major difference from VGG. In this case, the building block was modified into a bottleneck design due to concerns over the time taken to train the layers. This used a stack of 3 layers instead of the earlier 2. Therefore, each of the 2-layer blocks in Resnet34 was replaced with a 3-layer bottleneck block, forming the Resnet 50 architecture. This has much higher accuracy than the 34-layer ResNet model. The 50-layer ResNet achieves a performance of 3.8 billion FLOPS.

ResNet-50 is a convolutional neural network that is 50 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pre-trained network can classify images into 1000 object categories. As a

result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

Input layer: this layer of our ResNet50 model accepts RGB images of size 64 × 64 × 3 with 13 different classes, such as Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist. This layer only passes the input to the first convolution layer without any computation. Therefore, there are no learnable features and the number of parameters in this layer is 0.

Convolutional layer: in the ResNet50 model there are 4convolutional layers. The first convolutional layer of the model filters the 64× 64×3 input image by using 64 kernels with a size of 3× 3×3pixels. This layer has a depth of K=64, the output volume of this layer is 64×64×64. The product of the output volume gives a total number of neurons in the layer (first conv layer) which is 1792. All the convolutional layers of the ResNet50 model use ReLU non linearity as activation function ReLU is chosen because it is faster than other non-linearity such as tanh to train deep CNNs with gradient descent.

Pooling layer: There are five max-pooling layers, such as second, fourth, seventh, tenth and thirteenth convolutional layers that are used in the proposed model. The first max-pooling layer reduces the output of the second convolutional layer with a filter of size 5× 5. The second max- pooling layer takes as an input the output of the fourth convolutional layer and pools by using 5X 5 filters size. The third max-pooling layer takes as an input the output of the seventh convolutional layer.

Fully Connected (FC) layer: in this ResNet50 model there are three fully connected layers including the output layer. The first and second fully connected layers have 1024 neurons, the second and final fully connected layer which is the output layer of the model has four neuron. The first FC layer accepts the output of the thirteen convolutional layers after converting the 3D volume of data in to a vector value (Flattening). This layer computes the class core and the number of neurons in the layer predefined during the development of the model. It is the same as ordinary NN and as the name implies, each neuron in this layer is connected to all the numbers in the previous layer.

Output layer: the output layer is the last (the third FC layer) of the model and it has 4 neurons with a softmax activation function. Because the model is designed to classify 13 classes called Categorical classification.

4.6. Experimental Result

In this section, an attempt is made to construct model for content based image retrieval from medical image database and classification by using the deep learning algorithm like VGG16, VGG19 and ResNet50 and comparing all of them. In this section detailed implementation procedures, experimentation, analysis and the result are presented below.

4.6.1. Result Analysis Of Vgg16

The following two plots shows the classification accuracy and loss with respect to epochs by using classification accuracy metrics such as training and validation accuracy, training loss and validation loss of VGG16 pre-trained model that we have conducted experiment by making some changes to the original pre-trained model in order to able the model to classify well in our dataset.

When we see the training accuracy in the first epoch it is around 93% and slightly increases and passes 96% at epoch 5. From epoch 5 to 10, the training accuracy of the model gets higher with an accuracy of greater than 96% and after the 16th epoch, the accuracy gets higher than 97%. As we can see in the figure 9, the accuracy gets higher in the first few epochs, this is because of the dataset. The patterns of the images in our dataset are very visible to even human eyes and it is easy to differentiate by the VGG16 model. In general, as we can see in the following plots the validation accuracy line is almost in sync with the training accuracy line (see figure 9) and at the same time, the validation loss line is also in sync with the training loss (see figure 10). Even though the validation accuracy and validation loss lines are not linear, the model is not overfitting. In other words, the validation loss is decreasing, not increasing and also the validation accuracy is increasing, not decreasing.

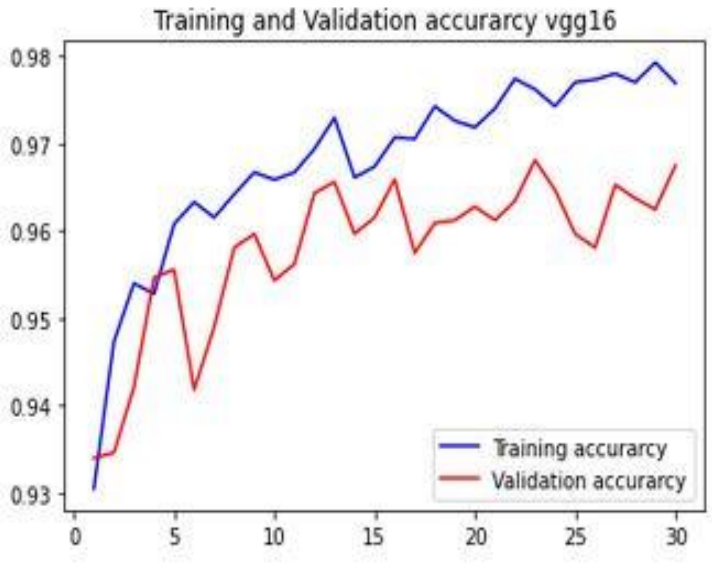


Figure 9. Training and validation accuracy for VGG16 Pre-trained model

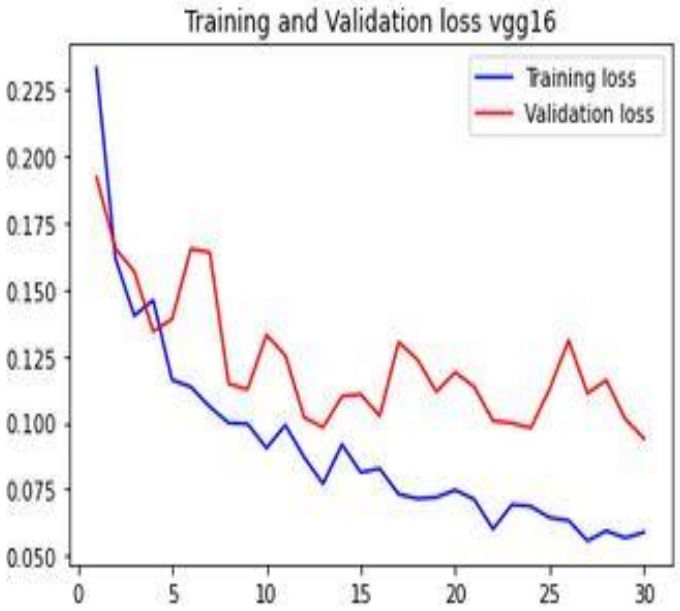
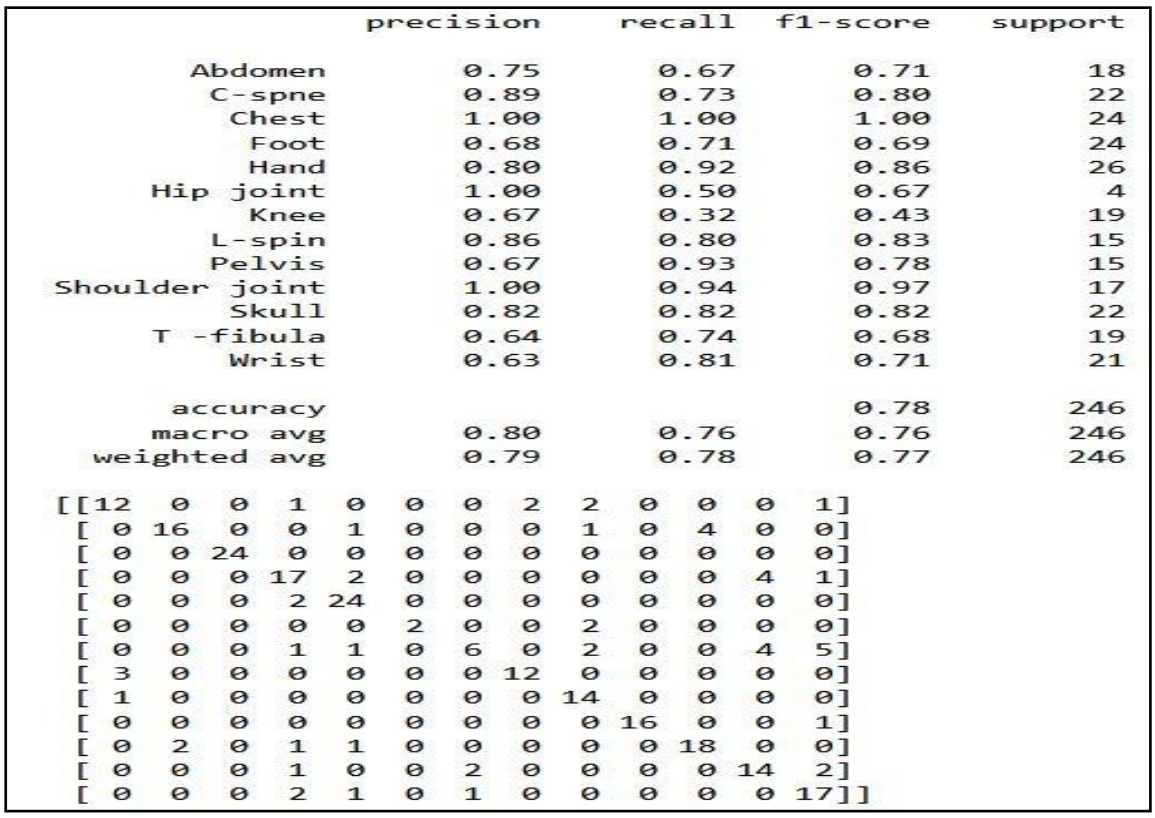


Figure 10. Training and validation loss for the VGG16 pre-trained model

The evaluation result of VGG16 in terms of accuracy, recall and precision based on the confusion matrix for the 13 classes, such as Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist are shown below in figure 11.



	precision	recall	f1-score	support
Abdomen	0.75	0.67	0.71	18
C-spne	0.89	0.73	0.80	22
Chest	1.00	1.00	1.00	24
Foot	0.68	0.71	0.69	24
Hand	0.80	0.92	0.86	26
Hip joint	1.00	0.50	0.67	4
Knee	0.67	0.32	0.43	19
L-spin	0.86	0.80	0.83	15
Pelvis	0.67	0.93	0.78	15
Shoulder joint	1.00	0.94	0.97	17
Skull	0.82	0.82	0.82	22
T -fibula	0.64	0.74	0.68	19
Wrist	0.63	0.81	0.71	21
accuracy			0.78	246
macro avg	0.80	0.76	0.76	246
weighted avg	0.79	0.78	0.77	246

```
[[12  0  0  1  0  0  0  2  2  0  0  0  1]
 [ 0 16  0  0  1  0  0  0  1  0  4  0  0]
 [ 0  0 24  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0 17  2  0  0  0  0  0  0  4  1]
 [ 0  0  0  2 24  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  2  0  0  2  0  0  0  0]
 [ 0  0  0  1  1  0  6  0  2  0  0  4  5]
 [ 3  0  0  0  0  0  0 12  0  0  0  0  0]
 [ 1  0  0  0  0  0  0  0 14  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 16  0  0  1]
 [ 0  2  0  1  1  0  0  0  0  0 18  0  0]
 [ 0  0  0  1  0  0  2  0  0  0  0 14  2]
 [ 0  0  0  2  1  0  1  0  0  0  0  0 17]]
```

Figure 11. Performance evaluation metrics results analysis of VGG16

The result of confusion matrix of VGG16 model classification accuracy shows that correctly classified instances of Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist are 12,16,24,17,24,2,6,12,14,16,18,14 and 17 respectively. Abdomen was misclassified more to L-spin and Pelvis (2, 2 respectively), C-spine was misclassified more to Shoulder Joint (4), Foot was misclassified more to T-fibula (4) and Knee was misclassified more to T-fibula and Wrist (4,5 respectively).

4.6.2. Result Analysis of Vgg19

The following two plots show the classification accuracy and loss of VGG19 with respect to epochs by using classification accuracy metrics such as training and validation accuracy. Training loss and validation loss of VGG19 pre-trained model that we have conducted experiment by making some changes to the original pre-trained model in order to able the model to classify well in our dataset.

When we see the training accuracy in the first epoch it is around 92% and slightly increases and passes 95% at epoch 5. In between epoch 5 to 10, the training accuracy of the model gets higher with an accuracy of greater than 96% and after the 19th epoch, the accuracy gets higher than 97%. As we can see in the graph, the accuracy gets higher in the first few epochs, this is because of the patterns of the images in our dataset where they are very visible to even human eyes and it is easy to differentiate by the VGG19 model. In general, as we can see in the following plots the validation accuracy line is almost in sync with the training accuracy line (see figure 12) and at the same time, the validation loss line is also in sync with the training loss (see figure 13). Even though the validation accuracy and validation loss lines are not linear, but it shows that the model is not overfitting. In other words, the validation loss is decreasing, not increasing and also the validation accuracy is increasing but not decreasing.

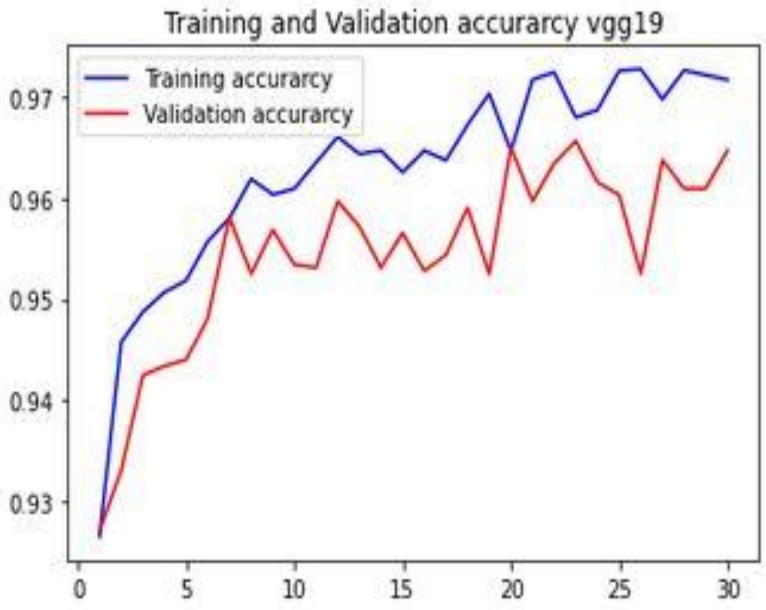


Figure 12. Training and validation accuracy for VGG19 Pre-trained model

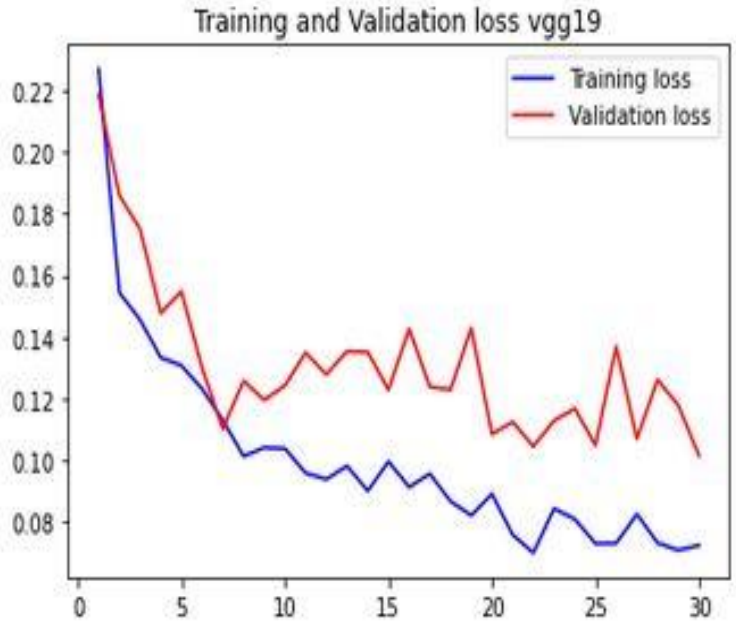


Figure 13. Training and validation loss for the VGG19 pre-trained model

The evaluation result of VGG19 in terms of accuracy, recall and precision based on the confusion matrix for the 13 classes, such as Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist are shown below in figure 14.

	precision	recall	f1-score	support
Abdomen	0.56	0.83	0.67	18
C-spne	0.79	0.68	0.73	22
Chest	1.00	0.83	0.91	24
Foot	0.67	0.67	0.67	24
Hand	0.80	0.77	0.78	26
Hip joint	0.67	0.50	0.57	4
Knee	0.58	0.74	0.65	19
L-spin	0.71	0.80	0.75	15
Pelvis	0.92	0.80	0.86	15
Shoulder joint	0.85	1.00	0.92	17
Skull	0.75	0.68	0.71	22
T -fibula	0.76	0.84	0.80	19
Wrist	0.77	0.48	0.59	21
accuracy			0.75	246
macro avg	0.76	0.74	0.74	246
weighted avg	0.76	0.75	0.75	246

```
[[15  0  0  0  0  0  0  3  0  0  0  0  0]
 [ 0 15  0  0  1  0  0  1  0  0  5  0  0]
 [ 3  0 20  0  0  0  0  0  0  1  0  0  0]
 [ 1  0  0 16  1  0  3  0  1  0  0  2  0]
 [ 0  0  0  3 20  0  1  0  0  0  0  0  2]
 [ 1  0  0  0  0  2  0  0  0  1  0  0  0]
 [ 0  0  0  1  0  0 14  0  0  1  0  2  1]
 [ 3  0  0  0  0  0  0 12  0  0  0  0  0]
 [ 2  0  0  0  0  0  1  0 12  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0 17  0  0  0]
 [ 0  4  0  0  1  1  0  1  0  0 15  0  0]
 [ 0  0  0  1  0  0  2  0  0  0  0 16  0]
 [ 2  0  0  3  2  0  3  0  0  0  0  1 10]]
```

Figure 14. Performance evaluation metrics results analysis of VGG19

VGG19 model classification based on confusion matrix results in correctly classified instances of 15,15,20,16,20,2,14,12,12,17,15,16 and 10 for classes Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist respectively. Abdomen was misclassified more to L-spin (3), C-spine was misclassified more to Skull(5), Skull was misclassified more to C-spin(4), Wrist was misclassified more to Abdomen, Foot, Hand and Knee (2,3,2 and 3 respectively).

4.6.3. Result Analysis of Resnet50

The following two plots shows the classification accuracy and loss of RESNET50 with respect to epochs by using classification accuracy metrics such as training and validation accuracy, training loss and validation loss of ResNet50 pre-trained model that we have conducted experiment by making some changes to the original pre-trained model in order to able the model to classify well in our dataset.

When we see the training accuracy in the first epoch it is more than 92% and slightly increases and passes 95% at epoch 5. In between epoch 5 to 10, the training accuracy of the model gets higher with an accuracy of greater than 96% and after the 14th epoch, the accuracy gets higher than 97%. As we can see in the figure 15, the accuracy gets higher in the first few epochs, this is because of the dataset. The patterns of the images in our dataset are very visible to even human eyes and it is easy to differentiate by the ResNet50 model. In general, as we can see in the following plots the validation accuracy line is almost in sync with the training accuracy line (see figure 15) and at the same time, the validation loss line is also in sync with the training loss (see figure 16). As depicted in the figure, the validation accuracy and validation loss lines are linear, which shows that the model is overfitting. In other words, the validation loss is increasing, not decreasing and also the validation accuracy is decreasing, not increasing.

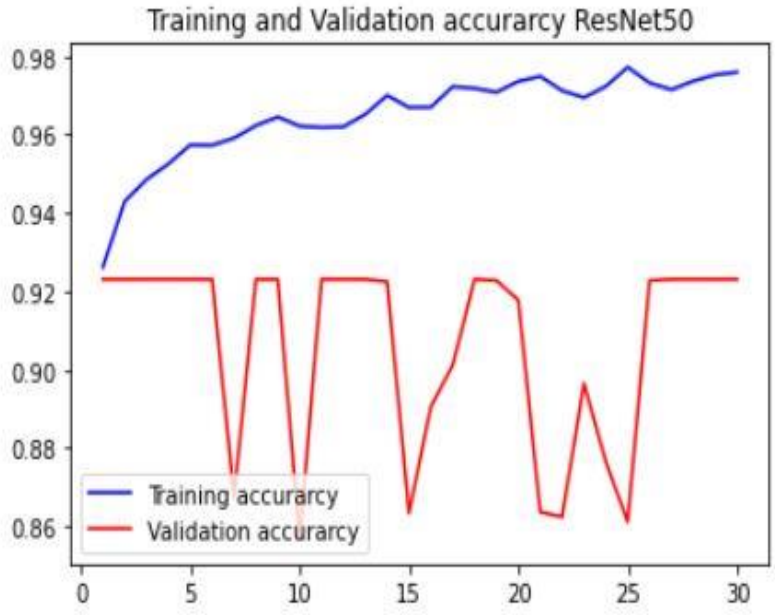


Figure 15. Training and validation accuracy for ResNet50 Pre-trained model

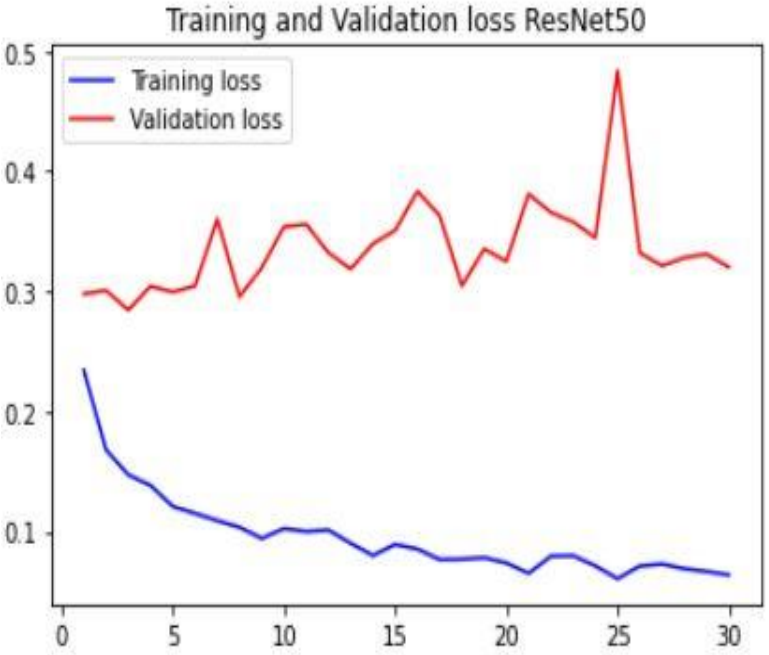
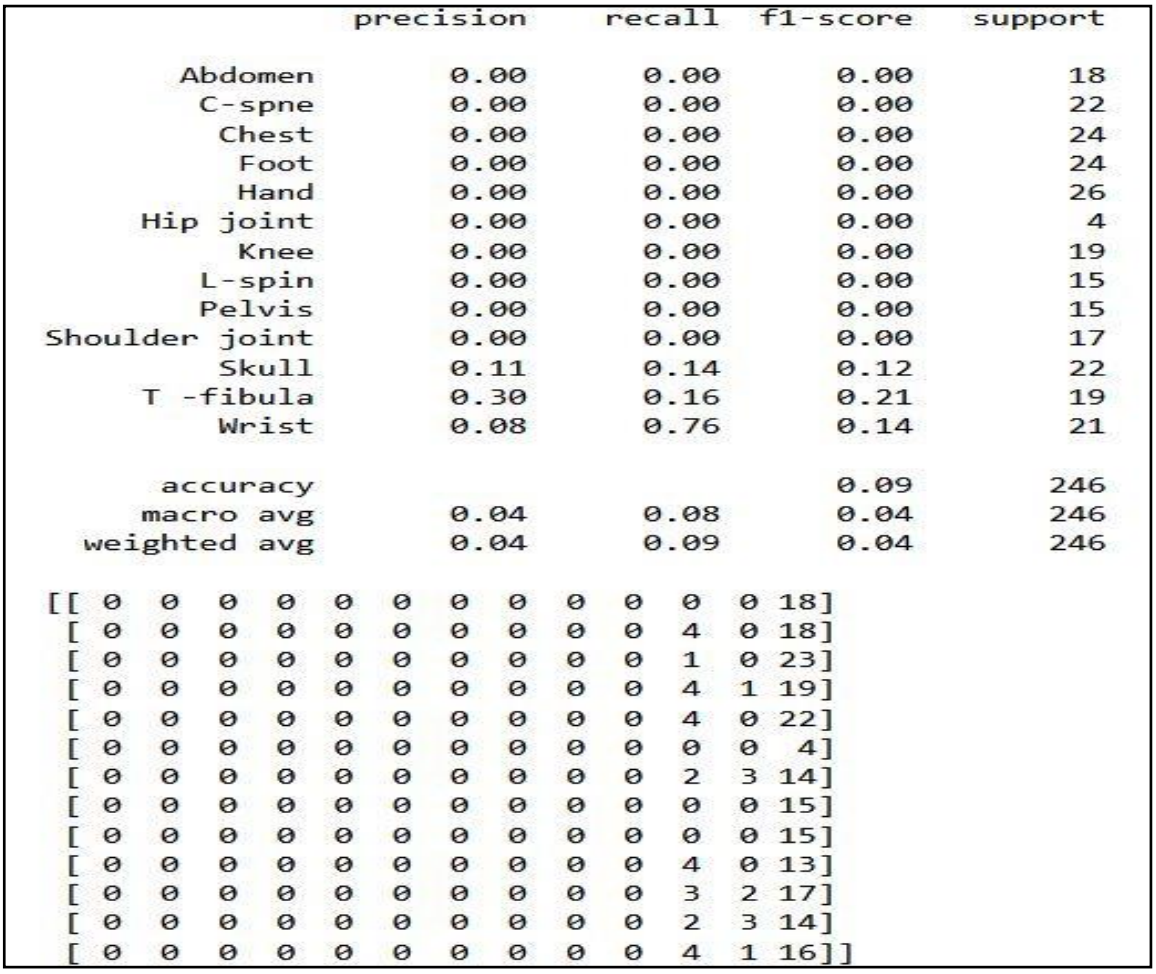


Figure 16. Training and validation loss for the ResNet50 pre-trained model

The below figure 17 presents evaluation result of ResNet50 in terms of accuracy, recall and precision based on the confusion matrix for the 13 classes, such as Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist.



	precision	recall	f1-score	support
Abdomen	0.00	0.00	0.00	18
C-spne	0.00	0.00	0.00	22
Chest	0.00	0.00	0.00	24
Foot	0.00	0.00	0.00	24
Hand	0.00	0.00	0.00	26
Hip joint	0.00	0.00	0.00	4
Knee	0.00	0.00	0.00	19
L-spin	0.00	0.00	0.00	15
Pelvis	0.00	0.00	0.00	15
Shoulder joint	0.00	0.00	0.00	17
Skull	0.11	0.14	0.12	22
T -fibula	0.30	0.16	0.21	19
Wrist	0.08	0.76	0.14	21
accuracy			0.09	246
macro avg	0.04	0.08	0.04	246
weighted avg	0.04	0.09	0.04	246

```
[[ 0  0  0  0  0  0  0  0  0  0  0  0 18]
 [ 0  0  0  0  0  0  0  0  0  0  4  0 18]
 [ 0  0  0  0  0  0  0  0  0  0  1  0 23]
 [ 0  0  0  0  0  0  0  0  0  0  4  1 19]
 [ 0  0  0  0  0  0  0  0  0  0  4  0 22]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  4]
 [ 0  0  0  0  0  0  0  0  0  0  2  3 14]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 15]
 [ 0  0  0  0  0  0  0  0  0  0  0  0 15]
 [ 0  0  0  0  0  0  0  0  0  0  4  0 13]
 [ 0  0  0  0  0  0  0  0  0  0  3  2 17]
 [ 0  0  0  0  0  0  0  0  0  0  2  3 14]
 [ 0  0  0  0  0  0  0  0  0  0  4  1 16]]
```

Figure 17. Performance evaluation metrics results analysis of ResNet50

The result of ResNet50 model classification based on confusion matrix shows that misclassified instances for classes Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder and Joint are 18, 18, 23, 19, 22, 4, 14, 15, 15, 13, 17 and14 respectively to wrist. Skull, T – fibula and Wrist was correctly classified are 3, 3 and 16 respectively.

4.6.4. Selecting the Best Performing Model

As we can see from the above result analysis of each pre-trained model of VGG 16, VGG 19 and ResNet50, the accuracy of VGG16, VGG 19 and ResNet50 model is 96.74%, 96.46%, and 92.30% respectively. These show the models are giving good result on training dataset. The test result shows that, VGG16 model can successfully retrieve and classify the given image as Abdomen, Chest, C-spine, Foot, Hand, Hip Joint, Knee, L-spin, Pelvis, Shoulder Joint, Skull, T-fibula and Wrist and retrieve with better accuracy than the VGG19 and ResNet50 pre-trained models. When we compute the difference between training accuracy and validation accuracy for each of the three experiments, there is very less and almost both training accuracy and validation accuracy is the same in the VGG16 model. These show that there is no overfitting in the models and we can say that the generalization ability of the VGG16 model is high. Hence, VGG16 model is used for medical image retrieval.

4.7. Developing Content Based Medical Image Retrieval (Cbmir)

Hence, once we are selection the best model using the training and testing of vgg 16. we have saved the model during training time in experiment and also we select the best model from the given model and running in python code what we save first using .out for image retrieval.

In this thesis work the learned features using CNN pre-trained model during training time for the extraction of features from each image by conv2d layers for the purpose of getting more information for the result and saved by .h5 for classification and .out for image retrieval and also used Euclidean distance and k: distances[k] for index closest. First to predict class labels and later used to model the feature space for similarity computation for the retrieval task based on the saved features using .out give an image from a dataset and it will output top K similar images from that dataset.

```
model = tf.keras.applications.VGG16(weights='imagenet', include_top=True)
feat_extractor = Model(inputs=model.input, outputs=model.get_layer("fc2").output)
pca_features = np.loadtxt("vgg16 X-ray images  new.out")
```

The model for training saved by .h5 and .out code is described below

vgg16_model.save('C:/Users/Muva/Desktop/Munir research/X-ray images in folders last reviewed/vgg16 X-ray images new.out')

vgg16_model.save('C:/Users/Muva/Desktop/Munir research/X-ray images in folders last reviewed/vgg16 X-ray images new.h5')

4.8. Evaluating Content Based Medical Image Retrieval

Content based medical image retrieval (CBMIR) is the process of retrieving images from a database or library of digital images according to the visual content of the images. Given users query via the graphical user interface (GUI), content based image retrieval is undertaken using the selected VGG16 model for relevant medical images retrieval.

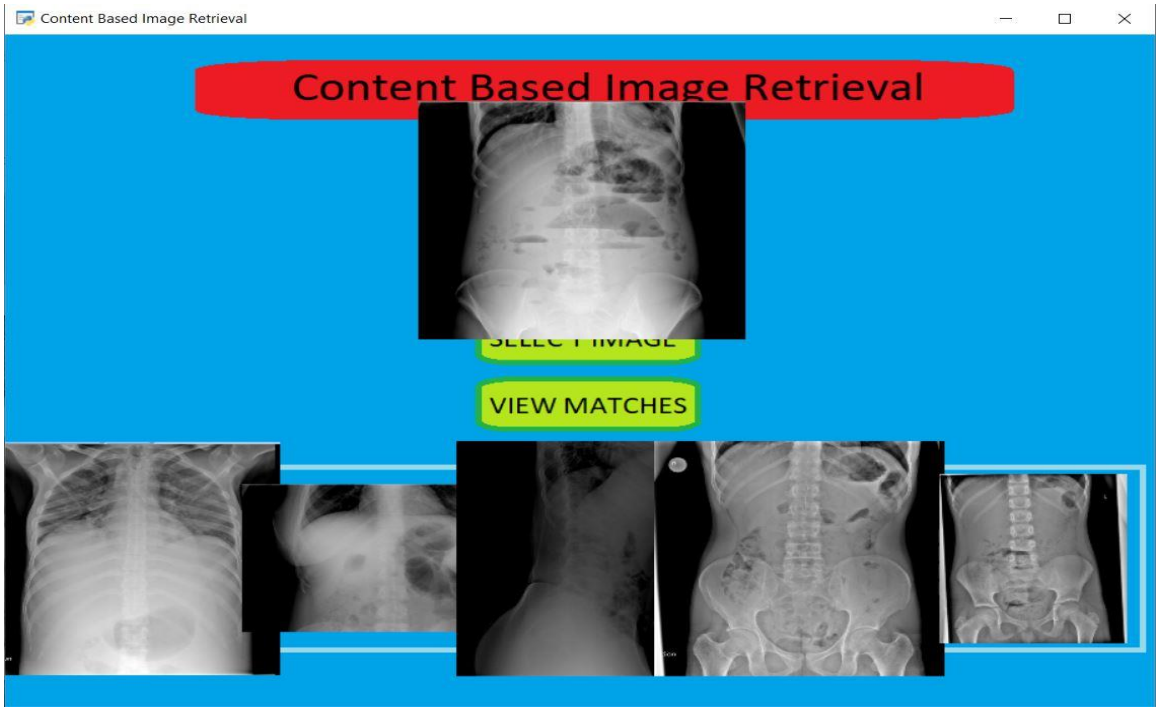
For evaluating the performance of the proposed content-based medical image retrieval, we selected 1 query images from each class with a total of 13 query images. The GUI accept only 1 query image in query search and also retrieved k best similar image for the query image from the medical database. Summary of the retrieval performance of the prototype is given below in table 5.

Table6.Performance of the proposed content-based medical image retrieval

Query	Total Relevant images	Total retrieved images	Retrieved Relevant images	Recall	Precision	F-measure
Abdomen image	18	19	17	0.94	0.89	0.96
C-spin images	22	23	21	0.95	0.91	0.92
Chest images	24	25	24	1.00	0.96	0.97
Foot images	24	25	22	0.91	0.88	0.89
Hand images	26	28	26	1.00	0.92	0.95
Hip joint	4	4	3	0.75	0.75	0.75
Knee images	19	21	18	0.94	0.85	0.89

L-spin images	15	16	14	0.93	0.87	0.89
Pelvis images	15	17	14	0.93	0.82	0.87
Shoulder joint	17	118	16	0.94	0.88	0.90
Skull images	22	23	22	1.00	0.95	0.97
T –fibula	19	20	18	0.94	0.90	0.91
Wrist images	21	22	20	0.95	0.90	0.92
Average				93.6	88.3	90.3

As described in above table 5, total relevant images in each class are different and also total retrieved images are different and retrieved relevant images are different for each class. In addition, we computed recall, precision and F-measure of each class. As shown in the above table 5, retrieval of images related to the result of all class is almost above 0.82 except Hip Joint measured by recall, precision and F-1 measure, on the other hand, for Chest Hand, and Skull classes of the CBMIR register 100% recall



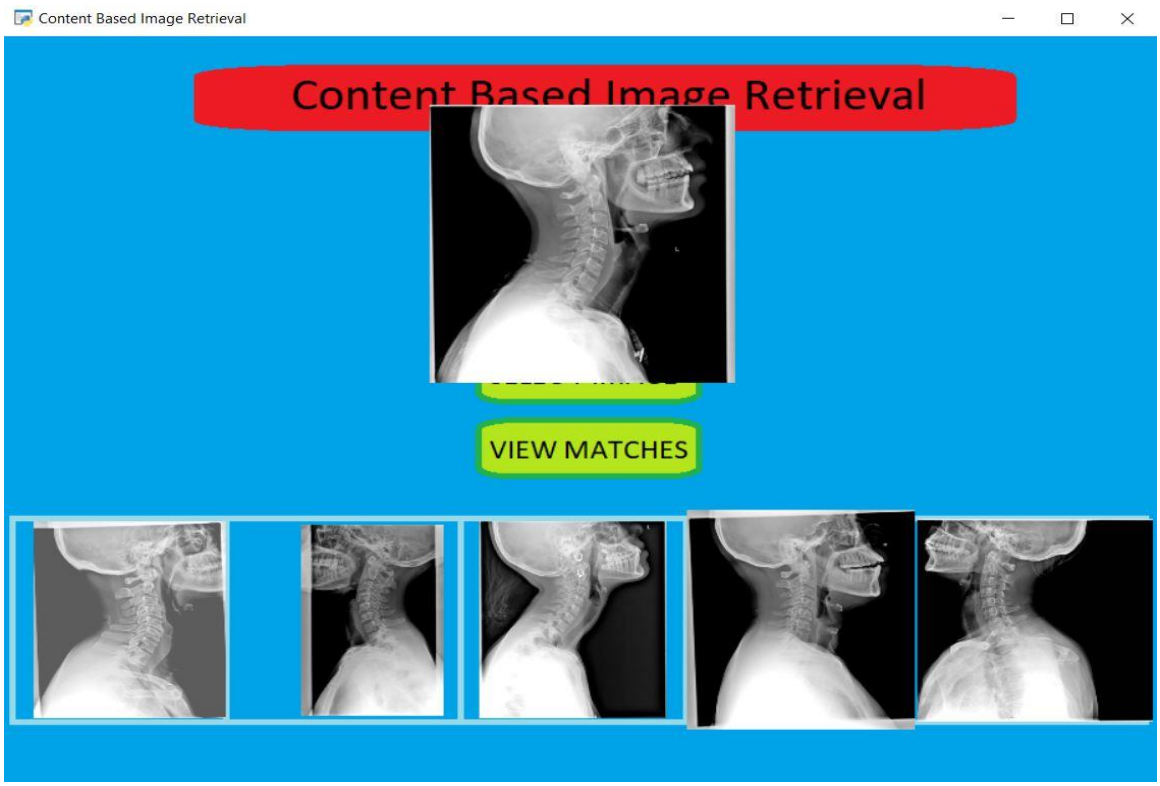


Figure 19 C-spine image retrieval

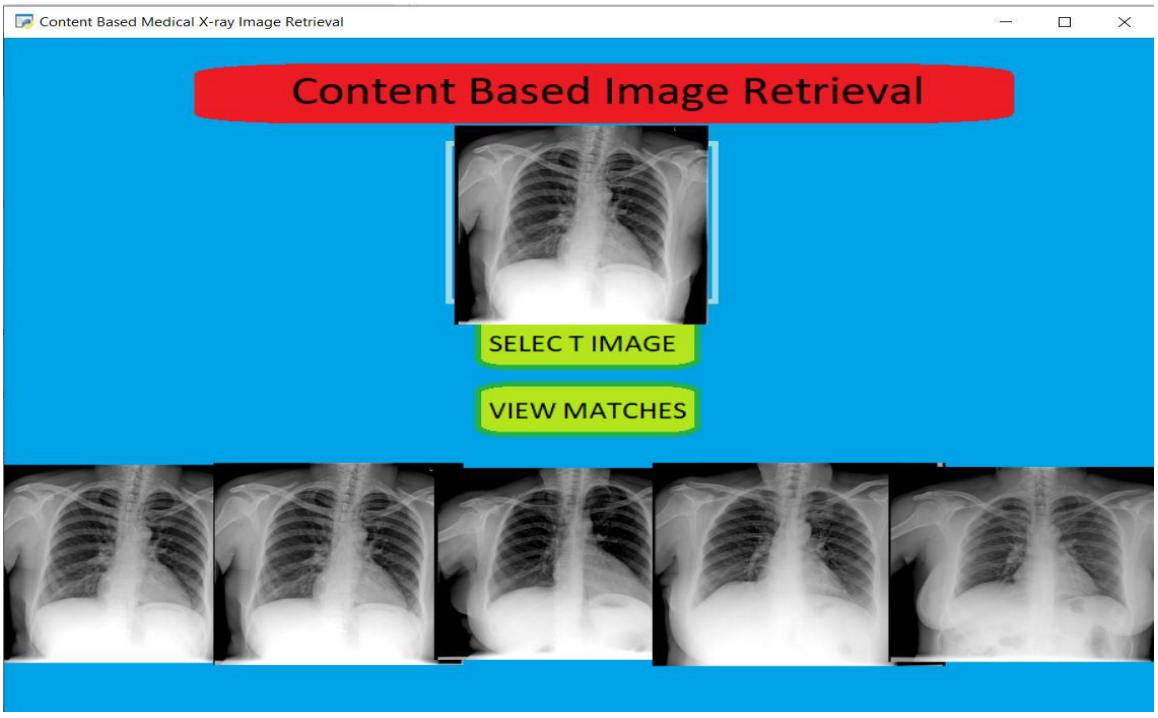


Figure 20. Chest image retrieval

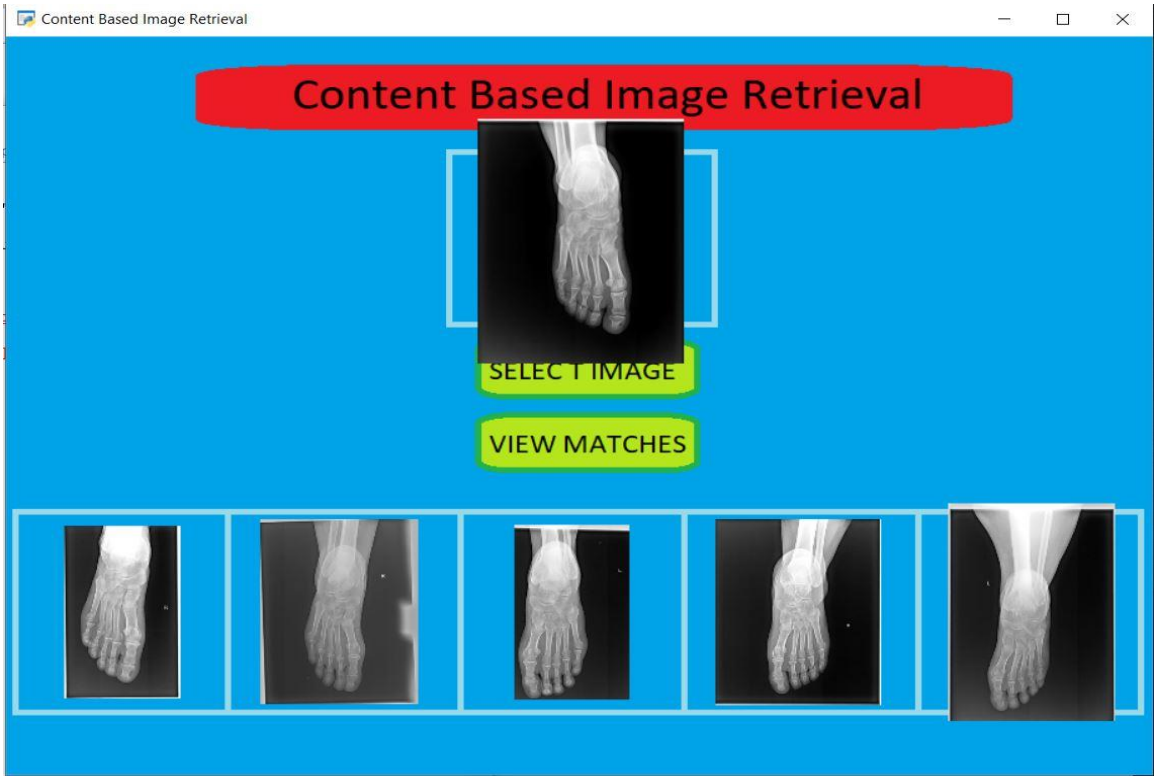


Figure 21. Foot image retrieval

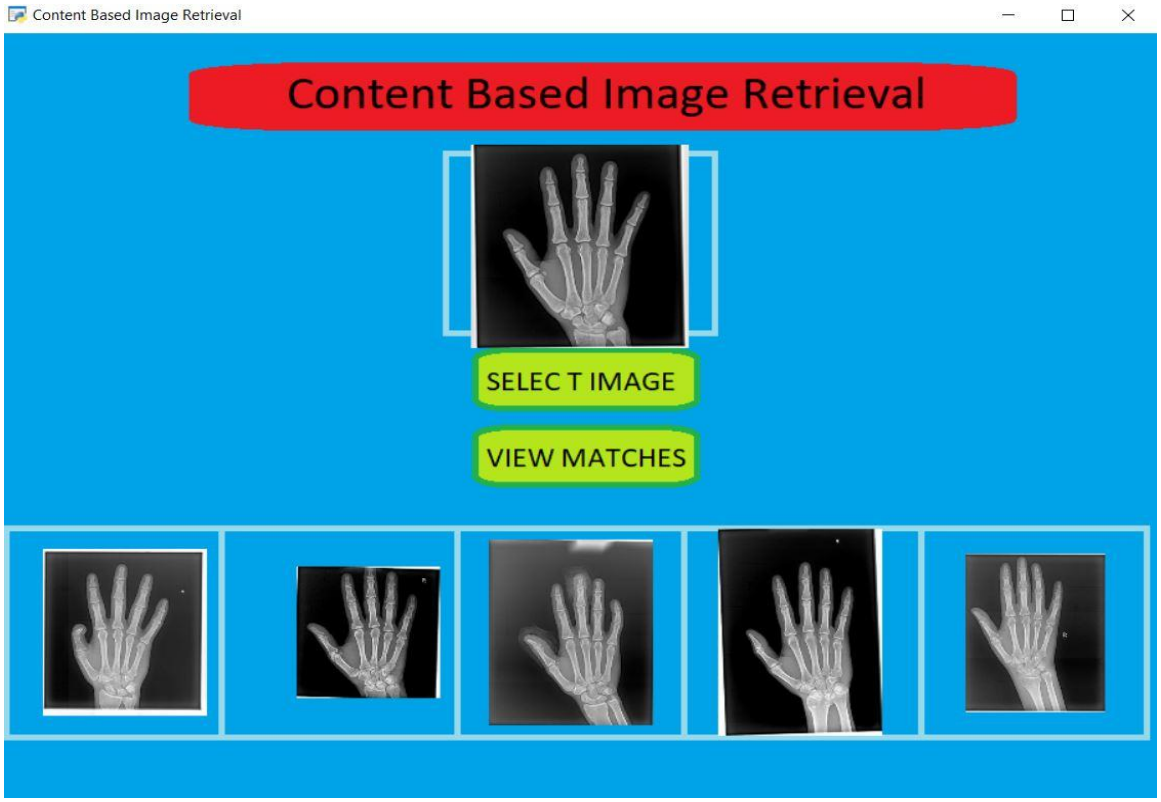


Figure 22. Hand image retrieval

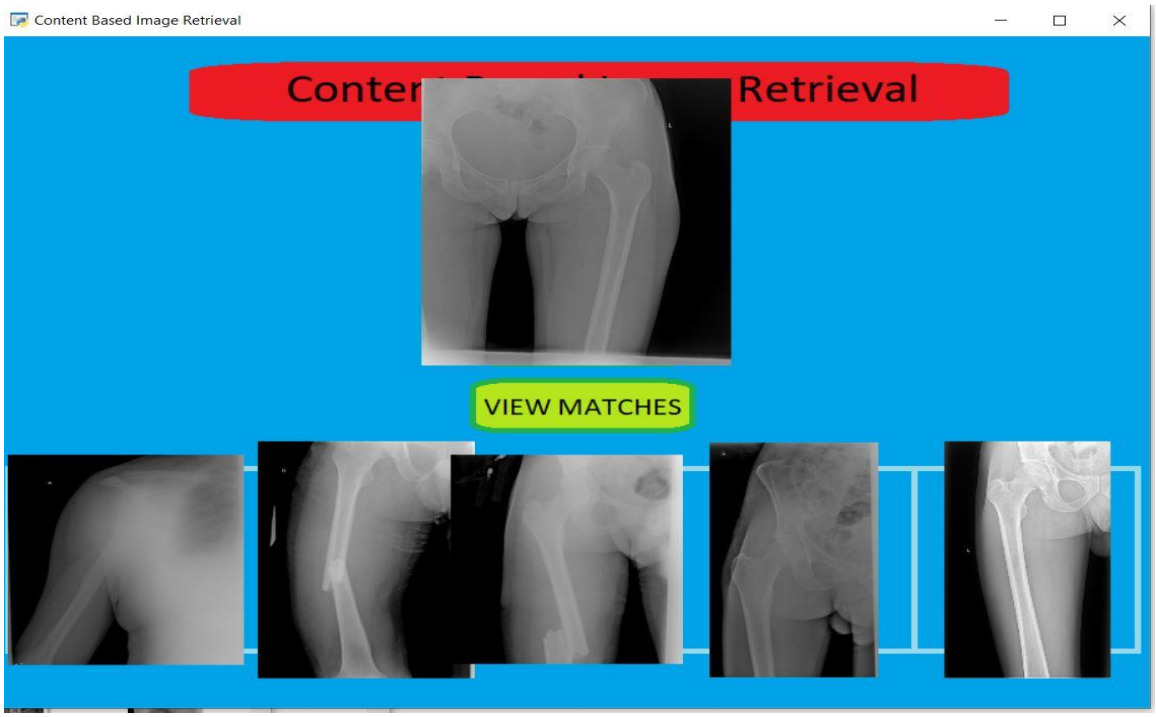


Figure 23. Hip-Joint image retrieval

Figure 24. Knee image retrieval

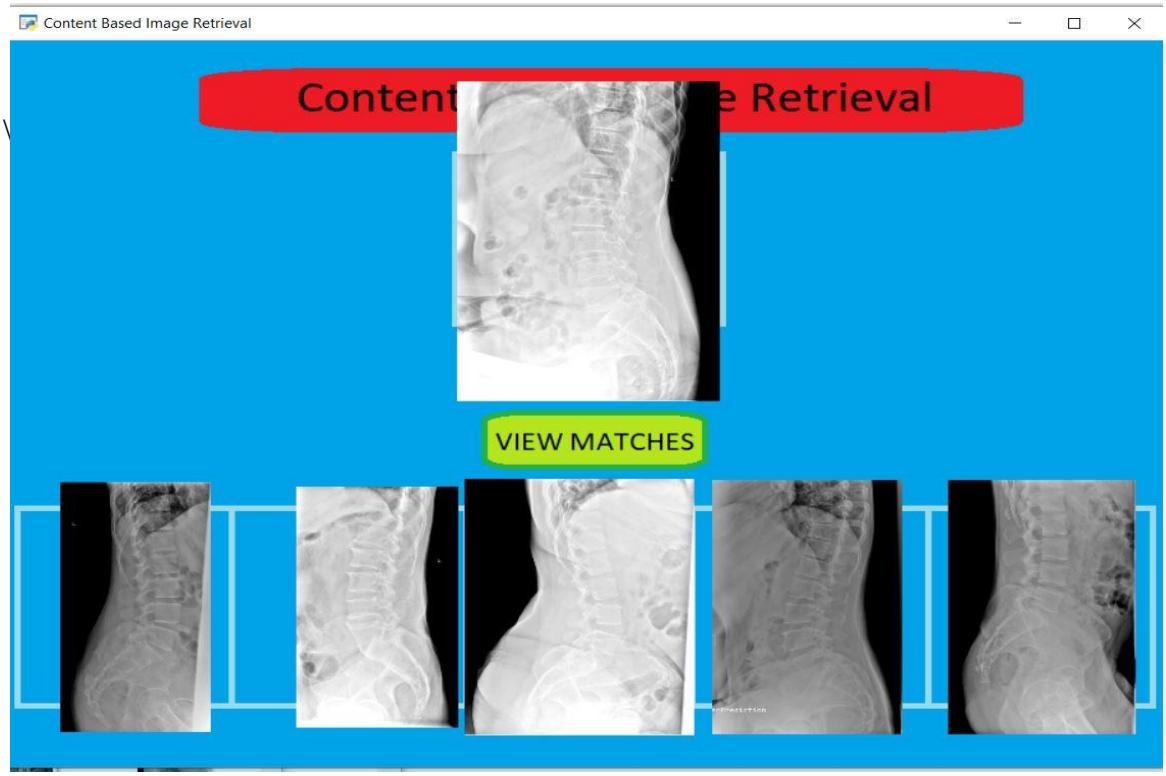


Figure 25. L-spine image retrieval

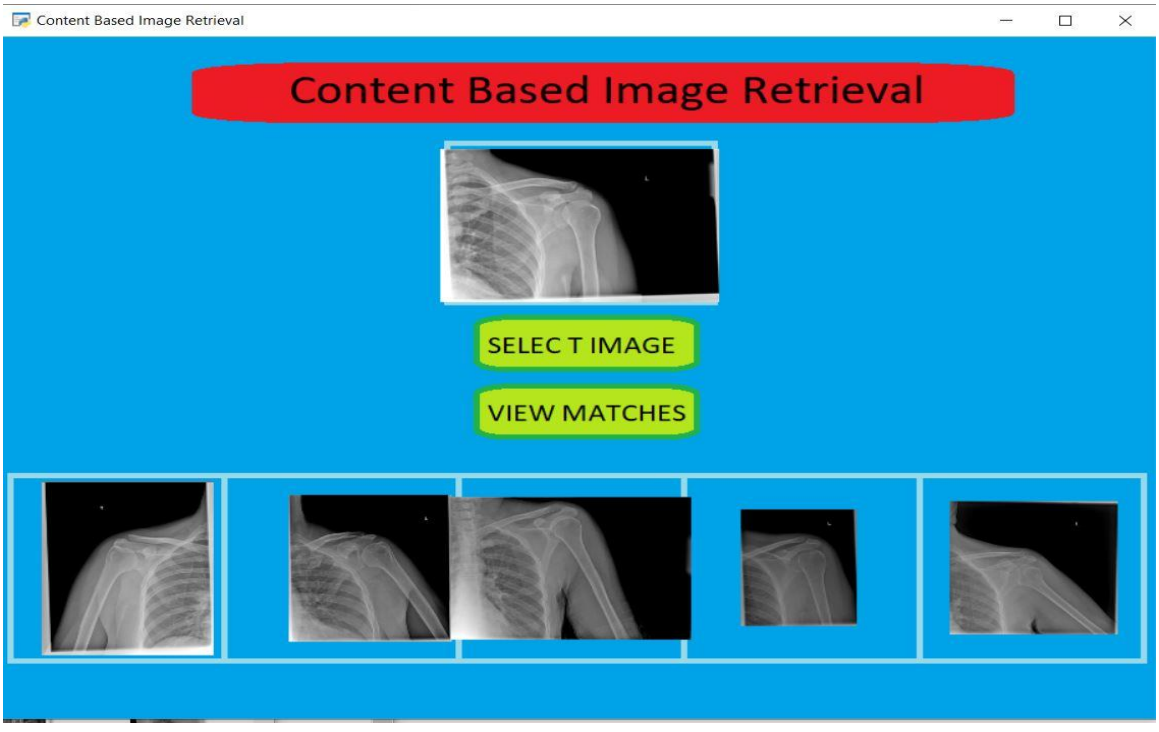


Figure 26. Shoulder Joint image retrieval

Figure 27. Pelvis image retrieval

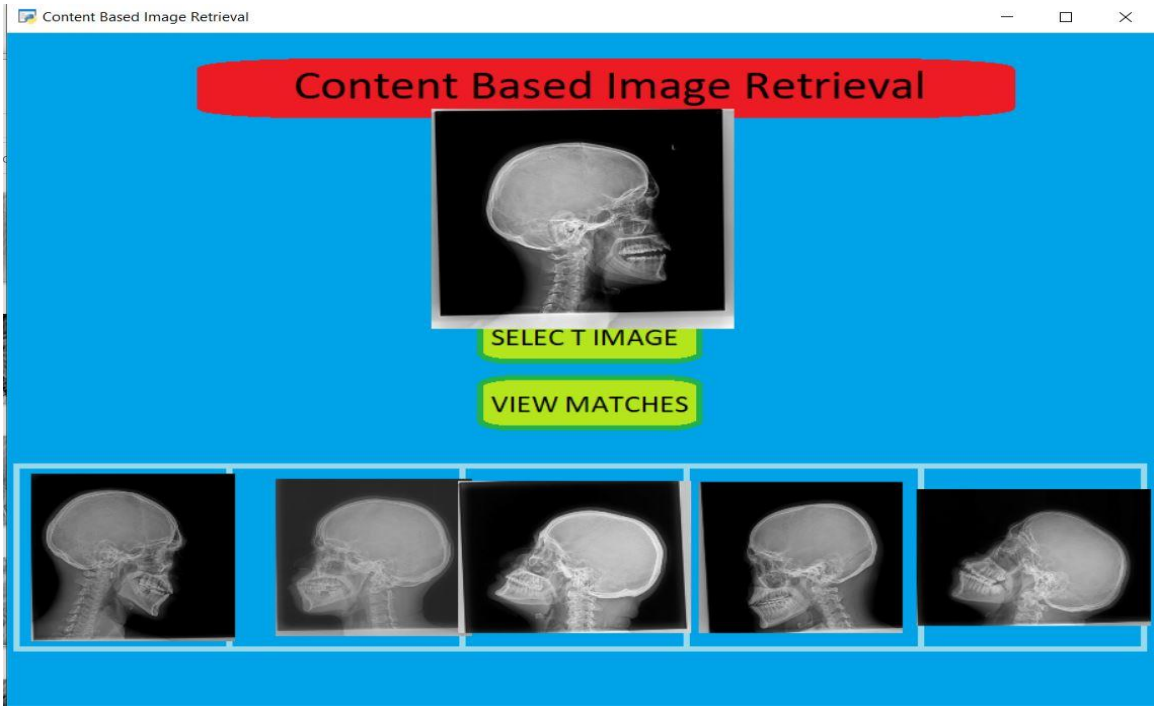


Figure 28. Skull image retrieval

Figure 29. T-Fibula image retrieval

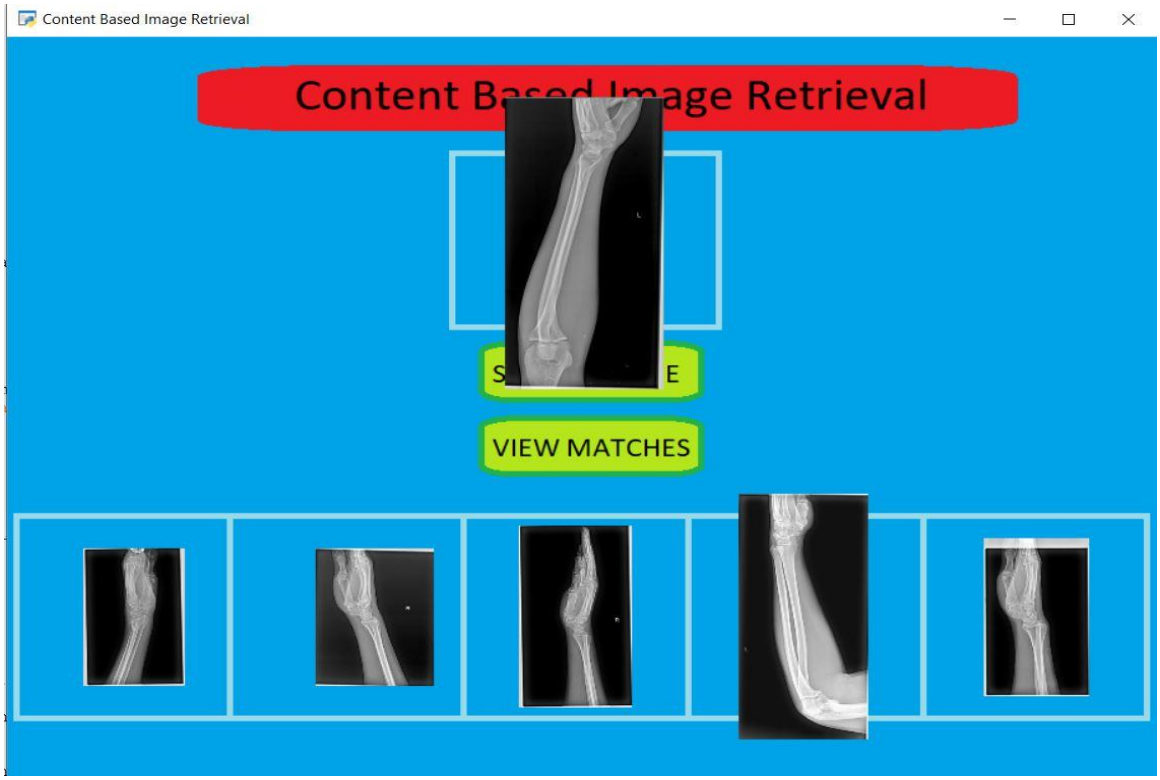


Figure 30. Wrist image retrieval

4.8. **Discussion**

In this study content-based medical image retrieval system based on a deep learning architecture is developed. Deep learning enables initially classify the query image, to reduce the search space while finding the relevant images. This improves the accuracy of the CBMIR system and also reduces the computational overhead. The proposed method is tested on the X-ray image data collection, consisting of X-ray images. The CBMIR system of medical image is built on the deep learning models for the retrieval and classification of X-ray image specific features using transfer learning based like VGG16, VGG19 and ResNet50. Models trained on standard X-ray image datasets. In this research work the distance of each query image measure by Euclidean distance content based image retrieval based on medical database. As presented in the previous sections, the experiments were conducted by using three different pre-trained models like VGG16, VGG19 and ResNet50. All of the experiments are conducted using the same hardware configuration. The VGG16, VGG19 and ResNet50 model was trained with a total of 4194 original images. Confusion metrics are used to measure the performance of the models and when we compared the performance of the models VGG16, VGG19 and ResNet50 models scored 96.74 %, 96.46% and 92.30% accuracy respectively. VGG16 model has been selected for developing content based medical image retrieval.

As we can see in the following figure 31 these show the models are giving best result on different pre-trained model.

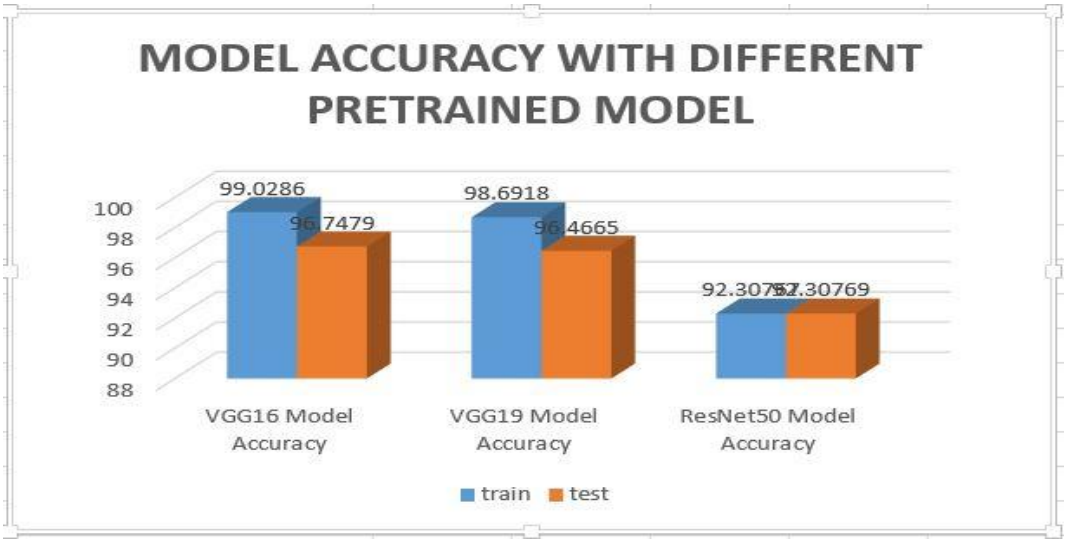


Figure 31. Accuracy of the three experiments using different learning rate

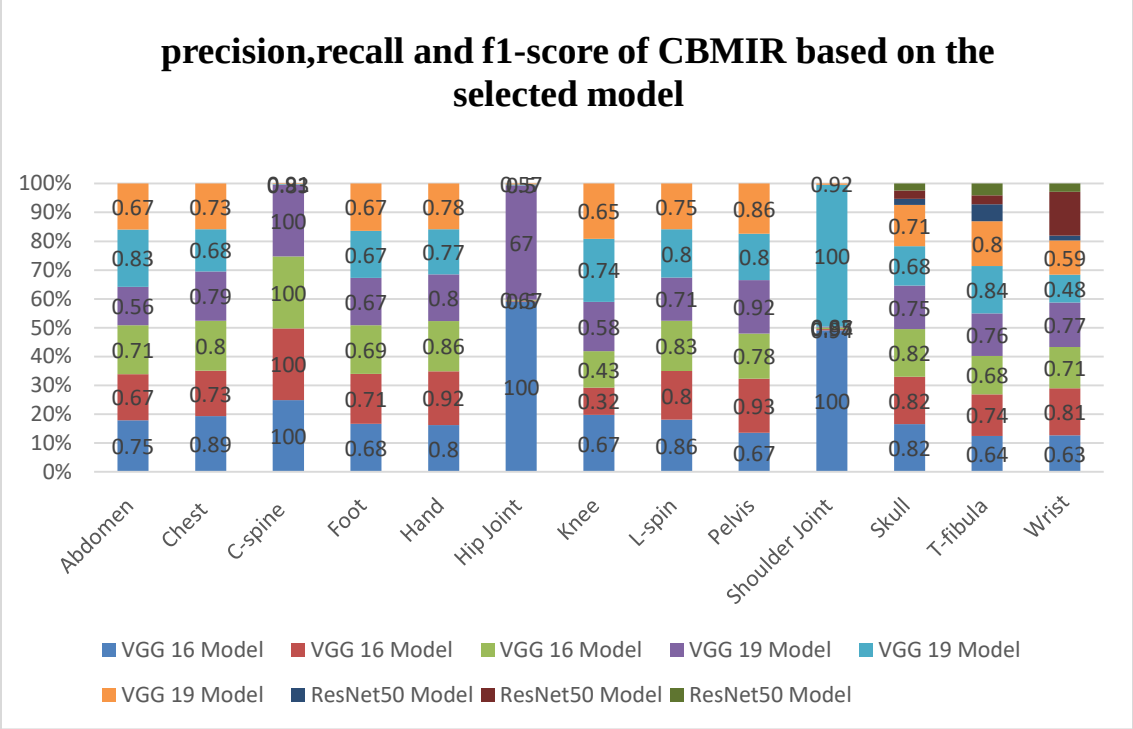


Figure 32.Precision, Recall and f1-score of CBMIR based on the selected model

Most deep learning algorithms especially computer vision for image classification problems are trained by using high performance computing machines with faster GPU, a huge number of images (in millions), and tens of millions of parameters. But we can train and get better results with small sized networks with fewer parameters, less hardware consumption, and fewer data. More accuracy results will be obtained if the images of the dataset are captured in a stable environmental condition, which is a stable distance from an object to the camera, proper light, and proper focus. The other point is that making preprocessing to the images by removing noise and unwanted features will increase the accuracy of the model.

Once the best model is selected, it is integrated with CBMIR for retrieval of medical images as per users query image. The retrieval result is also effective on the given query taken from each class. Accordingly, on the average the CBMIR system registers 93.6% recall, 88.3% precision and 90.3% F-measure. The result is promising to develop an applicable system that can search and retrieve relevant medical images from the database.

CHAPTER FIVE
CONCLUSION AND FUTURE WORK

5.1. Conclusion

This research was meant to a build model that could automatically retrieve and classify CBMIR using deep learning model of pre-trained algorithm. Now day's X-ray medical treatment is suffered from a severe problem; the short age of diagnostics tools in developing countries like Ethiopia has a devastating impact on their development and quality of medical. Therefore, there is an urgent need to retrieve and classify medical images with affordable and easy to use technological solutions .In order to make early identification of the retrieve we have proposed and implemented deep learning approach is implemented retrieve and classify medical X-ray image.

We have shown how transfer learning can be used when working with X-ray medical datasets. It was possible to design a model with very little data when compared to the amount usually required to train VGG16, VGG19 and ResNet50. The model was trained using the images taken from X-ray and achieved is 96.74 %, 96.46% and 92.30% accuracy respectively retrieving and classification ability for the X-ray medical image. This shows the promising ability of VGG16, VGG19 and ResNet50 to extract important features in the X-ray image which is required for retrieving and classification. We have also demonstrated that applying data augmentation on the training set improves the performance of the network when the dataset is very small. The effect of dropout and regularization to overcome overfitting also validated. Using the same model and performing transfer learning, it was possible to design a CNN model of pre-trained using only very limited X-ray image dataset.

The methodology involves image collection, image pre-processing especially for the image processing applying VGG16, VGG19 and ResNet50for classification purpose for extracting features from the image. All the input images will be passing through the image pre-processing steps before it proceeds feature extraction to the deep learning network techniques, whereas, for the deep learning all the input images will be passing through the image preprocessing before it proceeds to the VGG16, VGG19 and ResNet50. The methodology used in this study proves to be one of the simple's

ways of retrieving and classifying medical X-ray images. Consequently, it also proves to be one of the best in terms of accuracy because it works up to VGG16, VGG19 and ResNet50 models is 96.74 %, 96.46% and 92.30% accuracy respectively in terms of results.

After that, the best model is integrated with the CBMIR system for retrieving relevant medical images from the database as per users query. Effectiveness of the system shows that, it scores an average 93.6% recall, 88.3% precision and 90.3% F-measure which is a promising result.

5.2. Future Work

Image analysis technology has a paramount importance in the variety of retrieval and classification of medical images. In Ethiopia, little research has been conducted in this direction to support the health sector. Hence, the research may pave the way and initiate researchers to work more in the area. The medical image taken from medical center most of them have similar in shape, angle and also with more noise in the image like blurred and faded medical images. In the future works there is a need to use noise removal techniques to reduce the effect of noise and produce quality image for best result.

The image analysis especially by using neural network and deep learning techniques for classification and detection of medical images towards designing content based image retrieval can be further investigated. The model constructed in this study classifies only X-ray images but in medical imaging there are a variety of formats. In the future other research has to be done further using other type of medical images, such as MRI, CTscan, etc. which are not included in the current experiment.

As part of future work, there is a need to test the model with a larger dataset with more subclasses like CT scan, MRI, Mammogram, Ultrasound but we have done only with the x-ray images. And second the content based image retrieval doesn't display detail patient information like name, age, gender and other patient profile.

We also recommend to experiment other deep neural models to improve the performance of the CBMIR system and scale it for real world applications.

References

[1] l. K. Ashni, K. Jinman, M. Fulham, F. Weidong Cai and D. , "Content-Based Medical Image Retrieval: A Survey Of Applications To Multidimensional And Multimodality Data," vol. 26, p. 1025–1039, 2013.

[2] N. Seth and S. Jinda, "A REVIEW ON CONTENT BASED IMAGE RETRIEVAL," vol. 15, no. 14, January 2017.

[3] K. Singh and S. Ran, "Content Based Image Retrieval Using DWT and Modified K-Means," vol. 5, no. 5, pp. 861-868 , 2016.

[4] J. Behari and S. , "Concept Content Based Image Retrieval Performance using Combination of Color and Texture Features," vol. 6, no. 6, jun 2016.

[5] A. Depeursinge, B. Fischer, H. Müller and T. M. Deserno, "Prototypes for Content-Based Image Retrieval in Clinical Practice," vol. 5, p. 58–72, 27 jul 2011.

[6] A. Tafti and D. W. Byerly, X-ray Radiographic Patient, Jan,2022.

[7] K. and . E. A, "Current perspectives in medical image perception," vol. 1, no. 5, pp. 1205-1217, 28 feb 2010.

[8] NehaGhosh, A. Dr. Shikha and . M. Dr. Mahesh, "Color String Based CBIR for Image Retrieval," vol. 9, no. 8, August 2018.

[9] S. Khan and . K. Shamaila, "An Efficient Content based Image Retrieval: CBIR," vol. 152, no. 6, p. 0975 – 8887, 2016.

[10] Reshma, Chaudhari and M. P. A. , "Content Based Image Retrieval using Color and Shape Features," vol. 1, no. 5, November 2012.

[11] B. Ramamurthy, K.R. Chandran, V.R. Meenakshi and V. Shilpa, "CBMIR: Content Based Medical Image Retrieval System Using Texture and Intensity for Dental Images”," vol. 305, pp. 125-134, 2012.

[12] . F. Y. Fanid and M.A. Balafar, "Content Based Image Retrieval For Medical Images," vol. 4, no. 12, pp. 177-182, 12 September 2012.

[13] A R Mahajan, S D Zade and R. Pawan , "Content-Based Image Retrieval in Medical Images: Current Status and Future Directions," 2013.

[14] S. Jasmine and R.Dhivya, "A Survey on Local and Global Feature Extraction Techniques in Content Based Medical Image Retrieval," vol. 7, no. 5, march 2019.

[15] Q. Adnan , M. A. Syed , . A. Muhammad and . M. Muhammad, "Medical Image Retrieval using Deep Convolutional Neural Network," vol. 266, pp. 8-20, 29 Novembe 2017.

[16] Pilevar and H. Abdol , "CBMIR: Content-based Image Retrieval Algorithm for Medical Image Databases," vol. 1, no. 1, pp. 12-18, Jan-Apr 2011.

[17] Tena and . L. Shewatatek, "DEVELOPING A CONTENT-BASED IMAGE RETRIEVAL SYSTEM," March 2008.

[18] H. Shilema and H. Shilema , "Content Based Search System for Ethiopian Art Content on the Web," vol. 3, no. 2, pp. 16-20, 2020.

[19] V. Lingamuthu, "Content Based Image Retrieval using Colour,Gray ,Advanced Texture, Shape Features and Random Forest Classifier with Particle Swarm Optimization," 14 May 2021.

[20] R. Chhanda and . S. Krishnendu, "A New Approach for Clustering of X-ray Images," vol. 7, no. 4, July 2010.

[21] L. Afshan , R. Aqsa , . S. Umer, A. Jameel , . A. Nouman, I. R. Naeem , Z. Bushra , H. D. Saadat , . S. Muhammad and . K. Tehmina, "Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review," vol. 2019, 26 Aug 2019.

[22] Nascimento and . L. F. Marcio, "Brief history of X-ray tube patents," vol. 37, pp. 48-53, June 2014.

[23] A. Dhaliwal , . N. Singh, . R. Kapila and R. Rajput , "History of X-Rays in Dentistry," vol. 2, no. 1, pp. 21-25, 2012.

[24] Grinstaff, l. Hrvoje and W. Mark , "X-Ray Computed Tomography Contrast Agents," vol. 113, no. 3, march 2013.

[25] T. Dawood and V. M. Christopher , "X-ray Production," 3 August 2021.

[26] D. Mohd. , R. Ritika and S. Ratika , "A Survey: Content Based Image Retrieval Based On Color,Texture, Shape & Neuro Fuzzy," vol. 3, no. 5, pp. 839-844 , Sep-Oct 2013.

[27] A. m. Abdulrehman and A. Dr. Cyrus , "A Literature Survey of Image Descriptors in

Content Based Image Retrieval.," March 2016.

[28] M. H. Ibtihaal , H. A. Sadiq and . M. M. Basheera, "Content-Based Image Retrieval: A Review Of Recent Trends," vol. 8, no. 1, pp. 1-37, 02 Jun 2021.

[29] A. M. Abdulrehman , A. P. Dr. Cyrus and John Kamau, "A Literature Survey of Image Descriptors in Content Based Image Retrieval," vol. 7, no. 3, March 2016.

[30] T. A. p. Kotha and R. S. Dr. Jhansi , "Secure Content Based Image Retrieval System Using CNN and VGG-16," vol. 7, no. 11, pp. 842- 848, November 2020.

[31] A. Khalid EL, C. Youness and . I. Ali, "Automated Methods for Detection and Classification Pneumonia based on X-Ray Images Using Deep Learning".

[32] l. fei , . w. yong, w. fan-chuan , . z. yong-zheng and l. jie , "intelligent and secure content-based image retrieval for mobile users," vol. 7, pp. 119209- 119222, 6 september 2019.

[33] "ImageNet: VGGNet, ResNet, Inception, and Xception with Keras," 20 March 2017. [Online]. Available: https://pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/. [Accessed 30 04 04/30/2022].

[34] . P. Giulia, M. Massimo , R. Riccardo , . R. Luca and . F. Emanuele, "A Deep Learning-Based Approach for Automatic Leather Classification in Industry 4.0," vol. 12664, p. 662–674, 05 March 2021.

[35] . V. VIKAS, "Image Retrieval And Classification Using Local Feature Vectors," June 2011.

[36] S. Kopparthi, "Content based Image Retrieval using Deep Learning Technique with Distance Measures," pp. 251- 261, DECEMBER 2020.

[37] T. Satish , D.Jayadevappa and C.Gurur, "A Comparative Study of Content Based Image Retrieval Trends and Approaches," vol. 9, no. 3, 2015.

[38] M. Fazal and . B. Baharum, "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain," vol. 25, no. 2, pp. 207-218, July 2013.

[39] Ms. K. Arthi and Mr. J. Vijayaraghavan, "Content Based Image Retrieval Algorithm Using Colour Models," vol. 2, no. 3, March 2013.

[40] Z. Shaomin , Z. Lijia and Z. Tao , "Medical Image Retrieval Using Empirical Mode

Decomposition with Deep Convolutional Neural Network," vol. 2020, p. 12, 2020.

[41] Y. Rikiya , . N. Mizuho, . G. D. Richard Kinh and T. Kaori , "Convolutional neural networks: an overview and application in radiology," vol. 9, p. 611–629, 22 june 2018.

[42] M. Vaishnavi , G. Khyathi and . T. P. Sri, "Plant Leaf Disease Prediction," vol. 9, no. 7, pp. 1295-1305, July 2021.

[43] S. K. Lavanya, Vennela and .. K.Divya, "FIRE AND SMOKE RECOGNITION USING DEEP LEARNING," vol. 9, no. 7, july 2021.

[44] B. . H. Morteza, Tizhoosh, . Z. Shujin and M.E. , "Retrieving Similar X-Ray Images from Big Image Data Using Radon Barcodes with Single Projections," vol. 1, 2 Jan 2017.

[45] Subashini, . G. Sumathi and T.S., "A Content Based Approach to Medical X-Ray Image Retrieval using Texture Features," vol. 12, no. 7, pp. 3742- 3748, 2 F e b r u a r y 2 0 1 4.

[46] Manjunatha, . S. R. C and M. B., "An Efficient Chest X-Ray Image Retrieval using CBIR," vol. 10, no. 4, January 2017.

[47] . P. Sreekanth, . S. M.James and . P. P.V.G.D, "An Efficient Content-Based Medical Image Retrieval System For Clinical Decision Support In Brain Tumor Diagnosi," vol. 12, no. 9, pp. 2922-2929, 20 April 2021.

[48] S. D. K. Vijayakumar Bhandi, "Image Retrieval Using Features From Pre Trained Deep CNN," vol. 9, no. 6, pp. 687-693, JUNE 2020.

[49] C. Reshma and . P. A. M., "Content Based Image Retrieval Using Color and Shape Features," vol. 1, no. 5, pp. 386- 392, November 2012.

[50] M. . L. Thomas , O. G. Mark , . T. Christian, P. Bartosz , D Bastian and . S. Henning, "IRMA – Content-Based Image Retrieval in Medical Application," February 2004.

[51] François Chollet, Deep Learning with Python, Manning Publications, 2017.

[52] "Adobe Photoshop CS5 Tutorial".

[53] Mohammed Sunasra;, "medium.com," 11 Nov 2017. [Online]. Available: https://medium.com/@MohammedS/performance-metrics-for-classification-

problems-in-machine-learning-part-i-b085d432082b. [Accessed 8 10 2021].

[54] Bharathi;, "Analytics Vidhya," 24 June 2021. [Online]. Available: https://www.analyticsvidhya.com/b;pg/2021/06/confusion-matrix-for-multi-class-classification. [Accessed 8 October 2021].

[55] G., Tripathi, "Review on color and texture feature extraction techniques," 2014.

[56] J.Schmidhuber, Deep Learining in neural networks: An overview, Neural Networks, vol. 61, 2015, pp. 85-117.

[57] M.Arif Wani, Farooq Ahmad Ghat, Saduf Afzal, Asif Iqbal Khan, Advances in Deep Learning, vol. 57, P. A. o. S. W. Janusz Kacprzyk, Ed., Singapore: Springer, 2020, pp. 53-75.

[58] C., Francois, Deep Learning with Python, New York: Manning Publications, 2017.

[59] C.Aggarwal, Charu, Neural Networks and Deep Learning, Switzerland AG: Springer International Publishing AG,part of Springer Nature, 2018, pp. 339-349.

[60] L. Z. ,. a. T. Z. Shaomin Zhang, "Medical Image Retrieval Using Empirical Mode," vol. Volume 2020, p. 12, December 2020.

[61] A. Geron, "Hands On Machine Learning with Scikit-Learn, Keras and TensorFlow," 2019.

[62] C. Metz, "TensorFlow, Google's Open Source AI, Points to a Fast-Changing Hardware World," Wired, 2015.

APPENDICES

APPENDIX I

Approval Letter of X-ray image from Zewuditu Memorial Hospital

አዲስ አበባ ከተማ አስተዳደር ጤና ቢሮ
City Government of Addis Ababa Health Bureau

Ref.N.o. ጤ/ከ/ጥ/7936/227
Date 3/5/2014

TO:

- **ZEWUDITU MEMORIAL HOSPITAL**

Subject: Request to access Facilities to conduct approved research

This letter is to support **MUNIR ALI SEID** to conduct research hich is entitled as **"CONTENT BASED IMAGE RETRIEVAL FROM MEDICAL IMAGE DATABASE."** The study proposal was duly reviewed and approved by Addis Ababa Health Bureau procedures and submit an activity progress report to the Ethical Committee as required. Therefore we request the facility and staffs to provide support to the principal investigator.

With Regards
Ethical Clearance Committee

Cc

- **MUNIR ALI SEID**
- **To Ethical Clearance Committee**

APPENDIX II

Ethical letter from St. Mary's university

ቅድስት ማርያም ዩኒቨርስቲ
ድኅረ-ምረቃ ት/ቤት

St. Mary's University
School of Graduate Studies

☎+251-11-552-45 37/66 ✉1211, 18490 Fax 552 83 49 e-mails: sgs@smuc.edu.et, Addis Ababa, Ethiopia

Ref No: smu- 0637

Date: Dec 21, 2021

Addis Ababa Health Bureau

Addis Ababa

Dear Sir/Madam,

Mr.Munir Ali ID SGS/0396/2012A is a graduate student in Computer Science at St. Mary's University. Currently, he has been engaged in his thesis work entitled "***Content Based Image Retrieval for Medical Image Database***" Therefore, I would kindly request your esteemed organization to allow him to collect the required data for him thesis work. Any assistance rendered to him in this regard is highly appreciated.

With best regards,

Mosisa Kejela
Director ,SGS
Student Affairs

SMU