Survey of Databases Used in Image Processing and Their Applications

Shubhpreet Kaur, Gagandeep Jindal

Abstract- This paper gives review of Medical image database (MIDB) systems which have been developed in the past few years for research for medical fraternity and students. In this paper, I have surveyed all available medical image databases relevant for research and their use.

Keywords: Image database, Medical Image Database System.

1. INTRODUCTION

Medical imaging is the technique and process used to create images of the human for clinical purposes (medical procedures seeking to reveal, diagnose or examine disease) or medical science. As a discipline, it is part of biological imaging and incorporates radiology, nuclear medicine, investigative radiological sciences, endoscopy, (medical) thermography, medical photography and microscopy.

Shubhpreet Kaur *is currently pursuing masters degree program in Computer Science and engineering in Chandigarh Engineering College, Mohali, India. E-mail: shubh_86@ymail.com*

Gagandeep Jindal *is currently assistant processor in department Computer Science and Engineering in Chandigarh Engineering College, Mohali, India. E-mail: gaganpec@yahoo.com*

Measurement and recording techniques, such as electroencephalography, magnetoencephalography (MEG), Electrocardiography (EKG) and others, can be seen as forms of medical imaging. Image Analysis is done to ensure database consistency and reliable image processing.

Open source software for medical image analysis

Several open source software packages are available for performing analysis of medical images:

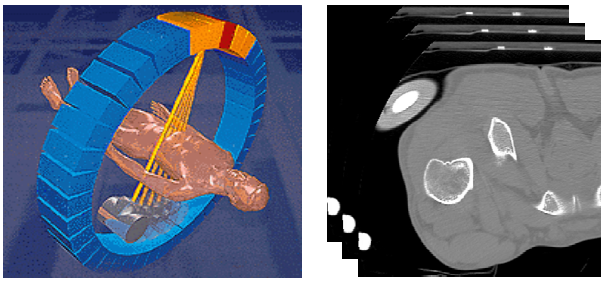
- ImageJ
- 3D Slicer
- ITK
- OsiriX
- GemIdent
- MicroDicom
- FreeSurfer

1.1 Images used in Medical Research

Here is the description of various modalities that are used for the purpose of research by medical and engineering students as well as doctors.

i) Computed Tomography (CT)

Computed Tomography, also known as computed axial tomography, or CAT scan is a medical technology that uses X-rays and computers to produce three-dimensional images of the human body. Unlike traditional X-rays, which highlight dense body parts, such as bones, CT provides detailed views of the body's soft tissues, including blood vessels, muscle tissue, and organs, such as the lungs.



a) b)

Fig. 1.1:a) CT scanner and b) CT scan Image.

ii) Magnetic Resonance Imaging(MRI)

Magnetic resonance imaging (MRI), nuclear magnetic resonance imaging (NMRI), or magnetic resonance tomography (MRT) is a medical imaging technique used in radiology to visualize detailed internal structures. MRI makes use of the property of nuclear magnetic resonance (NMR) to image nuclei of atoms inside the body. MRI provides good contrast between the different soft tissues of the body, which make it especially useful in imaging the brain, muscles, the heart, and cancers compared with other medical imaging techniques such as computed tomography (CT) or X-rays.

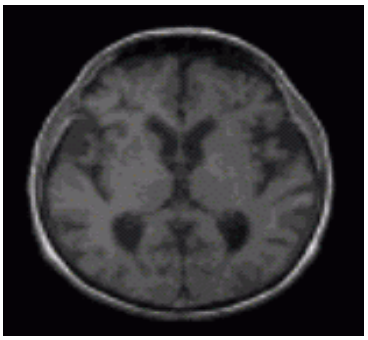


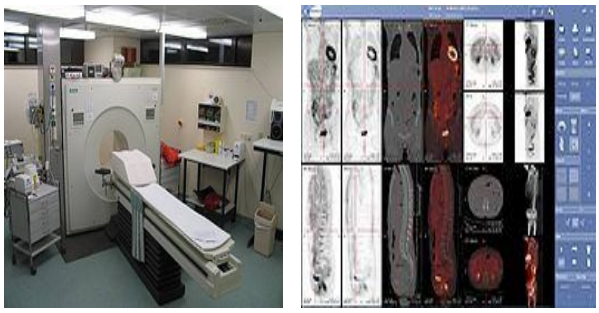
Fig.1.2: MRI Image.

iii) Single-photon emission computed tomography (SPECT)

SPECT, or less commonly, SPET is a nuclear medicine topographic imaging technique using gamma rays. SPECT imaging is performed by using a gamma camera to acquire multiple 2-D images (also called projections), from multiple angles thus, yielding a 3-D dataset. This dataset may then be manipulated to show thin slices along any chosen axis of the body, similar to those obtained from other tomographic techniques, such as MRI, CT, and PET.

iv) Positron emission tomography (PET)

PET is a nuclear medicine imaging technique that produces a three-dimensional image or picture of functional processes in the body. It is both a medical and research tool. It is used heavily in clinical oncology (medical imaging of tumors and the search for metastases), and for clinical diagnosis of certain diffuse brain diseases such as those causing various types of dementias. PET is also an important research tool to map normal human brain and heart function.

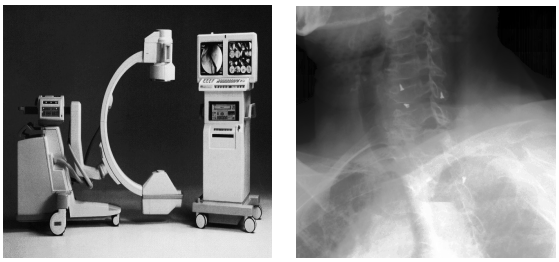


a) b)

Fig.1.3: a) Image of a typical positron emission tomography (PET) facility, b) A complete body PET / CT Fusion image**.**

v) Plain X-rays

X-rays are useful in the detection of pathology of the skeletal system as well as for detecting some disease processes in soft tissue. Some notable examples are the very common chest X-ray, which can be used to identify lung diseases such as pneumonia, lung cancer or pulmonary edema, and the abdominal X-ray, which can detect intestinal obstruction. X-rays may also be used to detect pathology such as gallstones or kidney stones which are often visible. Traditional plain X-rays are less useful in the imaging of soft tissues such as the brain or muscle.



a) b)

Fig.1.4: a) X-Ray machine, b) X-Ray of Neck.

1.2 WHY WE USE IMAGES FOR RESEARCH?

The images are the visual representations that help a person to retain the effects in his memory for a longer time and in the case of research work, they help the professionals as well as the practitioners to clearly go through the problem. Talking about the use of medical images, they help the physicians and nurses, allied health professionals, medical students, graduate nursing students and other post-graduate trainees in identifying the pathology, thus saving their valuable time, resulting in improved clinical decisions and lowering the medical cost.

2. WHERE ONE CAN GET MEDICAL IMAGES FOR DOING RESEARCH?

Hereby, we have discussed various MIDB's available world-wide, name of the database, their description and applications of these images.

2.1 LUNG IMAGE DATABASE CONSORTIUM (LIDC)

Description: Lung cancer is the leading cause of cancer death worldwide, both in men and women, with an estimate of over 164,000 new cases and over 156,000 deaths in 2000 in the United States alone. Preliminary clinical studies have shown that spiral CT scanning of the lungs can improve early detection of lung cancer in high-risk individuals. Image processing algorithms have the potential to assist in lesion detection on spiral CT studies, and to assess the stability or change in lesion size on serial CT studies. The use of such computer-assisted algorithms could significantly enhance the sensitivity and specificity of spiral CT lung screening, as well as lower costs by reducing physician time needed for interpretation.

Number and type of Images: LIDC have a collection of about 200 serial CT data.

Applications: This database serves as an important resource for researchers interested in developing improved methods for early detection and screening for lung cancer.

Specifically, the LIDC initiative aims were to provide:

- a reference database for the relative evaluation of image processing or CAD algorithms.
- a flexible query system that will provide investigators the opportunity to evaluate a wide range of technical parameters that may be important for research applications.

Various universities that makes access to this database are Cornell University, University of California, Los Angeles, University of Chicago , University of Iowa, University of Michigan for medical research and teaching.

2.2 NATIONAL BIOMEDICAL IMAGING ARCHIVE (National Cancer Institute, USA):

Description: NBIA is a searchable repository of medical images that provides the biomedical research community, industry, and academia with access to image archives. to be used in the development and

validation of analytical software tools that support:

- Lesion detection and classification
- Accelerated diagnostic imaging decision
- Quantitative imaging assessment of drug response

NBIA provides access to imaging resources that will improve the use of imaging in today's biomedical research and practice by:

- Increasing the efficiency and reproducibility of imaging cancer detection and diagnosis
- Ultimately enabling the development of imaging resources that will lead to improved clinical decision support.

Various database repository under NBIA are:

Database1: CT Colonography

The "CT Colonography" (CTC) image collection in DICOM format.

Collection Statistics are as under :

Modalities	CT
Number of Patients	825
Number of Studies	836
Number of Series	3,451
Number of Images	941,774

Database2: FDG-PET Lymphoma

FDG-PET Lymphoma is a collection consisting of Lymphoma cases using PET and CT modalities.

Modalities	PET, CT
Number of Patients	14
Number of Studies	40
Number of Series	77
Number of Images	28,462

Database3: Head-Neck Cetuximab (RTOG 0522 and ACRIN 4500)

The Head-Neck Cetuximab collection in NBIA consists of image data from ACRIN 4500/RTOG 0522, which was randomized phase III Trial of Radiation Therapy and Chemotherapy for stage III and IV Head and Neck carcinomas. Data was provided to NBIA through two independent channels:

- RTOG 0522: CT, Structures, RT Doses, RT Plans sent from ITC
- ACRIN 4500: Quantitative PET (PET/CT) sent from ACRIN

Modalities	PET, CT, RTSTRUCT, RTDOSE
Number of Patients	107
Number of Studies	354
Number of Series	1,387
Number of Images	202,320

Database4: IDRICONDUIT

Modalities	DX, CT, CR
Number of Patients	349
Number of Studies	1,411
Number of Series	4,494
Number of Images	416,004

Database5: I-SPY

The I-SPY collection is a demonstration project in which some images relating to the I-SPY breast cancer trial were collected and stored.

Modalities	MRI, HC
Number of Patients	6
Number of Studies	21
Number of Series	108
Number of Images	5,054

Applications: NBIA repository is used in the development and validation of analytical software tools that support:

- Lesion detection and classification
- Accelerated diagnostic imaging decision
- Quantitative imaging assessment of drug response

2.3 NATIONAL CENTRE FOR RESEARCH RESOURCES:

Description: The resource of National Centre for Research Resources are dedicated to the design of quantitative magnetic resonance (MR) acquisition and processing technology to assess tissue changes and alterations in function, metabolism, and physiology as the brain changes during neurodevelopment or neurodegeneration.

Applications: To accomplish the needs of a large community of clinicians and neuroscientists at several institutions, this technology development came into existence.

National Centre for Research Resources is supported by NIH/NCRR grant P41 RR015241.

2.4 F.M. Kirby Research Centre for functional brain Imaging at Kennedy Krieger institute (Baltimore, Maryland):

Description: The datasets are publicly accessible, in order to promote search in medical image processing and analysis. Their primary databases are possible through collaborative efforts in:

1. Diffusion Tensor Imaging
2. Image Analysis of Functional Anatomy
3. Kirby 21: Multi-Modal MRI Reproducibility

a. DTI Datasets

This database is developed by the Laboratory of Brain Anatomical MRI, these datasets show DTI data and three-dimensional fiber trajectories:

- *Human Brain Atlas*

  These atlases were developed from coronal, sagittal, and axial slices. The CMRM also has a human atlas template, a mouse atlas, a monkey atlas, and a CT-MRI atlas.
- *DTI Database*

  This human database contains raw and processed DTI data of the normal population. It is a public database, open to registered users. Its purpose is to facilitate research in DTI data processing and analysis, to study specific biological interests, or to be used as control data.
- *Developmental DTI Database*

  This database is still under construction.

b. Functional Anatomy Datasets

Developed by the Center for Imaging Science at Johns Hopkins University, these datasets show analyzed biological images.

c. Kirby 21: A Multi-Modal MRI Reproducibility Resource

This database consists of scan-rescan imaging sessions on 21 healthy volunteers. All data have been converted to NIFTI format and is publicly released.

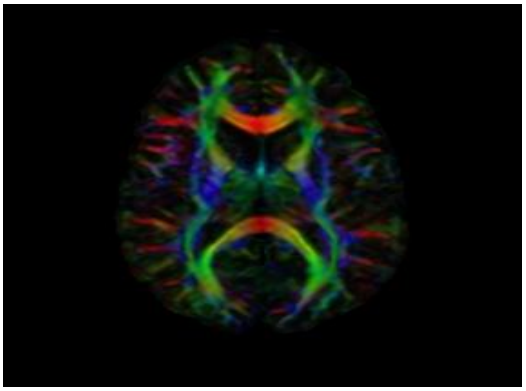


Figure.2.1: An axial slice from the CMRM DTI Human Atlas.

Applications: This dataset is intended to be a resource for statisticians and imaging scientists to be able to quantify the reproducibility of their imaging methods.

2.5 IMAGE SCIENCES INSTITUTES:

a. DRIVE: Digital Retinal Images for Vessel Extraction

Description: The DRIVE database has been established to enable comparative studies on segmentation of blood vessels in retinal images. The images for the DRIVE database were obtained from a diabetic retinopathy screening program in The Netherlands. The screening population consisted of 400 diabetic subjects between 25-90 years of age.

Number of Images:

Number	Description
40	images are selected.
33	do not show any sign of diabetic

retinopathy

7	show signs of early diabetic retinopathy

Each image has been JPEG compressed. The set of 40 images has been divided into a training and a test set, both containing 20 images.

b. SCR database: Segmentation in Chest Radiographs

Description: The automatic segmentation of anatomical structures in chest radiographs is of great importance for computer-aided diagnosis in these images. The SCR database has been established to facilitate comparative studies on segmentation of the lung fields, the heart and the clavicles in standard posterior-anterior chest radiographs.

Number of Images in the database:

Type of Images	154 nodule and 93 non-nodule images
Resolution of images	2048 x 2048 matrix size, 0.175mm pixel size
Density Range	Wide (12bit, 4096 gray scale)
Additional Information	patient age, gender, diagnosis (malignant or benign), X and Y coordinates of nodule, simple diagram of nodule location, degree of subtlety in visual detection of nodules

Applications: Useful for diagnostic training and testing

2.6 MEDPIX

Description: It is a free online Medical Image Database and Radiology Portal, provided by the Departments of Radiology and Biomedical Informatics, Uniformed Services University, Bethesda, Maryland, USA. MedPix is a fully web-enabled cross-platform database, integrating images and textual information. The primary target audience includes physicians and nurses, allied health professionals, medical students, graduate nursing students and other post-graduate trainees. The content material is organized by disease location, pathology category, patient profiles, and by image classification and caption.

Current MedPix Inventory:

Number of Registered users	37973
Images with Captions and Meta-	54192

data	
Peer Reviewed and Approved Teaching File Cases	11683
Approved Disease Topics	6992
Contains Dataset of	CT, MRI, CR, PET scans of ear, lungs, chest and thorax, prostate, liver, breast, skull and content, face, sinuses and neck, spine, Vascular and lymphatic images.

2.7 McKESSON CORPORATION**(MyPACS.net):**

Description: This database have relevant datasets for detecting ailments such as, Congenital, infection, non-infectious inflammatory disease, Benign Mass Cyst, Neoplasm, Trauma, Normal/Variants, latrogenic, Metabolic, Hematological, Vascular and all chest problems.

Applications: this database is helpful for medical researchers and doctors to cure the above ailments and can curb the problem at right time.

2.8 CORNELL UNIVERSITY (New York City) :

SIMBA Home: VIA-ELCAP Public Access

DICOM image data may be rapidly as well as directly downloaded in blocks of 10 cases.

Format of Image Data	DICOM
Repository	For lungs, heart, abdomen, liver, spine, etc

Applications: This database serves as an important resource for researchers interested in developing improved methods for early detection of diseases.

3. CONCLUSION

The Visible Human Dataset offers many additions to the original goal of a three-dimensional representation of a computer generated anatomical model of the human body and to the general study of human anatomy. In this paper, we have presented a large repository of medical image datasets that offers the opportunity to perform relevant research and their use in various areas.

REFERENCES

[1]http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC

[2]https://wiki.nci.nih.gov/display/CIP/NBIA+at+CBIIT+Image+Collections

[3] http://www.ncrr.nih.gov/

[4] http://mri.kennedykrieger.org/

[5] http://www.imago.uu.nl/

[6] http://www.isi.uu.nl/Portal/

[7] http://rad.usuhs.edu/medpix/parent.php3?mode=default#to

[8] http://www.mypacs.net/repos/mpv3_repo/static/m/About/index.html

[9] http://www.mckesson.com/en_us/McKesson.com/

[10] http://www.cornell.edu/