DDSM Utility v1.0

By
Anmol Sharma
Undergraduate Engineering Student
DAVIET Jalandhar
Punjab, India
Email: anmol.sharma293@gmail.com

Introduction

This is a tutorial for the DDSM Utility v1.0 written by me (Anmol Sharma, Undergrad Engineering Student at DAVIET Jalandhar) to greatly simplify downloading, converting, viewing and extracting annotations from the Digital Database for Screening Mammography (DDSM) database available here: http://marathon.csee.usf.edu/Mammography/Database.html

DDSM is a very famous mammogram database which researchers around the world use to test their mass detection and mass classification algorithms so as to ensure that the results are comparable to other authors or systems. However, an an open source utility or tool doesn't seem to exist to easily and efficiently download the database. The website where the database is hosted has not been updated from the last 15 (almost 16) years, and hence the deprecated and obsolete software that comes with it is essentially, unusable. None of the authors using the database has released code that can process DDSM images, with their research.

The utility is capable of automatically parsing ICS files which contains the LJPEG image sizes to facilitate LJPEG to LJPEG1 decompression, and further conversion into other known formats like PNG. I decided to write this utility when I was looking for ways to download and convert DDSM dataset, only to find out that the images are in a very old LJPEG format. The only utility available on the DDSM website (here, written for very old SunOS 5.6) was to convert it into LJPEG1 (another old, obsolete and unusable format), and that too was highly labor intensive because it required the user to manually run the script for every image he wanted to convert. Even if the user did that, converting LJPEG1 file to PNG was a nightmare, if not impossible.

Another utility available was written by Dr. Chris Rose of University of Manchester (here). It required the user to input names of IMAGE files manually, which then his script used to download the DDSM image in PNG format. This was again, ridiculously labor intensive, and didn't make sense if the user wants to download, say, 1000 images. The source code was not supplied, so making changes to the internal working was not possible. Moreover, if the user wanted to download the annotation files along with the image, he has to repeat the steps all over again. I wasn't getting paid by the hour, so this was downright ridiculous for me.

So I decided to write my own little utility to get the job done. However it was easier said than done. Eventually I figured out that the solution will have to span two operating systems (Linux and Windows), and two languages (C++ and MATLAB). This was due to the fact that I decided to reuse a small jpeg utility supplied by DDSM which only ran on UNIX based platforms (originally SunOS 5.6), and partially due to simplicity of MATLAB in reading, displaying and writing image files as compared to C++.

Contents

My utility contains the following files:

getDDSMNamesAndConvertToLJPEG1.cpp – C++ script to convert LJPEG to LJPEG1 uncompressed format.
openDDSMLJPEG1AndConvertToPNG.m – MATLAB script to open the uncompressed LJPEG1 files and then convert the files into PNG or any other user defined format.
openDDSMPngWithOverlay.m – MATLAB script to open the converted PNG files, read their corresponding OVERLAY files, read their annotations and display the boundary details by superimposing on the mammogram for simple viewing.
readBoundary.m – Supporting MATLAB script to read OVERLAY files (Original author Dr. Jayasree Chakraborty, Memorial Sloan Kettering Cancer Center)
jpeg – UNIX executable file to convert LJPEG to LJEPG1 (original author Dr. Chris Rose, University of Manchester)

Features of the DDSM Utility

-> Easily download multiple case files automatically.

-> Automatically convert LJPEG DDSM images into PNG, JPEG, TIF, GIF and other formats.

-> Read boundary information of each of the converted images and display them.

Initial Requirements

- A Linux based machine, preferably Ubuntu 12.04+, with GCC installed.
- A Windows based machine with MATLAB 8+ installed.
- WinSCP

Tested on:

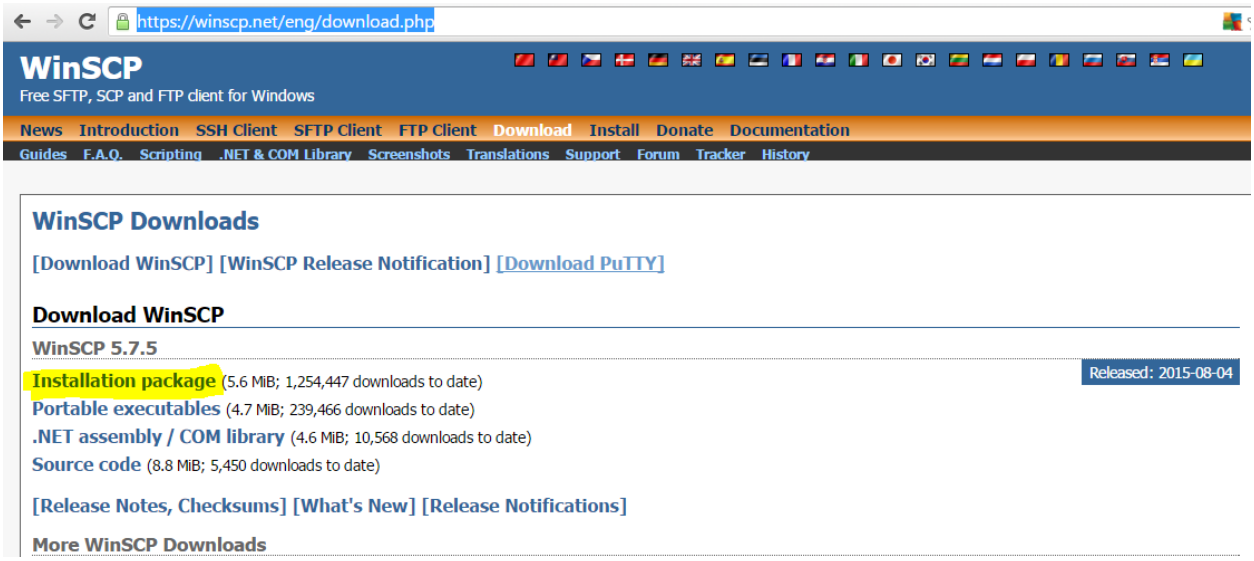
- Dell Inspiron 15 3537
- Intel i5 4200U
- 4GB RAM
- Ubuntu 14.04 with GCC 4.8
- Windows 10 with MATLAB R2015b

Steps

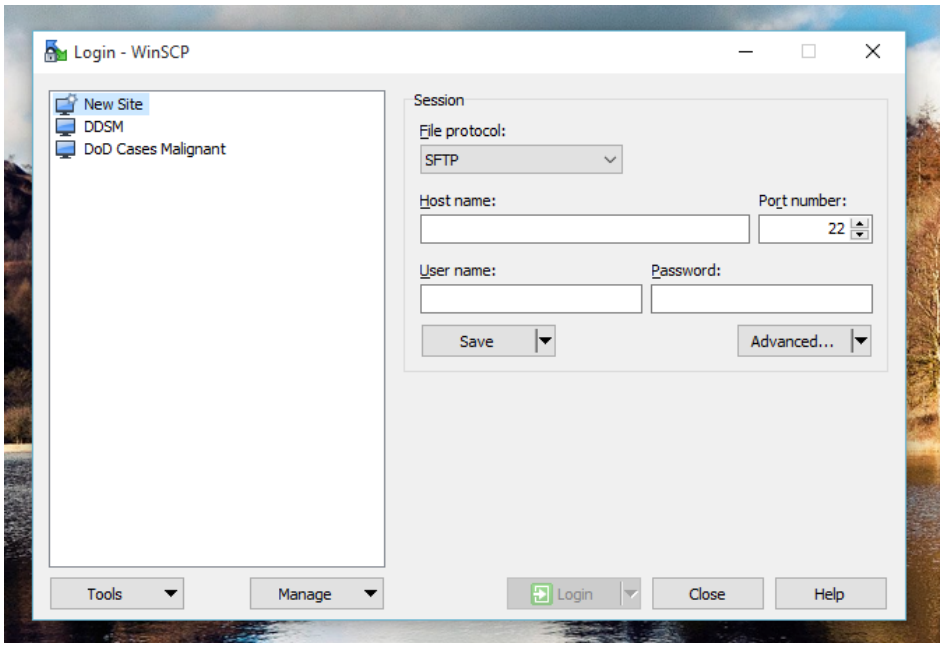
Follow the instructions and click on yellow highlighted regions in the figures.

Download the required DDSM cases using the FTP link from USF.

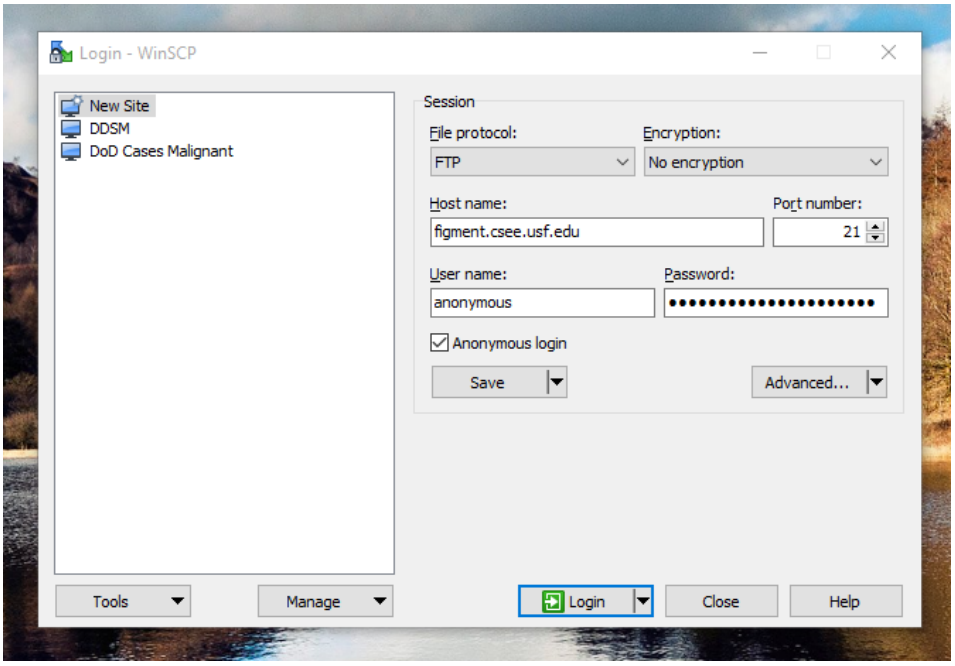
1. Download WinSCP (here).
    a. Download the latest version as shown in figure.



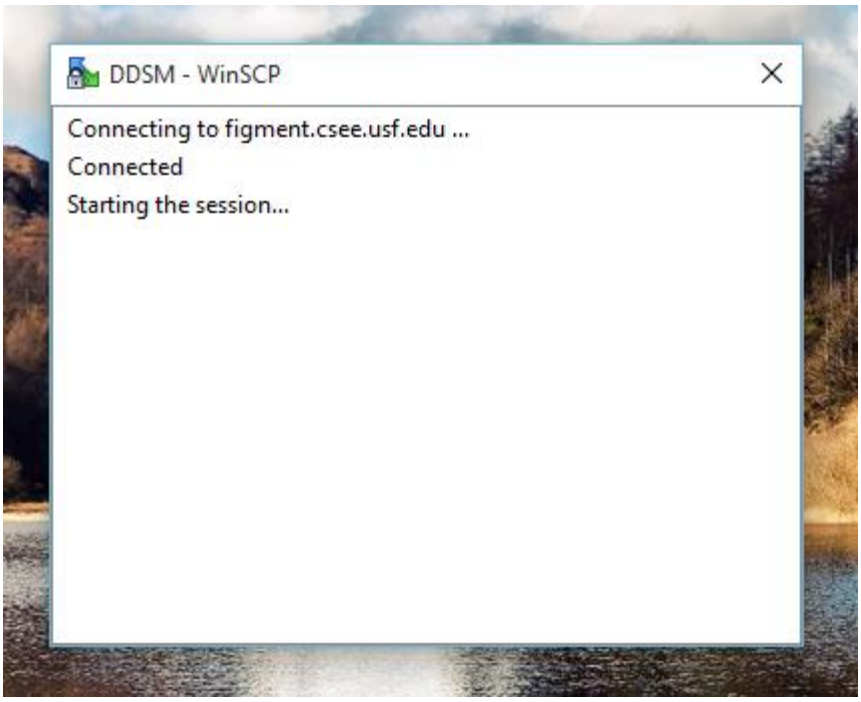
    b. Install the software, and open it.



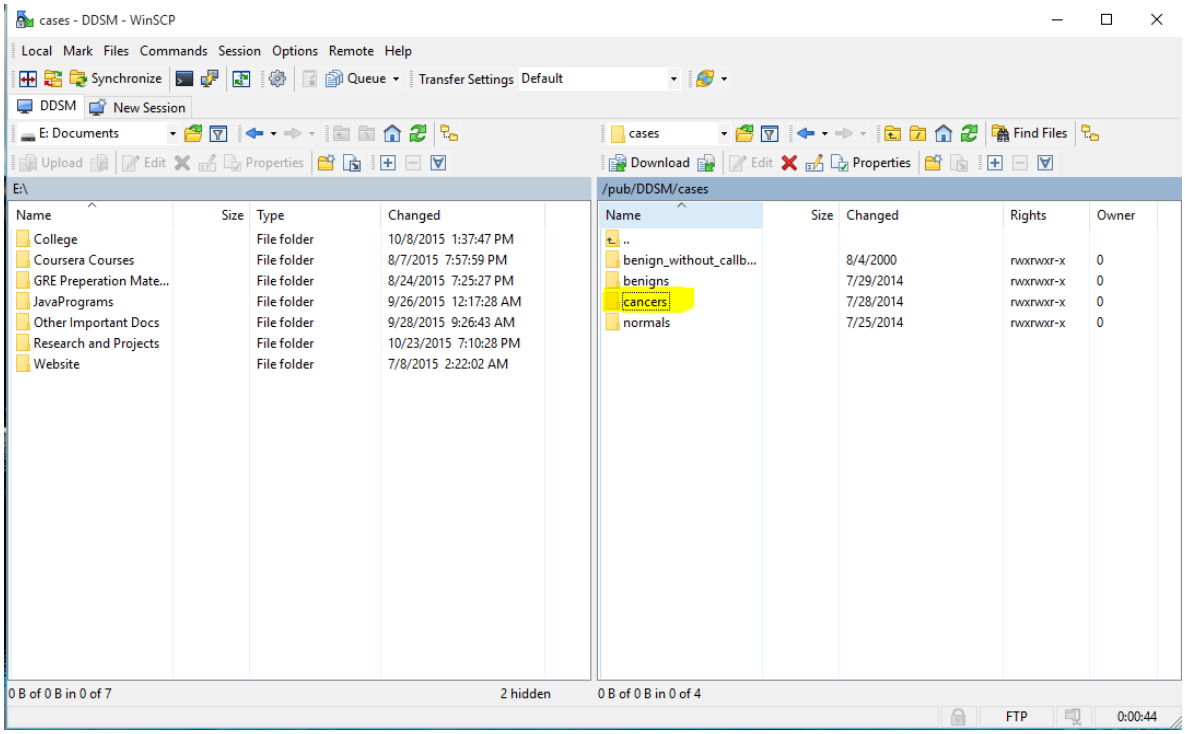
c. WinSCP will ask you for the website and login details for the FTP server. Enter the credentials as shown below.

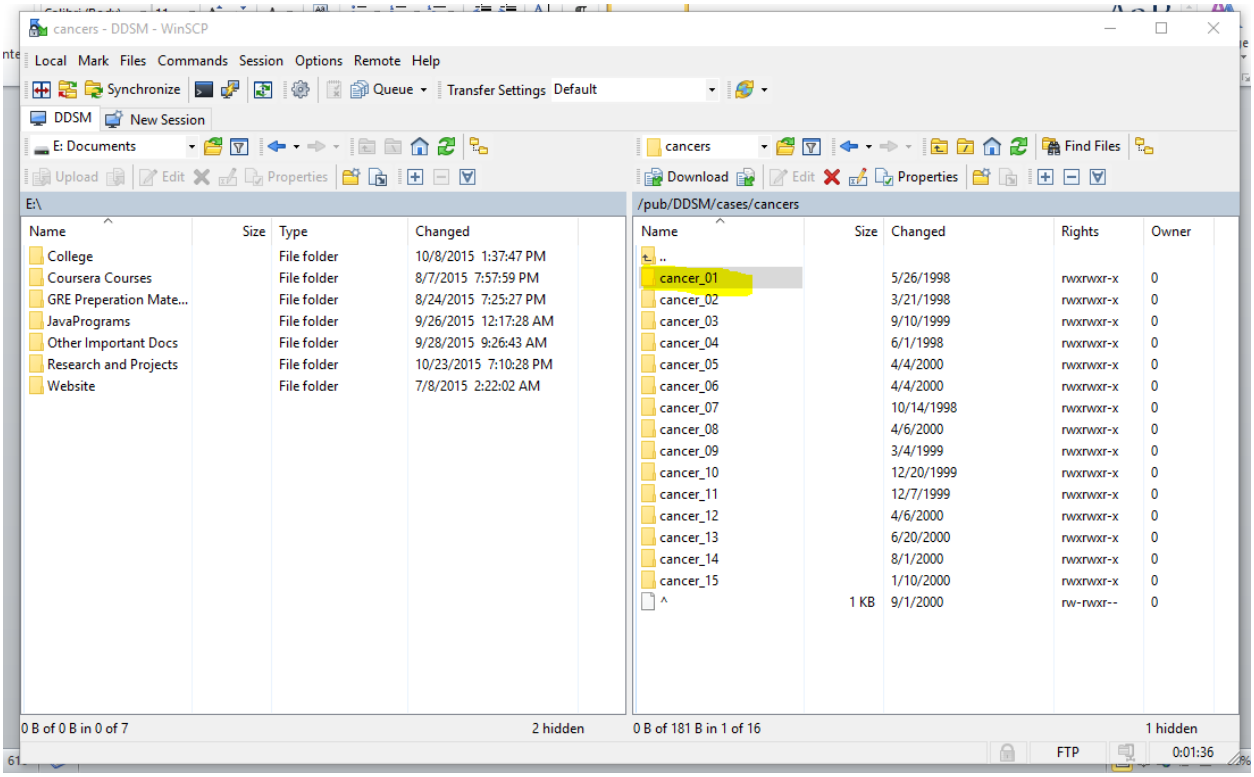


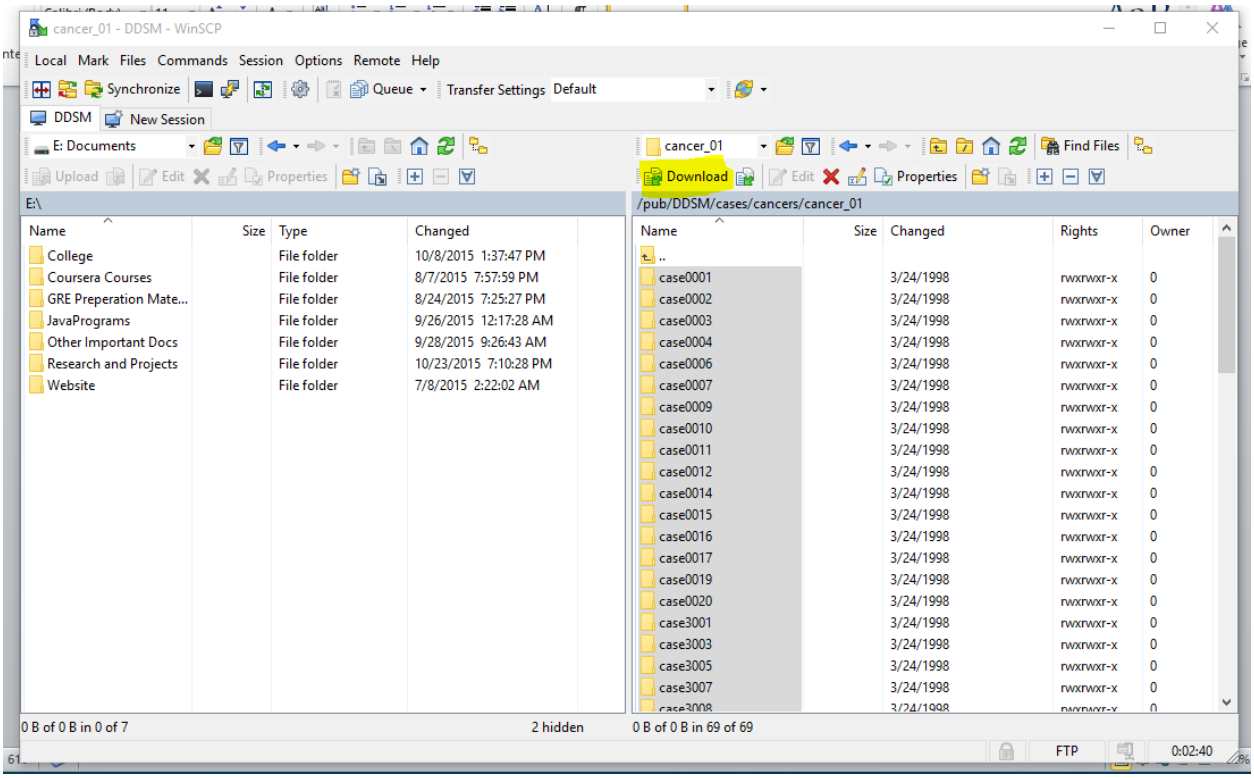
d. WinSCP will connect to the server.



e. Once connected, you will get the following screen. On the right, you have the FTP server parent directory, and on the left you have your own local computer's directory. Now you just need to select the files (cases) you want to download from USF's FTP server by selecting them from the right side, and then click on download button on the right to start the download to the left pane location.


cases - DDSM - WinSCP
Local Mark Files Commands Session Options Remote Help
Synchronize
Queue
Transfer Settings Default
DDSM
New Session
E: Documents
cases
Find Files
Upload
Edit
Properties
Download
E:\
/pub/DDSM/cases
Name
Size
Type
Changed
Rights
Owner
College
Coursera Courses
GRE Preperation Mate...
JavaPrograms
Other Important Docs
Research and Projects
Website
File folder
10/8/2015 1:37:47 PM
8/7/2015 7:57:59 PM
8/24/2015 7:25:27 PM
9/26/2015 12:17:28 AM
9/28/2015 9:26:43 AM
10/23/2015 7:10:28 PM
7/8/2015 2:22:02 AM
benign_without_callb...
benigns
cancers
normals
8/4/2000
7/29/2014
7/28/2014
7/25/2014
rwxrwxr-x
0
0 B of 0 B in 0 of 7
2 hidden
0 B of 0 B in 0 of 4
FTP
0:00:44


cancers - DDSM - WinSCP
Local Mark Files Commands Session Options Remote Help
Synchronize
Queue
Transfer Settings Default
DDSM
New Session
E: Documents
cancers
Find Files
Upload
Edit
Properties
Download
E:\
/pub/DDSM/cases/cancers
Name
Size
Type
Changed
Rights
Owner
College
Coursera Courses
GRE Preperation Mate...
JavaPrograms
Other Important Docs
Research and Projects
Website
File folder
10/8/2015 1:37:47 PM
8/7/2015 7:57:59 PM
8/24/2015 7:25:27 PM
9/26/2015 12:17:28 AM
9/28/2015 9:26:43 AM
10/23/2015 7:10:28 PM
7/8/2015 2:22:02 AM
cancer_01
cancer_02
cancer_03
cancer_04
cancer_05
cancer_06
cancer_07
cancer_08
cancer_09
cancer_10
cancer_11
cancer_12
cancer_13
cancer_14
cancer_15
^
1 KB
5/26/1998
3/21/1998
9/10/1999
6/1/1998
4/4/2000
4/4/2000
10/14/1998
4/6/2000
3/4/1999
12/20/1999
12/7/1999
4/6/2000
6/20/2000
8/1/2000
1/10/2000
9/1/2000
rwxrwxr-x
rw-rwxr--
0
0 B of 0 B in 0 of 7
2 hidden
0 B of 181 B in 1 of 16
1 hidden
FTP
0:01:36

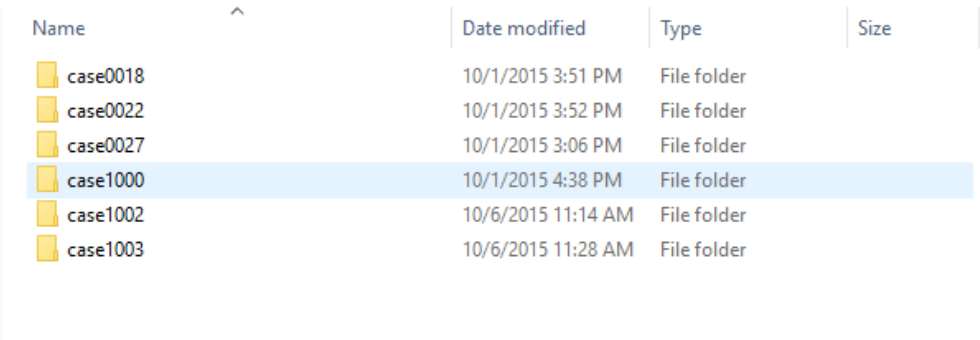


f. I pressed CTRL+A to select all cases, and then pressed Download button as highlighted.

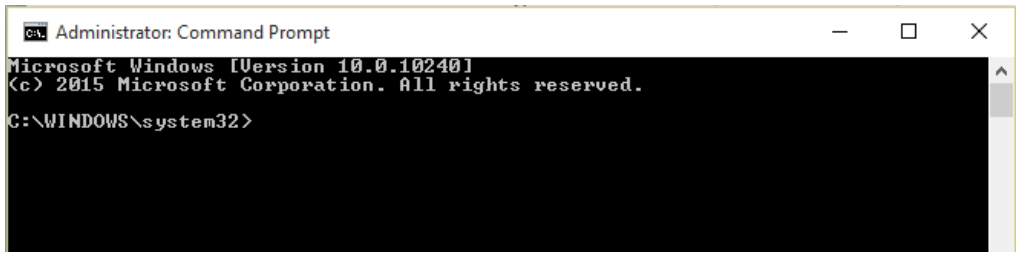
g. The download will begin and you just have to wait.

Copy all files from each of the case folder into one single folder using xcopy.

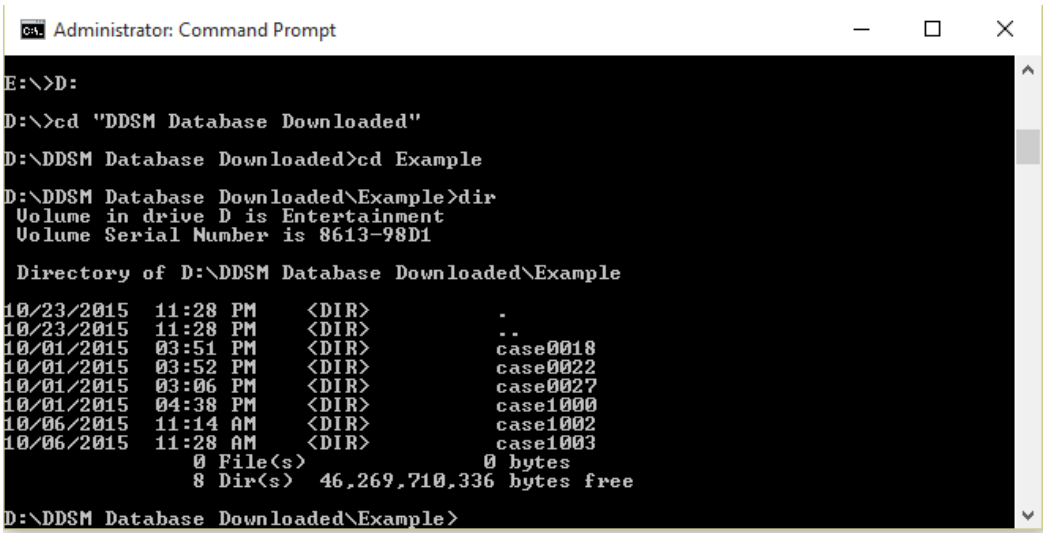
2. Suppose we have the following cases in a directory D:/DDSM Database Downloaded/Example/



a. Open Command Prompt with administrator privileges.



b. Go to your directory where you have your cases folders, using **cd** command.

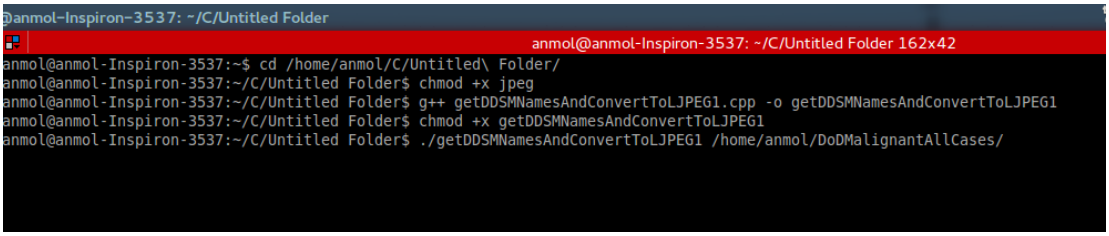


c. Let's create a folder called "All Files" at D:/DDSM Database Downloaded/All Files where we want to copy all the files present in the case folders.
d. Now go to your command prompt, and write the following commands one by one. Please change the source and destination of the command according to your own PC configuration. The format is:
for /R "source" %f in (.extension) do copy "%f" "destination"*
e. I used the following commands to copy all files inside case folder into All Files folder.

```
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.LJPEG) do copy "%f" "D:\DDSM
Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.OVERLAY) do copy "%f" "D:\DDSM
Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.ics) do copy "%f" "D:\DDSM
Database Downloaded\All Files\"
for /R "D:\DDSM Database Downloaded\Example\" %f in (*.16_PGM) do copy "%f" "D:\DDSM
Database Downloaded\All Files\"
```

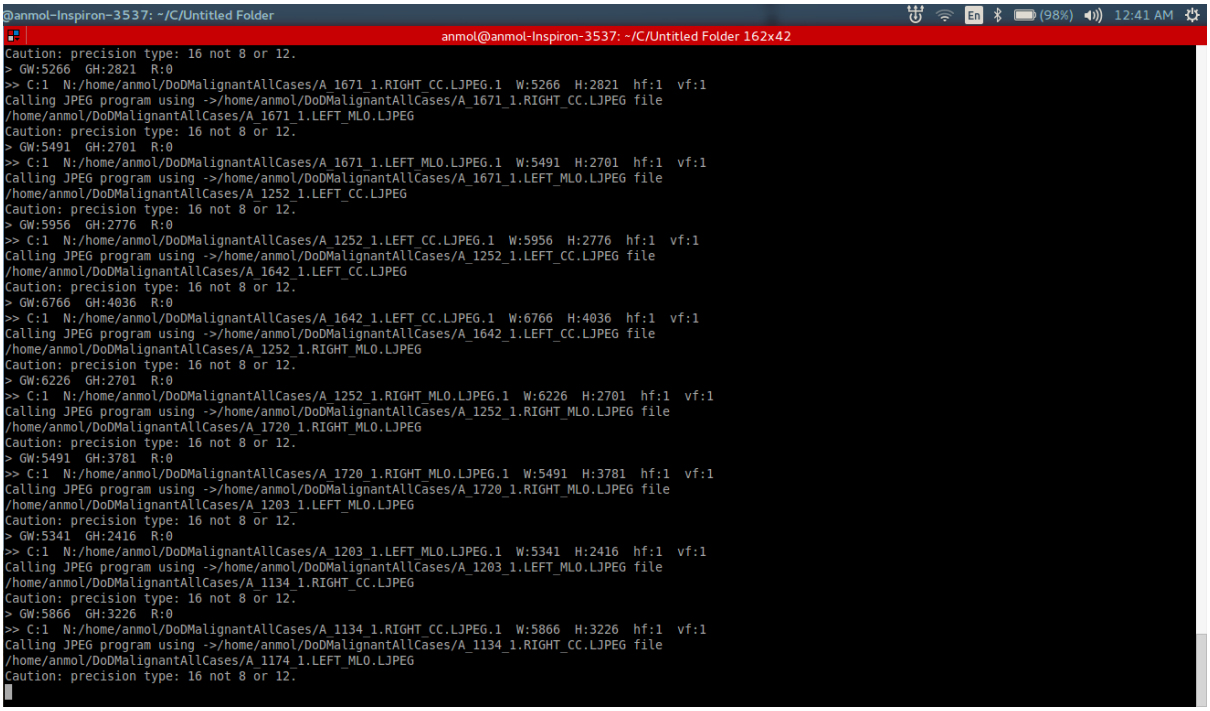
Convert all LJPEG images into LJPEG1 using the supplied C++ utility.

3. Now you need to login to an Ubuntu or Linux based machine. Once logged in, open terminal using CTRL+T, and navigate to the directory containing the **getDDSMNamesAndConvertToLJPEG1.cpp** and **jpeg.**
   In my case, I have my code in
   /home/anmol/C/Untitled Folder/
   So I navigate to that directory using the command:
   cd /home/anmol/C/Untitled Folder/
4. Please do the following steps strictly in order.
   a. Write:
      chmod +x jpeg



   b. Then, to compile the program write the following command:
      g++ getDDSMNamesAndConvertToLJPEG1.cpp –o getDDSMNamesAndConvertToLJPEG1 –std=c++0x
   c. Then write:
      chmod +x getDDSMNamesAndConvertToLJPEG1

   d. After that, run the script by:
      ./openDDSMNamesAndConvertToLJPEG1 <path_to_all_LJPEG_images>
      Example:
      ./openDDSMNamesAndConvertToLJPEG1 /home/anmol/DoDMalignantAllCases/



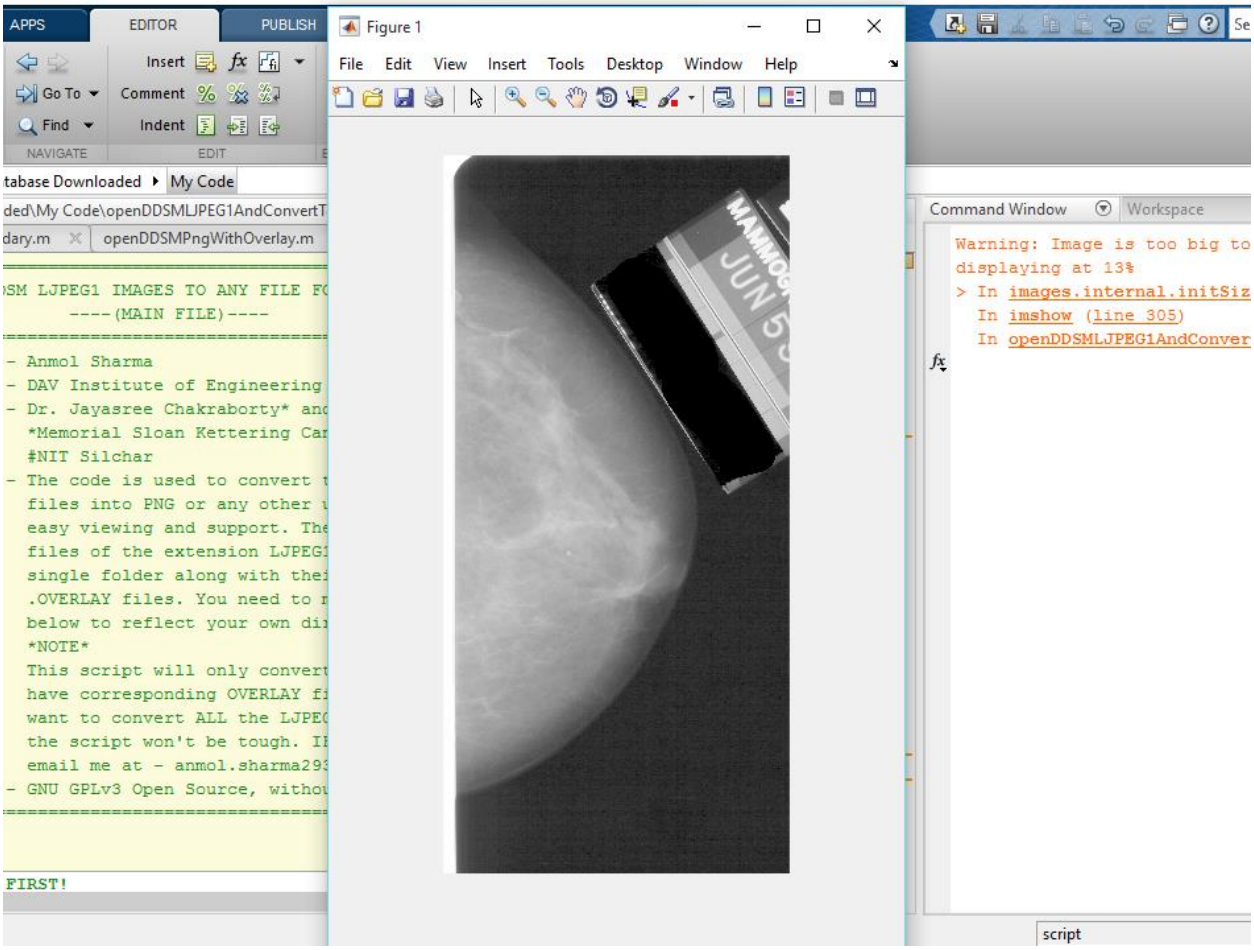
e. Now sit back and relax while the program does its job.
f. The LJPEG1 files will be saved in the same folder as that of LJPEG files.
g. You should copy the folder where you just saved your LJPEG1 files to a hard disk drive, from your Ubuntu home directory. This is because you need to have those files accessible in Windows for your next step, and you can't access Ubuntu's home directory from Windows. Hence, in my case, I copied the
   /home/anmol/DoDMalignantAllCases/
   (The folder which contains ALL files (LJPEG1, LJPEG, OVERLAY, ICS, 16_PGM)
   Folder to my D:/ drive with the same name, and hence the directory is:
   D:\DDSM Database Downloaded\DoDMalignantAllCases
h. The windows directory above will be used in the next step when we convert LJPEG1 to PNG.

Convert LJPEG1 images to PNG using MATLAB script (or JPEG, TIF, GIF, and many other formats).

5. Get back to Windows or if you have MATLAB on Linux, open it. Open MATLAB and open the script **openDDSMLJPEG1AndConvertToPNG.m**

a. Change the script directories to point to your directories where all the files have been copied in the earlier step (Your LJPEG1 directory). Change the following variables according to your directories:
b. `allFilesDirectory = 'D:\DDSM Database Downloaded\DoDMalignantAllCases\';`

c. `writePNGFilesHere = 'D:\DDSM Database Downloaded\DoDMalignantAllCasesPNG\';`
d. `imageOutputFileFormat = '.png'; % Notice the dot. Can be .tif, .jpg...`
e. Run the script and see the mammograms. Press any button to continue conversion process.



Read annotations corresponding to each converted PNG image automatically.

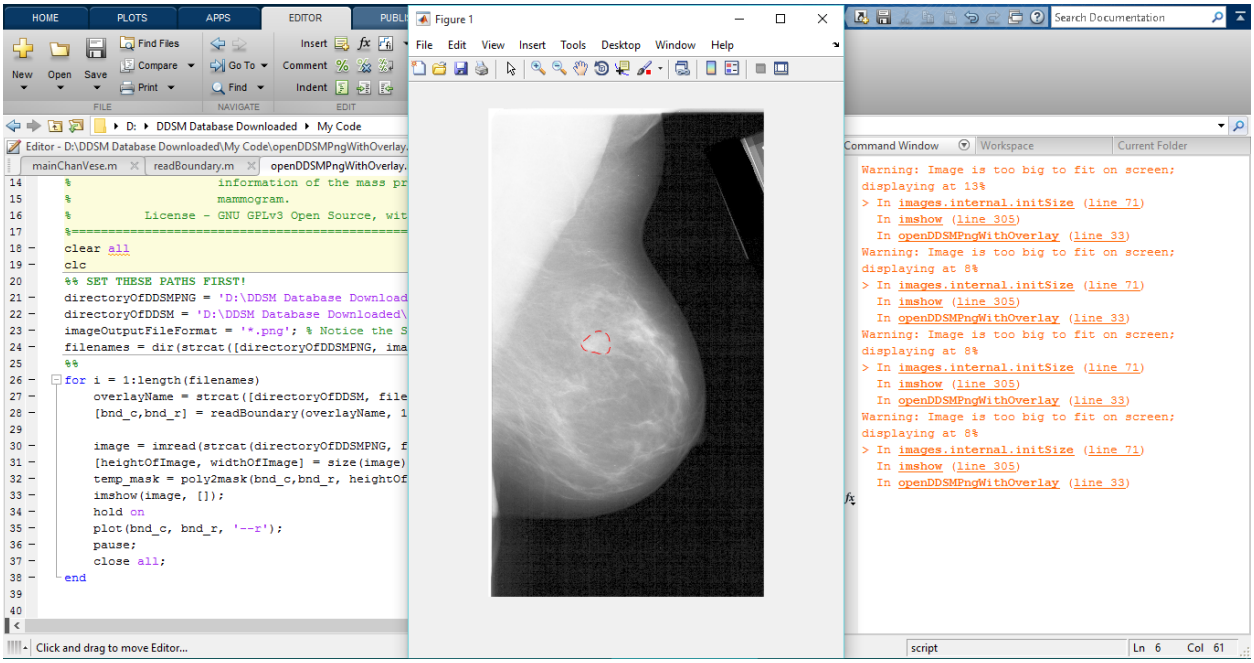
6. Open the **openDDSMPngWithOverlay.m** and change the directories accordingly to point to the newly saved PNG files directory. (Yellow highlighted)
7. Run the script to view the images with their mass boundaries superimposed.

```
%% ====================================================================%
%      OPEN AND VIEW PNG FORMAT DDSM IMAGES WITH THEIR ANNOTATIONS
%                          ----(MAIN FILE)----
% ====================================================================%
%            Author - Anmol Sharma (Undergraduate Engineering Student)
%       Affiliation - DAV Institute of Engineering & Technology
%       Supervisers - Dr. Jayasree Chakraborty* and Dr. Abhishek Midya#
%                     *Research Fellow, Memorial Sloan Kettering Cancer Center
%                     #Assistant Professor, NIT Silchar
%       Description - The code is used to view the PNG format DDSM images
%                     created using the openDDSMLJPEG1AndConvertToPNG.m
%                     script. This script opens the PNG file, and then also
%                     opens the corresponding OVERLAY file to get the boundary
%                     information of the mass present in that particular
%                     mammogram.
%           License - GNU GPLv3 Open Source, without any guarantees.
%====================================================================%
clear all
clc
%% SET THESE PATHS FIRST!
directoryOfDDSMPNG = 'D:\DDSM Database Downloaded\DoDMalignantAllCasesPNG\';
directoryOfDDSM = 'D:\DDSM Database Downloaded\DoDMalignantAllCases\';
imageOutputFileFormat = '*.png'; % Notice the STAR dot. Can be *.tif, *.jpg...
filenames = dir(strcat([directoryOfDDSMPNG, imageOutputFileFormat]));
%%
for i = 1:length(filenames)
    overlayName = strcat([directoryOfDDSM, filenames(i).name((1:(length(filenames(i).name) - 4))), '.OVERLAY']);
    [bnd_c,bnd_r] = readBoundary(overlayName, 1);

    image = imread(strcat(directoryOfDDSMPNG, filenames(i).name));
    [heightOfImage, widthOfImage] = size(image);
    temp_mask = poly2mask(bnd_c,bnd_r, heightOfImage,widthOfImage);
    imshow(image, []);
```