CHAPTER 3
DATA COLLECTION AND METHODOLOGY

CHAPTER 3

DATA COLLECTION AND METHODOLOGY

3.1. INTRODUCTION

According to Robert (1994) there are two sorts of research approaches namely qualitative and quantitative. A qualitative research method is inclusive of its description of the research problem in the form of text, exhaustive interviews with the people oriented towards the problem or the involvement of groups of research problem focused. The second approach called quantitative research method focuses on the thorough going use of statistical data and tools.

In qualitative research method, respondents are given at the most importance to express their views, ideas, and concepts without any hassle to retrieve the data best out of them. The researcher is able to get all the possible data to address the research problem through interviewing the people related to the problem. It is suggested by Yin to adopt the qualitative methods to examine the formerly unfamiliar areas which have not much supporting information to address the problem. The qualitative methods are unstructured due to the involvement of open questions to retrieve the answers. Because of this reason, it will be difficult for the researcher at the time of data analysis and it becomes a laborious task. Moreover, there is a high probability of subjectivity in the views from the respondents, which may be prejudiced, henceforth the results will come up with a higher margin of error.

The second type of research methodology called quantitative research is the methodical empirical examination of apparent phenomena by means of

statistical, mathematical or computational procedures. It is essential to use a huge number of cases demonstrating the population of focus towards the research problem. In the quantitative method, the respondents are selected randomly to overcome the pitfall of subjectivity. It is not like qualitative but uses structured procedures to retrieve and analyse the data. In the qualitative procedure, the data may be in the number format, statistical values, graphs, figures or combinations of these types and may be depicted in table format also. The statistical data retrieved in the quantitative method is represented in the form of tabulations. The research findings are decisive and usually descriptive in nature and they are used to commend a concluding sequence of actions.

3.2. METHODOLOGY

Even though the above two approaches are working well with research in healthcare, it is highly appropriate to use the quantitative method in the case of medical image processing. In reality, the researcher may be anxious to make the image best through any available software in ad hoc manner. But it is mandatory to be a rationale to choose the appropriate algorithm and software for image processing. Especially, in medical image processing, one should not go along with ad hoc attitude while choosing the algorithm.

Since medical image processing is a quantitative field, subjectivity should not be encouraged to avoid any misleading results which may encourage incorrect diagnosis and treatment.

3.3. PROTOCOLS AND WORKFLOWS FOR ACQUISITION AND PROCESSING

It is extremely appreciated to use clear and identical protocol for image acquisition and image processing to reach the optimum solution (Elizabeth, 2007). At the same time, clarity in the workflow will also enable the image processing in an efficient manner. A workflow defines each step of the procedure in processing from the time of image acquisition till arriving the final result and even for the result analysis. The understanding and correlation in the steps make the process flawless and successful. For instance, variations in spatial resolution of MRI from subject to subject, time to time, acquisition to acquisition and mode to mode will surely reflect in the segmentation or classification results.

It is ensured that the brain MRI data acquisition and processing have followed the same protocol and workflow to reach the desired optimum solution.

3.4. EVADING THE SUBJECTIVE IN SELECTING THE FEATURES

There are various features taken into consideration for MRI processing and choices are in the hands of researchers. For example, there will be different options available such as gray scale used for thresholding, size and shape of structuring element in morphology, different filtering techniques for spatial and frequency domain operations, and implementing global or local operations. The options must be chosen based on the rationale justification for the MRI data (Gokturk et al., 2001).

59

Therefore, two key aspects are considered in choosing the operations. Firstly, it is made clear that the chosen parameter pertinent to the spatial and gray scale features is to be improved or detached. It is decided that not to use the values described in the preceding published research papers which may contrast in crucial aspects such as spatial resolution.

Secondly, it is made sure that the choice of parameters is equally valid for all the collected well-defined Brain MR images processed inside the research carried out here.

3.5. DATA INTEGRITY

The first step of data integrity includes maintenance of the original data collected. It is important to take care of the format of the image to avoid any loss of a feature and enable lossless compression to retain the maximum quality.

Compression of MR image data is required because of the following reasons.

a) Increasing storage requirement

b) Transmission of data efficiently

c) Remote analysis of data

d) PACS (Picture archiving and communication systems)

e) Need for usage of cloud services

Because of the above reasons, lossless compression techniques are necessary for archival and transportation of brain MR image data, to decrease both space for storage and bandwidth requirements for fast

transmission of data (Zixiang et al., 2003). There are various compression techniques yielding a good level of compression but they do at the cost of quality reduction.

In the research, since the implementation of the algorithm is done in Apache Hadoop in AWS, images are transmitted over the internet using reversible KLT lossless algorithm proposed by Yodchanan (2008).

3.6. SOURCE OF BRAIN MR IMAGES

There were various hindrances like laws, policies and privacy issues for accessing and using the real-time medical images to do research on them. There are online databases from where the required real-time data for this research has been received. The brain MR image dataset was received from the TCGA Glioma Phenotype Research Group and cancerimagingarchive.net and they are cited in the bibliography.

Image features are commonly categorized into three core tissue types in brain MRI. They are,

1. White matter (WM),

2. Gray matter (GM), and

3. Cerebrospinal fluid (CSF).

To highlight the different aspects of normal and abnormal tissues in the brain, pulse sequences are used. The variations in the parameters such as time between the repetition of procedures (TR) and time consumed of echo (TE) are used to stress changes in anatomical images. For example, to stress on the contrast between gray and white matter, T1-Weighted images is

scanned with short TR and short TE where in the case to stress on the contrast between brain tissue and cerebrospinal fluid, T2-Weighted is scanned with long TR and long TE. The sequence will vary based on the information provide as well as the time consumption to acquire the image.

Figure 3.1. Sample MRI of a glioblastoma tumor in the parietal lobe.

There is one of the most common types of brain tumor called glioma which grows from glial cells. Glial cells are built with energy and nutrients. These glial cells are supporting nerve cells to provide insulation to the central nervous system and most frequently identified in the cerebral hemispheres.

Since the collection of primary data of brain MR images is a difficult task due to privacy and legal policies, the research used the authenticated secondary brain MR image data for analysis. There are 250 brain MR images received from TCGA Data Portal - Clinical and Genomic Data and Imaging Source Site (ISS) Groups collected MRI data from patients

doubted for a brain tumor. The scan has been don with the customized Siemens.

3.6.1. Hardware Description

- Place of MRI scanner: MGH/HST Athinoula A. Martinos Center for Biomedical Imaging (Setsompop et al., 2013).

- Device Used: customized Siemens 3T scanner (a modified 3T Skyra system).

- Data Acquisition method: A 64-channel, tight-fitting brain array coil (Keil et al., 2013).

3.6.2. Subjects

- Gender: Male and Female.

- Demographic Information is available for each image.

- Modalities: CT, MR

- Number of Studies: 224

- Number of Patients: 199

- Images Size (GB): 42.8

3.6.3. Summary of Imaging Protocols

Each dataset (in DICOM format) consists of an MPRAGE scan, a high resolution T2-SPACE scan.

3.6.4. Description of Protocol:

Protocol type used in this data acquisition is noninvasive imaging assessment. Magnetic resonance imaging (MRI) of the brain is a

noninvasive imaging assessment that produces cross-sectional images due to inherent contrast differences of tissues as a result of variable magnetic relaxation properties and magnetic susceptibilities. The American College of Radiology-American Society of Radiology-Society for Pediatric Radiology (ACR-ASNR-SPR) Practice Parameter for the Performance and Interpretation of Magnetic Resonance Imaging (MRI) of the Brain (Amended 2014, Resolution 39) outlines principles for performing high-quality MRI of the brains of adult and pediatric participants.

Table 3.2 Structural scans

Type	T1w	T2w
Description	3D MPRAGE	3D T2-SPACE
TR/TE (ms)	2530/1.15	3200/561
TI (ms)	1100	
Flip Angle	7.0 deg	
FOV (mm)	256x256	224x224
Voxel Size	1mm isotropic	0.7mm isotropic
BW (Hz/Px)	651	744
iPAT	2	2
Acquisition Time (min:sec)	6:02	6:48