Software Requirements Specification

For

Breast masses Segmentation and Classification from Mammograms

15/04/2023

Prepared by

Specialization	SAP ID	Name
B.Tech-CSE (CC&VT)	500083359	Chinmayi Parimal (B5-NH)

Department of Systemics
School of Computer Science
UNIVERSITY OF PETROLEUM & ENERGY STUDIES,
DEHRADUN- 248007. Uttarakhand

Table of Contents

1. INTRODUCTION

Breast Cancer is caused when the cells in the breast grow and divide in an uncontrolled way, creating a mass of tissue called tumor [1]. Symptoms of breast cancer can include, feeling a lump in the breast, experiencing a change in the size of breast and noticeable changes to the skin on the breast [2]. **Mammograms** can help with an early detection of breast cancer. The cancer originates and stems from the breast tissue, this process is called as **Metastasis**. It occurs when breast cells mutate (change) and grow out of control, creating a mass of tissue (tumor) [3]. Like other cancers, breast cancer can invade and grow into the tissues the surround the breast.

Breast cancer is one of the most common cancers among women, second only the skin cancer. It is most likely to affect women over the age of 50. Though rare, men can also develop this cancer. Approximately 2,600 men develop male breast cancer every year in the United States, making up less that 1% of all cases [4]. Transgender women are more likely to develop breast cancer compared to cisgender men. Additionally, transgender men are less likely to develop breast cancer compared to cisgender women.

1.1 Purpose of the Project

In this project, **Breast Masses Segmentation** is used to automate the breast cancer detection, as manual detection is an extremely time-consuming process. The main aim is to diagnose whether breast cancer of any class (Malign, Benign) is present in the mammograms or not. The proposed approach uses the traditional Deep Learning workflow including data preprocessing, segmentation of masses, testing and analysis. The required modifications will be performed on the mammograms using CLAHE (Contrast Limited Adaptive Histogram Equalization) for the enhancement of the dataset. The proposed model aims to improve the medical expertise where manually detecting breast cancer is still a very difficult task with this significant evolution in technology.

1.2 Target Beneficiaries

The target beneficiaries of the proposed model are oncologists from the healthcare sector. They use physical examinations- Checking for lumps, hardness, or tenderness in the breasts and lymph nodes. The following process can be inaccurate and extremely time-consuming. The proposed model increases the accuracy of the detection of breast masses which can be further used by oncologist to minimize the probability of false diagnosis.

1.3 Project Scope

The scope of the project is to improve the preprocessing techniques using CLAHE (Contrast Limited Adaptive Histogram Equalization), it works on the smaller areas of an image called Tiles rather than the complete image which results in rectifying the problem of noise enhancement. Moreover, for segmentation the model consists of Dual Cornet Algorithm as it mammographically achieves the best segmentation and classification simultaneously, outperforming recent state-of-the-art models.

1.4 References

[1] Gunjan Chugh, Shailendra Kumar, Nanhay Singh. (2020).
Survey on Machine Learning and Deep Learning Applications in Breast Cancer Diagnosis.
LLC part of Springer Nature 2021.

[2] Heyi Li, Dongdong Chen, William H. Nailon, Mike E. Davies. (2020).
Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography.
IEEE, and David Laurenson.

[3] Yong Joon Suh, Jaewon Jung, Bum-Joo Cho. (2020).
Automated Breast Cancer Detection in Digital Mammograms of Various Densities via Deep Learning.
Department of Breast and Endocrine Surgery, Hallym University Sacred Heart Hospital, Anyang 14068, Korea; Medical Artificial Intelligence Center, Hallym University Medical Center, Anyang 14068, Korea; Department of Ophthalmology, Hallym University Sacred Heart Hospital, Anyang 14068, Korea.

[4] Lian Zou, Shaode Yu, Tiebao Meng, Zhicheng Zhang, Xiakun Liang, Yaokin Xie. (2019). *A Technical Review of Convulational Neural Network-Based Mammographic Breast Cancer Diagnosis.*
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen, China. Cancer Center of Sichuan Provincial People's Hospital, Chengdu, China. Department of Radiation Oncology, 'e University of Texas Southwestern Medical Center, Dallas, TX, USA. Department of Medical Imaging, Sun Yat-sen University Cancer Center, Guangzhou, China. Medical Physics Division in the Department of Radiation Oncology, Stanford University, Palo Alto, CA, USA.

2.PROJECT DESCRIPTION

The overall design of the breast masses segmentation consists of three component:

The figure describes the overall steps involved in the proposed model.

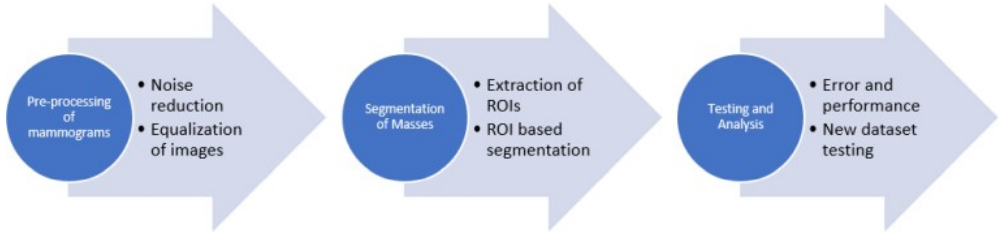

Fig.1 Standard Deep Learning model workflow

- Pre-processing of mammograms- Preprocessing data is a common first step in the deep learning workflow to prepare raw data in a format that the network can accept. For example, you can resize image input to match the size of an image input layer. You can also preprocess data to enhance desired features or reduce artifacts that can bias the network.

- Segmentation of Masses- Mass segmentation in the mammograms is a necessary and challenging task in the computer-aided diagnosis of breast cancer. Most of the existing methods tend to segment the mass by manually or automatically extracting mass-centered image patches.

- Test and analysis- The following step will handle the accuracy of the model and give the overall analysis of the algorithm and its working.

- Thread APIs are collections of programming interfaces for managing threads offered by the operating system or programming language. Threads are little processes that run independently within a programme and share memory. User threads and kernel threads are two types of threads. The programme manages user threads, while the operating system manages kernel threads..

- Thread synchronisation is the process of coordinating thread execution such that they do not interfere with one another. Various synchronisation methods, including as mutexes, semaphores, and condition variables, are used to accomplish this. Mutexes protect important areas of code, whereas semaphores limit access to shared resources. Changes in the state of a shared resource are

signalled by condition variables..

- Thread communication techniques are used to ease data sharing and control amongst threads. Message passing, shared memory, and pipes are examples of communication techniques. Message passing includes sending messages between threads, whereas shared memory requires threads sharing a portion of memory. Pipes are used to transport data from one process to another.

- Kubernetes is a container orchestration tool for deploying, managing, and scaling containerized applications. In the context of this project, Kubernetes can be used to manage and scale the application's many components, including the numerous microservices and APIs. Docker containerization allows the programme to be quickly deployed and executed on any platform that supports Docker. Kubernetes can then be used to manage container deployment and scalability. For example, based on the workload, Kubernetes may automatically scale the number of containers running the microservice that does picture segmentation.

- Kubernetes can also be used to manage communication between the application's various microservices and APIs. Kubernetes, for example, can employ service discovery to route requests from one microservice to another based on their addresses and availability. Furthermore, Kubernetes includes capabilities like automatic failover, self-healing, and load balancing that can increase the application's stability and availability. In the event of a failure or outage, Kubernetes can automatically restart or move failed containers to other nodes.

- MPI (Message Passing Interface) can be be used for parallel processing in this project, especially when working with large datasets. MPI allows several processes to communicate with one another via messages, allowing for task parallelization. When working with massive datasets, the usage of MPI can dramatically increase application performance.

2.1 Reference Algorithm

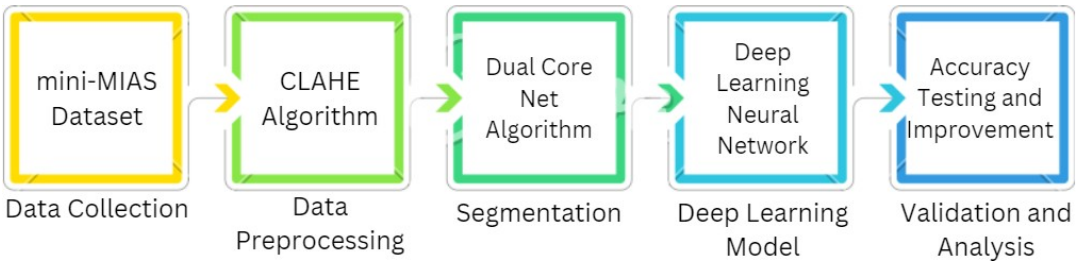

Fig.2 Proposed Model Workflow

- Data Collection - The proposed model works upon mini MIAS dataset which is publicly available on the platform named Kaggle. The dataset consists of anonymously distributes mammograms of three breast cancer classes, benign,

malign and Normal. Moreover, the model works upon another dataset which was also available publicly consists of ROIs (Region of interest) of MIAS dataset.

- Data Pre-Processing – In this project, enhancement of the mammograms is performed by using CLAHE (Contrast Limited Adaptive Histogram Equalization) which helps in the edges enhancement and improves the overall apparent sharpness. After which erosion is applied on the enhanced image to remove all the small anomalies and removes pixels on object boundaries.

- Segmentation – For this step the model uses Dual CoreNet Algorithm (Dual Convolutional Neural Network) which decouples the differentiation of benign and Malignant classes into dual problem: Segmentation and Classification. In the classification task, each input ROI samples (with surrounding tissues) will be classified into cancer category or not; whereas in the segmentation task, each pixel is labeled as either 0 or 1 so that mass pixels can be accurately identified within the tight bounding box ROI.

- Deep Learning Neural Network Model – It will help the model to detect the main problem i.e., breast cancer from the processed data through some hidden layers performing major functions by which the model can achieve high accuracy.

- Validation and Analysis – This step contains the conclusion, results and observations of the model. It will help in the comparison of the proposed model with previously used model and highlights the points were those models are lacking. The major point of this step is to mention the improvements which are carried by the technique.

2.2 Characteristic of Data

The Mammographic Image Analysis Society (MIAS) is an organization of UK research groups interested in the understanding of mammograms and has generated a database of digital mammograms. Films taken from the UK National Breast Screening Program have been digitized to 50micron pixel edge with a Joyce-Loebl scanning microdensitometer, a device linear in the optical density range 0-3.2 and representing each pixel with an 8-bit word. The database contains 322 digitized films and is available on 2.3GB 8mm (ExaByte) tape. It also includes radiologist's "truth"-markings on the locations of any abnormalities that may be present. The database has been reduced to a 200micron pixel edge and padded/clipped so that all the images are 1024x1024. Mammographic images are available via the Pilot European Image Processing Archive (PEIPA) at the University of Essex. The Dataset is extracted from a secondary named Kaggle.

2.3 SWOT Analysis

- Strength: To perform image segmentation over breast masses using supervise learning, we are performing Deep Learning techniques over mammographic patches of breast implemented in python programming language using SkLearn project. The

strength is derived by the fact that a comparatively less work has done in the respective field using DL. Also, on proper expected completion, the model will be more efficient and time consuming.

- Weakness: During the training phase, our general low-end machines will take very long time to train the model. Till the time model isn't ready to give optimum output, the DL model is comparatively less trustworthy.

- Threat: The model is planned and made for the assistance of the specialist/doctors and could not be directly used by the general public for the respective treatments.

- Opportunity: This could turn out to be a big step in the field of health sciences, as all the previous work is basically focused over Machine Learning, and take a lot 8k manual and hectic work and that's the reason this could turn out to be a relief.

2.4 Project Features

❖ In this model enhancement of the mammograms is performed by using CLAHE (Contrast Limited Adaptive Histogram Equalization) which is one of the best technique for the edges enhancement and improves the overall apparent sharpness.

❖ It uses Dual CoreNet Algorithm (Dual Convolutional Neural Network) which decouples the differentiation of benign and Malignant classes into dual problem: Segmentation and Classification.

2.5 User classes and characteristic

User classes the proposed model are oncologists from the healthcare sector. They use physical examinations- Checking for lumps, hardness, or tenderness in the breasts and lymph nodes. The following process can be inaccurate and extremely time-consuming. The proposed model increases the accuracy of the detection of breast masses which can be further used by oncologist to minimize the probability of false diagnosis.

2.6 Assumption and Dependencies

❖ Major assumption is regarding the dataset that it contains three different classes benign, malign and normal. As the dataset contains anonymous unlabeled mammograms

2.7 Challenges faced

Some of the common challenges that may be faced during deployment of the project are:

1. AWS resource configuration: One of the difficulties in launching an image processing project on AWS is setting the appropriate AWS resources, such as EC2 instances, S3 buckets, and IAM roles. Setting up these resources can be time-consuming and necessitates AWS experience.

2. Image processing projects can be resource-intensive, and as the demand grows, scaling and load balancing become critical. To ensure best performance, setting up auto-scaling groups

and load balancers requires experience and careful planning.

3. Data storage and retrieval: It can be difficult to store and retrieve huge amounts of image data on AWS. A proper storage architecture must be established to ensure that data is conveniently accessible and retrievable.

4. Security and access control: Because image processing projects may necessitate access to sensitive data, security and access control are critical. It can be difficult to configure the requisite security groups, network access control lists, and IAM roles.

5. Interaction with additional services: Image processing projects may necessitate interaction with additional AWS services such as Lambda, SNS, and SQS. Configuring these integrations necessitates careful preparation and knowledge.

3. SYSTEM REQUIREMENTS

3.1 User Interface

- MIAS dataset is used to train the model.
- Libraries used-
    1. OpenCV

2. Numpy
3. Matplotlib
4. OS
5. Tensorflow
6. Torch
7. Keras
8. Torchvision.transforms
9. Dataset

3.2 Software Interface

- The software used is VS Code to implement the code.

3.3 Database Interface
No use of database.

3.4 Protocols
No protocols have been used in our proposed model.

4. NON- FUNCTIONAL REQUIREMENTS

4.1 Performance Requirements

- The proposed model should be able to diagnose each and every kind of image, the dataset is containing.
- The proposed model should have high accuracy.
- The proposed technique should be better than the existing models

4.2 Security requirements.

- There is no specific security requirement for our proposed model

4.3 Software Quality Attributes

- The output of the proposed model should not be deceptive, as it can cause a drawback in the treatment and can also lead to various severe circumstances (including death).

5. OTHER REQUIREMENTS

- There are no other requirements for our proposed model.

6. MAINTENANCE AND SUPPORT

6.1 Bug Fixing

Bug fixing will be an important part of the maintenance and support strategy. To reduce the possibility of defects, the application will be rigorously tested before launch. However, if any defects are discovered after deployment, they will be addressed as soon as possible. The following steps will be taken during the bug fixing process:

1. Finding the bug.
2. Replicating the bug.
3. Investigating the bug to determine its origin.
4. Resolving the bug.
5. Test the repair to confirm that no new bugs are introduced.
6. Implementing the patch in the production environment.

6.2 Feature Updates

To stay up with changes in technology and user needs, the breast mass segmentation and classification application will need to be updated. The following steps will be taken throughout the feature update process:

1. Recognising the new feature or enhancement.
2. Assessing the new feature's or improvement's feasibility and impact.
3. Creating a plan for putting the new feature or enhancement into action.
4. Creating and testing the new feature or enhancement.
5. Implementing the new feature or enhancement in the production environment.

6.3 Technical Support

Users of the breast mass segmentation and classification programme will receive technical support to help them handle any problems they may experience. Technical help will be available via email, phone, and online chat. The following actions will be taken during the technical support process:

1. Receiving and recognising the user's request for technical assistance.
2. Gathering information about the problem.
3. Analysing the problem to determine its root cause.
4. Providing a solution or workaround to the user.
5. Documenting the problem and solution for future reference.

APPENDIX A: GLOSSARY

- Pre-processing of mammograms- Preprocessing data is a common first step in the deep learning workflow to prepare raw data in a format that the network can accept. For example, you can resize image input to match the size of an image input layer. You can also preprocess data to enhance desired features or reduce artifacts that can bias the network.

- Segmentation of Masses- Mass segmentation in the mammograms is a necessary and challenging task in the computer-aided diagnosis of breast cancer. Most of the existing methods tend to segment the mass by manually or automatically extracting mass-centered image patches.

- Test and analysis- The following step will handle the accuracy of the model and give the overall analysis of the algorithm and its working.