A Minor Project Synopsis on

**CANCER DETECTION AND CLASSIFICATION USING CNN**

Submitted to Manipal University, Jaipur

Towards the partial fulfillment for the Award of the Degree of

**BACHELORS OF TECHNOLOGY**

In Information Technology

2019-2020

By

Aniruddh Yadav

Avanish Singhal

179302021

179302037



Under the guidance of

Dr. Pratishtha Mathur

**Department of Information Technology**

**School of Computing and Information Technology**

**Manipal University Jaipur**

**Jaipur, Rajasthan**

**Introduction**

One of the common disease plaguing the world has been the prevalence of cancer and lacking of methods for quick detection and a proper cure for it. Using advances in the modern world particularly in the domain of technology and the computer age, the usage and adoption of these methods is needed in the modern medical era.

Machine Learning algorithms are now being used for processing medical imagery and pathological tools. Manual detection of a cancer cell is a tiresome task and also involves human error, and hence computer aided mechanisms are applied to obtain better results as compared to manual pathological detection systems.

The project will aim to solve the common problems faced by the community in the order of significantly reducing the time taken for such a diagnosis. Another advantage of this project is the continuous improvement by the machine itself to give more accurate results over time.

Deep learning is extensively utilized in the medical imaging field as it does not require prior expertise in a related field and also reduces costs. Additionally, the machines itself find new patterns in the data which had earlier gone undetected.

The utilization of data science and machine learning approaches in medical fields proves to be prolific as such approaches may be considered of great assistance in the decision making process of medical practitioners.

**Motivation**

A common problem facing the medical community is the early detection of cancer. Whereas a same medical practitioner if diagnoses cancer in 12hrs, a machine well versed in the diagnosis can do it in a matter of seconds.

Suppose, let us consider one example of detecting breast cancer in woman’s. The tumors are classified in two types based on its characteristics and cell level behavior: benign and malignant. Person detected with a malignant tumor, it is recommended to undergo treatment to cure those cancerous cells. With the advent of machine learning techniques, specifically in the direction of deep neural networks that can learn from the images labeled with the type that each image represents, it is now possible to recognize one type of tumor from another based on its ultrasonic image automatically with high accuracy.

If the doctor misclassifies the tumor as benign instead of malignant, while in the reality the tumor is malignant and chooses not to recommend patient to undergo treatment, then there is a huge risk of the cells metastasizing in to larger form or spread to other body parts over time. This can lead to a life threatening situation for the patient. In the statistical terminology, this would be considered as the doctor making ‘Type 1’ error, where the patient has malignant tumor, yet she is not identified as having it.

**PROS CONS**

1) it will result in faster computational time as compared to a human.

2)The continuous evaluation of results will improve the accuracy to detect cancer.

3)It will open up a new pathway in the medical industry.

4)It will result in significant reduction of human error in the detection of cancer.

5)It will result in a significant reduction of cost in detection of cancer.

1)Without a medical professional, the reliance of the computed data cannot be cross validated.

2)Without a proper dataset, the machine will face the difficulties in analyzing patterns and giving fairly accurate results.

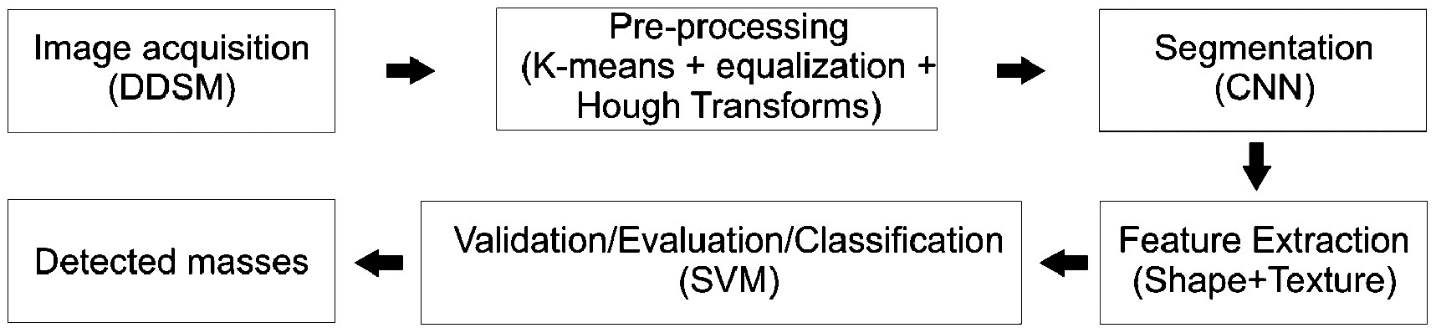
3)The software and hardware requirements of the system is high. So a normal personal computer may face difficulties.

4)If in a case, there are inaccuracies in the classification algorithm, it may result in inaccuracies and lossy data.

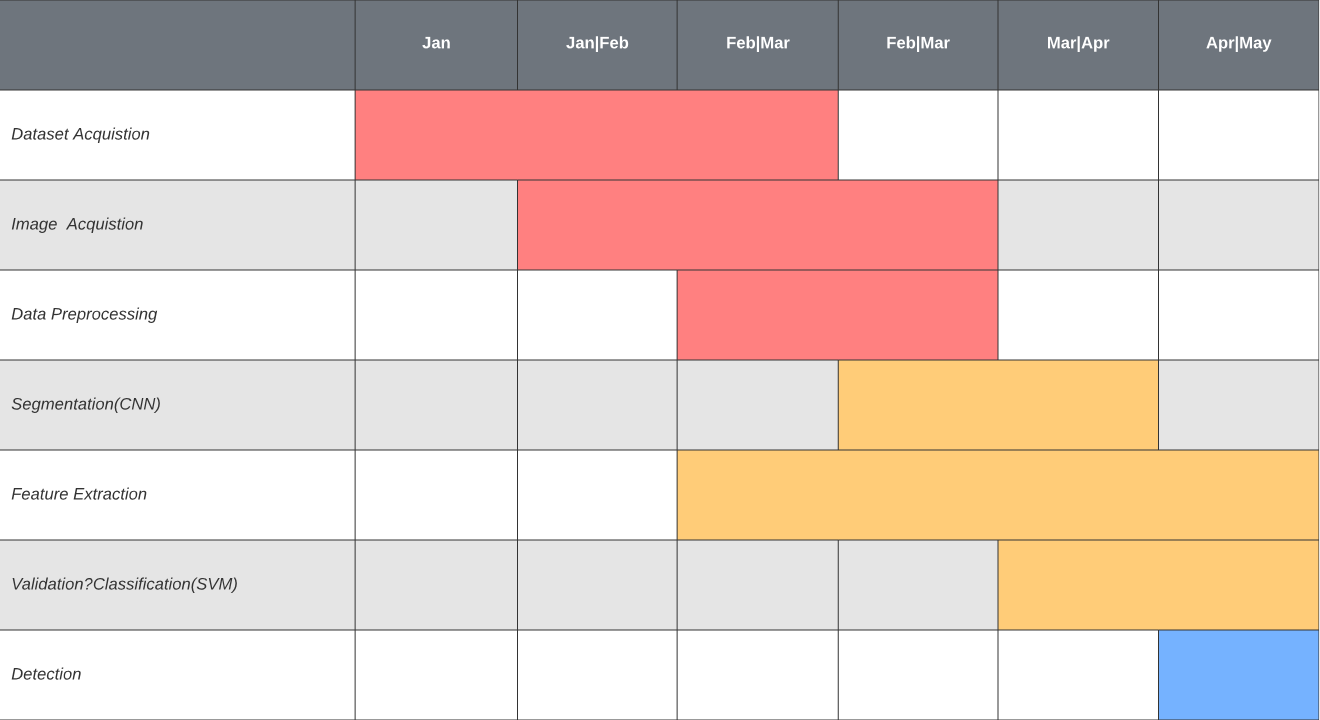
**METHODOLOGY**

Google Tensor Flow will be used to implement the machine learning algorithms in this study, with the aid of other scientific computing libraries: matplotlib, numpy, and scikit-learn. The machine learning algorithms were trained to detect cancer in the initial stages using the Kaggle dataset.

Another analysis of the different datasets for different detection of cancers is on the Broad Institute.  This methodology comprises the following stages: acquisition, pre-processing, segmentation of regions of interest, feature extraction, feature selection, and classification of the regions of interest as mass or non-mass.



**Gantt Chart:**



**Facilities Required for Project**

**HARDWARE**

* Intel Core(TM) i5-6300HQ
* CPU @ 3.40GHz x 4
* 8GB of DDR4 RAM
* NVIDIA GeForce GTX 960M 4GB DDR5

**SOFTWARE**

* Python 3.7
* Anaconda Navigator
* Spyder
* TensorFlow
* Keras
* Cuda 10.0
* CUDNN 7.0
* Python Libraries (NumPy, Pandas, Matplotlib, Scipy)

**Bibliography**

<https://ieeexplore.ieee.org/abstract/document/8389034>

<https://ieeexplore.ieee.org/document/7727519>

<https://towardsdatascience.com/how-to-fight-breast-cancer-with-deep-learning-ab28e42e4250>

<https://www.sciencedirect.com/science/article/pii/S0010482511001132>

<https://www.kaggle.com/paultimothymooney/breast-histopathology-images>