GUJARAT TECHNOLOGICAL UNIVERSITY

CERVICAL CANCER DETECTION USING MACHIINE LEARNING AND RASPBERRY PI

A Detailed Report to be submitted
For
Project – II (2181005) Semester VIII
In
Bachelor of Engineering (EC)

Guided By

Dr. Falgun Thakkar

Associate Professor,
GCET, Vallabh Vidyanagar

By

Dhruvin Patel	**140110111033**
Harshit Patel	**140110111034**
Nirav Patel	**140110111041**

Team Id: 27987

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

G H PATEL COLLEGE OF ENGINEERING & TECHNOLOGY

BAKROL ROAD, VALLABH VIDYANAGAR – 388 120

May – 2018

Certificate

Date:__/__/2018

This is to certify that the dissertation entitled ***"Cervical Cancer Detection using Machine Learning & Raspberry Pi"*** has been carried out by ***Dhurvin Patel (140110111033), Harshit Patel (140110111034) & Nirav Patel (140110111041)*** under my guidance in the fulfilment of the degree of ***Bachelor of Engineering*** in ***Electronics & Communication (8th Semester)*** of Gujarat Technological University, Ahmedabad during academic year 2017-18.

Internal Guide
Dr. Falgun Thakkar

Head of the Department
Dr. Hitesh Shah

ACKNOWLEDGEMENT

We would like to thank our erudite and experienced guide Dr. Falgun Thakkar of Department of Electronics & Communications Engineering who was not just guide for our project but also our career. We would like to thank to G. H. Patel College of Engineering & Technology for providing us with the required resources and made this project possible.

We would like to thanks Dr. Nirav Asarwala (Oncologist) and Dr. Sadhana Saraiya (Pathologist) of Shri Krishna Medical Hospital, Karamsad for providing us their expert guidance which was extremely essential for the project.

Furthermore, we would like to thank the course coordinator Prof. Samir Trapasiya, Head of the Department Dr. Hitesh Shah and all the faculty members of Department of Electronics & Communications as well as non-teaching staff for providing us support and knowledge require to complete this project.

Lastly, we would like to express my heedful gratitude to our god, our parents and family members for their endless support.

Date: **Dhruvin Patel**

GCET, Vallabh Vidyanagar **Harshit Patel**

Nirav Patel

ABSTRACT

Today cancer is one of the most prevailing disease when it comes to case of incidence as well as death. In women the second most prevailing cancer after breast cancer is the cervical cancer. Our project aims to provide a comprehensive solution Detection of Cervical Cancer by means of software as well as hardware. As this cancer is curable if detected in early stages such device would help lowering the incidence and morbidity rates due to cervical cancer in impoverished areas. This would also help health providers in providing faster and more accurate treatment to the women in those impoverished areas.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

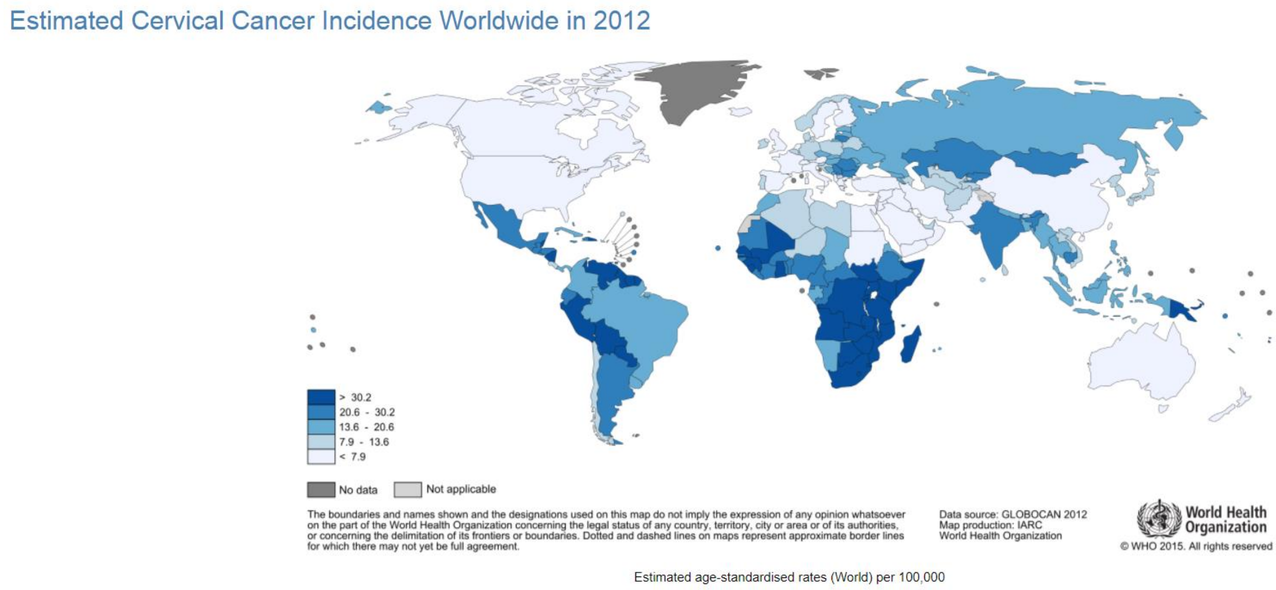
Today the world is growing very rapidly and with this rapidly growing world new deadly diseases are also growing day by day. Cancer being one of the most prevailing and deadly disease in the world is killing thousands of humans and animals every year. Especially in low resource area i.e. economically and socially backward areas like Africa, South America and some parts of Asia cancer the incidence and morbidity rates are really high.

In males prostate cancer is most prevailing when it comes to incidence and morbidity rate. When it comes to female breast cancer is most prevailing while cervical cancer is the second most prevailing type of cancer. This project aims to provide a comprehensive solution to health providers and the direct users i.e. the patients which can **"detect cervical cancer using modern age technologies like machine learning and raspberry pi"**. Thus help world in reducing incidence and morbidity rates caused due to cervical cancer and also help women of all ages live a healthier and happier life.

1.1 Motivation

- According to statistics collected by World Health Organization (WHO), there were nearly 530,000 new cases of cervical cancer in the year 2012[1].
- It caused second highest mortality rate in female cancer patients [1].
- More than 270,000 females died from cervical cancer every year in the world, more than 85% of which occurred in developing countries [1].

- In India over 62,000 women died in the year 2015 due to cervical cancer, accounting for 24% of total cancer related deaths of women in India, as estimated by Indian Council of Medical Research (ICMR) [2].
- Cervical cancer is the most common cancer among females in Gujarat.



Estimated Cervical Cancer Incidence Worldwide in 2012

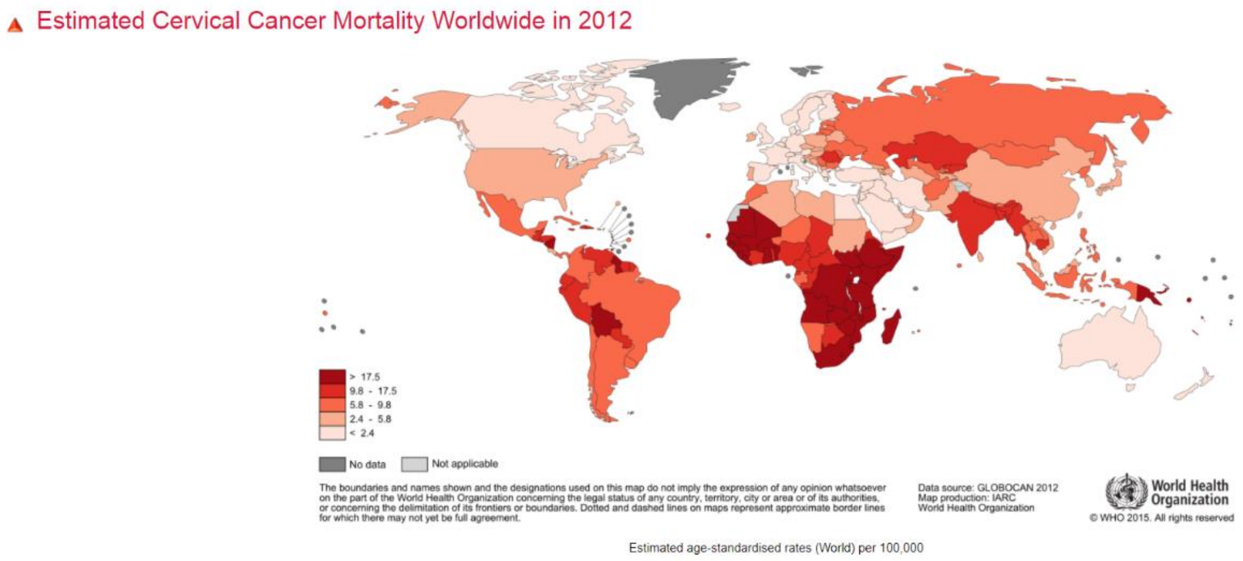
The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Data source: GLOBOCAN 2012
Map production: IARC
World Health Organization

© WHO 2015. All rights reserved

Estimated age-standardised rates (World) per 100,000

Figure 1.1.1



Estimated Cervical Cancer Mortality Worldwide in 2012

The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

Data source: GLOBOCAN 2012
Map production: IARC
World Health Organization

© WHO 2015. All rights reserved

Estimated age-standardised rates (World) per 100,000

Figure 1.1.2

This accounts for 28-30 % of all cancers in women. Few statistical graphs are shown below which are taken from an article by World Health Organization (WHO).

1.2 Objective of Project

- A software based system which can detect cervical cancer
- A similar hardware based approach using advanced controller for low resource areas

1.3 Present Solution & Limitation

- Government and UN based voluntary groups are unable to reach each and every low resource area of world.
- Health providers are facing trouble in taking high end machines and expertise in low resource areas.
- Many times when this cancer is in initial stage then it can be treated in a single visit but health providers are facing difficulty in identifying type of cervix and thus are sometimes unable to judge which therapy to use for treatment.

1.4 Literature Review

1.4.1 Automated Cervical Cancer Detection Using Pap smear Images

By: Payel Rudra Paul, Mrinal Kanti, Bhowmik and Debotosh Bhattacharjee

Cervical cancer is the most common cancer among the women. Pap smear screening is the most effective test for detecting the cervical precancerous. But this process requires a long time to complete and also may be an erroneous procedure. In this paper, an automated cervical cancer detection method is presented. This method introduces adaptive median filter to remove impulse noises from the Pap smear images and then uses bi-group enhancer to discriminate the nuclei pixels from other object pixels. Then, segmentation methodology is presented to separate the nucleus regions from the cervical smear images. Two clustering-based classifiers, minimum distance and K-nearest neighbor classifiers, have been used in the classification phase for verifying the performance. The technique was evaluated using 158 Pap smear images from DTU/HERLEV Pap smear benchmark database. The accuracy of the detection method is 92.37 and 98.31 % for minimum distance and K-nearest neighbor classifiers, respectively.

Result:

Table 1.4.1 Results of Paper 8

Classifier	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
Minimum distance Classifier	75	34	0	9	89.29	100	92.37
K-nearest neighbor Classifier	82	34	0	2	97.62	100	98.31

The method has been applied in 158 cervical Pap smear images, in which 54 normal cell and 104 abnormal cell images are there. To fully explore the system performance, training and testing has been conducted. For training purpose, 40 cell images are used, which include 20 normal cell and 20 abnormal cell images. For testing purpose, 118 cell images are used, which includes 34 normal cell and 84 abnormal cell images. The results of the classification are given in Table. As we can see from Table, within 84 abnormal cells, 75 cells and 82 cells were correctly classified using minimum distance classifier and K-nearest neighbor classifier, respectively, and within 34 normal cells, all cells were correctly classified using both of the classifier. The classification accuracy of our method using minimum distance classifier in differentiating the cancerous and normal cells is 92.37 % with sensitivity 89.29 % and specificity 100 %. On the other hand, K-nearest neighbor gives impressive accuracy of 98.31 % with sensitivity 97.62 % and specificity 100 %. The future scope of this work includes effort to improve the classification performance with the selection of more features from the nucleus and also from the cytoplasm and to classify the abnormal cells in stage wise.

1.4.2 Cervical Cancer Detection and Classification Using Texture Analysis

By: M.K. Soumyu, K. Sneha and C. Arunvinodh

Cervical cancer is one of the deadliest cancer among women. The main problem with cervical cancer is that it cannot be identified in its early stages since it doesn’t show any symptoms until the final stages. Therefore the accurate staging will help to give the accurate treatment volume to the patient. Some diagnosing tools like X-ray, CT, MRI, etc. can be used with image processing techniques to get the staging of disease. Transform features such as contourlet and Gabor features mainly based on energy are used for the prediction of output. Second-order statistical features based on contrast, correlation, energy and homogeneity are significantly used to predict outcome from pre-treatment MR images of cervical cancer tumors. This paper proposes a classification technique using Magnetic Resonance Images (MRI) to obtain the staging of cervical cancer patients.

Algorithm:

Step 1: Input the MR image.
Step 2: Enhance the MR image using Adaptive gamma correction method.
Step 3: The enhanced image is then segmented using Otsu’s segmentation technique.
Step 4: Features are now extracted using 3 methods:

1. Grey level co-occurrence matrix (GLCM) features are extracted
2. Contourlet transforms feature extraction
3. Gabor filter feature extraction

Step 5: Classification of the image is done using Support vector machine (SVM) classifier.
Step 6: Staging of each image is the resultant output.

Result:

The purpose of this study was to investigate whether features derived from MR images of patients with locally advanced cervical cancer could be used to predict the stag-ing of disease. Nonlinear SVM classification models were constructed based on both second-order texture features and transform features of the tumors. Transform features such as contourlet and Gabor features mainly based on energy are used for the prediction of output. However, second-order statistical features based on contrast, correlation, energy and homogeneity are significantly used to predict outcome from pre-treatment MR images of cervical cancer tumors. The best performing transform based models had accuracies around 81% for the axial T1-weighted MR images, 82% for axial T2-weighted and 83% for sagital T2-weighted MR images performed somewhat better than models based solely on clinical factors. Thus, texture features outperformed transform features as well as statistical features for staging prediction, and can compete with predictions based on tumor volume. As a suggestion to the future work, it is better to predict the treatment volume according to the staging will help the radiologist for a better treatment planning.

1.4.3 Pap smear Image based Detection of Cervical Cancer

By: Sreedevi M T, Usha B S, Sandya S

In this paper, a new approach is proposed for the early detection of cervical cancer using Pap smear images. Regular Pap smear screening is the most successful attempt of medical science and practice for the early detection of cervical cancer. Manual analysis of the cervical cells is time consuming, laborious and error prone. This paper presents an algorithm for classifying cervical cells as normal or abnormal. It is tested on 80Papsmear images and the experimental results show that the algorithm is on par with the results obtained by earlier work and gives satisfactory results in terms of sensitivity (100%) and specificity (90%).

Cervical cancer is a malignant disease that develops in the cells of the cervix or the neck of the uterus. These cells do not suddenly change into cancer. Instead, the normal cells of the cervix first gradually develop precancerous changes which later turn into cancer. Cancerous cells show increasing nucleus area when compared to normal cells. This characteristic feature can be used to do a first level of classification of the cervical cells as normal or abnormal

Cervical cancer if detected early has very good prognosis. Cervical screening using Pap smear images is one of the most effective ways of detecting and diagnosing the disease even at an early pre-cancerous stage. During mass screening program there will be huge number of samples to be analyzed and diagnosed and the current manual screening methods are time consuming and restricts the capabilities of the cyto-technicians in diagnosing more samples in shorter time. Therefore there is a need for a support system for faster analysis of samples. The main methods used by Martin were Hard C-means(HCM), Fuzzy C-means(FCM) and Gustafson-Kessel clustering(GK) for classification.

Result:

In this paper, we have proposed an approach for classification of cervical cell as normal or abnormal using area of the nucleus as a feature. The experimental results show that this method gives good classification and achieves sensitivity of 100% and specificity of 90% with acceptable overall error rate of 5%. The classification results are validated with the benchmark database prepared by Martin. The proposed method serves as a basis for first level classification of Pap smear images for detection of cervical abnormality using area of the nucleus as the parameter.

1.4.4 Feature Extraction of Cervical Pap smear Images Using Fuzzy Edge Detection Method

By: K. Hemalatha and K. Usha Rani

In Medical field Segmentation of Medical Images is significant for disease diagnose. Image Segmentation divide an image into regions precisely which helps to identify the abnormalities in the Cancer cells for accurate diagnosis. Edge detection is the basic tool for Image Segmentation. Edge detection identifies the discontinuities in an image and locates the image intensity changes. In this paper, an improved Edge detection method with the Fuzzy approach is proposed to segment Cervical Pap Smear Images into Nucleus and Cytoplasm. Four important features of Cervical Pap Smear Images are extracted using proposed Edge detection method. The accuracy of extracted features using proposed method is analyzed and compared with other popular Image Segmentation techniques.

Proposed Methodology:

The theoretical structure of the proposed Fuzzy Edge Detection Method (FEDM) for feature extraction of Cervical Pap Smear Images based on Fuzzy Logic approach is presented. Cervical Pap Smear Images sometimes contains menstrual discharge, vaginal discharge, air artifacts, etc., which may lead to wrong classification from normal to abnormal cells. Hence, Image Pre Processing is necessary to obtain better results. Four important features of Cervical Pap Smear images are automatically extracted using proposed Fuzzy Edge Detection Method (FEDM). The Proposed method is implemented on seven Cervical Pap Smear Images using MATLABR2015a tool for the experiment. Seven images are related to different cervical cell classes like Mild Dysplasia, Moderate Dysplasia, Severe Dysplasia and Carcinoma in situ, etc.

Result:

In this study, Fuzzy Edge Detection Method (FEDM) is proposed. The proposed method automatically extracted the Nucleus Size, Cytoplasm Size, Nucleus Grey Level and Cytoplasm Grey Level from the Cervical Pap Smear Images. The accuracy of extracted features is analyzed using Correlation Coefficient. The performance of proposed method is compared with other popular methods. The proposed method performed better than existing methods. The extracted features may be used for Classification of Cervical Pap Smear Images in future using different techniques.

1.4.5 Segmentation of Cervical Cell Nucleus using Intersecting Cortical Model optimized by Particle Swarm Optimization

By: Jing Rui Tang, Nor Ashidi Mat Isa, Ewe Seng Ch’ng

Changes in the morphology of cervical cell nucleus are one of the most important features to be observed during Pap-smear screening. In this study, Intersecting Cortical Model (ICM) was employed to segment the nucleus from cervical cell images. The four unknown parameters in ICM were optimized by Particle Swarm Optimization (PSO). Two hundred and fifty test images were randomly selected from Herlev dataset. The segmented results were compared with Otsu thresholding, Expectation Maximization technique, region growing and Fuzzy C-Means clustering technique. Analyses revealed that ICM produced the best segmentation result, with Zijdenbos Similarity Index (ZSI) of 0.914, Peak Signal to Noise Ratio (PSNR) of 62.946 dB, Misclassification Error (ME) of 0.056 and Relative Foreground Area

Error (RAE) of 0.132. Wilcoxon Signed-rank Test reported ICM significantly outperformed the four comparison techniques, with *p*-values less than 0.05 for all the performance metrics.

Result:

Table 1.4.2 Results of Paper 11

Techniques	Quantitative Analyses			
	ZSI	PSNR(db)	ME	RAE
ICM	0.914	62.946	0.056	0.132
Otsu	0.857	60.336	0.090	0.179
EM	0.819	59.085	0.140	0.240
RE	0.645	57.446	0.155	0.472
FCM	0.859	60.365	0.089	0.176

In this paper, the Intersecting Cortical Model (ICM) optimized by the Particle Swarm Optimization (PSO) was employed to segment the nucleus from cervical cell images. A total of 250 randomly selected images from Herlev dataset were used as test images. The segmented results were compared with four segmentation techniques, namely Otsu thresholding, Expectation Maximization, region growing and Fuzzy C-Means clustering techniques. Both qualitative and quantitative analyses reported that ICM yielded the best segmentation results, with the highest Zijdenbos Similarity Index (ZSI) and Peak Signal to Noise Ratio (PSNR) values. The segmented images from ICM also possessed the lowest Misclassification Error (ME) and Relative Foreground Area Error (RAE). Wilcoxon Signed-rank Test further verified that ICM significantly outperformed the other four techniques. Findings from this study revealed that ICM has great potential to be used for cervical cell nucleus segmentation due to its superior performance when compared to the other segmentation techniques.

1.4.6 Pap-smear benchmark Data for Pattern Classification

By: Jan Jantzen, Jonas Norup, George Dounias, Beth Bjerregaard

This case study provides data and a baseline for comparing classification methods. The data consists of 917 images of Pap-smear cells, classified carefully by cyto-technicians and doctors. Each cell is described by 20 numerical features, and the cells fall into 7 classes. A basic data analysis includes scatter plots and linear classification results, in order to provide domain knowledge and lower bounds on the acceptable performance of future classifiers. Students and Researchers can access the database on the Internet, and use it to test and compare their own classification methods. The term Pap-smear refers to samples of human cells stained by the so-called Papanicolau method. A specimen of cells is smeared onto a glass slide and coloured, making it easier to examine the cells under a microscope for any abnormalities indicating a pre-cancerous stage.

Result:

From this basic analysis we have gained qualitative and quantitative insight. We have established

- That class’s overlap, but 1 and 2 are separated more or less from the rest;
- That there is a gradual, successive transition through classes 4, 5, 6, and 7;
- That class 3 is somewhat ’odd’, and mixes with especially classes 5 and 6;
- That the separation into normal {1, 2, 3} and abnormal {4, 5, 6, 7} therefore is unfortunate;
- That an acceptable overall error OE% must be less than 6.4% for the 2-class problem, and less than 7.9% for the
7-class problem; and
- That a false negative rate FN% down to 1.2% for the 2-class problem is feasible, it could even go as low as 0%, but

There is a tradeoff with the false positive rate FP%. These results are achieved using simple methods. The sensitivity of the results has not been investigated in detail, that is, whether the results change a lot if there is only a small change in the underlying data. On the other hand, the results can be regarded as quite reliable, since we have avoided scaling of data and feature selection. It is usual practice to scale or standardize data before modelling, but by virtue of the linear model, this was not necessary. With 20 features, feature selection is usually necessary. We have, implicitly in our plots selected only two features, but the linear classifier operates on the full set of 20 features, and thus avoids another degree of complexity.

The objective of this case study is to provide a good database for benchmarking of classification methods. We have provided data, which have been selected and examined as carefully as possible in the hospital. The data are now on the Internet for public use, and the present paper provides a first, basic analysis so that students and researchers can get a head start, and avoid making any fundamental mistakes. The next step is to apply nonlinear classification methods. We are interested in hearing about new results, and we would like to include references to other studies in the database, in the hopes that the quality and accuracy of future classification results will increase.

1.5 Block Diagram

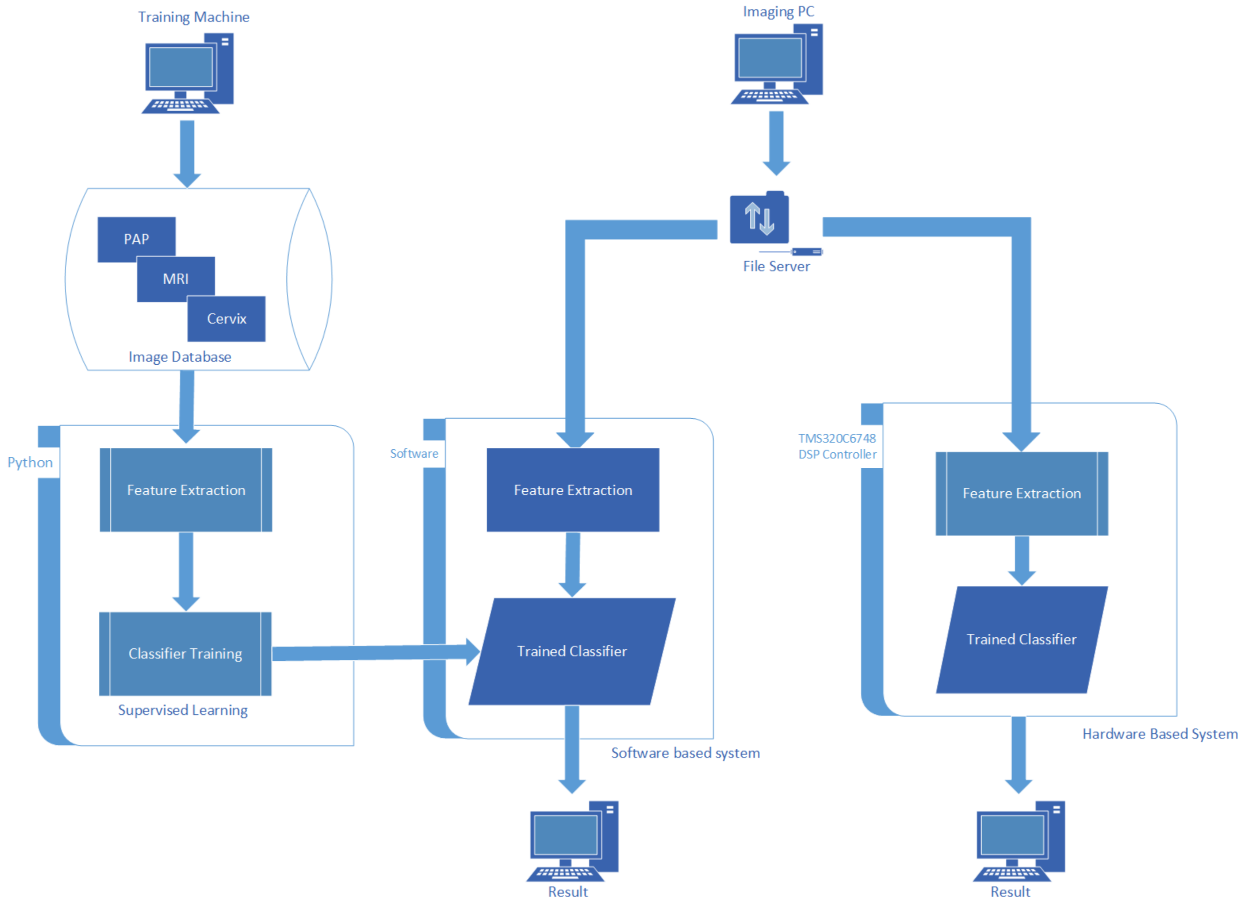


Figure 1.5.1

Our project is divided into the following major blocks:

- Software Block
  - ➢ In software based approach we will use Python 3 to train our hypothesis on our training data which we collected from multiple sources online and offline.
- Hardware Block
  - ➢ This trained hypothesis function would be coded into our advanced controller which would then be used to take input from image sources and give output accordingly.

1.6 Innovation

- We are using ML in classifying the type of cervix which is a major decisive factor when it comes to selection of suitable therapy for treatment.
- We are also using a Hardware based approach which would be really useful in low resource area.

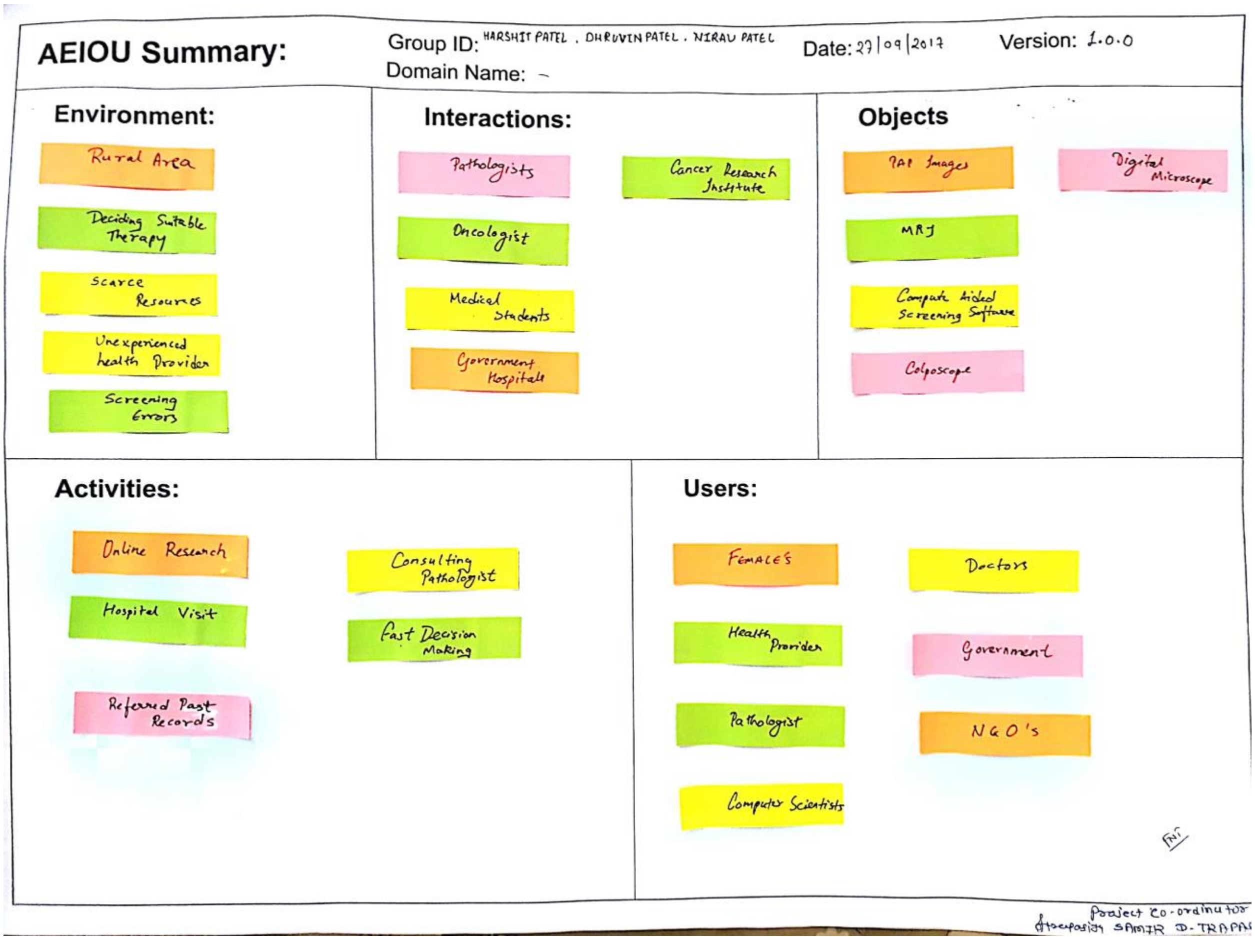
1.7 Social Impact

- Routine cervical cancer screening and early treatment can prevent up to 80% of cervical cancers if abnormalities of the cervix are identified at stages when they can be easily treated.
- WHO recommends screening for all women aged 30–49 years to identify precancerous lesions, which are usually asymptomatic.
- Today, women worldwide in low-resource settings are benefiting from programs where cancer is identified and treated in a single visit.

CHAPTER 2

DESIGN ENGINEERING CANVAS

2.1 AEIOU Canvas



AEIOU Summary:

Group ID: HARSHIT PATEL . DHRUVIN PATEL . NIRAV PATEL

Domain Name: -

Date: 27|09|2017

Version: 1.0.0

Environment:

- Rural Area
- Deciding Suitable Therapy
- Scarce Resources
- Unexperienced health Provider
- Screening Errors

Interactions:

- Pathologists
- Oncologist
- Medical Students
- Government Hospitals
- Cancer Research Institute

Objects

- PAP Images
- MRI
- Compute Aided Screening Software
- Colposcope
- Digital Microscope

Activities:

- Online Research
- Hospital Visit
- Referred Past Records
- Consulting Pathologist
- Fast Decision Making

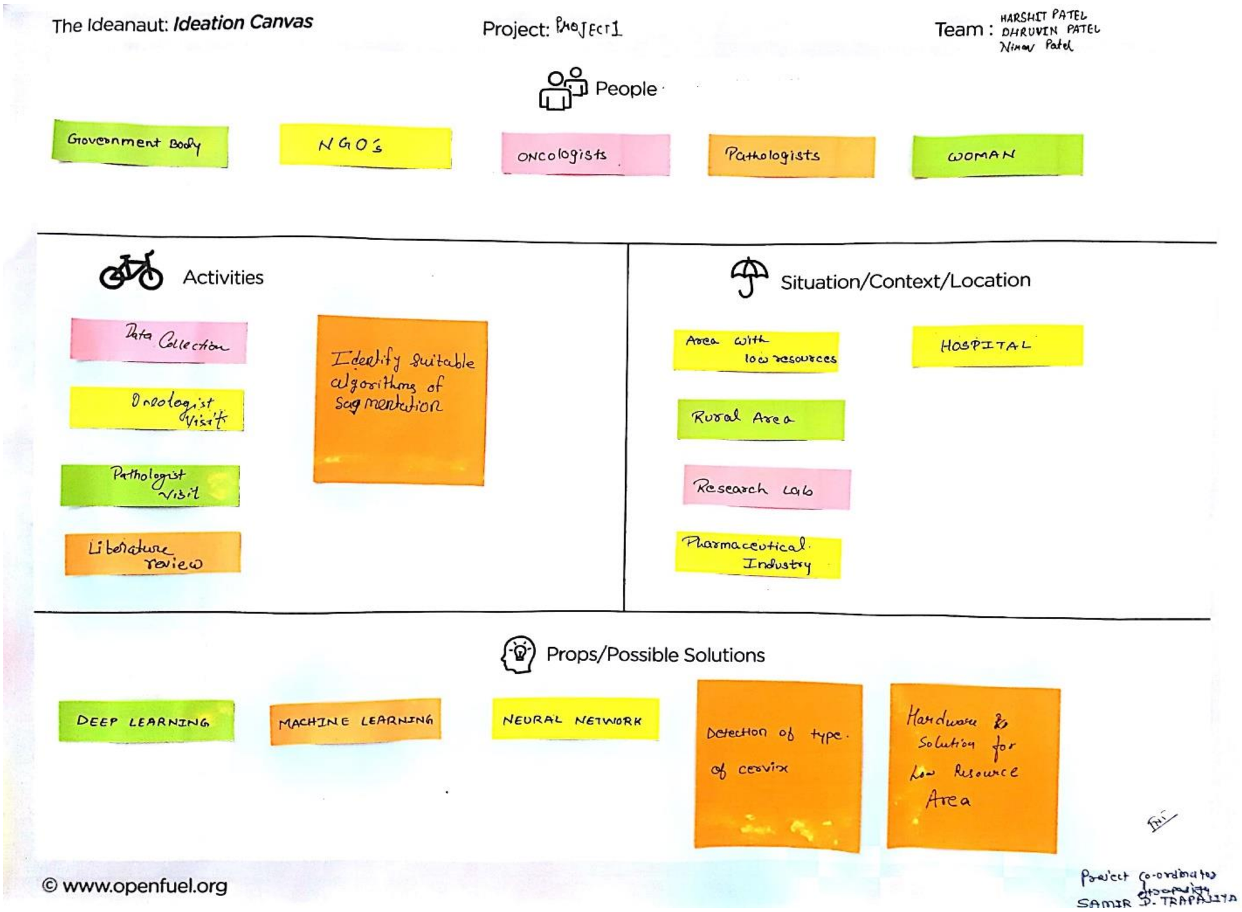
Users:

- Females
- Health Provider
- Pathologist
- Computer Scientists
- Doctors
- Government
- NGO's

Project Co-ordinator
[illegible] SAMIR D. TRAPA

Figure 2.1.1

2.2 Ideation Canvas



The Ideanaut: *Ideation Canvas*

Project: PROJECT1

Team : HARSHIT PATEL
DHRUVIN PATEL
Nirav Patel

People

- Government Body
- NGO's
- ONCologists
- Pathologists
- WOMAN

Activities

- Data Collection
- Oncologist Visit
- Pathologist Visit
- Literature review
- Identify suitable algorithms of segmentation

Situation/Context/Location

- Area with low resources
- HOSPITAL
- Rural Area
- Research Lab
- Pharmaceutical Industry

Props/Possible Solutions

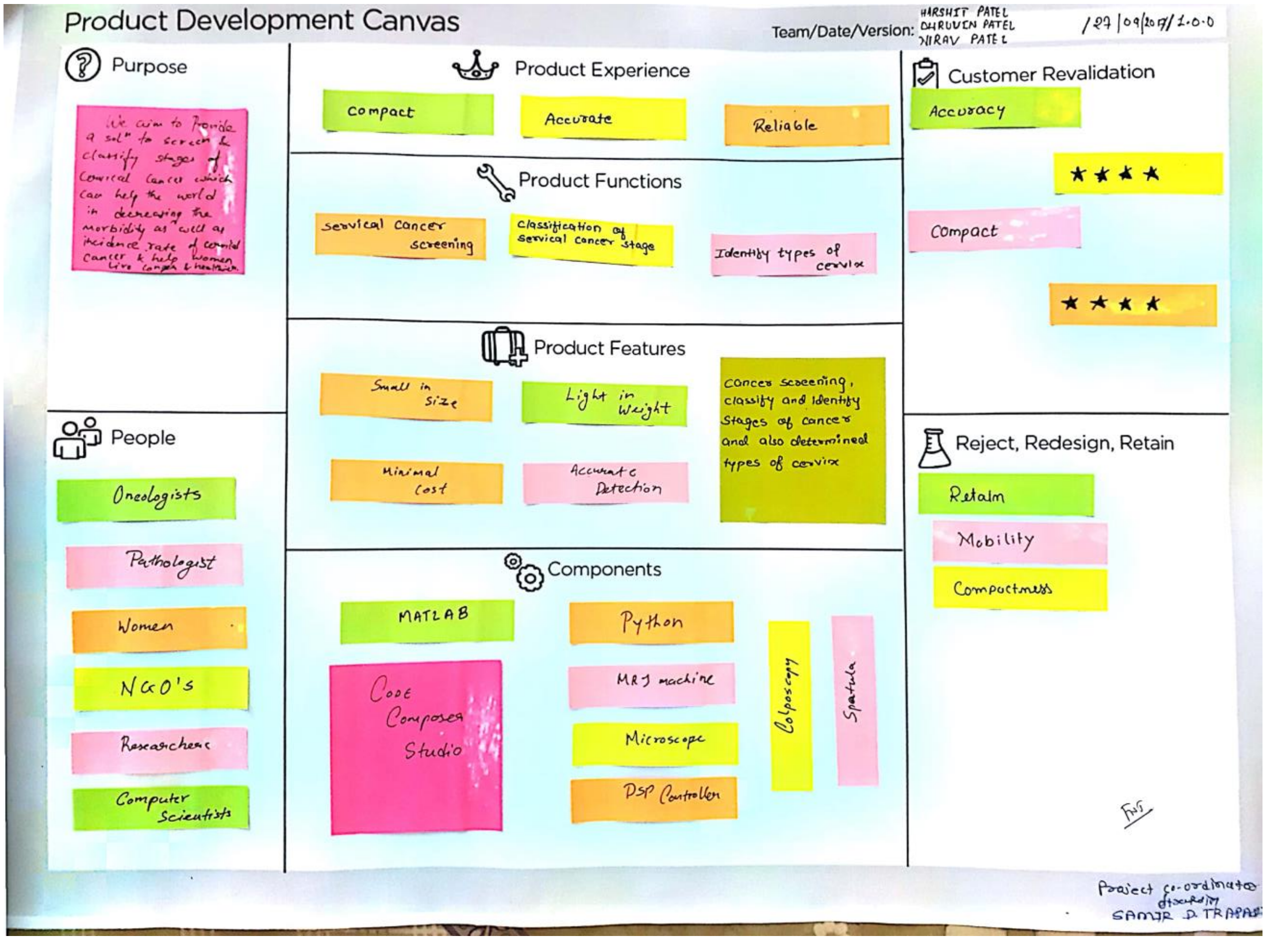
- DEEP LEARNING
- MACHINE LEARNING
- NEURAL NETWORK
- Detection of type of cervix
- Hardware & Solution for Low Resource Area

© www.openfuel.org

Project Co-ordinator
SAMIR D. TRAPASIYA

Figure 2.2.1

2.3 Product Development Canvas



Product Development Canvas

Team/Date/Version: HARSHIT PATEL
DHRUVIN PATEL
NIRAV PATEL
/27/09/2017/1.0.0

Purpose

We aim to Provide a sol^{n} to screen & classify stages of Cervical Cancer which can help the world in decreasing the morbidity as well as incidence rate of cervical cancer & help women live longer & healthier.

Product Experience

compact

Accurate

Reliable

Product Functions

servical cancer screening

classification of servical cancer stage

Identify types of cervix

Product Features

Small in size

Light in weight

Minimal cost

Accurate Detection

cancer screening, classify and Identify stages of cancer and also determined types of cervix

Components

MATLAB

Code Composer Studio

Python

MRI machine

Microscope

DSP Controller

Colposcopy

Spatula

People

Oncologists

Pathologist

Women

NGO's

Researchers

Computer Scientists

Customer Revalidation

Accuracy

★★★★

Compact

★★★★

Reject, Redesign, Retain

Retain

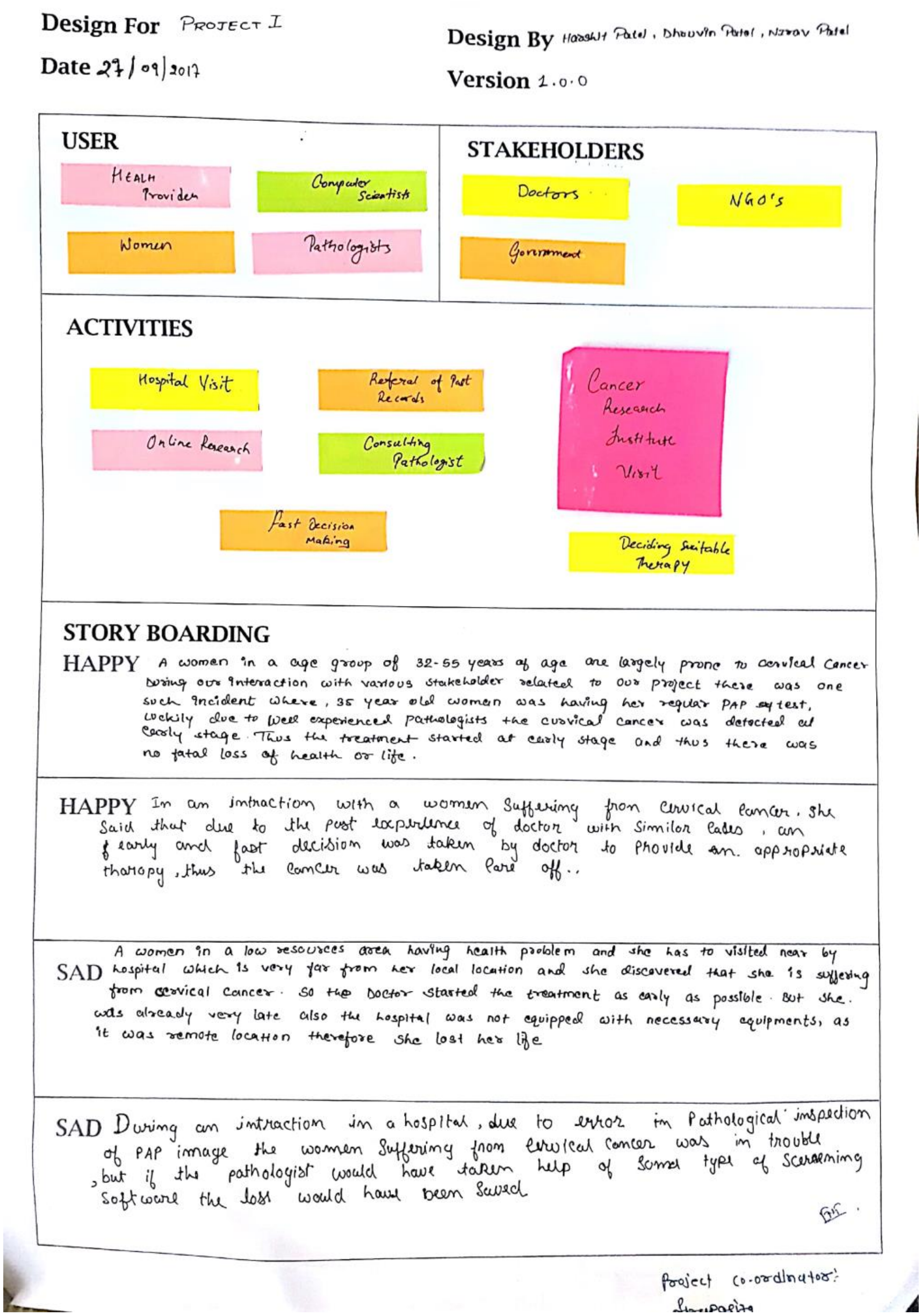
Mobility

Compactness

Project co-ordinator
SAMIR D TRAPAS

Figure 2.3.1

2.4 Empathy Mapping Canvas



Design For PROJECT I

Date 27/09/2017

Design By Harshit Patel, Dhruvin Patel, Nirav Patel

Version 1.0.0

USER

Health Provider

Computer Scientists

Women

Pathologists

STAKEHOLDERS

Doctors

NGO's

Government

ACTIVITIES

Hospital Visit

Referal of Past Records

Cancer Research Institute Visit

Online Research

Consulting Pathologist

Fast Decision Making

Deciding Suitable Therapy

STORY BOARDING

HAPPY A women in a age group of 32-55 years of age are largely prone to cervical cancer During our interaction with various stakeholder related to our project there was one such incident where, 35 year old women was having her regular PAP test, luckily due to well experienced pathologists the curvical cancer was detected at early stage. Thus the treatment started at early stage and thus there was no fatal loss of health or life.

HAPPY In an intraction with a women Suffering from cervical cancer. She Said that due to the post experience of doctor with Similar cases, an early and fast decision was taken by doctor to provide an appropriate tharopy, thus the cancer was taken care off..

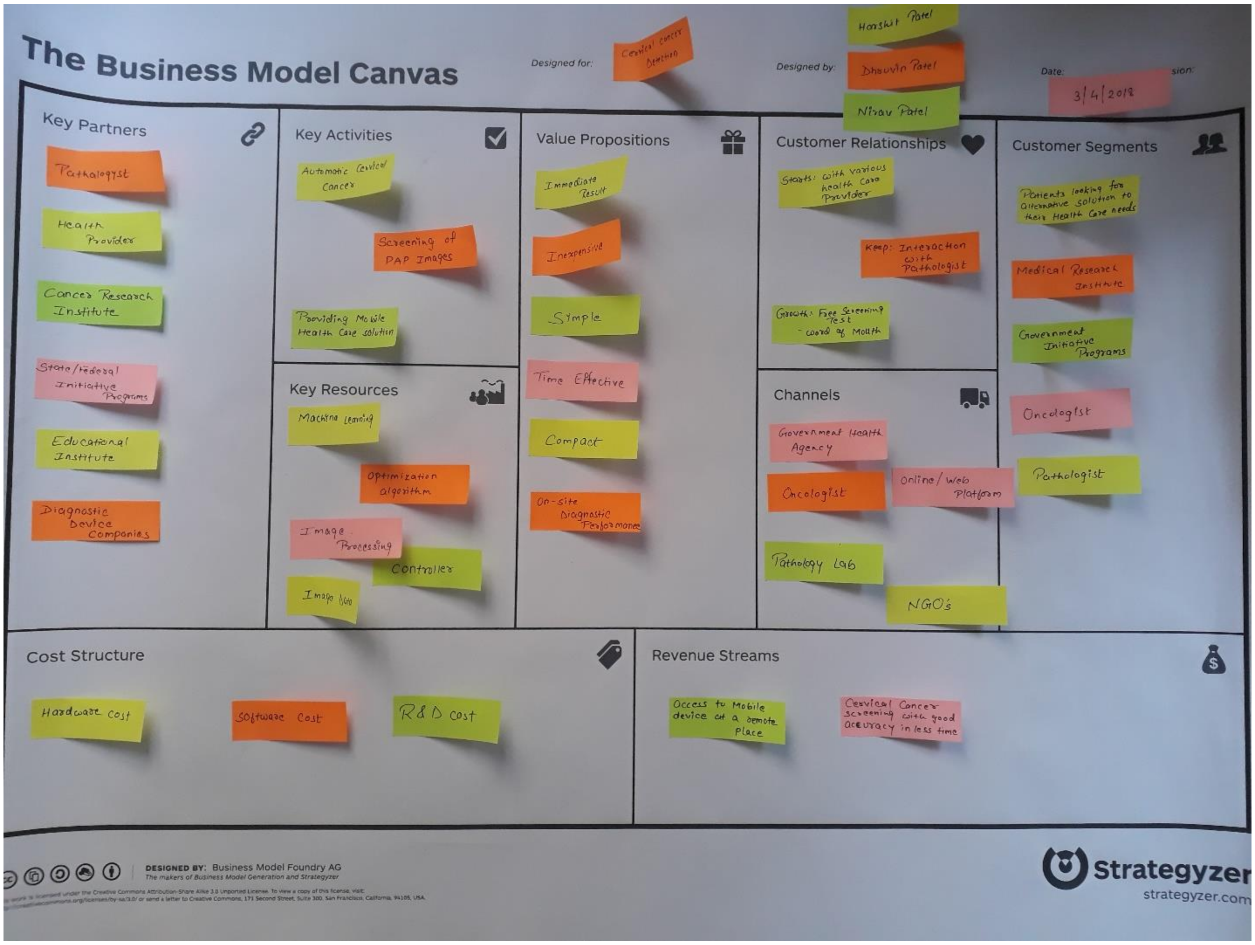
SAD A women in a low resources area having health problem and she has to visited near by hospital which is very far from her local location and she discovered that she is suffering from cervical cancer. So the Doctor started the treatment as early as possible. But she was already very late also the hospital was not equipped with necessary equipments, as it was remote location therefore she lost her life

SAD During an intraction in a hospital, due to error in Pathological inspection of PAP image the women Suffering from cervical cancer was in trouble, but if the pathologist would have taken help of some type of Screening Software the loss would have been Saved

Project co-ordinator:

Figure 2.4.1

2.5 Business Model Canvas



The Business Model Canvas

Designed for: Cervical cancer Detection

Designed by: Harshit Patel, Dhruvin Patel, Nirav Patel

Date: 3/4/2018

Key Partners
- Pathalogyst
- Health Provider
- Cancer Research Institute
- State/federal Initiative Programs
- Educational Institute
- Diagnostic Device Companies

Key Activities
- Automatic Cervical Cancer
- Screening of PAP Images
- Providing Mobile Health Care solution

Key Resources
- Machine Learning
- Optimization algorithm
- Image Processing
- Controller
- Image Data

Value Propositions
- Immediate Result
- Inexpensive
- Simple
- Time Effective
- Compact
- On-site Diagnostic Performance

Customer Relationships
- Starts: with various health Care Provider
- Keep: Interaction with Pathologist
- Growth: Free Screening Test - word of Mouth

Channels
- Government Health Agency
- Oncologist
- Online/Web Platform
- Pathology Lab
- NGO's

Customer Segments
- Patients looking for alternative solution to their Health Care needs
- Medical Research Institute
- Government Initiative Programs
- Oncologist
- Pathologist

Cost Structure
- Hardware cost
- Software Cost
- R&D cost

Revenue Streams
- Occess to Mobile device at a remote Place
- Cervical Cancer screening with good accuracy in less time

DESIGNED BY: Business Model Foundry AG
The makers of Business Model Generation and Strategyzer

Strategyzer
strategyzer.com

Figure 2.5.1

2.6 Business Model Canvas Report

2.6.1 HOW

- Key Partners

- Health Providers
- Pathologist
- Cancer Research Center
- State/Federal Initiative programs
- Educational institution
- Diagnostic device companies

- Key Resources

- Machine Learning
- Optimization algorithms
- Image processing
- Raspberry Pi
- Image Data

2.6.2 WHAT

-Value proposition

- Immediate Results
- Inexpensive
- Simple
- Time Effective
- Compact
- On-site diagnostic performance

2.6.3 WHO

- **Customer Segments**
  - Patients looking for alternative solution to their healthcare needed
  - Medical research institute
  - Governments initiative programs
  - Oncologist
  - Pathologist

- **Customer Relationships**
  - State: With various healthcare providers
  - Keep: Interaction with Pathologist
  - Growth: Free Screening Test, Word of Mouth
- **Channels**
  - Government Health Agency
  - Oncologist
  - Online/Web portal
  - Pathology Lab
  - NGO's

2.6.4 HOW MUCH

- **Revenue Streams**
  - Access to Mobile device at a remote place
  - Cervical cancer Screening with good accuracy in less time
- **Cost Structure**
  - Hardware cost
  - Software cost
  - R&D cost

CHAPTER 3

SOFTWARE IMPLEMENTATION

3.1 Papsmear Images

Pap smear images are microscopic cell images of squamous epithelial cells collected on cytological slides. Few samples of such images are given below.

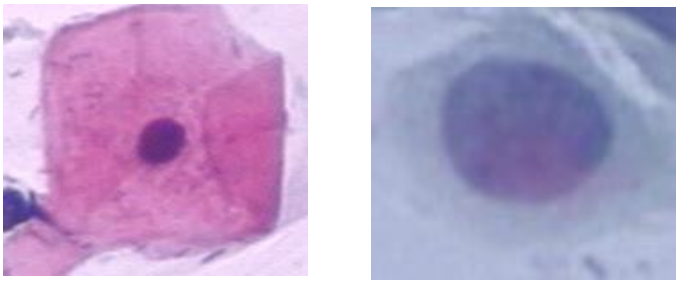


Figure 3.1.1 cervical cells

3.2 Image Segmentation

Image segmentation is basically used to find region of interest in image. After comparing many image segmentation techniques we have decided to use Intersecting Cortical Model method for training and Otsu’s thresholding for application of our project.

3.1.1 Intersecting Cortical Model

Deriving from several models of the cortex, particularly based on PCNN, ICM is especially designed for image processing. As compared to PCNN, reduction of the unknown parameters in ICM decreases the computational complexity while preserving the superior performance of PCNN in image processing. ICM is inspired from the research outcomes of the mammalians’ visual cortex neurons cells. The artificial neuron model is then derived from the simulation of mammalian visual activities. The state oscillators of all the neurons in ICM are represented by a two-dimensional array, *G* with initial values equal to zero. The threshold oscillators, represented by a two dimensional array, α have initial values equals to zero. The ICM is defined by the following equations:

$$\boldsymbol{G_{ij}[n+1] = fG_{ij}[n] + T_{ij} + X_{ij}\{[Z[n]]\}} \qquad \textbf{(1)}$$

$$\boldsymbol{Z_{ij}[n+1] = \{\, 1, if\ G_{ij}[n+1] > \propto_{ij}\ [n]\ \&\ 0, Otherwise} \qquad \textbf{(2)}$$

$$\boldsymbol{\propto_{ij}\ [n+1] = g \propto_{ij}\ [n] + hZ_{ij}[n+1]} \qquad \textbf{(3)}$$

$\boldsymbol{T_{ij}}$ is the stimulus, which is the input image scaled with largest pixel value equals to 1.0. $\boldsymbol{\alpha_{ij}}$ is the threshold. $\boldsymbol{Z_{ij}}$ is the firing state of the neuron, which is the output image. Parameters $\boldsymbol{f}$, $\boldsymbol{g}$ and $\boldsymbol{h}$ are scalars. Both $\boldsymbol{f}$ and $\boldsymbol{h}$ are less than 1 and $\boldsymbol{g}$ always less than $\boldsymbol{f}$ so that the threshold value will fall below the state of neuron of pulses. The neuron at two-dimensional position described with *i* and *j* pixel locations has state $\boldsymbol{G_{ij}}$ at $\boldsymbol{\alpha_{ij}}$. $\boldsymbol{X_{ij}}$ is the connection function for inter-neuron communication. To find the optimum values for parameters *f*, *g*, *h* and the number of iteration, n, Particle Swarm Optimization is used for every test image.

3.2.2 Otsu's Thresholding

Otsu's method is one of the most successful methods for image thresholding. Otsu's method is utilized to consequently perform clustering based image thresholding or, the decrease of a gray level picture to a binary picture. The calculation expect that the picture contains two classes of pixels following bi-modular histogram , it then calculates the optimum threshold separating the two classes so that their combined spread is minimal, or equivalently because the sum of pairwise squared distances is constant, so that their inter-class variance is maximal.

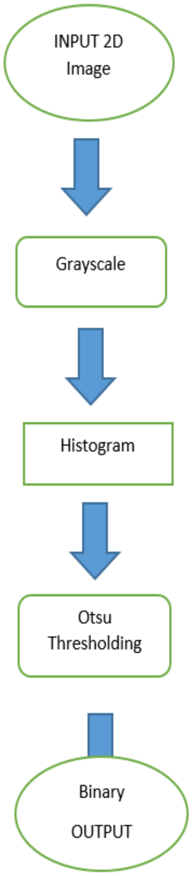


Figure 3.2.1 Otsu's Thresholding Algorithm

3.3 Particle Swarm Optimization

Introduced by Kennedy and Eberhart in 1995 [22], PSO is an evolutionary computational technique inspired from the movement and intelligence behavior of swarms. Mathematically, the particles are manipulated according to the following equations:

$$v_{id} = w * v_{id} + c_1 * r_1 * (p_{ibest} - x_{id}) + c_2 * r_2 * \left(p_{gbest} - x_{id}\right) \quad (4)$$

$$x_{id} = x_{id} + v_{id} \quad (5)$$

Where w is the inertia weight, c_1 and c_2 represent positive random constants within the range [0, 1]. The i-th particle in the D-dimensional space is represented as $X_i = (x_{i1}, x_{i2},\ldots, x_{iD})$ with velocity $V_i = (v_{i1}, v_{i2,} \ldots, v_{iD})$. The fitness function used is the mutual information between the input gray scale image and the segmented image. PSO search for the optimal solution of the particle itself and global optima solution in the population, represented by p_{ibest} and p_{gbest} respectively. The velocity, v_{id} and the location of the particles, x_{id} will be updated according to equation 4 and 5 until converge.

3.4 Feature Engineering

3.4.1 Red Ratio Mean

It is an important parameter used in ratio image processing. The formula for red ratio image is given below:

$$R = \frac{100 * R * 256}{(1 + R + G) * (1 + R + G + B)} \qquad (6)$$

The mean of this image is called Red Ratio Mean

3.4.2 Blue Ratio Mean

Similarly, the equation for blue ratio image is given below:

$$B = \frac{100 * B * 256}{(1 + R + G) * (1 + R + G + B)} \qquad (7)$$

3.4.3 Green Ratio Mean

The equation for green ratio image is given below:

$$G = \frac{100 * G * 256}{(1 + R + B) * (1 + R + G + B)} \qquad (8)$$

3.4.4 Image Mean

This is nothing but mean of pixel values and it is given by:

$$Image\ Mean = \frac{1}{m * n * 3} \sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=1}^{3} p(i, j, k) \qquad (9)$$

3.4.5 Roughness Index

The surface roughness index is defined based on the differences of heights between neighboring pixels in a particular window. It is computed by normalizing logarithmic value of N with the logarithmic value of N_{max} of the same window size as shown below:

$$RI = \begin{cases} \frac{\log(N)}{\log(N_{max})}, for\ N > 0 \\ 0, \quad for\ N = 0 \end{cases} \qquad (10)$$

Where, $N_{max} = 2 * H * L * (L - 1)$

Here, H = max intensity and L = window size.

3.4.6 Entropy

Image entropy is a quantity which is used to describe the `business' of an image, i.e. the amount of information which must be coded for by a compression algorithm.

$$E = -\sum_{i} P_i * \log P_i$$

Where, P_i is probability of pixel value i.

3.4.7 Variance

The variance measures how far each number in the set is from the mean. Variance is calculated by taking the differences between each pixel in the set and the mean, squaring the differences (to make them positive) and dividing the sum of the squares by the number of values in the set.

3.4.8 Local Binary Pattern Mean

Local binary pattern is a type of visual descriptor used for classification in computer vision. It is very powerful feature for texture classification. Local binary pattern mean is nothing but mean of the local binary pattern of the input image.

3.4.9 Local Binary Pattern Variance

Local binary pattern is variance calculated over local binary pattern of input image.

3.4.10 Nucleus Area

The area of nucleus of a cell is calculated as no. of white pixels in thresholded image as white pixels would represent nucleus of cell in image. It is measured in pixel^2.

3.4.11 Cytoplasm Area

The area of cytoplasm of a cell is calculated as no. of white pixels in inverted thresholded image because now white pixels would represent cytoplasm of cell in image. It is also measured in pixel^2.

3.4.12 Nucleus to Cytoplasm Ratio

This is a very important feature in cervical cancer classification as the most of the time cancerous cell have larger Nucleus compared to Cytoplasm which is reflected by Nucleus to Cytoplasm Ratio.

3.4.13 Perimeter

This is calculated by applying edge detection methods. We have applied canny edge detection algorithm and from the edge image we calculate perimeter by getting total no. of white pixels (as they represent the edge).

3.5 Machine Learning Model

3.5.1 Decision Tree Classifier

Decision tree learning uses a decision tree (as a predictive model) to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modelling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees.

We used scikit-learn python library to implement Decision tree classifier. We got an accuracy of 74.59% and f1 score of 81.85%.

3.5.2 Neural Networks

Artificial neural networks (ANNs) or connectionist systems are computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" (i.e. progressively improve performance on) tasks by considering examples, generally without task-specific programming.

We implemented a neural network of with 500 hidden layers of 100 neurons each. We kept learning rate $\alpha = 0.00005$ and used combination of different activation function and optimizer to get best possible results

a) Using SGD Optimizer and logistic activation:

We got an accuracy of 73.51% and f1 score of 84.73%.

b) Using LBFGS Optimizer and tanh activation:

We got an accuracy of 63.78% and f1 score of 71.96%.

c) Using ADAM Optimizer and identity activation:

We got an accuracy of 74.59% and f1 score of 84.98%

3.5.3 k Nearest Neighbors

In pattern recognition, the *k*-nearest neighbors algorithm (*k*-NN) is a non-parametric method used for classification and regression.[1] In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression:

- In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If *k* = 1, then the object is simply assigned to the class of that single nearest neighbor.
- In *k-NN regression*, the output is the property value for the object. This value is the average of the values of its *k* nearest neighbors.

k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification. The *k*-NN algorithm is among the simplest of all machine learning algorithms.

We got an accuracy of 82.16% and f1 score of 89.03%.

3.5.4 Support Vector Machine

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier . An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

We got an accuracy of 73.51% and f1 score of 84.73%.

3.5.5 Perceptron

In machine learning, the perceptron is an algorithm for supervised learning of binary classifiers (functions that can decide whether an input, represented by a vector of numbers, belongs to some specific class or not). It is a type of linear classifier, i.e. a classification algorithm that makes its predictions based on a linear predictor function combining a set of weights with the feature vector. The algorithm allows for online learning, in that it processes elements in the training set one at a time.

We got an accuracy of 77.29% and f1 score of 85.21%.

3.5.6 Random Forest Classifier

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set.

We got an accuracy of 85.40% and f1 score of 90.65%.

So, we decided to use this model in our final software as well as on our hardware device.

3.6 Python GUI

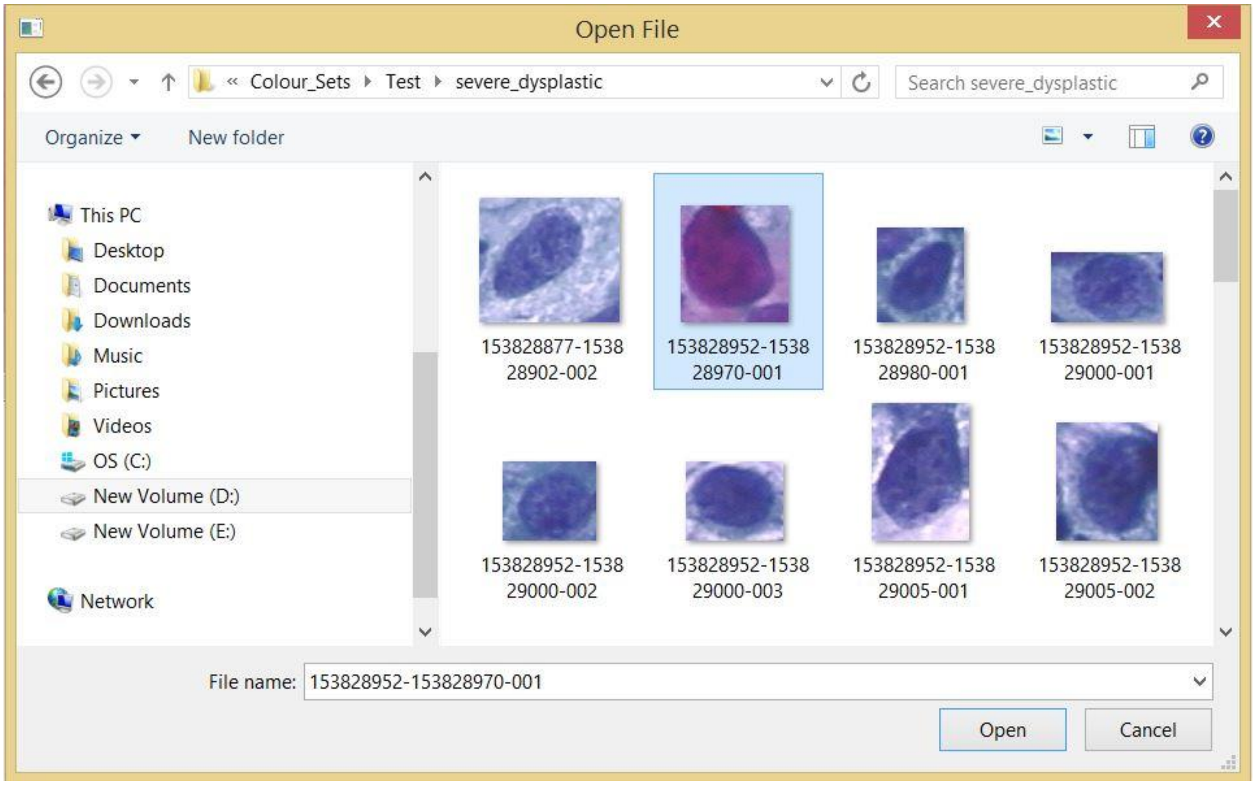
We used PyQt5 python library for generating required software for this project.

PyQt is a Python binding of the cross-platform GUI toolkit Qt, implemented as a Python plug-in. PyQt is free software developed by the British firm Riverbank Computing. It is available under similar terms to Qt versions older than 4.5; this means a variety of licenses including GNU General Public License (GPL) and commercial license, but not the GNU Lesser General Public License (LGPL). PyQt supports Microsoft Windows as well as various flavours of UNIX, including Linux and MacOS (or Darwin).[4]

PyQt implements around 440 classes and over 6,000 functions and methods including:

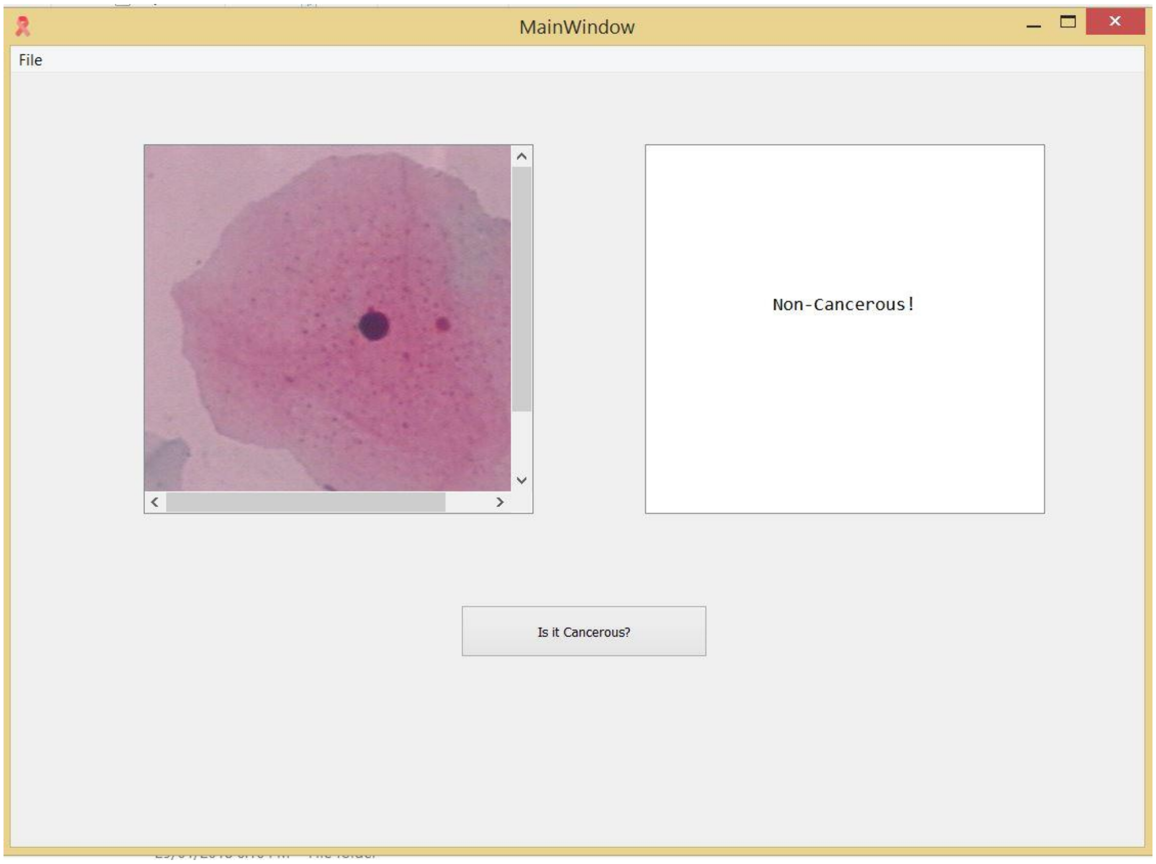
- a substantial set of GUI widgets
- classes for accessing SQL database
- Q-Scintilla, Scintilla-based rich text editor widget
- data aware widgets that are automatically populated from a database
- an XML parser
- SVG support
- Classes for embedding ActiveX controls on Windows.

Images from the generated software is given below:



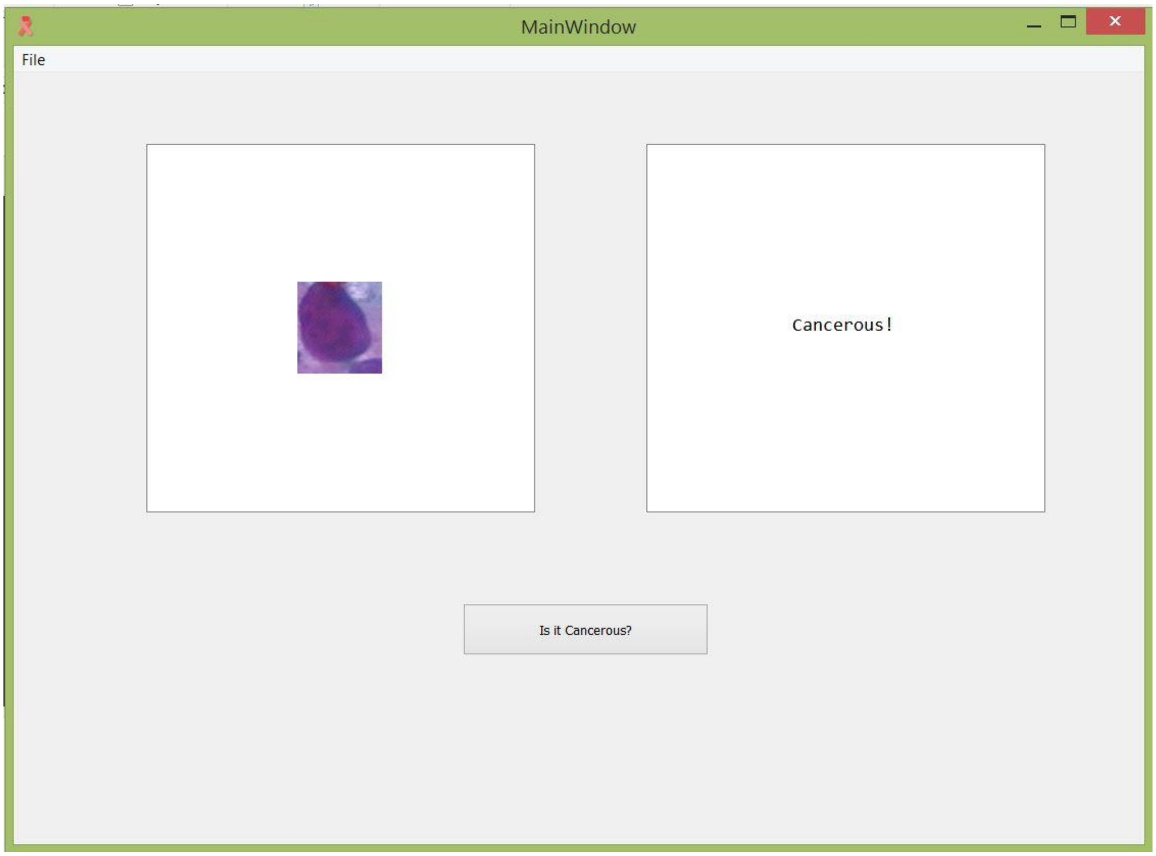
File name: 153828952-153828970-001

Figure 3.6.1 software window I



Non-Cancerous!

Figure 3.6.2 software window II



Cancerous!

Figure 3.6.3 software window III

CHAPTER 4

HARDWARE IMPLEMENTATION

One of the innovation we brought in this project is adding a hardware component in order to make the application more mobile enabling telemedicine application. Thus, reducing costs occurring in primary screening of cancer. This will enable the government and NGO ran programs to wider no. of women and will ultimately help in the goal of reducing incidence and mortality rate due to cervical cancer.

4.1 Raspberry Pi

Raspberry Pi® is an **ARM** based credit card sized **SBC**(Single Board Computer) created by Raspberry Pi Foundation. Raspberry Pi runs Debian based **GNU/Linux** operating system Raspbian and ports of many other OSes exist for this SBC.

We used Raspberry Pi 3 Model B for the project as it comes with all necessary peripherals required ranging from Bluetooth to WiFi, USB ports to LAN ports, etc.

Its Specifications are given below:

Table 4.1.1 Raspberry Pi B+ Specifications

Processor	Broadcom BCM2837
CPU Core	Quadcore ARM Cortex-A53 64 bit
Clock Speed	1.2 GHz
RAM	1 GB
Network Connectivity	1 x 10/100 Ethernet (RJ45 Port)
Wireless Connectivity	802.11 WiFi and Bluetooth 4.1
USB Ports	4 x USB 2.0

4.2 Setting Up Environment

Raspberry Pi like every other computer requires an operating system to work on. There are various OS available for Raspberry Pi to name a few Raspbian Jessie, Raspbian Stretch, Windows IoT core, Ubuntu mate, Pidora, Snappy Ubuntu, SARPi, etc.

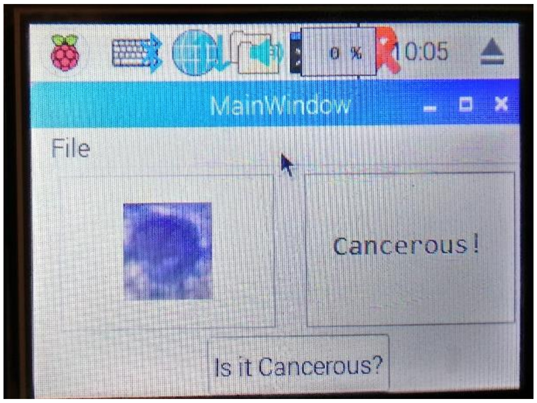
Out of all listed above Raspbian Stretch is the latest OS for raspberry pi which is from Raspberry Pi community and it is widely used. We used raspbian stretch in the project because the community support for raspbian stretch was remarkably high compared to other OS.

The Python GUI build using PyQt5 can also work on linux environment because python is a portable code. Before that we need to install all the libraries we used on training machine in raspberry pi as well. The steps are given below:

- Update pip package manager
- Install numpy, scipy and matplotlib
- Install scikit-learn
- Install scikit-image
- Install PyQt5

4.3 Working Model

In the working model we attached a touchscreen display to the raspberry pi so that the user can interact with the device easily using stylus. The screen shots of the final working environment are given below:



Cancerous!

Figure 4.3.1 final product

CHAPTER 5

CONCLUSION

In this fast growing world there are lots of tragic diseases which causes thousands of death every year. To fight such diseases NGO's and Government are working together to bring new technologies like telemedicine, etc. to provide health care in each and every section of world. With this we conclude that with our portable software as well as hardware devices the incidence and morbidity rates of cervical cancer would come down. Also with the portable device which we make we would be able to provide health care in farthest corners of world by means of telemedicine. This project can lead to a new set of devices in medicine that are not just connected to internet but also are smart and thus are open to future possibilities.

APPENDICES

APPENDIX 1

College: G. H. Patel College of Engineering & Technology

Department: Electronics & Communication Engineering

Discipline: BE

Semester: 8th

Project Name: Cervical Cancer Detection using Machine Learning & Raspberry Pi

Team ID: 27987

Form 1 – APPLICATION FOR GRANT FOR PATENT

Applicants & Inventors

Sr. No	N a m	Nationality	Address	Mobile No.	E m a
1	Patel Harshitkumar Prakashbhai	Indian	Electronics & Communication Engineering , G. H. PATEL COLLEGE OF ENGINEERING & TECHNOLOGY, V V NAGAR ,Gujarat Technological University.	8511800258	harshit.p.patel96@gmail.com
2	Patel Nirav Rajendrakumar	Indian	Electronics & Communication Engineering , G. H. PATEL COLLEGE OF ENGINEERING & TECHNOLOGY, V V NAGAR ,Gujarat Technological University.	9824655606	niravrmk@gmail.com
3	Patel Dhruvinkumar Pinkeshkumar	Indian	Electronics & Communication Engineering , G. H. PATEL COLLEGE OF ENGINEERING & TECHNOLOGY, V V NAGAR ,Gujarat Technological University.	7600201001	d2dhruvinpatel@gmail.com
4	Thakkar Falgunkumar Navinchandra	ndian		9904355301	falgunthakkar@gcet.ac.in

I/We, the applicant(s) hereby declare(s) that:

Following are the attachments with the applications:

Form 2: PROVISIONAL/COMPLETE SPECIFICATION

1. Title of the project/invention: Cervical Cancer Detection using Machine Learning and Raspberry pi
2. Preamble to the description: Provisional
3. Description
    a. Field of Project/ Invention/ Application: Computer Science, Electronics, Machine Learning
    b. Prior Art/ Background of the project/ Invention: Prior to this project no work was done on generating a hardware.
    c. Summary of the project/ Invention: The project provides a smart compact standalone device which can detect cancer using pre-trained machine learning model.
    d. Objects of project/ Invention: Raspberry Pi
    e. Drawings: None
    f. Description of project/ Invention(full detail of project):

        The project is proposed to detect the cervical cancer in the women of rural and remote area of India. In this, the image database is taken from standard sources like DICOM and local hospitals. Nowadays machine learning is very popular tool to compare the features from different group of natural and medical images. Thus, here machine learning will be used to obtain the similarity of features of images under test. This classifier will be trained in python environment. Further the trained algorithm will also be implemented on Raspberry Pi. Such a device will reduce the cost incurred in telemedicine application.

    g. Examples: None
    h. Claims (Not required for provisional application)/ Unique features of project: A Standalone Hardware.
4. Claims: None
5. Date and Signature:

6. Abstract of the project/ Invention:

   The project is proposed to detect the cervical cancer in the women of rural and remote area of India. In this, the image database is taken from standard sources like DICOM and local hospitals. Nowadays machine learning is very popular tool to compare the features from different group of natural and medical images. Thus, here machine learning will be used to obtain the similarity of features of images under test. This classifier will be trained in python environment. Further the trained algorithm will also be implemented on Raspberry Pi. Such a device will reduce the cost incurred in telemedicine application.

Form 3: STATEMENT OF UNDERTAKING UNDER SECTION 8

Name of the applicant(s):

I/We, Patel Harshitkumar Prakashbhai, Patel Nirav Rajendrakumar, Patel Dhurvinkumar Pinkeshbhai and Thakkar Falgunkumar Navinchandra.

Hereby declare:

Name, Address and Nationality of the joint applicant:

(i) That I/We have not made any application for the same/substantially the same victim invention outside India.
(ii) That the rights in the application(s) has/have been assigned to
(iii) That I/We undertake that upto the date of grant of the patent by the Controller, I/We would keep him informed in writing the details regarding corresponding applications for patents field outside India within three months from the date of filing such application.

Dated this 9 day of April 2018

To be authorized by the applicant or
His authorized registered patent agent:
____

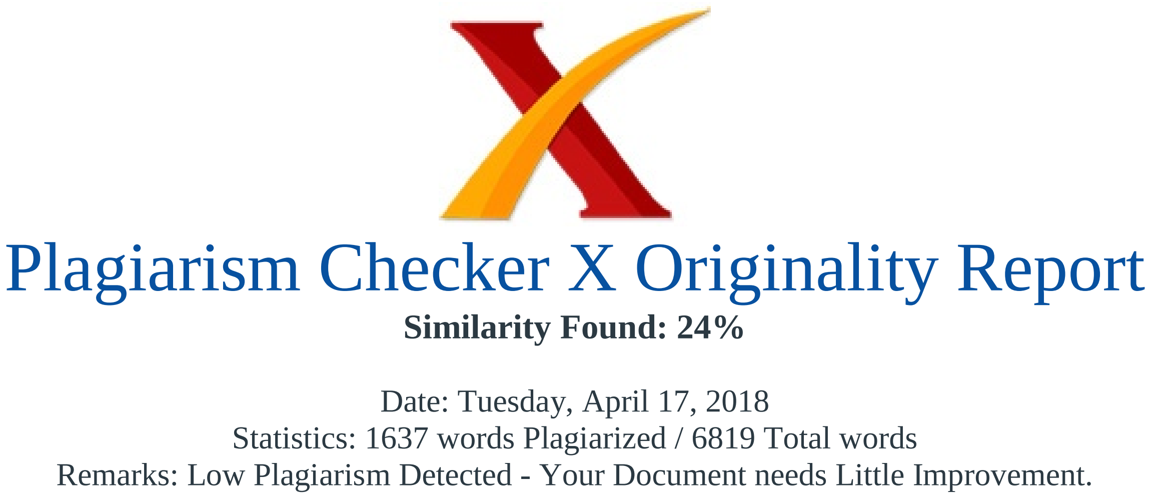
Name of the natural person who has Signed:

Patel Harshitkumar Prakashbhai,
Patel Nirav Rajendrakumar,
Patel Dhruvinkumar Pinkeshbhai,
Thakkar Falgunkumar Navinchandra

To,
The Controller of Patents,
The Patent Office, At Mumbai

APPENDIX 2

PLAGIARISM REPORT



Plagiarism Checker X Originality Report

Similarity Found: 24%

Date: Tuesday, April 17, 2018
Statistics: 1637 words Plagiarized / 6819 Total words
Remarks: Low Plagiarism Detected - Your Document needs Little Improvement.

APPENDIX 3

BIBLIOGRAPHY

1. Automatic Detection of Cervical Cancer Cells by a Two-Level Cascade Classification System by Jie Su, Xuan Xu, Yongjun He and Jinming Song

2. Cervical Cancer Detection and Classification Using Texture Analysis by M.K. SOUMYA, K. SNEHA and C. ARUNVINODH

3. Cervical Cancer Screening and ClassificationUsing Acoustic Shadowing by N.Sakthi Priya

4. Automatic detection system of cervical cancer cell using color intensity classification by Eko supriyanto, Nur Azureen M. Pista, Lukman hakim Ismail, Bustanur rosidi and Tati Lafitah Mengko

5. Automated cervical cancer screening through image analysis by Patrik Malm

6. COMPUTER AIDED DIAGNOSIS FOR DETECTION AND STAGE IDENTIFICATIONOF CERVICAL CANCER BY USING PAP SMEAR SCREENING TEST IMAGES by S. Athinarayanan, M.V. Srinath and R. Kavitha

7. Papsmear Image based Detection of Cervical Cancer by Sreedevi M T, Usha B S and Sandya S

8. Automated Cervical Cancer Detection Using Pap Smear Images Payel Rudra Paul, Mrinal Kanti Bhowmik and Debotosh Bhattacharjee

9. Pap-smear Benchmark Data For Pattern Classification by Jan Jantzen, Jonas Norup, George Dounias and Beth Bjerregaard

10. Feature Extraction of Cervical Pap Smear Images Using Fuzzy Edge Detection Method by K. Hemalatha and K. Usha Rani

11. Segmentation of Cervical Cell Nucleus using Intersecting Cortical Model optimized by Particle Swarm Optimization by Jing Rui Tang, Nor Ashidi Mat Isa and Ewe Seng Ch’ng

12. Roughness Index and Fractal Dimension for Surface Information Extraction by Lea-Tien TAY and Hse-Tzia TENG