**A Preliminary Project Report**

**On**

“

**“BREAST CANCER DETECTION USING**

**ARTIFICIAL INTELLIGENCE”**

**Submitted to the**

**Savitribai Phule Pune University**

**In the partial fulfillment for the award of the degree of**

**Bachelor of Engineering**

**In**

**Information Technology**

**By**

**ANKIT KUMAR SINGH**

**KARAN NAINWAL**

**PANKAJ MEHRA**

**RAKESH KUMAR**

**Under the guidance of**

**Prof. Ashwini Sapkal**

****

****

**CERTIFICATE**

This is to certify that the project report entitled **“BREAST CANCER DETECTION USING ARTIFICIAL INTELLIGENCE”** being submitted by **KARAN NAINWAL (B120228529)** is a record of bonafide work carried out by him under the supervision and guidance of **Prof. Ashwini Sapkal**  in partial fulfillment of the requirement for **BE(IT)** course of Savitribai Phule Pune University, Pune in the academic year 2016-2017.

Prof. Sangeeta Jadhav Prof. Ashwini Sapkal

(*Head of Department) (Professor)*

*Information Technology* *Information Technology*

Dr. B.P. Patil

(*Principal)*

*AIT, Pune*

**ACKNOWLEDGEMENT**

It is our proud privilege and duty to acknowledge the kind of help and guidance received from several people in preparation of this project report. It would not have been possible to prepare this report in this form without their valuable help, cooperation and guidance.

First and foremost, we wish to record our sincere gratitude to **Prof. Ashwini Sapkal**  for his constant support and encouragement in preparation of this report and for making available the required literary resources.

Our sincere thanks to **Dr. (Mrs.) Sangeeta Jadhav,** Head of Department, Information Technology, AIT, Pune for her valuable suggestions and guidance throughout the period of this project ideation. Last but not the least, my heartiest gratitude goes to my project partners for being there throughout this period.

KARAN NAINWAL

(B120228529)

**CONTENTS**

* ABSTRACT………………………………………………………………………...05
* LIST OF ABBREVIATIONS………………………………………………………06
* LIST OF FIGURES…………………………………………………………………06
* LIST OF TABLES…………………………………………………………………. 06

**SRNO. CHAPTER PAGE NO.**

1. INTRODUCTION………………………………………………………….07
   1. Need…………….……………………………………………………...07
   2. The Proposed System……………….………………………………….08
   3. Improving the System…………………………………………………..10
2. LITERATURE SURVEY…………………………………………………..12
3. PROBLEM STATEMENT………………………………………………….14
4. PROJECT REQUIREMENTS……………………………………………...16
5. SYSTEM ANALYSIS……………………………………………………...17
6. PROJECT PALN………………………………………………………...…18
7. CONCLUSION…………………………………………………………… 19
8. REFERENCES…………………………………………………………..... 20

ABSTRACT

Breast Cancer is one of the main reasons of the death among women due to diseases. It accounts for the second-most deaths in woman every year. Mammography is the most popular detection method and there are several CAD (Computer Aided Design) techniques for the detection using mammogram images.

Based on World Health Organization (WHO) statistics, cancer was responsible for 13% of all deaths in the world by 2004, an estimated 12 million people dying from cancer in 2030.

Mammography offers high-quality images at a low radiation dose, and is currently the only widely accepted imaging method used for routine breast cancer screening.

One of the difficulties with mammography is that mammograms generally have low contrast. This makes it difficult for radiologists to interpret the results. So a need of CAD was experienced to provide a second opinion to the radiologist using Machine Learning techniques.

The new system uses three algorithms for the detection of cancer using mammogram images. We have used Back propagations, SVM (Support Vector Machine), and the advanced algorithm SVM with MGSA (Mixed Gravitational Search Algorithm).

The system uses the above mentioned three algorithms and compares the performance and efficiency of each algorithm on a particular dataset of mammogram images and detects whether the cancer is present on a given sample,

Firstly the BP and SVM are applied and then the SVM is improved using MGSA.

The database used is the Wisconsin Breast Cancer Database and the languages used are Java and Python with different machine learning libraries. The system aims to provide a detection of the mammogram images and provide a comparison between different algorithms.

In a country like India where the medical facilities are slow in their delivery there is a shortage of work force and medical experts so there is a urgent need to use the Information Technology for fast delivery off medical services.

## LIST OF ABBREVIATIONS

## 

**Sr. No.**   **Short**   **Full**

1. BP Backpropagation
2. SVM Support Vector Machine
3. AI Artificial Intelligence
4. NN Neural Network
5. MGSA Mixed Gravitational Search Algorithm

## 

**LIST OF FIGURES**

**Sr. No. Figure Name Page No.**

1. Working of a Detection System……………………………………..9
2. Improving SVM with MGSA………………………………………11

**LIST OF TABLES**

**Sr No. Table Name Page No.**

1. Comparison of BP ……………………………………………….12
2. Comparison of SVM………………………………………………13
3. Planning and Schedule table………………………………………18

**Chapter-1**

**INTRODUCTION**

According to the World Health Organization, breast cancer is the most common cancer among women worldwide, claiming the lives of hundreds of thousands of women each year and affecting countries at all levels of modernization.

There are several imaging techniques for examination of the breast, including magnetic resonance imaging, ultrasound imaging, and X-ray imaging.

Mammography is a specific type of imaging that uses a low-dose X-ray system to examine the breast, and is currently the most effective method for detection of breast cancer before it becomes clinically palpable.

Mammography offers high-quality images at a low radiation dose, and is currently the only widely accepted imaging method used for routine breast cancer screening.

* 1. **Need**

One of the difficulties with mammography is that mammograms generally have low contrast. This makes it difficult for radiologists to interpret the results.

Double reading of mammograms has been advocated to reduce the proportion of missed cancers. The basic idea of double reading is to have two radiologists read the same mammograms.

There are following benefits and needs of the CAD:

* **Better Result -** CAD provides a second opinion to the radiologist and it eliminates the need of second opinion form other radiologist. It also improves the detection capability of a system as whole
* **Cost Efficiency -** The cost of a CAD system is comparatively low when compared to the cost of radiologist as it requires special technical skills.

Also, these systems have low maintenance cost .

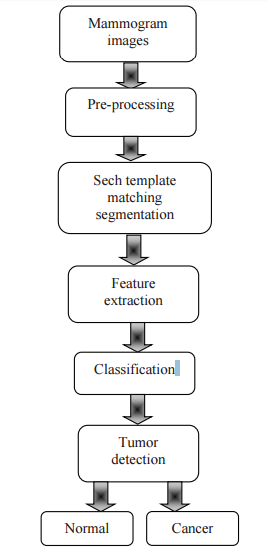
* **Faster Result –** Using the several machine learning algorithms and advanced hardware, faster detection and diagnosis of the mammogram images is possible.
* **Sustainable -** These systems are quite sustainable as they are very much adaptive to new changes in the algorithm and new techniques. So these systems can be modified very quickly.
  1. **The Proposed System – Introduction**

The system uses three algorithms to detect the present of cancer in a given sample of mammogram image :

* Back Propagation
* Support Vector Machine(SVM)
* SVM with MGSA (Mixed Gravitational Search Algorithm)

The basic system has the following steps involved only the algorithm used changes in each case :

* Image Preprocessing
* Segmentation
* Feature Extraction
* Classification
* Tumor Detection



*Fig 1:Working of a Detection system*

We have worked on the following ten features of a mammogram image:

a) radius (mean of distances from center to points on the perimeter)

b) Texture (standard deviation of gray-scale values)

c) Perimeter

d) Area

e) Smoothness (local variation in radius lengths)

f) Compactness (perimeter^2 / area - 1.0)

g) Concavity (severity of concave portions of the contour)

h) Concave points (number of concave portions of the contour)

i) Symmetry

j) Fractal dimension ("coastline approximation" - 1)

We have used the Wisconsin Breast Cancer Database provided by the UCI Repository.

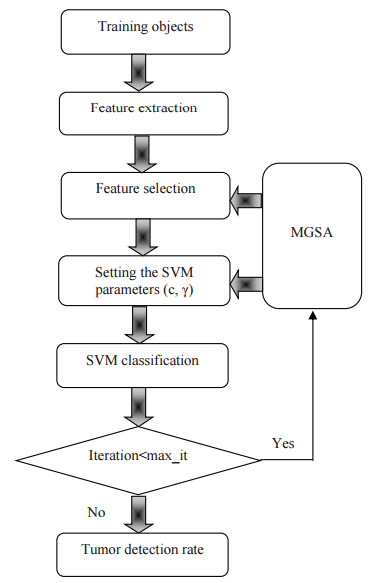
* 1. **Improving the System**

We apply the Back propagation and SVM to the database and compare these algorithms on the basis of efficiency and accuracy. Then we found the need to improve the SVM as the existing system with SVM has low efficiency so SVM with MGSA is introduced in the system.

If all of these extracted features (such as mean, variance, entropy, Correlation and etc.) from images are used, the processing time is increased and the training of the SVM would be complicated.

The aim is to optimize the evaluation function that is the precision of the SVM classifier. After training, the selected features and tuned SVM are used for detecting target objects.

MGSA improves two parameters of the SVM. These parameters are the kernel Parameter, gamma (ϒ, (for the radial basis functions (RBF) kernel and penalty parameter (C).



*Fig 2: Improving SVM with MGSA*

All the three algorithms are applied and the results of all are compared and the detection is done and the result is displayed. Also the result of each feature is also displayed and the final result is also displayed.

**Chapter-2**

**LITERATURE SURVEY**

A survey was done on various CAD techniques to detect Breast Cancer and out of different algorithms. Back Propagation and SVM was selected due to better efficiency and accuracy of results. A comparative study was done on these algorithms on the basis of efficiency and accuracy.

During the survey a need to improve the SVM algorithm so the MGSA was introduced to improve feature selection in SVM.

Machine learning methods have been demonstrated to generate powerful classifiers. However, in many instances, there could be a risk of over fitting the data if these methods were not properly validated on independent datasets or tested using statistical resampling methods.

|  |  |  |
| --- | --- | --- |
| Method | Error | Number of epochs |
| BP | 0.989 ± 0.045 | 1000 |
| MPANN | 0.981 ± 0.005 | 5100 |
| Fogel et. Al | 0.981 ± 0.464 | 200000 |

*Table 1: Comparison of BP*

Introduction of MGSA in the SVM is used to o decrease the number of features and to improve the SVM classification accuracy. Then, the selected features and the SVM classifier (tuned) are used for detecting Breast Cancer.

The experiments also proved that the system is able to optimize the SVM parameters and feature selection for the breast cancer detection.

|  |  |  |
| --- | --- | --- |
|  | PERFORMANCE | |
| Method | SVM | MGSA-SVM |
| Classification Rate (%) | 86 | 93.1 |
| Number of features | 24 | 12 |

*Table 2:Comparison of SVM*

Both the BP and SVM with some optimization have been proved to be the best choice for Cancer Detection in mammogram images. When compared with other algorithms these were proved to be the best on the basis of efficiency and accuracy.

**Chapter-3**

**PROBLEM STATEMENT**

**3.1 PROBLEM STATEMENT**

Develop an Artificial Intelligent system to detect cancer in mammogram images to provide support to the manual diagnosis system and also provide comparison between different algorithms on the basis of accuracy and efficiency.

**3.2 PROJECT OVERVIEW**

The development of the proposed system would bring into consideration the following user scenario wherein:-

1. The user (radiologist) will login into the system.
2. The radiologist will provide the preprocessed data of the digital mammograms.
3. The input can be in unformatted form or in form of feature matrix.
4. The user will then command the system to perform detection.
5. The system will then display the result. The result will be in two parts :
6. Yes or No : Cancer Present
7. Analysis of every feature
8. Now the radiologist can use the above result as a second opinion.

The system also provides research opportunities to the developer in the following ways:

* Stores the result of every detection and stores the data of each algorithm.
* This data can be further used to compare the different algorithm on regular basis.
* On the basis of comparison the best method can be selected and that method is selected as primary methods
* Remarks from the radiologist are also collected so as to improve the system.

The system can also be useful to the patient in following ways:

* The data of all the patients is stored in the database.
* Data is saved in both the data form also in digital form.
* The data can also be used to track the health of a patient
* Patient can directly collect their reports from the online system.

**Chapter-4**

**PROJECT REQUIREMENTS**

For the Breast Cancer Detection system to function effectively, following hardware and software requirements must be met.

* **HARDWARE REQUIREMENTS**

For effective functioning of the portal, you’ll need a computer that meets these minimum requirements:

* GHz Intel Pentium Processor (or equivalent).
* 1GB of system memory (2GB recommended).
* At least 2 GB free disk space.
* **SOFTWARE REQUIREMENTS**

Following are the software requirements for the web portal:

* A computer running one of the following operating systems:
* Microsoft Windows XP or later.
* Fedora Core 10.0 or later.
* MacOS 10.4 or later.
* Java Runtime Environment (JRE) version 5 or later.
* Python
* TCP/IP port 3306 free for MySQL service.(To Store Reports)

**Chapter -5**

**SYSTEM ANALYSIS**

1. **Proposed Architecture**

The system will be provided with following specifications:

* A user Interface for to upload images.
* Option to upload the preprocessed data directly.
* The user can check for the result.
* Classification of the image ,whether cancer is present or not
* Specific data for the features can also be provided.
* Comparison between different method of detection
* Selecting the best method on the basis of comparison on regular basis.
* Storing the reports and results on the database.

1. **High Level Design of The Project** (Future Advancements)

* The system can be implemented online with a central database.
* Different hospitals can access the system remotely.
* Patients can directly access there reports.
* Data and results can be used for research and in optimization of results.
* In the current system the image are not preprocessed but we can implement the system with preprocessing also

**Chapter-6**

**PLANNING AND SCHEDULING**

|  |  |  |
| --- | --- | --- |
| ***Sr. No.*** | ***Period*** | ***Activity*** |
| 1. | November-December | Ideation |
| 2. | January | User and system requirement analysis |
| 3. | February | System Design |
| 4. | March-April | Coding and Development |
| 5. | May | Testing and Deployment |

*Table 3:Planning and scheduling*

**Chapter-7**

**CONCLUSION**

The proposed system can be implemented using Back propagation and SVM. The system now can be implemented after literature survey and requirement analysis.

Now, on the basis of required information we can start the design and coding phase of the system using different platforms and languages.

This system will be very useful for the detection of Breast Cancer , it will provide a quick second opinion to the radiologist ,It will also prove to be cost efficient and will be able to manage the reports and data of the patient and will also be helpful in tracking the health of different patients.

The system can be utilized by different diagnosis laboratory and can provide a second opinion .It will make the process efficient and faster in a country like India where the medical facilities are very slow in their deliveries.

This system also provides research opportunities to the developers also as it stores the data and the result and provides a comparison between different methods of detection.

**REFERENCES**

* **Base Paper**

Detection of cancer tumors in mammography images using support vector machine and mixed gravitational search algorithm, 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC2016), Higher Education Complex of Bam, Iran, **2016,** **Esmat Rashedi**, **Fatemeh Shirazi.**

* An Evolutionary Artificial Neural Network Approach for Breast Cancer Diagnosis, Third International Conference on Knowledge Discovery and Data Mining,**2010,** **Lijuan Liu**, **Mingrong Deng.**
* “GSA: a Gravitational Search Algorithm,” Information Sciences. J, vol. 179, pp. 2232–2248, June **2009**,**E. Rashedi** , **H. Nezamabadi-pour** , **S. Saryazdi,**
* Breast Cancer Detection Using BA-BP Based Neural Networks and Efficient Features,**2011,** **Alireza khosravi** , **Jalil Addeh** , **Javad Ganjipour.**
* <http://wikipedia.org>