

Manipal Academy of Higher Education

Impressions@MAHE

Manipal College of Health Professions, Manipal
Theses and Dissertations

MAHE Student Work

Spring 7-1-2022

**DEVELOPING AND TESTING OF A MACHINE LEARNING
PREDICTIVE MODEL FOR THE EARLY DETECTION OF BENIGN
AND MALIGNANT BREAST LESIONS**

RINSY R

Follow this and additional works at: <https://impressions.manipal.edu/mcph>



Part of the [Medicine and Health Sciences Commons](#)

**DEVELOPING AND TESTING OF A MACHINE LEARNING
PREDICTIVE MODEL FOR THE EARLY DETECTION OF BENIGN
AND MALIGNANT BREAST LESIONS**



A dissertation submitted in partial fulfillment of the award of

MSc. Health Informatics(Healthcare IT Management)

Degree to

MANIPAL ACADEMY OF HIGHER EDUCATION

By

RINSY R [201198001]

Under the Guidance of

Ms. DOLA SAHA

Assistant Professor

Department of Health Information Management

MCHP, MAHE, Manipal

JULY 2022

ABSTRACT

Background: Breast cancer is the uncontrolled growth of cells in the breast called tumor. This tumor can be of two forms benign or cancerous tumor. A cancerous tumor is malignant, as it can grow and spread to other parts of the body. A benign tumor can grow but does not spread. According to WHO 2.3 million women were diagnosed with breast cancer, leading to 6,85 000 deaths globally in the year of 2020 ^(1,2). If the breast cancer is identified in the early stage then the treatment can be highly effective. Machine learning is one of the trending technologies that have been widely used for breast cancer prediction and diagnosis. Since prediction and diagnosis of cancer is important for a healthy life, and the number of deaths due to cancer are increasing each year, accuracy in prediction plays a significant role. Machine Learning can be an answer to similar cases, whereby machines can be trained to offer effective predictions ⁽³⁾.

Objectives: To design and determine a machine learning predictive model for the early detection of benign and malignant breast lesions using various machine learning algorithms. To evaluate and identify the most effective algorithm with respect to specificity, sensitivity and accuracy for the early detection and differential diagnosis of benign and malignant breast lesions

Methodology: A machine learning predictive model was developed for the early detection and differential diagnosis of benign and malignant breast lesions. The data are obtained from the secondary data base i-e Wisconsin (diagnostic) dataset. The model was developed with the help of five machine learning classification algorithms. The machine learning algorithms which used are Logistic Regression, Linear support vector classification, Support vector classification- radial basis function, Decision tree classifier and Random forest classifier. Then the quality of prediction from a classification algorithm is measured with the help of confusion matrix. The quality of prediction is determined by the parameters Accuracy, Precision, Sensitivity/Recall, Specificity/ Support and F1 Score. The model development and quality calculation are performed with the help of python programming language.

Data analysis:

1. Feature analyses are performed using python programming language.
2. Sensitivity/Recall, Specificity/Support, Accuracy, Precision and F1 score are calculated using the Confusion Matrix.

Results: The result shows that the Support Vector Classifier Radial Basis Function(SVC - RBF) is the best algorithm for the early detection of benign and malignant breast lesions. SVC-RBF have the accuracy of 0.99.

Conclusion: The developed SVC RBF model proves to be the most efficient one. The SVC RBF model gives better results in analyzing breast cancer dataset with high accuracy, precision, Sensitivity/recall, Specificity/Support and F1 score. SVC RBF model can be used for the differential diagnosis of the benign and malignant breast lesions. This is in line with the various other studies done earlier which recommend SVM as the best algorithm for the early detection of breast cancer.