# A Comparative Study of Machine Learning Algorithms Applied to Cancer Diagnosis Specific to Breast Cancer

**Ankur Gupta[1*], Rajdeep Singh[2], Krishan Kumar[3], Piyush Kumar[4], Arvind Kumar Shukla[5]**

[1,2] School of Computer Science and Applications, IFTM University, Moradabad, U.P., India

[3]Department of Computer Science, Gurukula Kangri (Deemed to be University), Haridwar, 249404, Uttarakhand, India

[4]Department of Radiation Oncology, Shri Ram Murti Smarak Institute of Medical Sciences, Bareilly, U.P., India

[5] Department of CS & IT, Mahatma Gandhi Central University, Motihari, Bihar (India)

Corresponding Author Email: Ankur Gupta

## Abstract

Improvement of the fact at which patients' survival and treatment results break is that which we see out of very early detection of breast cancer. Presently we see that traditional diagnostic methods do in to play to that of human subjectivity, tiredness as a factor in performance and also variable interpretation which is a large issue because they mostly use radiology and pathology for evaluation. These issues put into light the need for data based, automated approaches which in turn will give out very reliable and unbiased results. To have an objective performance evaluation we pre-processed the breast mass image data which included features from fine-needle aspirates. We used the same settings for training and testing each model to ensure evaluation consistency. We found that ensemble models which include Random Forest and Gradient Boosting performed the best out of the studied algorithms with over 98.8% accuracy. Also, these models had better recall and AUC which in turn showed their performance in differentiation between benign and malignant cases. Also, we saw that which statistical analysis, confusion matrices, and descriptive visualizations. What we found is that ensemble-based machine learning does in fact out perform traditional models in the case of breast cancer which we put forth as a reliable and scalable solution for early detection. In health care settings this type of predictive modelling may improve the accuracy of cancer screen out comes, reduce in diagnostic errors, and support clinical decision making.

***Keywords:*** *Machine Learning, Feature Importance, ROC Curve, Classification, Breast Cancer, and Comparative Analysis*

## I. Introduction

Each year breast cancer reports as a very large cause of death from cancer which also has the highest incidence and is the greatest health issue in the field of oncology [1], [5]. We see that early and accurate diagnosis is what really improves outcomes which in turn see to better results in terms of survival. Although we have seen great technical growth in the field of diagnostics, we still see mainly the use of mammography, ultrasounds, and histopathological analysis which at their root still require the very subjective manual interpretation by radiologists and pathologists which in turn brings in issues of variable diagnosis between readers and also in some cases diagnostic error which may be a result of human fatigue [6]. Also, these issues play a large role in the quality of diagnosis in very busy clinical settings and in health care structures which are limited in resource [7].

In Fig. 1 we present a case of breast cancer which displays the most tell tissue feature of cancer a spiculated mass. What we see is that when a tumor produces fibrotic tissue which extends out like spikes the result is spiculated margins which in turn point to invasive growth. Thus, these imaging features stress the importance of accurate image interpretation.