# ***Breast Cancer Prediction through Machine Learning***

by

Vaishnavi Tandon

Date:5/01/23

*"First forget inspiration. Habit is more dependable. Habit will sustain you whether you're inspired or not. Habit will help you finish and polish your stories. Inspiration won't. Habit is persistence in practice."*

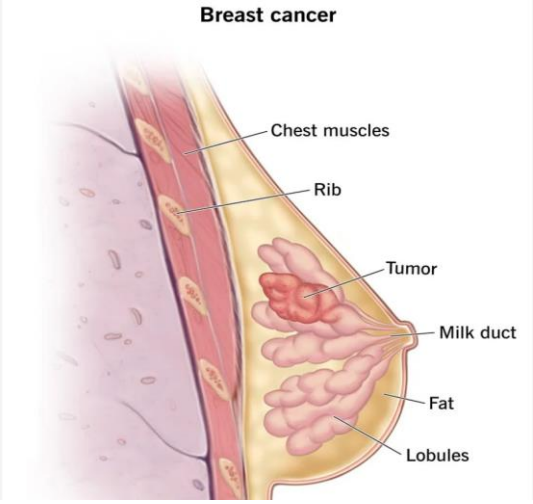
*—Octavia Butler*

## ***Abstract:***

*Machine learning is a growing technology which enables computers to learn automatically from past data. Machine learning uses various algorithms for building mathematical models and making predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender system, and many more. One of the major use of machine learning in Health care. Breast cancer is cancer that forms in the cells of the breasts. Breast cancer is the most common cancer diagnosed in women all over the world. Breast can occur both in both men and women, but it's in far more common in women. Substantial support for breast cancer awareness and research funding has helped create advances in the diagnosis and treatment of breast cancer. The ability to diagnose this issue quickly and accurately is crucial for sparing patients' lives and limiting further damage. To assist patients in determining whether they have breast cancer or not, creating an intelligent learning-based diagnostic tool will make it easier for them to diagnose a sickness and receive timely treatment.*

## ***Problem Statement:***

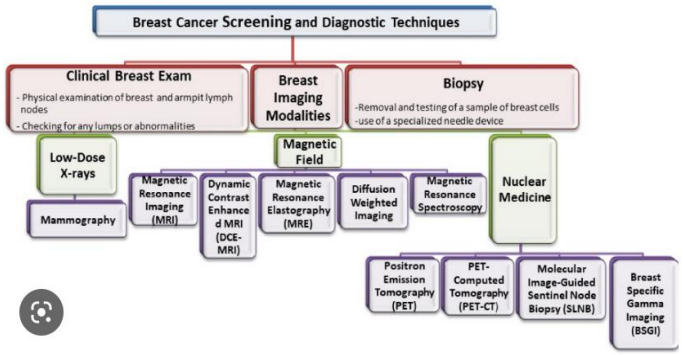
*Breast cancer develops when cells in your breast multiply and expand out of control, resulting in a lump of tissue known as a tumor. Breast cancer symptoms can include feeling a lump, noticing a change in breast size, or noticing changes to the skin around your breasts. The second most frequent malignancy in women after skin cancer is breast cancer. Over-50-year-old women are the ones most likely to be affected. Breast cancer develops when cells in your breast multiply and expand out of control, resulting in a lump of tissue known as a tumor. Breast cancer symptoms can include feeling a lump, noticing a change in breast size, or noticing changes to the skin around your breasts. The second most frequent malignancy in women after skin cancer is breast cancer. Over-50-year-old women are the ones most likely to be affected***.**



*Conventional techniques frequently result in inaccurate diagnoses and take longer because of human error. Additionally, it requires time for assessment and is a very expensive, computationally complicated, and intricate method to disease diagnosis. Researchers worked to build several non-invasive smart healthcare systems based on accurate machine learning algorithms in an effort to address the problems with conventional invasive-based procedures for the early detection of breast cancer. These systems use the fundamental data provided by the patient to determine whether or not they are likely to get breast cancer. Making a system that is intelligent and capable of detecting breast cancer with the highest degree of accuracy feasible overcomes the problems of expensive tests and erroneous test results.*

## ***Market/Customer/ Business need assessment***

*The cost of diagnostic tests is currently prohibitive. Therefore, if a machine learning diagnostic system is employed instead, a significant amount of resources can be saved, which lowers the cost of diagnosis. As a result, there will be a greater market demand for such systems, and the revenue raised from diagnostic testing can go toward supporting the development of such AI systems.*




*Everyone wants the services to be more precise and timely in this age of the internet and technology. These highly accurate AI-based tools and systems will save time and money while making it simple for patients to receive a timely diagnosis of their ailment.*

## ***Target Specifications and characterization***

Targets or purposes for using such tools:

- Identify the patient's illness with the greatest degree of accuracy.

- To encourage the use of contemporary technologies and instruments over conventional procedures and to lower the cost of diagnostics.

- To support the AI industry and emerging technologies.

The targets/objectives described above can be achieved by:

- Obtaining accurate information from a variety of trustworthy sources.
- Conducting surveys to determine the relationships between the data collected Analyzing the data to identify patterns.
- Correctly preprocessing the data, using it to plot numerous insights, applying multiple Machine Learning algorithms to the data, and selecting the optimal model based on the accuracy score.
- Developing a user-friendly, cost-efficient diagnosis system.
- Suggest using the system rather than the outdated, ineffective alternatives.
- Invest the money raised from the system's success in the creation of more AI technologies.

## ***External Search:***

*Dataset: Breast Cancer Prediction from Kaggle*

*Link to dataset:* https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset

*Description of the dataset:*

*This dataset from Kaggle has a total of six columns, the final of which contains the results of the prediction that needs to be made. Based on the patient-inserted health information, we are requested to determine whether the patient has breast cancer or not. It includes many details that are crucial for determining if the patient may or may not have breast cancer, including*

- *Mean radius*
- *Mean texture*
- *Mean perimeter*
- *Mean area*
- *Mean smoothness*
- *Diagnosis-0-false,1-True*

*Link to my GitHub: https://github.com/vaish9499/Feynn_labs/blob/main/ml.py*

*Here, I've utilized a Naive Bayes Classifier to make outcomes predictions. Given the values of the features, the aim in the supervised learning method naive bayes for classification is to determine the class of the observation (data point). A naive bayes classifier (p(yi | x1, x2,..., xn)) determines the likelihood of a class given a set of feature values. To extract more valuable information from a dataset, we can utilize different classification algorithms in machine learning.*

## ***Mammography:***

*Mammography is specialized medical imaging that uses a low-dose x-ray system to see inside the breasts. A mammography exam, called a mammogram, aids in the early detection and diagnosis of breast diseases in women.*

*An x-ray exam helps doctors diagnose and treat medical conditions. It exposes you to a small dose of ionizing radiation to produce pictures of the inside of the body. X-rays are the oldest and most often used form of medical imaging.*

*Three recent advances in mammography include digital mammography, computer-aided detection and breast tomosynthesis.*

*Digital mammography, also called full-field digital mammography (FFDM), is a mammography system in which the x-ray film is replaced by electronics that convert x-rays into mammographic pictures of the breast. These systems are similar to those found in digital cameras and their efficiency enables better pictures with a lower radiation dose. These images of the breast are transferred to a computer for review by the radiologist and for long term storage. The patient's experience during a digital mammogram is similar to having a conventional film mammogram.*

*Computer-aided detection (CAD) systems search digitized mammographic images for abnormal areas of density, mass, or calcification that may indicate the presence of cancer. The CAD system highlights these areas on the images, alerting the radiologist to carefully assess this area.*

*Breast tomosynthesis, also called three-dimensional (3-D) mammography and digital breast tomosynthesis (DBT), is an advanced form of breast imaging where multiple images of the breast from different angles are captured and reconstructed ("synthesized") into a three-dimensional image set. In this way, 3-D breast imaging is similar to computed tomography (CT) imaging in which a series of thin "slices" are assembled together to create a 3-D reconstruction of the body.*

*Although the radiation dose for some breast tomosynthesis systems is slightly higher than the dosage used in standard mammography, it remains within the FDA-approved safe levels for radiation from mammograms. Some systems have doses very similar to conventional mammography.*

*Large population studies have shown that screening with breast tomosynthesis results in improved breast cancer detection rates and fewer "call-backs," instances where women are called back from screening for additional testing because of a potentially abnormal finding.*

*Breast tomosynthesis may also result in:*

- *earlier detection of small breast cancers that may be hidden on a conventional mammogram*
- *fewer unnecessary biopsies or additional tests*
- *greater likelihood of detecting multiple breast tumors*
- *clearer images of abnormalities within dense breast tissue*
- *greater accuracy in pinpointing the size, shape and location of breast abnormalities*

## ***Benefits***

- *Screening mammography reduces the risk of death due to breast cancer. It is useful for detecting all types of breast cancer, including invasive ductal and invasive lobular cancer.*
- *Screening mammography improves a physician's ability to detect small* ***tumors****. When cancers are small, the woman has more treatment options.*
- *The use of screening mammography increases the detection of small abnormal tissue growths confined to the* ***milk ducts*** *in the breast, called* ***ductal carcinoma in situ (DCIS)****.*
- *No radiation stays in your body after an x-ray exam.*
- *X-rays usually have no side effects in the typical diagnostic range for this exam.*

## ***Risks***

- *There is always a slight chance of cancer from excessive exposure to radiation. However, given the small amount of radiation used in medical imaging, the benefit of an accurate diagnosis far outweighs the associated risk.*
- *The radiation dose for this procedure varies.* See the *Radiation Dose in X-Ray and CT Exams* page for more information about radiation dose.
- *False Positive Mammograms. Five percent to 15 percent of screening mammograms require more testing such as additional mammograms or ultrasound. Most of these tests turn out to be normal. If there is an abnormal finding, a follow-up or biopsy may have to be performed. Most of the biopsies confirm that no cancer was present. It is estimated that a woman who has yearly mammograms between ages 40 and 49 has about a 30 percent chance of having a false-positive mammogram at some point in that decade and about a 7 percent to 8 percent chance of having a breast biopsy within the 10-year period.*
- *Women should always tell their doctor and x-ray technologist if they are pregnant.* See the *Safety in X-ray, Interventional Radiology and Nuclear Medicine Procedures* page for more information about pregnancy and x-rays.

## ***Business model***

*Finding a way to diagnose this illness in the shortest amount of time and for the least amount of money is crucial given the rise in the number of patients with breast cancer and the hazards they face. By taking into account the aforementioned elements, it is crucial to create a computationally intelligent system that can accept patient data, analyze it using algorithms, and determine if the patient has a probability of developing breast cancer or not. By creating such a model, the price of this diagnostic test decreases, and the cash generated from it may be used to create further AI-based business models similar to this one by funneling the profits from its success into the creation and improvement of additional AI based tools and systems.*

## ***Concept Generation***

*The Concept Generation has their following reasons:*

- *To increase the ability to diagnose Breast Cancer with the greatest degree of precision.*
- *To assist the underprivileged and provide timely, low-cost disease diagnoses*
- *Making people aware of the most recent and cutting-edge technologies, such as AI, IoT, etc., and their potential to improve as well as make life easier for people.*
- *Developing the AI industry by investing the profits from this model in the development of more models based on machine learning, deep learning, data science, and artificial intelligence.*

## ***Final Product Type***

*The previous dataset for a patient who has experienced breast cancer is the proposed model for the investigation. It determines whether or not to forecast a breast cancer based on a collection of dataset attributes and provides a minimum level of accuracy.*

## ***Conclusion:***

*Machine learning has the potential to transform conventional methods of operation, not just in the medical field but in all other fields as well. AI and machine learning are applied in practically all fields nowadays. The dataset needed to study these use-cases is continually growing in this age of the internet and growing amounts of data. Deep learning is also being applied in various applications with the growing amount of data. Machine learning algorithms will also be trained to continuously increase their accuracy and deliver accurate and exact information in the form of results as all industries advance. People can become more aware of AI's potential and capabilities by using successful implementations of these applications.*

***References:***


- ***https://www.radiologyinfo.org/en/info/mammo***
- ***https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset***
- ***https://www.kaggle.com/datasets/merishnasuwal/breast-cancer-prediction-dataset***