| | |
|---|---|
| | *Please fill in the following – and title your PDF in the same order, for example:* **Khan, Aleksandra, 123456789, EPP, 23-24.pdf** |
| Student Surname | Chowdhury |
| First Names | Musa |
| Student Number | 97113521 |
| Course Name and short code (example: Fashion, FAS) | Y3 C18301 Dissertation 23-24 |
| Academic year | 23-24 |
| Unit Title | Dissertation |
| Unit Code | C18301 |
| Year, Term | Year 3, Semester 1 |
| | *Please provide the following:* |
| Six keywords (separated by commas) | AI . MACHINE LEARNING, Mammogram, breast cancer , methods of detection |
| Your LEAD QUESTION | Evolution of Diagnostic methods for breast cancer and the new integration of Ai technology into breast cancer diagnosis |
| Alternative Dissertations: place a link to the practical element (and password if needed) here – and again on the title page | File name : Breast cancer Musa c<br><br>https://www.dropbox.com/scl/fi/ultgnwcylkwr2pnteve0z/breast-cancer-Musa-c.ipynb?rlkey=859l5z27c2bulyuomr2k2113s&dl=0<br><br>datasheet<br>https://www.dropbox.com/scl/fi/5hsx4pz2ncjx50olnbknd/wisc_bc_data.csv?rlkey=9suqa1kddzyxci035xskj7qs7&dl=0 |

By submitting this document:

Musa Chowdhury project base

1. I certify that this assignment is my/our own work and that I am familiar with Ravensbourne's Plagiarism Policy. I also understand that plagiarism is a serious academic offence.
2. I certify that this assignment has not been previously submitted for assessment on this programme.
3. Where material has been used from other sources it has been properly acknowledged.
4. I confirm that I have retained an electronic copy of this assignment and understand that written assignments may be submitted to the JISC Plagiarism Detection Service, I must therefore be able to produce electronic copies of written assignments.
5. I understand that Ravensbourne is at liberty to delete submitted work 12 months after assessment.
6. I also understand that Ravensbourne may wish to use my work (or copy) for future academic purposes in accordance with Ravensbourne's regulations.

Ravensbourne University London

**Context page**

Introduction

Clinical Breast Examination (CBE)

How is the clinical breast examination conducted

Mammogram

How does mammograms work to detect breast cancer

The Early Adoption and Emergence of Mammography

What's the next major advancement for breast cancer detection ?

What is machine learning?

What is Artificial neural network (ANN)

What is Convolutional Neural Network (CNN)

Why I will use the ANN algorithm rather than the CNN algorithm.

Personal Breast cancer detection project using

Machine learning

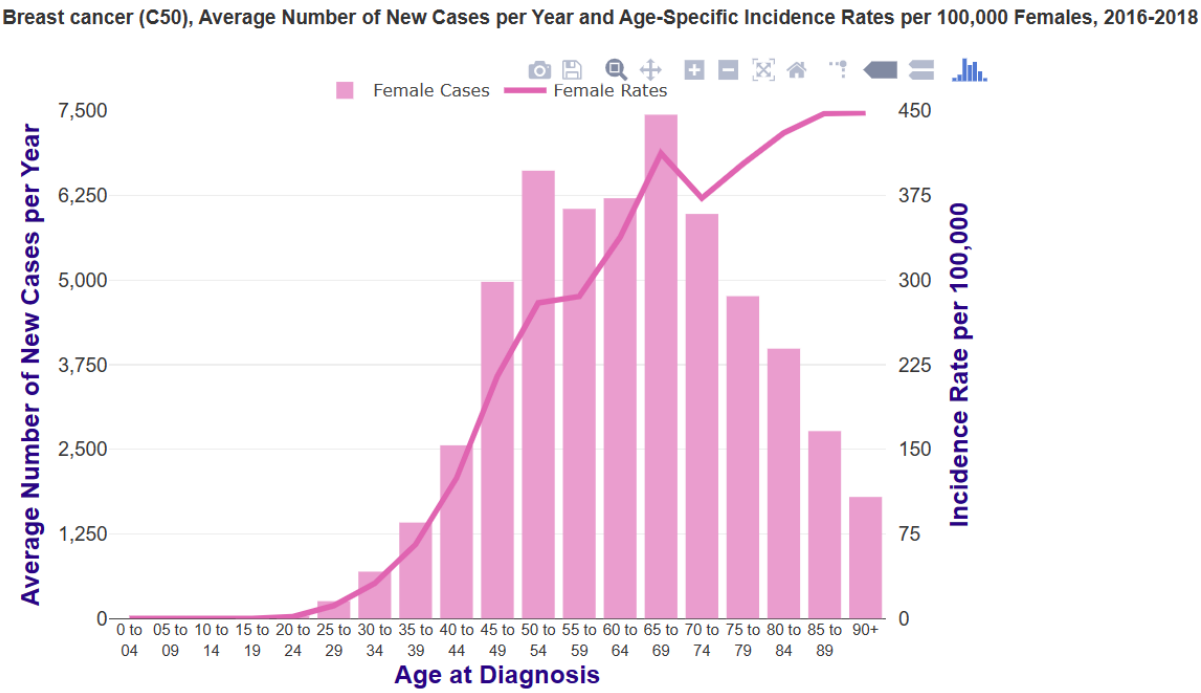
Interviews

Bibliography

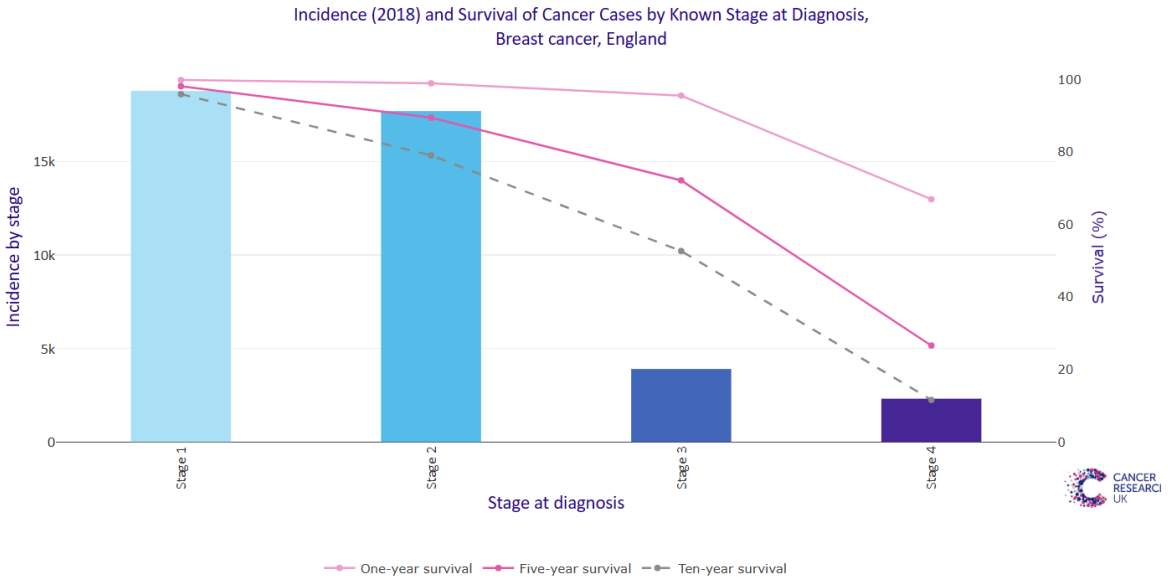
## Introduction

Breast cancer diagnosis has undergone a remarkable evolution in recent decades, with diagnostic methods becoming increasingly sophisticated and precise.

From the years 2016 – 2018 research from cancer research uk shows there was around 55,000 new cases of breast cancer in the uk. That's around 150 every day. It also shows breast cancer is the most common form of cancer in the uk. Projections show that rate could increase by 1% by 2028 – 2040. With this increase it puts more people in the uk at risk of death as studies conducted from the years 2017 to 2019 breast cancer in 2nd most common type of cancer that results in death. Thankfully within the last yr the death count has decreased drastically by 18% overall according to research from 2017-2019. show



Early diagnosis is essential to enhance the results of treatment and, eventually, saving lives. As early diagnosis will reduce death massively. A study conducted showing the survival rate of breast cancer patients by stages show, earlier identification of breast cancer will reduce death rates massively.




Cancer Research UK (no date) *Early diagnosis data hub*, *Early Diagnosis*. Available at: https://crukcancerintelligence.shinyapps.io/EarlyDiagnosis/

As shown by the study survival rates drastically drop as breast cancer reaches later stages, stages 3 and 4. The research carried out by breast cancer uk goes from the year 2018 to 2020.

The three main techniques for diagnosing breast cancer will be examined in this dissertation, along with their historical development, advantages, and progressive advancements in diagnostic accuracy. The first technique used by health professionals , clinical breast examination (CBE), served as the foundation for the early identification of breast cancer. Breast cancer diagnosis began when doctors used touch and their observational skills to find palpable anomalies. Then came the revolutionary period of mammography, which provided a more sophisticated method that could identify tiny lesions that were missed by hand. Mammography became a cornerstone in breast cancer screening, dramatically boosting the prospects of early detection and intervention.

Breast magnetic resonance imaging (MRI) became a useful adjunctive tool as technology progressed. MRI overcame the limits of mammography by offering detailed images of breast tissue, particularly in situations when the results were equivocal or the breast tissue was dense. This breakthrough demonstrated the value of a multifaceted diagnostic strategy, where each technique adds something special to the whole diagnostic environment.

But with the addition of cutting-edge technologies, the progression of breast cancer diagnosis is not static and is set to continue evolving. The integration of artificial intelligence (AI) and machine learning in breast cancer diagnostics is a paradigm shift that this dissertation will examine. With the

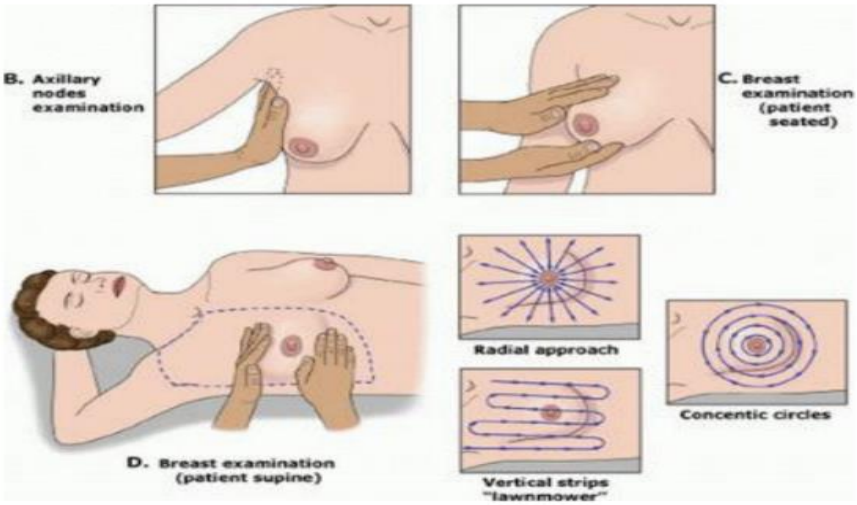
potential to transform breast cancer diagnosis, these technologies present a huge opportunity as we approach a new chapter in the history of healthcare. By developing a new test diagnosis approach with AI and machine learning, I hope to shed light on the promising future in which data-driven insights will improve decision-making speed, improve diagnostic accuracy, and ultimately aid in the ongoing fight against breast cancer.

## Clinical breast examination (CBE)

In the annals of breast cancer diagnosis, the Clinical Breast Examination (CBE) is the foundation. Developed in an era before advanced imaging capabilities, CBE was the first technique used by medical experts to identify palpable breast abnormalities. The medical community's practical approach, which included manual palpation and visual inspection, cleared the way for the first attempts to recognise breast cancer symptoms.

## How is the clinical breast exam conducted ?

A qualified healthcare professional thoroughly examines and feels the breasts using both visual and physical means during a CBE in order to look for any irregularities or changes that might point to the existence of breast cancer. Examining the look of the breasts, including symmetry, skin texture, and the existence of any obvious anomalies such dimpling or colour changes, usually comes first in the inspection process. The physician then feels the breasts and underarms using palpation, methodically looking for any palpable lumps, abnormalities, or alterations in the substance of the tissue. In order to ensure a thorough assessment of breast health, the CBE approach depends on the healthcare provider's knowledge to identify small abnormalities.




Armando E. Giuliano, J.O.-C.F.A. (2016) *Breast disorders*, *Obgyn Key*. Available at: https://obgynkey.com/breast-disorders-2/.

In this image above it shows how the examination would've been carried out. The area labelled B: The cervical, supraclavicular, infraclavicular, and axillary nodal basins are examined as part of the regional lymph node exam while the patient is still seated. C: The patient can remain seated while undergoing a bimanual examination of the breasts. Those who have huge, protruding breasts will find this extremely helpful. D: The patient is placed in a supine posture with her ipsilateral arm lifted

above her head to complete the breast exam. The region to be checked should go from the sternum medially to the midaxillary line laterally, and from the clavicle superiorly to the rib cage inferiorly. A methodical approach guarantees a comprehensive examination of the breast. Vertical strips, a radical approach, or concentric circles can all be used to achieve this. Even though clinical breast examination was one of the first diagnosis tools its still used to detect breast cancer till this day, followed by other exams later on.

CBE was had multiple benefits a few include:

Cost-Effectiveness and Accessibility: CBE's cost-effectiveness and accessibility are two of its greatest benefits. It functions as an easily accessible screening tool, especially in environments with limited resources where access to advanced imaging technologies may be restricted.(4/5)

Comprehensive Evaluation: CBE provides a comprehensive evaluation of breast health that goes beyond the identification of palpable lumps. Healthcare practitioners can assess nipple anomalies, changes in skin texture, and other clinically relevant indicators by using a hands-on approach.(4/5)

Direct Patient-Physician Interaction: By encouraging direct communication between medical staff and patients, the approach may be able to treat psychological aspects of diagnosis. This level of personal involvement helps to make healthcare more patient-centred. (4/6)

Clinical breast examination is very useful but does have its flaws. This is why clinical breast examination isn’t the sole diagnosis method anymore. The flaws include Practitioner variability is introduced by the subjectivity of manual inspection, which affects the consistency of outcomes. Variations in clinical knowledge and interpretation might result in discrepancies in the accuracy of diagnosis.

Limited Sensitivity for Small Lesions: CBE's ability to identify small or deeply seated lesions is limited. This restriction presents a problem because early-stage tumours might not be palpable, which could lead to misleading negative results. (4/7)

Impact of Breast Density: A number of variables, including breast density, may have an effect on how well CBE detects abnormalities. When breasts are dense, there is a higher chance of false negative results. (4/7)

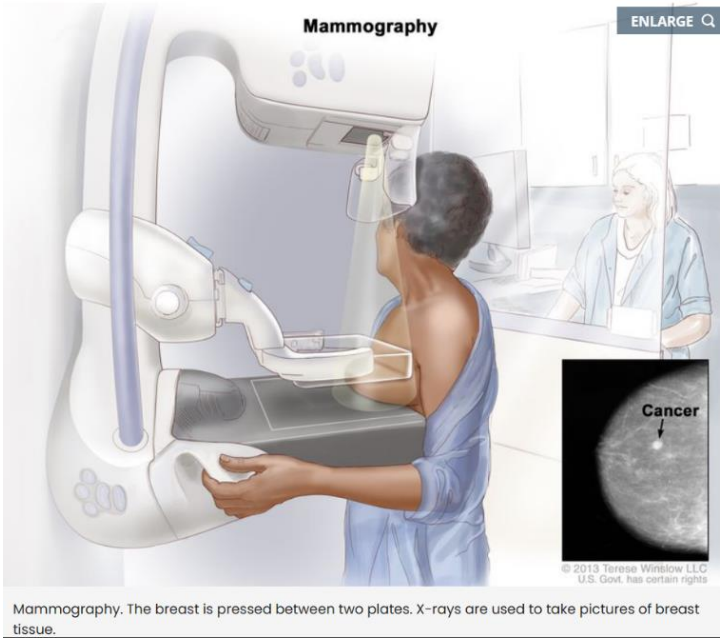
Going through all those limitations, we see Its limits highlight the need for additional diagnostic techniques, even as its advantages include affordability, accessibility, and a comprehensive approach to breast health assessment. With the goal of advancing our understanding of clinical breast screening, this dissertation lays the groundwork for future research into the development of breast cancer diagnostics and the application of cutting-edge technology for improved precision and early detection. (4/7)

## **Mammogram**

Mammography, a key component in the diagnosis of breast cancer, is a revolutionary advancement in medical imaging technology that has greatly improved breast cancer detection and treatment. This was the second major advancement in breast cancer diagnosis after clinical breast examination. According to an article on National cancer institute mammography is the most common diagnosis method used for breast cancer, this article was updated on the 26$^{th}$ of June 2023 so is fairly knew .(8)

## how does mammography work to detect cancer?

It uses  low-dose X-rays, this diagnostic technique enables medical personnel to see abnormalities in the breast, such as tumours or microcalcifications, in the breast tissue, frequently before they become palpable. To provide sharper and more precise images, the breast is usually pressed between two plates during the operation as shown in the image below . Over time, mammography has changed, showcasing significant advancements in technique and technology.




*Breast cancer screening  National Cancer Institute*. Available at: https://www.cancer.gov/types/breast/patient/breast-screening-pdq

There is 3 main mammograms:

Film mammography is an x-ray picture of the breast. (8)

Digital mammography (DM) is a computer picture of the breast. (80

Digital breast tomosynthesis (DBT) uses x-rays to take a series of pictures of the breast from many different angles. A computer is used to make 3-D pictures of the breast from these x-rays. (8)

## The Early Adoption and Emergence of Mammography:

Mammography has been used to diagnose breast cancer since the middle of the 20th century. Mammography was initially used in a documented medical setting in the 1960s. the 1960 was called the age of technical progress as stated on an article written by national library of medicine. The article states this era had major contributors include Gould, Wolfe, Gross, and their associates. The outcome of industry and medical cooperation was the creation of xeromammography. Howard and Gould reported on the advancements in imaging that came about as a result of the xeromammography technology in 1960, later on John Wolfe shared his extensive knowledge with this new technique in 1966 at the Emory University (Atlanta) 5th Conference on Mammography. (9)

Later on in 1965 in France Charles Gross established the initial mammography unit. This device was notable for having a molybdenum x-ray tube with a 0.7 mm focus spot, which allowed for excellent differential contrast and improved imaging of breast tissue, including microcalcifications, fat, and parenchyma. Additionally, the apparatus included a powerful compression mechanism. Acknowledging the potential of mammography, Gross committed his career to highlighting its ability to identify latent tumours. (10)

The following years with the use of high-definition intensifying screens and films, Price and Butler were able to significantly lower the radiation levels in mammography in 1970. Kodak and Dupont did make major contributions to this effort. Reducing radiation levels was a massive step as it ensured patient safety as less radiation given of by the mammogram means less potential health risks associated with ionising radiation. Later on in the same decade Sickles, Doi, and Genant in 1977 highlighted the necessity for ongoing advancements in mammography, arguing that technical know-how and the ability to recognise subtle indications be crucial in the detection of malignant tumours. This was the time, the idea of mobile mammography devices became widespread. (11)

In more recent times In accordance with the opinion of experts in breast diagnostics, the US National Institutes of Health gave priority to funding the advancement of digital mammography in 1991. Mammography was one area where digital technological developments were already noticeable in radiology as stated in an article from National library of medicine. Then in the early 2000s Food and Drug Administration (FDA) advised a comparison research with at least 520 women, which resulted in the Senographe 2000 D's approval. Senographe 2000 D's was a digital mammography. This was the mammography representing the advancement in breast image technology. This Senographe unlike earlier film/screen systems had an advanced electronic detector and a computerised control unit.

*12 GE Digital senographe DS mammography Block Imaging* . Available at: https://www.blockimaging.com/equipment/mammography/ge-digital-senographe-ds-mammography

These days, a number of businesses are involved in the development and marketing of digital mammography equipment, computer-aided diagnostic (CAD) systems, and breast tomosynthesis—the latter of which was approved by the FDA in 2011.

As show the advancement made has helped detection of breast cancer become more accurate with the use of x rays. Now tumours or bumps not spotted by the eye can be easily picked up by the mammogram. Despite this mammography has drawbacks and critiques that point to areas that need to be improved and alternative methods that may be more effective in detecting and diagnosing breast cancer.

Breast density is one of the main issues raised when talking about mammograms since women with dense breast tissue have noticeable effects on mammography performance. Dense regions that resemble possible tumours and show white on mammograms might lead to false negative results and reduced sensitivity. This restriction raises concerns regarding the usefulness of mammography, especially in populations whose breast density is higher than where it was tested. 8

Other concerns are related to the radiation exposure linked to mammography. As stated before in this dissertation radiation is minimal. However the cumulative exposure is increased over time by repeated examinations. This becomes especially important when thinking about screening younger women, as there's a chance that radiation-induced cancer could be something to take into account. 12

## What's the next major advancement for breast cancer detection ?

With their extraordinary potential to completely change the landscape of screening and diagnosis, artificial intelligence (AI) and machine learning have emerged as revolutionary forces in the field of breast cancer detection. With the help of these technologies, healthcare workers can make better decisions by analysing large datasets of mammograms, clinical notes, and histopathological images using complex algorithms.

The two primary forms of breast cancer are benign and malignant, and the most important part of cancer detection is the ability to discern between the two. This is where machine learning might be quite helpful. Supervised machine learning approaches prevent physician misdiagnosis and provide higher degree of accuracy categorization when all dependent elements are considered. Mammograms are mainly accuracy only predicting false negatives 1 out of 8 times according to an article from American cancer society. (14). My goal in this venture is to create an artificial intelligence system that surpasses the accuracy of a mammography in order to minimise the number of incorrect diagnoses.

## What is machine learning?

A subset of artificial intelligence (AI), machine learning (ML) enables computer systems to automatically learn from experience and get better at it without needing to be explicitly programmed. Massive volumes of health-related data, such as mammograms, patient histories, genetic data, and pathology reports, are analysed by ML algorithms in the context of breast cancer detection in order to find patterns and provide predictions with previously unheard-of efficiency and accuracy. (15)

These algorithms exhibit strong classification performance, which motivates numerous academics to use them to tackle difficult problems. With an accuracy of nearly 88%, a convolutional neural network (CNN) was utilised in to identify and classify invasive ductal carcinoma in breast histopathology images. (16)

## What is Artificial neural network (ANN)

I have decided to employ artificial neural networks (ANN) for my project. An AI approach called an artificial neural network (ANN) trains a computer to comprehend and process data in a manner similar to that of a human brain. ANNs often consist of three primary layers, which are as follows: The input, layer's function is to receive an input value so in this context takes in data about the patients data. Then the real work is done in the hidden layer. The hidden layer, which usually consists of several hidden layer is the brains of the operation the depiction and extraction of features is the main purpose of hidden layers. Every hidden layer in the network gains the ability to extract more intricate and abstract elements from the input data as it passes through it. These characteristics are depictions of connections or patterns in the data that are pertinent to the current task. Through the integration of these acquired features among several concealed layers, the network is capable of simulating complex interactions and generating more precise forecasts or categorisations. and values are output by the output layer. (17)

The hidden layer withing the ANN algorithm servers many benefits but also has its disadvantages. Some of the benefits include:

Improved Learning Capacity, by enabling AI systems to spot complex patterns and correlations in the data the hidden layers can enhance learning capacities. Another pro to hidden layer is that its adaptable. Adaptability means the neural networks with hidden layers may learn from changing inputs and adjust to a variety of data circulations. (17)

Some of the cons include:

Computational complexity, training and inference neural networks can get more complex when there's multiple hidden layers added and made accessible. Another potential disadvantage is overfitting. Overfitting happens when a model performs very well on training data but poorly on unknown data. (17)

## What is Convolutional Neural Network (CNN)

Another deep learning algorithm that is commonly used for evaluating visual data, like photographs or movies, is the Convolutional Neural Network (CNN). CNNs are able to automatically learn and extract information from input images because they are inspired by the structure and operation of the human visual cortex. The main features of CNN algorithm includes the convolutional layers, these layers process the input image by applying convolutional filters. Every filter runs over the input image, looking for features and patterns like textures, edges, and forms. The network can capture the spatial hierarchies of features in an image thanks to convolutional layers. (18)

Activation Functions is the second aspect within CNN algorithm, by adding non-linearity to the network, activation functions enable it to represent intricate feature interactions. Sigmoid and ReLU (Rectified Linear Unit) functions are examples of common activation functions.(19) Layers for pooling is the third aspect of CNN. Layers for pooling reduce the spatial dimensions of the data while preserving the most significant characteristics by down sampling the feature maps generated by the convolutional layers. Two popular pooling techniques used in CNNs are max pooling and average pooling. CNN also has fully connected layers, these layers resemble conventional artificial neural networks in that they link each neuron in one layer to every other layer's neuron. In order to carry out classification or regression tasks based on the retrieved features, fully connected layers are usually used towards the end of the network. (19)

## Why I will use the ANN algorithm rather than the CNN algorithm.

To diagnose breast cancer, medical imaging data must be accurately and efficiently analysed to find anomalies that may indicate malignancy. Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) are two exceptionally potent machine learning techniques; nevertheless, ANNs have certain advantages when it comes to breast cancer diagnosis.

ANN's flexibility in handling diverse data sources is one of its main advantages. The diagnosis of breast cancer frequently entails the integration of many data sources, such as patient histories, genetic information, and mammograms. Because ANNs are so good at processing and combining this kind of diverse data, it is possible to provide a thorough analysis that fully reflects the complex character of breast cancer. In contrast to CNNs, which are designed to process grid-like data, such pictures, ANNs are able to process a variety of input data formats and use the combined information to improve diagnostic precision. (21)

ANNs also is very good at solving problems involving feature extraction and pattern recognition, which are essential for diagnosing breast cancer. Via the independent learning of hidden layers of ANNs, pertinent features are extracted from the input data to capture complex patterns suggestive of malignant tumours. Since small irregularities may indicate the existence of malignancy, the capacity to automatically identify discriminative features from raw input data is especially helpful in the diagnosis of breast cancer. Additionally, ANNs can simulate complicated non-linear correlations between the probability of breast cancer and these collected variables, improving the accuracy of diagnoses. (22)

ANNs have superior benefits that make it more appropriate for breast cancer diagnosis, even if CNNs have impressive abilities in image processing and object recognition. With their flexibility to adapt to heterogeneous data, their skill at feature extraction, and their incorporation of clinical context, ANNs are a vital weapon in the fight against breast cancer, helping to improve patient outcomes and improve diagnostic accuracy.

## Breast cancer detection project

There are two primary methods utilised in machine learning. Learning both supervised and unsupervised. A machine learning technique called supervised learning uses fully defined data to train the model. Unsupervised data is not fully labelled and is mostly used to comprehend relationships within datasets as well as to identify unknown patterns in the data. As the supervised learning algorithm can anticipate new data based on its prior training, I will use that for this project.

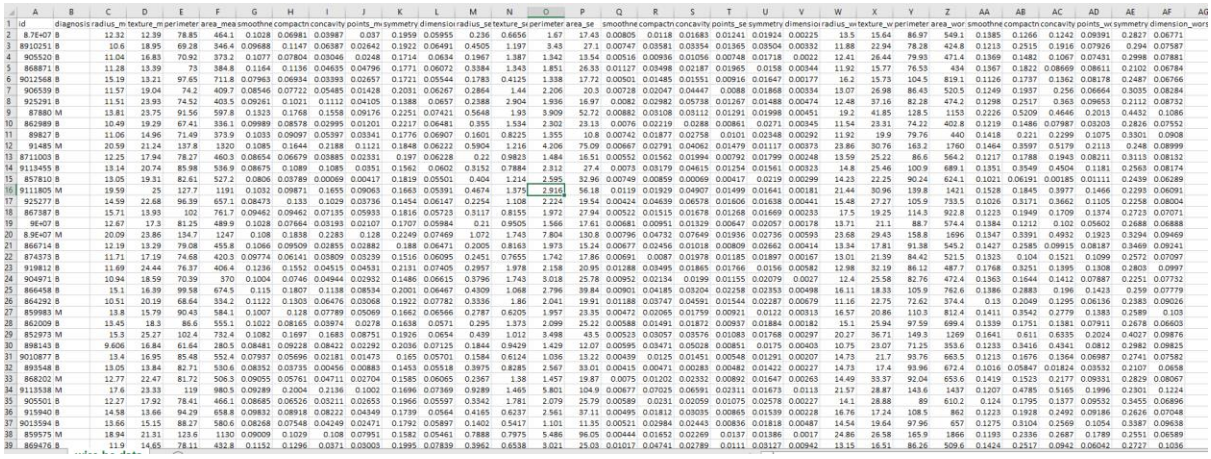
There is also different models within machine learning like decision trees , sequential model etc. for this project I will use the sequential model as it provides a simplified framework for building neural networks and is considered a core architecture in machine learning. The sequential approach is essential to the development of advanced testing devices in the field of breast cancer detection technology, which aims to increase the success of early diagnosis and treatment outcomes. When it comes to breast cancer detection, the sequential approach makes it easier to build neural networks that can recognise complex patterns and characteristics that point to malignancy. Through a series of layer stacks, each with learnable parameters, the model gains the ability to derive hierarchical representations from the input data. This makes it possible to spot minute anomalies in medical imaging, which helps in breast cancer early detection. (23)

The input layer, the dense layer, and the output layer are the three primary layers of the sequential model. The layer of input would require If the result from the first data sample is less than the input, it will be utilised as the input for the following step. I just mentioned the input for the first stage of this model because the output from the previous step would be utilised as the input for the subsequent stages.

The following layer, known as the dense layer, is a type of neural network layer that gives the input data a linear transformation. A layer is considered "dense" if it is fully connected to every unit in the layer above. This indicates that the dense layer output is a weighted sum of the inputs from each thick layer unit, which receives input from each layer unit that came before it. Training creates the weights of the links between units in a dense layer.

The goal of the training procedure is to identify the weight values that most accurately reflect the link between the input data and the desired output. Once the weights are learned, the dense layer can be utilised for predicting new, untrained data. I made the choice to include four dense layers because I believe that more layers will help the model understand more intricate, non-linear correlations in the data, which will improve task accuracy. On the other hand, an excessive number of dense layers could have the opposite effect, making the model harder to train and possibly leading to overfitting, a condition in which the model performs better on training datasets but not as well on test training sets.

The output layer, which is the last layer, is where the last prediction is created. The type and configuration of that specific output layer will depend on the sort of task involved and how the model is expected to behave. For example, in a classification job, the output layer might consist of a dense layer with a sigmoid activation function, where each unit in the dense layer represents a different class. The sigmoid activation function would transfer the dense layer's output to a value between 0 and 1, representing the estimated probability that each class's input contains the input.



| id | diagnosis | radius_m | texture_m | perimeter | area_mea | smoothne | compactn | concavity | points_m | symmetry | dimensio | radius_se | texture_se | perimeter | area_se | smoothne | compactn | concavity | points_se | symmetry | dimensio | radius_wo | texture_w | perimeter | area_wor | smoothne | compactn | concavity | points_wo | symmetry | dimension_worst |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 8.7E+07 | B | 12.32 | 12.39 | 78.85 | 464.1 | 0.1028 | 0.06981 | 0.03987 | 0.037 | 0.1959 | 0.05955 | 0.236 | 0.6656 | 1.67 | 17.43 | 0.00805 | 0.01118 | 0.01683 | 0.01241 | 0.01924 | 0.00225 | 13.5 | 15.64 | 86.97 | 549.1 | 0.1385 | 0.1266 | 0.1242 | 0.09391 | 0.2827 | 0.06771 |
| 8910251 | B | 10.6 | 18.95 | 69.28 | 346.4 | 0.09688 | 0.1147 | 0.06387 | 0.02642 | 0.1922 | 0.06491 | 0.4505 | 1.197 | 3.43 | 27.1 | 0.00747 | 0.03581 | 0.03354 | 0.01365 | 0.03504 | 0.00332 | 11.88 | 22.94 | 78.28 | 424.8 | 0.1213 | 0.2515 | 0.1916 | 0.07926 | 0.294 | 0.07587 |
| 905520 | B | 11.04 | 16.83 | 70.92 | 373.2 | 0.1077 | 0.07804 | 0.03046 | 0.0248 | 0.1714 | 0.0634 | 0.1967 | 1.387 | 1.342 | 13.54 | 0.00516 | 0.00936 | 0.01056 | 0.00748 | 0.01718 | 0.0022 | 12.41 | 26.44 | 79.93 | 471.4 | 0.1369 | 0.1482 | 0.1067 | 0.07431 | 0.2998 | 0.07881 |
| 868871 | B | 11.28 | 13.39 | 73 | 384.8 | 0.1164 | 0.1136 | 0.04635 | 0.04796 | 0.1771 | 0.06072 | 0.3384 | 1.343 | 1.851 | 26.33 | 0.01127 | 0.03498 | 0.02187 | 0.01965 | 0.0158 | 0.00344 | 11.92 | 15.77 | 76.53 | 434 | 0.1567 | 0.1822 | 0.08669 | 0.08611 | 0.2102 | 0.06784 |
| 9012568 | B | 15.19 | 13.21 | 97.65 | 711.8 | 0.07963 | 0.06934 | 0.03393 | 0.02657 | 0.1721 | 0.05544 | 0.1783 | 0.4125 | 1.338 | 17.72 | 0.00501 | 0.01485 | 0.01551 | 0.00916 | 0.01647 | 0.00177 | 16.2 | 15.73 | 104.5 | 819.1 | 0.1126 | 0.1737 | 0.1362 | 0.08178 | 0.2487 | 0.06766 |
| 906539 | B | 11.57 | 19.04 | 74.2 | 409.7 | 0.08546 | 0.07722 | 0.05485 | 0.01428 | 0.2031 | 0.06267 | 0.2864 | 1.44 | 2.206 | 20.3 | 0.00728 | 0.02047 | 0.00447 | 0.0088 | 0.01868 | 0.00334 | 13.07 | 26.98 | 86.43 | 520.5 | 0.1249 | 0.1937 | 0.256 | 0.06664 | 0.3035 | 0.08284 |
| 925291 | B | 11.51 | 23.93 | 74.52 | 403.5 | 0.09261 | 0.1021 | 0.1112 | 0.04105 | 0.1388 | 0.0657 | 0.2388 | 2.904 | 1.936 | 16.97 | 0.0082 | 0.02982 | 0.05738 | 0.01267 | 0.01488 | 0.00474 | 12.48 | 37.16 | 82.28 | 474.2 | 0.1298 | 0.2517 | 0.363 | 0.09653 | 0.2112 | 0.08732 |
| 87880 | M | 13.81 | 23.75 | 91.56 | 597.8 | 0.1323 | 0.1768 | 0.1558 | 0.09176 | 0.2251 | 0.07421 | 0.5648 | 1.93 | 3.909 | 52.72 | 0.00882 | 0.03108 | 0.03112 | 0.01291 | 0.01998 | 0.00451 | 19.2 | 41.85 | 128.5 | 1153 | 0.2226 | 0.5209 | 0.4646 | 0.2013 | 0.4432 | 0.1086 |
| 862989 | B | 10.49 | 19.29 | 67.41 | 336.1 | 0.09989 | 0.08578 | 0.02995 | 0.01201 | 0.2217 | 0.06481 | 0.355 | 1.534 | 2.302 | 23.13 | 0.0076 | 0.02219 | 0.0288 | 0.00861 | 0.0271 | 0.00345 | 11.54 | 23.31 | 74.22 | 402.8 | 0.1219 | 0.1486 | 0.07987 | 0.03203 | 0.2826 | 0.07552 |
| 89827 | B | 11.06 | 14.96 | 71.49 | 373.9 | 0.1033 | 0.09097 | 0.05397 | 0.03941 | 0.1776 | 0.06907 | 0.1601 | 0.8225 | 1.355 | 10.8 | 0.00742 | 0.01877 | 0.02758 | 0.0101 | 0.02348 | 0.00292 | 11.92 | 19.9 | 79.76 | 440 | 0.1418 | 0.221 | 0.2299 | 0.1075 | 0.3301 | 0.0908 |
| 91485 | M | 20.59 | 21.24 | 137.8 | 1320 | 0.1085 | 0.1644 | 0.2188 | 0.1121 | 0.1848 | 0.06222 | 0.5904 | 1.216 | 4.206 | 75.09 | 0.00667 | 0.02791 | 0.04062 | 0.01479 | 0.01117 | 0.00373 | 23.86 | 30.76 | 163.2 | 1760 | 0.1464 | 0.3597 | 0.5179 | 0.2113 | 0.248 | 0.08999 |
| 8711003 | B | 12.25 | 17.94 | 78.27 | 460.3 | 0.08654 | 0.06679 | 0.03885 | 0.02331 | 0.197 | 0.06228 | 0.22 | 0.9823 | 1.484 | 16.51 | 0.00552 | 0.01562 | 0.01994 | 0.00792 | 0.01799 | 0.00248 | 13.59 | 25.22 | 86.6 | 564.2 | 0.1217 | 0.1788 | 0.1943 | 0.08211 | 0.3113 | 0.08132 |
| 9113455 | B | 13.14 | 20.74 | 85.98 | 536.9 | 0.08675 | 0.1089 | 0.1085 | 0.0351 | 0.1562 | 0.0602 | 0.3152 | 0.7884 | 2.312 | 27.4 | 0.0073 | 0.03179 | 0.04615 | 0.01254 | 0.01561 | 0.00323 | 14.8 | 25.46 | 100.9 | 689.1 | 0.1351 | 0.3549 | 0.4504 | 0.1181 | 0.2563 | 0.08174 |
| 857810 | B | 13.05 | 19.31 | 82.61 | 527.2 | 0.0806 | 0.03789 | 0.00069 | 0.00417 | 0.1819 | 0.05501 | 0.404 | 1.214 | 2.595 | 32.96 | 0.00749 | 0.00859 | 0.00069 | 0.00417 | 0.0219 | 0.00299 | 14.23 | 22.25 | 90.24 | 624.1 | 0.1021 | 0.06191 | 0.00185 | 0.01111 | 0.2439 | 0.06289 |
| 9111805 | M | 19.59 | 25 | 127.7 | 1191 | 0.1032 | 0.09871 | 0.1655 | 0.09063 | 0.1663 | 0.05391 | 0.4674 | 1.375 | 2.916 | 56.18 | 0.0119 | 0.01929 | 0.04907 | 0.01499 | 0.01641 | 0.00181 | 21.44 | 30.96 | 139.8 | 1421 | 0.1528 | 0.1845 | 0.3977 | 0.1466 | 0.2293 | 0.06091 |
| 925277 | B | 14.59 | 22.68 | 96.39 | 657.1 | 0.08473 | 0.133 | 0.1029 | 0.03736 | 0.1454 | 0.06147 | 0.2254 | 1.108 | 2.224 | 19.54 | 0.00424 | 0.04839 | 0.06578 | 0.01606 | 0.01638 | 0.00441 | 15.48 | 27.27 | 105.9 | 733.5 | 0.1026 | 0.3171 | 0.3662 | 0.1105 | 0.2258 | 0.08004 |
| 867387 | B | 15.71 | 13.93 | 102 | 761.7 | 0.09462 | 0.09462 | 0.07135 | 0.05933 | 0.1816 | 0.05723 | 0.3117 | 0.8155 | 1.972 | 27.94 | 0.00522 | 0.01515 | 0.01678 | 0.01268 | 0.01669 | 0.00233 | 17.5 | 19.25 | 114.3 | 922.8 | 0.1223 | 0.1949 | 0.1709 | 0.1374 | 0.2723 | 0.07071 |
| 9E+07 | B | 12.67 | 17.3 | 81.25 | 489.9 | 0.1028 | 0.07664 | 0.03193 | 0.02107 | 0.1707 | 0.05984 | 0.21 | 0.9505 | 1.566 | 17.61 | 0.00681 | 0.00951 | 0.01329 | 0.00647 | 0.02057 | 0.00178 | 13.71 | 21.1 | 88.7 | 574.4 | 0.1384 | 0.1212 | 0.102 | 0.05602 | 0.2688 | 0.06888 |
| 8.9E+07 | M | 20.09 | 23.86 | 134.7 | 1247 | 0.108 | 0.1838 | 0.2283 | 0.128 | 0.2249 | 0.07469 | 1.072 | 1.745 | 7.804 | 130.8 | 0.00796 | 0.04732 | 0.07649 | 0.01936 | 0.02736 | 0.00593 | 23.68 | 29.43 | 158.8 | 1696 | 0.1347 | 0.3391 | 0.4932 | 0.1923 | 0.3294 | 0.09469 |
| 866714 | B | 12.19 | 13.29 | 79.08 | 455.8 | 0.1066 | 0.09509 | 0.02855 | 0.02882 | 0.188 | 0.06471 | 0.2005 | 0.8163 | 1.973 | 15.24 | 0.00677 | 0.02456 | 0.01018 | 0.00809 | 0.02662 | 0.00414 | 13.34 | 17.81 | 91.38 | 545.2 | 0.1427 | 0.2585 | 0.09915 | 0.08187 | 0.3469 | 0.09241 |
| 874373 | B | 11.71 | 17.19 | 74.68 | 420.3 | 0.09774 | 0.06141 | 0.03809 | 0.03239 | 0.1516 | 0.06095 | 0.2451 | 0.7655 | 1.742 | 17.86 | 0.00691 | 0.0087 | 0.01978 | 0.01185 | 0.01897 | 0.00167 | 13.01 | 21.39 | 84.42 | 521.5 | 0.1323 | 0.104 | 0.1521 | 0.1099 | 0.2572 | 0.07097 |
| 919812 | B | 11.69 | 24.44 | 76.37 | 406.4 | 0.1236 | 0.1552 | 0.04515 | 0.04531 | 0.2131 | 0.07405 | 0.2957 | 1.978 | 2.158 | 20.95 | 0.01288 | 0.03495 | 0.01865 | 0.01766 | 0.0156 | 0.00582 | 12.98 | 32.19 | 86.12 | 487.7 | 0.1768 | 0.3251 | 0.1395 | 0.1308 | 0.2803 | 0.0997 |
| 904971 | B | 10.94 | 18.59 | 70.39 | 370 | 0.1004 | 0.0746 | 0.04944 | 0.02932 | 0.1486 | 0.06615 | 0.3796 | 1.743 | 3.018 | 25.78 | 0.00952 | 0.02134 | 0.0199 | 0.01155 | 0.02079 | 0.0027 | 12.4 | 25.58 | 82.76 | 472.4 | 0.1363 | 0.1644 | 0.1412 | 0.07887 | 0.2251 | 0.07732 |
| 866458 | B | 15.1 | 16.39 | 99.58 | 674.5 | 0.115 | 0.1807 | 0.1138 | 0.08534 | 0.2001 | 0.06467 | 0.4309 | 1.068 | 2.796 | 39.84 | 0.00901 | 0.04185 | 0.03204 | 0.02258 | 0.02353 | 0.00498 | 16.11 | 18.33 | 105.9 | 762.6 | 0.1386 | 0.2883 | 0.196 | 0.1423 | 0.259 | 0.07779 |
| 864292 | B | 10.51 | 20.19 | 68.64 | 334.2 | 0.1122 | 0.1303 | 0.06476 | 0.03068 | 0.1922 | 0.07782 | 0.3336 | 1.86 | 2.041 | 19.91 | 0.01188 | 0.03747 | 0.04591 | 0.01544 | 0.02287 | 0.00679 | 11.16 | 22.75 | 72.62 | 374.4 | 0.13 | 0.2049 | 0.1295 | 0.06136 | 0.2383 | 0.09026 |
| 859983 | M | 13.8 | 15.79 | 90.43 | 584.1 | 0.1007 | 0.128 | 0.07789 | 0.05069 | 0.1662 | 0.06566 | 0.2787 | 0.6205 | 1.957 | 23.35 | 0.00472 | 0.02065 | 0.01759 | 0.00921 | 0.0122 | 0.00313 | 16.57 | 20.86 | 110.3 | 812.4 | 0.1411 | 0.3542 | 0.2779 | 0.1383 | 0.2589 | 0.103 |
| 862009 | B | 13.45 | 18.3 | 86.6 | 555.1 | 0.1022 | 0.08165 | 0.03974 | 0.0278 | 0.1638 | 0.0571 | 0.295 | 1.373 | 2.099 | 25.22 | 0.00588 | 0.01491 | 0.01872 | 0.00937 | 0.01884 | 0.00182 | 15.1 | 25.94 | 97.59 | 699.4 | 0.1339 | 0.1751 | 0.1381 | 0.07911 | 0.2678 | 0.06603 |
| 852973 | M | 15.3 | 25.27 | 102.4 | 732.4 | 0.1082 | 0.1697 | 0.1683 | 0.08751 | 0.1926 | 0.0654 | 0.439 | 1.012 | 3.498 | 43.5 | 0.00523 | 0.03057 | 0.03576 | 0.01083 | 0.01768 | 0.00297 | 20.27 | 36.71 | 149.3 | 1269 | 0.1641 | 0.611 | 0.6335 | 0.2024 | 0.4027 | 0.09876 |
| 898143 | B | 9.606 | 16.84 | 61.64 | 280.5 | 0.08481 | 0.09228 | 0.08422 | 0.02292 | 0.2036 | 0.07125 | 0.1844 | 0.9429 | 1.429 | 12.07 | 0.00595 | 0.03471 | 0.05028 | 0.00851 | 0.0175 | 0.00403 | 10.75 | 23.07 | 71.25 | 353.6 | 0.1233 | 0.3416 | 0.4341 | 0.0812 | 0.2982 | 0.09825 |
| 9010877 | B | 13.4 | 16.95 | 85.48 | 552.4 | 0.07937 | 0.05696 | 0.02181 | 0.01473 | 0.165 | 0.05701 | 0.1584 | 0.6124 | 1.036 | 13.22 | 0.00439 | 0.0125 | 0.01451 | 0.00548 | 0.01291 | 0.00207 | 14.73 | 21.7 | 93.76 | 663.5 | 0.1213 | 0.1676 | 0.1364 | 0.06987 | 0.2741 | 0.07582 |
| 893548 | B | 13.05 | 13.84 | 82.71 | 530.6 | 0.08352 | 0.03735 | 0.00456 | 0.00883 | 0.1453 | 0.05518 | 0.3975 | 0.8285 | 2.567 | 33.01 | 0.00415 | 0.00471 | 0.00283 | 0.00482 | 0.01422 | 0.00227 | 14.73 | 17.4 | 93.96 | 672.4 | 0.1016 | 0.05847 | 0.01824 | 0.03532 | 0.2107 | 0.0658 |
| 868202 | M | 12.77 | 22.47 | 81.72 | 506.3 | 0.09055 | 0.05761 | 0.04711 | 0.02704 | 0.1585 | 0.06065 | 0.2367 | 1.38 | 1.457 | 19.87 | 0.0075 | 0.01202 | 0.02332 | 0.00892 | 0.01647 | 0.00263 | 14.49 | 33.37 | 92.04 | 653.6 | 0.1419 | 0.1523 | 0.2177 | 0.09331 | 0.2829 | 0.08067 |
| 9113538 | M | 17.6 | 23.33 | 119 | 980.5 | 0.09289 | 0.2004 | 0.2136 | 0.1002 | 0.1696 | 0.07369 | 0.9289 | 1.465 | 5.801 | 104.9 | 0.00677 | 0.07025 | 0.06591 | 0.02311 | 0.01673 | 0.0113 | 21.57 | 28.87 | 143.6 | 1437 | 0.1207 | 0.4785 | 0.5165 | 0.1996 | 0.2301 | 0.1224 |
| 905501 | B | 12.27 | 17.92 | 78.41 | 466.1 | 0.08685 | 0.06526 | 0.03211 | 0.02655 | 0.1966 | 0.05597 | 0.3342 | 1.781 | 2.079 | 25.79 | 0.00589 | 0.0231 | 0.02059 | 0.01075 | 0.02578 | 0.00227 | 14.1 | 28.88 | 89 | 610.2 | 0.124 | 0.1795 | 0.1377 | 0.09532 | 0.3455 | 0.06896 |
| 915940 | B | 14.58 | 13.66 | 94.29 | 658.8 | 0.09832 | 0.08918 | 0.08222 | 0.04349 | 0.1739 | 0.0564 | 0.4165 | 0.6237 | 2.561 | 37.11 | 0.00495 | 0.01812 | 0.03035 | 0.00865 | 0.01539 | 0.00228 | 16.76 | 17.24 | 108.5 | 862 | 0.1223 | 0.1928 | 0.2492 | 0.09186 | 0.2626 | 0.07048 |
| 9013594 | B | 13.66 | 15.15 | 88.27 | 580.6 | 0.08268 | 0.07548 | 0.04249 | 0.02471 | 0.1792 | 0.05897 | 0.1402 | 0.5417 | 1.101 | 11.35 | 0.00521 | 0.02984 | 0.02443 | 0.00836 | 0.01818 | 0.00487 | 14.54 | 19.64 | 97.96 | 657 | 0.1275 | 0.3104 | 0.2569 | 0.1054 | 0.3387 | 0.09638 |
| 859575 | M | 18.94 | 21.31 | 123.6 | 1130 | 0.09009 | 0.1029 | 0.108 | 0.07951 | 0.1582 | 0.05461 | 0.7888 | 0.7975 | 5.486 | 96.05 | 0.00444 | 0.01652 | 0.02269 | 0.0137 | 0.01386 | 0.0017 | 24.86 | 26.58 | 165.9 | 1866 | 0.1193 | 0.2336 | 0.2687 | 0.1789 | 0.2551 | 0.06589 |
| 869476 | B | 11.9 | 14.65 | 78.11 | 432.8 | 0.1152 | 0.1296 | 0.0371 | 0.03003 | 0.1995 | 0.07839 | 0.3962 | 0.6538 | 3.021 | 25.03 | 0.01017 | 0.04741 | 0.02789 | 0.0111 | 0.03127 | 0.00942 | 13.15 | 16.51 | 86.26 | 509.6 | 0.1424 | 0.2517 | 0.0942 | 0.06042 | 0.2727 | 0.1036 |

Stednick, Z.S. (no date) *breast cancer data.cvs*, *GitHub*. Available at: https://github.com/stedy/Machine-Learning-with-R-datasets/blob/master/wisc_bc_data.csv

Before I start the project to make a breast cancer diagnosis machine, I first had to get a dataset. This dataset will be used within the ANN algorithm to find its accuracy.

In figure 1 these lines import the libraries and functions required for Python neural network modelling, machine learning, data processing, and visualisation.

```
import pandas as pd
import numpy as np # Numpy is used for numerical computing in Python
import sklearn as sk # Scikit-learn is a machine learning library in Python that provides tools for data mining and data analysis.
import seaborn as sns # used for data visulatation
from matplotlib import pyplot as pl
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from pandas.plotting import scatter_matrix
from keras.models import Sequential
from keras.layers import Dense
from keras.utils.vis_utils import  plot_model
```

Figure 1

In figure 2 I load up my data file and tell the program to show the first 5 rows, this is just to make sure the right file is loading up and is being read properly

```
data = pd.read_csv("wisc_bc_data.csv")
```

```
data.head(5)
```

Python

| | id | diagnosis | radius_mean | texture_mean | perimeter_mean | area_mean | smoothness_mean | compactness_mean | concavity_mean | points_mean | ... | radius_worst | texture_worst | perimeter_worst | area_worst | sm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 87139402 | B | 12.32 | 12.39 | 78.85 | 464.1 | 0.10280 | 0.06981 | 0.03987 | 0.03700 | ... | 13.50 | 15.64 | 86.97 | 549.1 | |
| 1 | 8910251 | B | 10.60 | 18.95 | 69.28 | 346.4 | 0.09688 | 0.11470 | 0.06387 | 0.02642 | ... | 11.88 | 22.94 | 78.28 | 424.8 | |
| 2 | 905520 | B | 11.04 | 16.83 | 70.92 | 373.2 | 0.10770 | 0.07804 | 0.03046 | 0.02480 | ... | 12.41 | 26.44 | 79.93 | 471.4 | |
| 3 | 868871 | B | 11.28 | 13.39 | 73.00 | 384.8 | 0.11640 | 0.11360 | 0.04635 | 0.04796 | ... | 11.92 | 15.77 | 76.53 | 434.0 | |
| 4 | 9012568 | B | 15.19 | 13.21 | 97.65 | 711.8 | 0.07963 | 0.06934 | 0.03393 | 0.02657 | ... | 16.20 | 15.73 | 104.50 | 819.1 | |

5 rows × 32 columns

Figure 2

```
sns.countplot(x="diagnosis", data=data) #shows use the visual on how many malignant and beigin patients we have in the dataset
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f3009e34580>
```

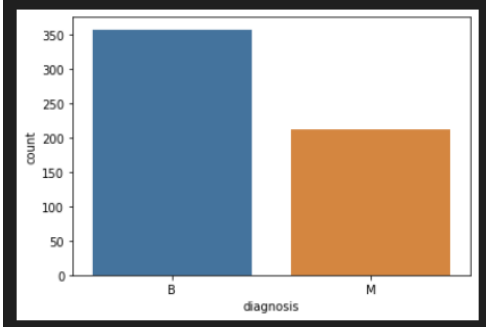



Figure 3

```
data['diagnosis'] = np.where(data['diagnosis']=='M',1,0) # this line converts all the 'M' to 1 and all the 'B' to 0
```

```
data #re-run the dataset to make sure all the 'm' and 'b' have converted to '1' and '0'
```

Figure 4

| | id | diagnosis | radius_mean | texture_mean | perimeter_mean | area_mean | smoothness_mean | compactness_mean | concavity_mean | points_mean | ... | radius_worst | texture_worst | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 87139402 | 0 | 12.32 | 12.39 | 78.85 | 464.1 | 0.10280 | 0.06981 | 0.03987 | 0.03700 | ... | 13.50 | 15.64 | 86.97 | 549.1 |
| 1 | 8910251 | 0 | 10.60 | 18.95 | 69.28 | 346.4 | 0.09688 | 0.11470 | 0.06387 | 0.02642 | ... | 11.88 | 22.94 | 78.28 | 424.8 |
| 2 | 905520 | 0 | 11.04 | 16.83 | 70.92 | 373.2 | 0.10770 | 0.07804 | 0.03046 | 0.02480 | ... | 12.41 | 26.44 | 79.93 | 471.4 |
| 3 | 868871 | 0 | 11.28 | 13.39 | 73.00 | 384.8 | 0.11640 | 0.11360 | 0.04635 | 0.04796 | ... | 11.92 | 15.77 | 76.53 | 434.0 |
| 4 | 9012568 | 0 | 15.19 | 13.21 | 97.65 | 711.8 | 0.07963 | 0.06934 | 0.03393 | 0.02657 | ... | 16.20 | 15.73 | 104.50 | 819.1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 564 | 911320502 | 0 | 13.17 | 18.22 | 84.28 | 537.3 | 0.07466 | 0.05994 | 0.04859 | 0.02870 | ... | 14.90 | 23.89 | 95.10 | 687.6 |
| 565 | 898677 | 0 | 10.26 | 14.71 | 66.20 | 321.6 | 0.09882 | 0.09159 | 0.03581 | 0.02037 | ... | 10.88 | 19.48 | 70.89 | 357.1 |
| 566 | 873885 | 1 | 15.28 | 22.41 | 98.92 | 710.6 | 0.09057 | 0.10520 | 0.05375 | 0.03263 | ... | 17.80 | 28.03 | 113.80 | 973.1 |
| 567 | 911201 | 0 | 14.53 | 13.98 | 93.86 | 644.2 | 0.10990 | 0.09242 | 0.06895 | 0.06495 | ... | 15.80 | 16.93 | 103.10 | 749.9 |
| 568 | 9012795 | 1 | 21.37 | 15.10 | 141.30 | 1386.0 | 0.10010 | 0.15150 | 0.19320 | 0.12550 | ... | 22.69 | 21.84 | 152.10 | 1535.0 |

569 rows × 32 columns

```
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.25, random_state=50)
```

Figure 5

```
print(' training data is:', x_train.shape)
print('testing data is:', x_test.shape)
```

In figure 5 the data was divided into training and testing sets by myself. In order to help the model identify more intricate patterns in the data and improve the accuracy of its predictions, I decided to use 75% of the dataset for training. After that, I tested using 25% of the dataset.

```
Classifcation_Model.compile(optimizer='rmsprop', loss='binary_crossentropy', metrics=['accuracy']) #line is just used to set up classification
```

Figure 6

```
Classifcation_Model.add(Dense(units= 20, activation='relu', input_dim=30))
Classifcation_Model.add(Dense(units= 15, activation='relu'))
Classifcation_Model.add(Dense(units= 10, activation='relu'))
Classifcation_Model.add(Dense(units= 8, activation='relu'))
Classifcation_Model.add(Dense(units= 5, activation='relu'))
Classifcation_Model.add(Dense(units= 1, activation='sigmoid'))
```

```
print(Classifcation_Model.summary())
```

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 20)                620

dense_1 (Dense)              (None, 15)                315

dense_2 (Dense)              (None, 10)                160

dense_3 (Dense)              (None, 8)                 88

dense_4 (Dense)              (None, 5)                 45

dense_5 (Dense)              (None, 1)                 6
```

in Figure 6 I set up the classification for the sequential model "Classifcation_Model.add(Dense(units=20, activation='relu', input_dim=30))" This line inserts the categorization model's initial layer. With 20 neurons (units), it is a fully linked layer (Dense) that uses the ReLU activation function. In this scenario, there are 30 features in the input data, as indicated by the input_dim option.

"Classifcation_Model.add(Dense(units=15, activation='relu'))" shows the model gains a hidden layer thanks to this line. Using the ReLU activation function, it is an additional completely linked layer made up of fifteen neurons. The next line does the same but reduces the neurons to 10 and that happens for another2 lines.

The output layer is added to the model by the Classifcation_Model.add(Dense(units= 1, activation='sigmoid')) line. This single neuron layer exhibits sigmoid activation. This line will only give 1 result.

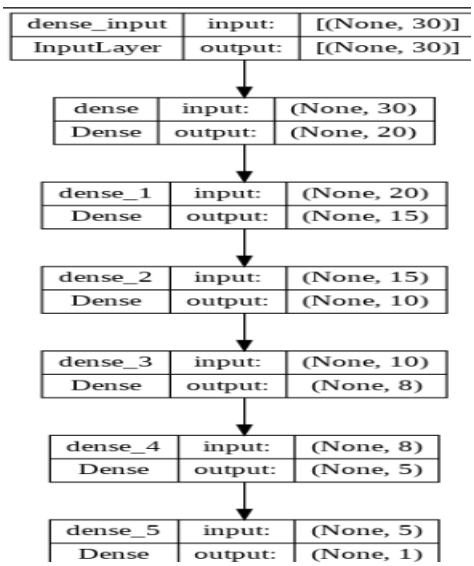



Figure 7 shows the same thing as what has just been mentioned but in a diagram form.

Figure 7

```
history = Classifcation_Model.fit(x_train, y_train, batch_size=5,   epochs=100)
```

Figure 8

```
Epoch 1/100
86/86 [==============================] - 2s 3ms/step - loss: 2.1045 - accuracy: 0.6315
Epoch 2/100
86/86 [==============================] - 0s 2ms/step - loss: 0.3881 - accuracy: 0.8662
Epoch 3/100
86/86 [==============================] - 0s 2ms/step - loss: 0.3568 - accuracy: 0.8756
Epoch 4/100
86/86 [==============================] - 0s 3ms/step - loss: 0.3121 - accuracy: 0.8756
Epoch 5/100
86/86 [==============================] - 0s 2ms/step - loss: 0.2707 - accuracy: 0.8920
Epoch 6/100
86/86 [==============================] - 0s 3ms/step - loss: 0.2701 - accuracy: 0.9038
Epoch 7/100
86/86 [==============================] - 0s 2ms/step - loss: 0.2320 - accuracy: 0.9108
Epoch 8/100
86/86 [==============================] - 0s 2ms/step - loss: 0.2326 - accuracy: 0.9014
Epoch 9/100
86/86 [==============================] - 0s 3ms/step - loss: 0.2518 - accuracy: 0.9038
Epoch 10/100
86/86 [==============================] - 0s 3ms/step - loss: 0.2347 - accuracy: 0.8991
Epoch 11/100
86/86 [==============================] - 0s 2ms/step - loss: 0.2428 - accuracy: 0.8991
Epoch 12/100
86/86 [==============================] - 0s 2ms/step - loss: 0.2485 - accuracy: 0.9085
Epoch 13/100
...
Epoch 99/100
86/86 [==============================] - 0s 1ms/step - loss: 0.1979 - accuracy: 0.9225
Epoch 100/100
86/86 [==============================] - 0s 1ms/step - loss: 0.1882 - accuracy: 0.9319
```

Figure 8 is the first time the machine will be tested to see if its accurate. So the classification we made previously is called. Batch_size=5 tells the program to update. So in this case the program will update its weight after processing each batch of 5 samples. epochs=100: suggests the number of epochs, or the total number of forward and backward passes of the training dataset through the neural network, is indicated by this parameter. The model will go through 100 iterations on the training set in this instance. While training for additional epochs may lead to better model performance, improper regularisation raises the danger of overfitting.

As we can clearly see in figure 8 the training set gave use aroung 88 – 93% accuracy. Now I know it works on our training data set I can use it on unseen test dataset to see whether we get similar results or if the program fails.

```
x_test #calling the test data
```

Figure 9

| | radius_mean | texture_mean | perimeter_mean | area_mean | smoothness_mean | compactness_mean | concavity_mean | points_mean | symmetry_mean | dimension_mean | ... | radius_worst | texture | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 356 | 9.876 | 17.27 | 62.92 | 295.4 | 0.10890 | 0.07232 | 0.01756 | 0.01952 | 0.1934 | 0.06285 | ... | 10.42 | 23.22 | 67.08 | |
| 556 | 11.500 | 18.45 | 73.28 | 407.4 | 0.09345 | 0.05991 | 0.02638 | 0.02069 | 0.1834 | 0.05934 | ... | 12.97 | 22.46 | 83.12 | |
| 283 | 13.640 | 15.60 | 87.38 | 575.3 | 0.09423 | 0.06630 | 0.04705 | 0.03731 | 0.1717 | 0.05660 | ... | 14.85 | 19.05 | 94.11 | |
| 495 | 12.880 | 18.22 | 84.45 | 493.1 | 0.12180 | 0.16610 | 0.04825 | 0.05303 | 0.1709 | 0.07253 | ... | 15.05 | 24.37 | 99.31 | |
| 364 | 12.180 | 17.84 | 77.79 | 451.1 | 0.10450 | 0.07057 | 0.02490 | 0.02941 | 0.1900 | 0.06635 | ... | 12.83 | 20.92 | 82.14 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 382 | 14.480 | 21.46 | 94.25 | 648.2 | 0.09444 | 0.09947 | 0.12040 | 0.04938 | 0.2075 | 0.05636 | ... | 16.21 | 29.25 | 108.40 | |
| 94 | 14.220 | 23.12 | 94.37 | 609.9 | 0.10750 | 0.24130 | 0.19810 | 0.06618 | 0.2384 | 0.07542 | ... | 15.74 | 37.18 | 106.40 | |
| 227 | 20.180 | 19.54 | 133.80 | 1250.0 | 0.11330 | 0.14890 | 0.21330 | 0.12590 | 0.1724 | 0.06053 | ... | 22.03 | 25.07 | 146.00 | |
| 25 | 13.800 | 15.79 | 90.43 | 584.1 | 0.10070 | 0.12800 | 0.07789 | 0.05069 | 0.1662 | 0.06566 | ... | 16.57 | 20.86 | 110.30 | |
| 203 | 12.630 | 20.76 | 82.15 | 480.4 | 0.09933 | 0.12090 | 0.10650 | 0.06021 | 0.1735 | 0.07070 | ... | 13.33 | 25.47 | 89.00 | |

143 rows × 30 columns

```
prediction = Classifcation_Model.predict(x_test)
prediction = [1 if y > 0.5  else 0 for y in prediction] ## if the prediction value is over 0.5 print out 1 for malignant if its less than 0.5 print out 0
prediction
```

in figure 9 we call our test data set. we later ask it to predict whether the data presented to it is malignant or benign. So if the prediction is over 0.5 it means its Malignant so it show print out a 1 else it should print 0.

```
len(prediction) #the length of the test dataset
```

```
143
```

```
type(prediction) #list out the results
```

```
list
```

Figure 10

```
type(y_test)
```

```
pandas.core.frame.DataFrame
```

```
pred = pd.DataFrame(prediction)
```

```
type(pred)
```

```
pandas.core.frame.DataFrame
```

```
pred = pred.values
```

```
y_test_ndarray = y_test.values
```

Figure 10. here we verify the length we previously set for the test data is the same and make sure it still prints out as a list.

Figure 11

```
total = 0 #just a base number it will change when prgram is ran
correct = 0 # just a place holder so it starts from 0
wrong =0

for i in range(len(pred)): #Every prediction the model makes is iterated over in this loop.
    total = total+1 #increase total count by 1
    if(y_test_ndarray[i]== pred[i]): #check to see if ndarray and pred is match if yes put the place holder 1
        correct = correct+1 # if its correct increase  correct by 1
    else:
        wrong = wrong+1 #if wring increas wrong by 1
print('Total = ' + str(total))
print('Correct = ' + str(correct) + ' - Parcentage correct: ' + str(round((correct/total)*100, 2))+'%')
print('Wrong = ' + str(wrong) + ' Parcentage wrong: ' + str(round((wrong/total)*100, 2)) + '%')
```

```
Total = 143
Correct = 127 - Parcentage correct: 88.81%
Wrong = 16 Parcentage wrong: 11.19%
```

In figure 11 we do the final steps for the test dataset and we make a small program to run through all the test data, in this case 143 cases, and work out which patients had malignant tumours or benign. As shown in the screenshot the test dataset had 88% accuracy. This is lower than the training set.

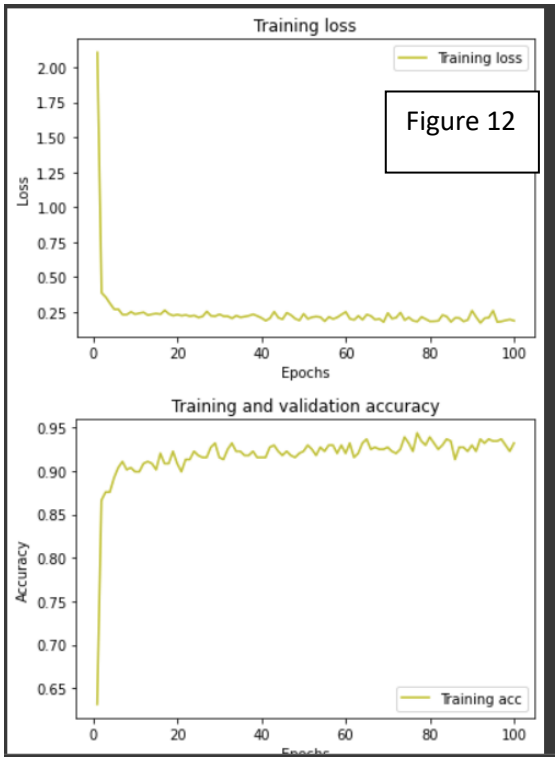



Figure 12

figure 12.

This is a graph showing the decrease of mistakes and increase in accuracy of the training set as the program was running. This shows the diagnosis tool has potential to be even better with a few minor tweaks.

From the results shown it tells me the diagnosis machine needs more work and maybe more dense layers to increase accuracy. As right now its giving back similar accuracy reading as our current diagnosis methods. However this does show with enough time and work put behind artificial intelligence we could get a very accurate breast cancer detection system.

## The ethical issue

Musa Chowdhury project base

It is crucial to discuss the ethical ramifications of artificial intelligence (AI) technology as they are incorporated into healthcare systems more and more, especially in the area of breast cancer detection. Even though AI has the potential to significantly improve early diagnosis and treatment outcomes, there are a number of ethical issues that need to be properly considered and handled.

The ethical implications of integrating AI technology into healthcare are multifaceted and require careful consideration. Key ethical concerns encompass patient privacy and data security , transparency and explain ability of AI algorithms , addressing algorithmic bias and ensuring fairness, obtaining informed consent from patients , upholding healthcare professionals' ethical responsibilities , ethically allocating resources and establishing regulatory and governance frameworks . These ethical dimensions are critical in ensuring the responsible and just deployment of AI in healthcare, safeguarding patient well-being and trust while advancing the field.

I conducted a two interviews to find out the general view people have about AI in the healthcare industry.

Interview 1 – anonymous – mathematic tutor

Question - Have you heard about the use of artificial intelligence and machine learning in breast cancer diagnosis before?

Answer – yes, I have heard about Ai being used in the health care industry and in breast cancer screening.

Question - How confident are you in the accuracy of AI-assisted breast cancer screening compared to traditional methods?

Answer – I don’t know the stats but I would assume the accuracy levels will be high as ai’s currently are taking over our everyday life and doing stuff much better than humans.

Question - What factors contribute to or diminish your trust in AI technology for medical diagnosis?

Answer – Ai is used everywhere in cars like tesla and in our homes and phones with SIRI and ALEXA.

Question - Are there specific privacy or ethical concerns you have regarding the use of AI in healthcare?

Answer – yes with all these cybercrimes it would be easier now for hackers to get access to sensitive data. In addition ai detection could still malfunction and could accidently misdiagnose someone what happens then?

Interview 2 – anonymous – student studying pharmacology

Question - Have you heard about the use of artificial intelligence and machine learning in breast cancer diagnosis before?

Answer – I have never researched it before, I've heard about AI being used in healthcare. However not directly about ai being used in breast cancer diagnosis.

Question - How confident are you in the accuracy of AI-assisted breast cancer screening compared to traditional methods?

Answer – I'm neither confident or not confident in Ai's ability. if I was shown some stats and real life result I would become more confident.

Question - What factors contribute to or diminish your trust in AI technology for medical diagnosis?

Answer – the fact ai is being used everywhere from self service checkouts to cars and even homes with Alexa and google home I have trust it will be used properly and not maliciously in health care

Question - Are there specific privacy or ethical concerns you have regarding the use of AI in healthcare?

Answer – the only ethical concern I have isn't directly related to the ai machine. With this new age we see a lot of hackers targeting the NHS for file and data. So with ai taking over breast cancer diagnosis if a hacker can control it they can steal private data.

From these interviews I see the concept of artificial intelligence is known by many, however the use of ai in breast cancer detection isn't that well known. This is expected as ai usage in breast cancer detection isn't something that will publicly be discussed as "hot scoop." Even if they don't know about ai being used for breast cancer detection they still have concerns about privacy and how that could get breached. We also see there's concern about misdiagnosis.

In conclusion breast cancer diagnosis has come a long way from CBE. From detecting breast cancer with the naked eye to using machines that gave off too much radiation rasing health concerns to now the modern day using mammograms and mri to find tumours. There is still space to grow and increase accuracy in detection reducing the risk of getting breast cancer. Ai is the future and will most probably be used to detect breast cancer in the nearby future. As shown by my attempt at making a breast cancer detection machine ai can be used to detect early signs of tumours. With a bit more work and a greater training and test size the accuracy level could increase drastically.

***Bibliography***

*1 National Cancer Institute (NCI) (2023) National Institutes of Health. Available at: https://www.nih.gov/about-nih/what-we-do/nih-almanac/national-cancer-institute-nci (Accessed: 03 November 2023).*

*2 Cancer Research UK Early diagnosis data hub, Early Diagnosis. Available at: https://crukcancerintelligence.shinyapps.io/EarlyDiagnosis/*

*Picture : cbe*

*3 Armando E. Giuliano, J.O.-C.F.A. (2016) Breast disorders, Obgyn Key. Available at: https://obgynkey.com/breast-disorders-2/ (Accessed: 23 January 2024).*

*Cbe benfut and limi*

*4 Burke W, S.D.B.C. ACS Breast Cancer Screening Guidelines, ACS Breast Cancer Screening Guidelines | American Cancer Society. Available at: https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/american-cancer-society-recommendations-for-the-early-detection-of-breast-cancer.html (Accessed: 23 January 2024).*

*Cbe*

*5 D. Shockney, L. (2024) Clinical breast exam, National Breast Cancer Foundation. Available at: https://www.nationalbreastcancer.org/clinical-breast-exam/*

*6 Huang, N. et al. (2022) The efficacy of clinical breast exams and breast self-exams in detecting malignancy or positive ultrasound findings, Cureus. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8942605/ (Accessed: 23 January 2024).*

*7 KJ;, G.P. (no date) Screening for breast cancer with mammography, The Cochrane database of systematic reviews. Available at: https://pubmed.ncbi.nlm.nih.gov/23737396/ (Accessed: 23 January 2024).*

*8 Breast cancer screening (no date) National Cancer Institute. Available at: https://www.cancer.gov/types/breast/patient/breast-screening-pdq (Accessed: 24 January 2024).*

*9 Wolfe JN, Albert S, Belle S, et al. Breast parenchymal patterns and their relationship to risk for having or developing carcinoma. Radiol Clin North Am. 1983;21:127–136.*

*10 Gold RH. The evolution of mammography. Radiol Clin North Am. 1992;30:1–19.*

*11 Sickles EA. Mammographic features of 300 consecutive nonpalpable breast cancers. AJR Am J Roentgenol. 1986;146:661–663.*

*12 Smith-Bindman, R., Lipson, J., Marcus, R., Kim, K. P., Mahesh, M., Gould, R., ... & Miglioretti, D. L. (2009). Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer. Archives of Internal Medicine, 169(22), 2078-2086.*

*13 GE Digital senographe DS mammography (no date) Block Imaging . Available at: https://www.blockimaging.com/equipment/mammography/ge-digital-senographe-ds-mammography (Accessed: 24 January 2024).*

*14 Limitations of Mammograms (no date) How Accurate Are Mammograms? Available at: https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/limitations-of-mammograms.html#:~:text=Overall%2C%20screening%20mammograms%20miss%20about,when%20in%20fact%20they%20do. (Accessed: 30 January 2024).*

*15 A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," in IBM Journal of Research and Development, vol. 3, no. 3, pp. 210-229, July 1959, doi: 10.1147/rd.33.0210.*

*16 Alghodhaifi, H., Alghodhaifi, A., Alghodhaifi, M.: Predicting Invasive Ductal Carcinoma in breast histology images using Convolutional Neural Network. In: 2019 IEEE National Aerospace and Electronics Conference (NAECON), pp. 374–378 (2019)*

*17 Editorial Team, L. (2023) Hidden Layer, Lark. Available at: https://www.larksuite.com/en_us/topics/ai-glossary/hidden-layer#what-is-a-hidden-layer? (Accessed: 01 February 2024).*

*18 Convolutional Neural Network (CNN) in Machine Learning (2023) GeeksforGeeks. Available at: https://www.geeksforgeeks.org/convolutional-neural-network-cnn-in-machine-*

*learning/#:~:text=Convolutional%20Neural%20Network(CNN)%20%3A,layers%2C%20and%20fully%20connected%20layers. (Accessed: 01 February 2024).*

*19 Gurucharan, M. (2022) Basic CNN architecture: Explaining 5 layers of Convolutional Neural Network, upGrad blog. Available at: https://www.upgrad.com/blog/basic-cnn-architecture/ (Accessed: 01 February 2024).*

*20 Stednick, Z.S. (no date) breast cancer data.cvs, GitHub. Available at: https://github.com/stedy/Machine-Learning-with-R-datasets/blob/master/wisc_bc_data.csv*

*21 Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural networks, 61, 85-117.*

*22 Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.*

*23 Chollet, F. (2018*). Deep Learning with Python. Manning Publications.