A Project Report on

"A STUDY OF BREAST CANCER DETECTION USING MACHINE LEARNING."

Submitted to:
In partial fulfilment for the award of the degree of
Master of Business Administration

Submitted by:
ANIRBAN CHAKRABORTY

Under the Guidance of:

S. MUKHERJEE

DECLARATION

I here-by declare that the project with title **"A STUDY OF BREAST CANCER DETECTION USING MACHINE LEARNING."** has been completed by me in partial fulfilment of **MASTER**

OF BUSINESS ADMINISTRATION degree examination and this has not been submitted for any other examination and does not form the part of any other course undertaken by me.

Anirban Chakraborty

Date:25/05/25 learner's signature

ACKNOWLEDGEMENT

With immense pride and sense of gratitude, I take this golden opportunity to express
my sincere regards to Indian Institute of Technology Patna
I am extremely thankful to my Project Guide _____NALIN BHARATI_____for her guidance
throughout the project. I tender my sincere regards to the __S MUKERJEE_____for
giving me guidance, suggestions and invaluable encouragement, which helped me
in the completion of the project.
I would like to thank all those who helped me in making this project complete and successful.

Anirban Chakraborty

Date:22/5/25 Learner's name

Index

1. Introduction

Amongst many forms of treatment and screening, breast cancer (BC) – the unmonitored production of abnormal breast cells – is still the second most diagnosed and leading cause of cancer death among women in the United States. In 2019, there have been 268,600 women diagnosed with BC where 41,760 have died, according to the American Cancer Association. In the European Union, in 2020, breast cancer was the leading cancer diagnosis among women that accounted for 13.3 % of all cancer diagnoses ([1], p. 1). The early detection of abnormal breast tissue is necessary for an accurate breast cancer diagnosis. By detecting breast tumours early, it can improve survival rates, because the tumours can be easily treated in earlier stages.

For women diagnosed with breast cancer in the EU, 7.4 % of women that are diagnosed with initial primary breast cancer (PBC) will need a second primary breast cancer (SBC) diagnosis in the span of 10 years. For breast cancer recurrence, it is the highest within the first five years with 10.4 %. The European Society for Medical Oncology has guidelines for breast cancer patients that suggest routine visits every 3–4 months in the first 2 years after the PBC. This will aid in the early detection of SBC among breast cancer survivors by improving the rate of survival from 27 % to 47 %

(Sialon et al., [2], p. 933). In predicting the probability of SBC with patient-level data from PBC tumours or from the patient themselves, it has the potential to be a useful tool for doctors to create informed decisions regarding at-risk patient's follow-up, detect new, growing tumours, or prevent new BC incidence.

In the US, the majority of BC responds well to treatment and have good prognoses and survival rate. At the same time, a considerable proportion of BC are considered triple negative breast cancer (TNBC) – for the fact that BC is a heterogeneous disease with multiple types and many subtypes. The TNBC is a specific subtyping of BC that is portrayed by its lack of expression to the three most targeted biomarkers for BC treatment and accounts for 15 %–20 % of all diagnosed BCs annually ([1], p. 2). The TNBC are also depicted as more clinically aggressive in behaviour, poor prognosis, higher recurrence rate, and grim survivor rate. There is a higher need for accurate algorithm developments for identifying and distinguishing positive TNBC tumours to non-TNBC tumours, to prioritize special treatment for TNBC and standard treatment for non-TNBC.

With the various diagnostic tests – such as breast biopsy aspiration cytology – that can be used for diagnoses, mammography is the standard method for BC detection. However, there are instances where a biopsy is necessary when the mammography is insufficient. Correspondingly, the detection rate through mammography is about 60 %–70 % accurate (Malakouti et al., [3], p. 1). These biopsies take the

breast tissue to examine the cells to see whether the cancer is malignant or benign. The identification is crucial in determining the types of cells involved in BC, the aggressiveness of the cancer, and whether it has hormone receptors – as it affects treatment options. In post-biopsy, the breast tissue is sent to a specialist to analyze the size, consistency, and the arrangement of the malignant and noncancerous cells in the breast tissue. Any discrepancies in opinions facilitate more biopsies and further examinations (Malakouti et al., [3], p. 1).

Machine learning techniques have become more integrated and applied in medical health. It supports case-based reasoning in medical decisions – by not having the limitations of traditional diagnostic methods – to increase the diagnostic accuracy and prognosis decisions – by reducing human error that leads to patient mortality – to improve the health of BC patients (Malakouti et al., [3], p. 1). As the machine learning algorithm improves in accuracy and response time, it helps automate decision making processes that will enhance patient survival rates through their beneficial usage as support systems. In existing studies for machine learning in BC predictions, the algorithm is used to predict and classify benign or malignant BC tumours or recurrent cancer with characteristics from clinical datasets.

With the improvements of medicine and treatments, there have been an increase in women surviving breast cancer. With the increase in cases of patients entering remission, there has been a growing importance in survivorship issues.

There has been reported pain at different stages after surgery with 15 % experiencing moderate to severe pain 1 year post surgery and 34 % of patients have indicators of neuropathic pain – which could last for several years that reduces their quality of life (Letsch et al., [4], p. 399–400). In order to prevent post-surgical pain, these patients would need to be identified as high risk to post-surgical pain in order to provide the required medical and psychosocial intervention. Also, being able to identify those not at risk, it is important to see which patient can be dismissed from persistent pain to prevent unnecessary therapeutic intervention (Letsch et al., [4], p. 400). This would be another method in where machine learning can be implemented in order to predict which patient would be at risk for persistent pain.

Outside of the scope of machine learning being used for breast cancer, it also has its uses in predicting other diseases. A chronic disease that affects people of all age groups is diabetes mellitus (DM). The factors that have an influence on the development of DM are age, family history, relative diseases, pregnancy, variable glucose levels, blood pressure, etc., and diabetes can be separated into four broad categories: type-1, type-2, gestational diabetes, and prediabetes. In recent developments, machine learning has been slowly integrated to predict DM. “Some of the commonly used algorithms include Logistic Regression (LR), XGBoost (XGB), gradient boosting (GB), decision trees (DTs), Extra Trees, Random Forest (RF), and light gradient boosting machine (LGBM) (Ahamed et al., [5], p. 1).”

After seeing how machine learning can be applied to diabetes mellitus, machine learning can be used as a method of predicting type 2 diabetes (T2D). T2D is considered as one of the fastest growing, life-threatening chronic diseases. Within the past 40 years, T2D has increase by four times with the chance of death never decreasing and symptoms increasing with age from 20 years to 60 (Anasanti et al., [6], p. 832). The mortality rate of T2D is caused by the latency in diagnosis.

The symptoms of T2D can be very similar to other common sicknesses that can cause people to be unaware of the fact that they may have T2D. In order to diagnose T2D, blood tests are used to properly diagnose T2D. The issue with this is that blood tests are becoming costly for developing countries. This is how machine learning can be used as an alternative by predicting health risk factors and symptoms. Machine learning can be a lower cost opportunity to effectively predict diseases (Anasanti et al., [6], p. 832).

Just as diabetes has been a fast-growing disease, heart disease has become another fast-growing disease. In the medical field, machine learning techniques are not only reserved in predicting breast cancer. Machine learning techniques can be applied in detection of heart disease (HD). With high blood pressure, diabetes, cholesterol fluctuations, exhaustion, etc., HD is a widely common disease among people ([7,8], p. 2). The capabilities of early diagnosis have been widely sought after with data analytical tools being provided to health care workers to identify any early signs of HD. For preventative measures, there can be multiple

assigned tests for potential HD patients as a precaution to reduce the likelihood of developing HD, and if there can be any reliable methods of predicting HD, the use of machine learning can be pivotal in saving the lives of the patient.

Considering how machine learning is used to benefit women in detecting the early onset of breast cancer, machine learning is also used in predicting early symptoms of another disease that is common among women. Endometriosis is a complex gynecological disorder that is common among 176 million women worldwide, 8.5 million women affected in North America. The impact of endometriosis is high among different age groups of women: 5–10 % of women who are of reproductive age, 20–30 % of women with subfertility, and 40–60 % of women with infertility – alongside with chronic pelvic pain (Akter et al., [9], p. 2); for those with pelvic pain, 70 % of teen are later diagnosed with endometriosis. For all 600,000 hysterectomies in the US every year, endometriosis is the leading cause and can severely impair the mental and physical quality of life for patients. Endometriosis also has an impact on an economic scale. It has a detrimental effect on work performance for women and is a burden on workdays from the loss. The cost in the health care such as outpatient visits, hospitalization, and medications allotted for an estimated $22 billion each year.

As standard practice, the process of endometriosis diagnosis involves a laparoscopy – an invasive procedure. Coupled with a lack of definitive clinical diagnostic approach and a simple molecular diagnostic approach, the latency of a laparoscopy

is an average 4–11 years (Akter et al., [9], p. 3). This would mean that an early intervention is paramount in reducing the suffering and costs that would come from this disease. Another method that is less invasive compared to laparoscopy is endometrial biopsy – which is useful in reducing diagnostic latency. This is due to the fact that endometriosis patients have an altered methylome (DNA methylation) and transcriptome (RNA-seq), and the differences in DNA methylation and gene expression leads to identifying biomarkers to develop less invasive diagnostic techniques for endometriosis (Akter et al., [9], p. 3).

With the discovery of relevant biological patterns from microarray expression or next generation sequencing data, there has been continuous advancements by implementing machine learning tools. Supervised and unsupervised machine learning methods have been used in microarray expression data. For unsupervised machine learning, studies use this method in evaluating clustering techniques such as hierarchical clustering and K-means clustering to identify the genes that share similar expressions. In supervised machine learning methods, it is used to evaluate the application of disease against healthy classification tasks in decision trees, Random Forests, artificial neural networks (ANN), support vector machines (SVM) and Bayesian networks. Also, the availability of transcriptomics and methylomics data have been increasing with time; this allows for more opportunities in clinical diagnostics (Akter et al., [9], p. 3).

In recent years, machine learning techniques have been applied extensively in various fields, including healthcare and bioinformatics, to improve diagnostic accuracy and reduce human error. For instance, neural networks have shown significant potential in classifying sleep-wake stages in neonatal patients using EEG data. Abbasi et al. [10] implemented a multilayer perceptron neural network to classify EEG-based neonatal sleep-wake stages, demonstrating impressive accuracy in this domain. Similarly, ensemble learning methods were applied by Abbasi, Jamil, and Chen [11] for EEG-based neonatal sleep stage classification, further improving performance through combined classifier outputs. Expanding these machine learning applications, a more recent study by Arslan et al. [12] introduced a deep features-based approach utilizing a modified ResNet50 and gradient boosting for visual sentiment classification, showcasing the versatility of machine learning algorithms across diverse fields.

Main contributions to this paper are:

-

Enhanced Diagnostic Accuracy: The study applies machine learning classifiers to improve the accuracy of breast cancer diagnosis, addressing the limitations of mammography (70 % accuracy) and biopsy-related human error.

-

Feature Selection Impact: The research explores how feature selection affects the performance of eight classifiers—Logistic

Regression, Random Forest, Extra Trees, XGB, LGBM, CatBoost, SVC, and Gaussian Naïve Bayes—in predicting breast cancer diagnoses.

- Classifier Performance Evaluation: The study evaluates classifier performance using raw accuracy, precision score, recall score, and F1-score, highlighting the effectiveness of specific classifiers like Logistic Regression (91.67 % accuracy) and LGBM (90.74 % accuracy with feature selection).

- Improved LGBM Classifier: A significant improvement is observed in the LGBM classifier's performance after applying feature selection, demonstrating the importance of selecting relevant features.

- Common Features Across Classifiers: Tumour Size, Age, Metastasis, and Inv-Nodes were identified as the most influential features across all classifiers, reinforcing their significance in breast cancer prediction.

- Contribution to Bioinformatics: The study advances the application of machine learning in bioinformatics, suggesting the potential to improve diagnostic accuracy not only for breast cancer but also for other diseases like heart disease, diabetes, and endometriosis.

-

Support for Medical Professionals: The research emphasizes the growing need for integrating machine learning into healthcare to assist medical professionals in diagnosing and treating diseases with higher precision and reduced latency.

The rest of the paper is organized as follows. Materials and Methodology presented in Section 2. We present the results in Section 3. Feature Selections methods are presented in Section 4. Discussion in Section 5, and Future Work and Improvements in Section 6, and finally the conclusion in Section 7.

2. Materials and Methodology

The preparation of the dataset and the methodology employed for classification using various machine learning classifiers are detailed in the following section. This includes data preprocessing steps, feature selection techniques, and the application of multiple classifiers to assess the predictive performance on the dataset. Each classifier's approach and evaluation metrics are thoroughly explained to provide insights into the accuracy, precision, recall, and F1-scores achieved during the classification process.

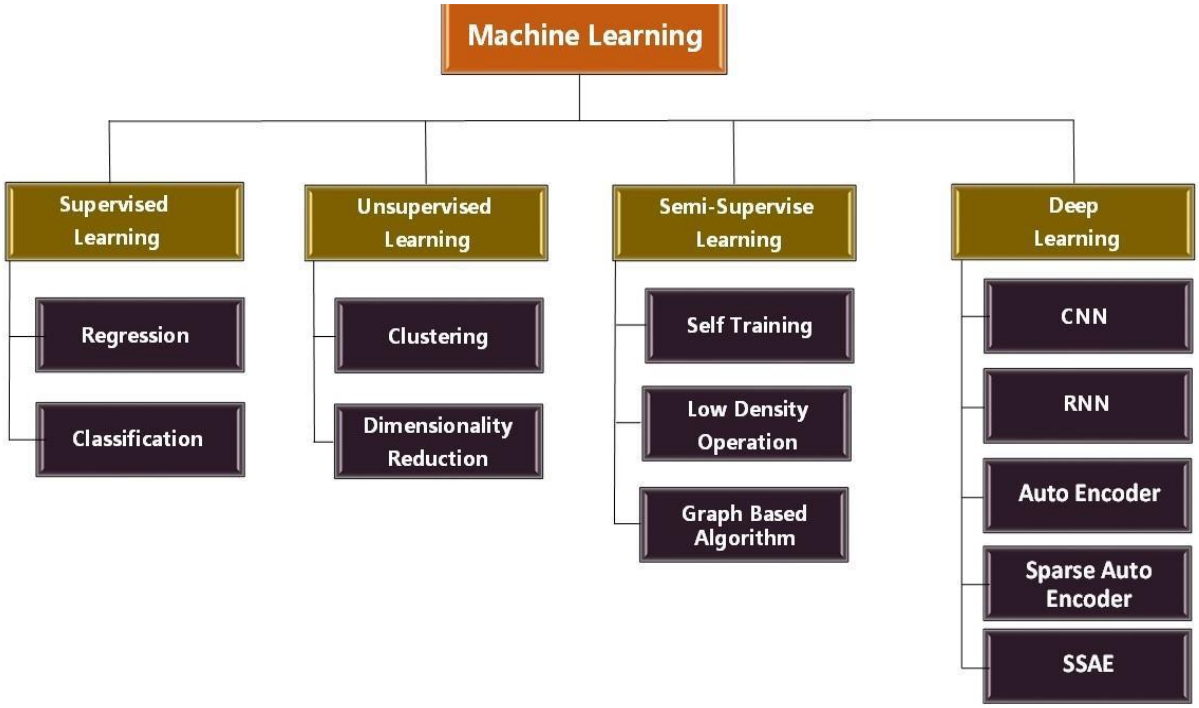
2.1. Preparation of breast cancer dataset

The breast cancer data was taken from (Breast Cancer Prediction (kaggle.com)) by Fatemeh Mahavar. The only preprocessing that was done on the data was removing a line that contained empty data. It was line 42 of the datasheet or number 41 of the unique identification of the patient. In the dataset, there are total of 11 features: the unique

identification for each patient (S/N), the year of diagnosis (Year), the patient's age at time of diagnosis (Age), if the patient is pro- or postmenopausal at time of diagnosis – where a 0 value is when a patient has researched menopause and a value of 1 is when a patient has not reached menopause – (Menopause), the size of the excised tumor in centimetre's (Tumour Size), the number of axillary lymph node that contain metastatic (Inv-Nodes), whether the cancer occurs on the left or right side of the breast (Breast), whether the cancer has spread across to other parts of the body (Metastatic), where of the four sections on the breast has cancer with the nipple as the centre (Breast Quadrant), if there is any history of the patient having cancer or a family history of cancer – where a value of 1 means there is history of cancer and a value of 0 is no history of cancer (History), and the diagnosis result whether the patient has a benign or malignant tumour in the dataset (Diagnosis result).

LITERATURE REVIEW

1. This chapter gives the recent research work and contributions done in the field of breast cancer detection with machine learning techniques, and explains the various methods use to detect breast cancer. 2.2 Overview on Machine Learning Algorithms Machine Learning is a subset of Artificial Intelligence that uses statistical learning algorithms to build systems that have the ability to automatically learn and improve from experiences without being explicitly programmed. Deep learning is a type of machine learning and artificial intelligence (AI) that imitates the way humans gain certain types of knowledge. While traditional machine learning algorithms are linear, deep learning algorithms are stacked in a hierarchy of increasing complexity and abstraction. At its most basic sense, machine learning uses programmed algorithms that learn and optimize their operations by analysing input data to make predictions within an acceptable range. With the feeding of new data, these algorithms tend to make more accurate predictions. Although there are some variations of how to group machine learning algorithms, they can be divided into three broad categories according to their purposes and the way the underlying machine is being taught. These three categories are: supervised, unsupervised and semi-supervised. There also exists a fourth category known as reinforcement ML. Figure 2.1 shows an illustration of the classification of machine learning algorithms.



2.2.1 Supervised Machine Learning Algorithms

In this type of algorithms, a model gains knowledge from the data that has predefined examples of data with both input and expected output to compare its output with the correct input. Classification problem is one of the standard formulations for supervised learning task where the data is mapped into a class after looking at numerous input-output examples of a function. Supervised learning is a branch of ML which deals with a given dataset consisting of multiple data along with their corresponding classes. It can be used both for decision trees and artificial neural networks. In decision trees it can be used to determine which attributes of the data given provides the most relevant information. In artificial neu ral networks, the models are trained on the given dataset and classifications of an unknown sample of data are being carried out.

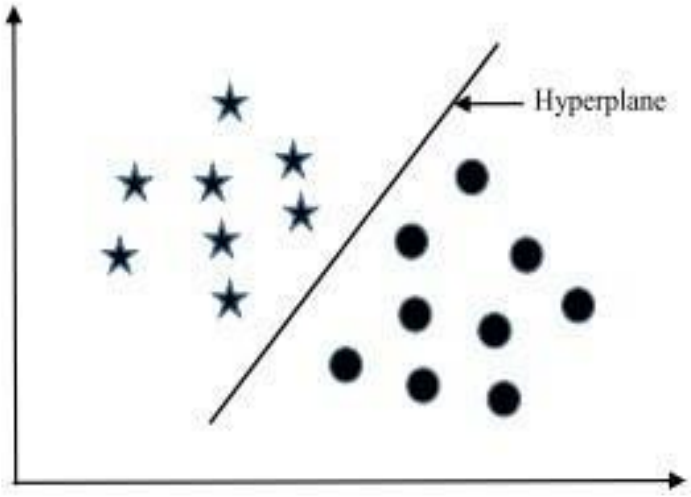
1.Logistic Regression

Logistic regression Logistic regression (LR) is a powerful and well-established method for supervised classification [4]. It can be considered as an extension of ordinary regression and can model only a dichotomous variable which usually represents the occurrence or non-occurrence of an event. LR helps in finding the probability that a new instance belongs to a certain class. Since it is a probability, the outcome lies between 0 and 1. Therefore, to use the LR as a binary classifier, a threshold needs to be assigned to differentiate two classes. For example, a probability value higher than 0.50 for an input instance will classify it as ‟class A‟; otherwise, ‟class B‟.

2. Support Vector Machine (SVM)

Support vector machine (SVM) algorithm can classify both linear and non-linear data. It first maps each data item into an n-dimensional feature space where n is the number of features. It then identifies the hyper plane that separates the data items into two classes while maximizing the marginal distance for both classes and minimizing the classification errors. The marginal distance for a class is the distance between the decision hyper plane and its nearest instance which is a member of that class. Figure 2.2 shows an illustration of the support Vector machine. The SVM has identified a hyper plane (actually a line) which maximizes the separation between the „star‟ and „circle‟ classes. More formally, each data point is plotted first as a point in an n-dimension space (where n is the number of features) with the

value of each feature being the value of a specific coordinate. To perform the classification, we then need to find the hyper plane that differentiates the two classes by maximum margin.



2: A simplified illustration of how the support vector machine works

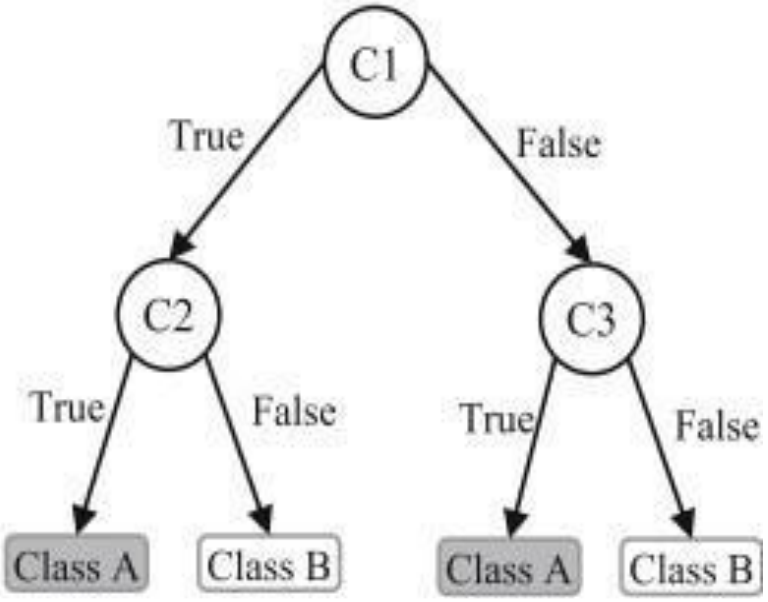
3 Decision Tree (DT)

Decision tree (DT) is one of the earliest and prominent machine learning algorithms. A decision tree tests and corresponds outcomes for classifying data items into a tree-like structure. The nodes of a decision tree normally have multiple levels where the first or top-most node is called the root node. All internal nodes (i.e., nodes having at least one child) represent tests on input variables or attributes.

Figure 2.3 shows an illustration of the Decision Tree. Each variable (C1, C2, and C3) is represented by a circle and the decision outcomes (Class A and Class B) are shown by rectangles. In order to successfully classify a sample to a

class, each branch is labelled with either „True‟ or „False‟ based on the outcome value from the test of its ancestor node.

Depending on the test outcome, the classification algorithm branches towards the appropriate child node where the process of test and branching repeats until it reaches the leaf node. The leaf or terminal nodes correspond to the decision outcomes. DTs have been found easy to interpret and quick to learn, and are a common component to many medical diagnostic protocols. When traversing the tree for the classification of a sample, the outcomes of all tests at each node along the path will provide sufficient information to conjecture about it’s class

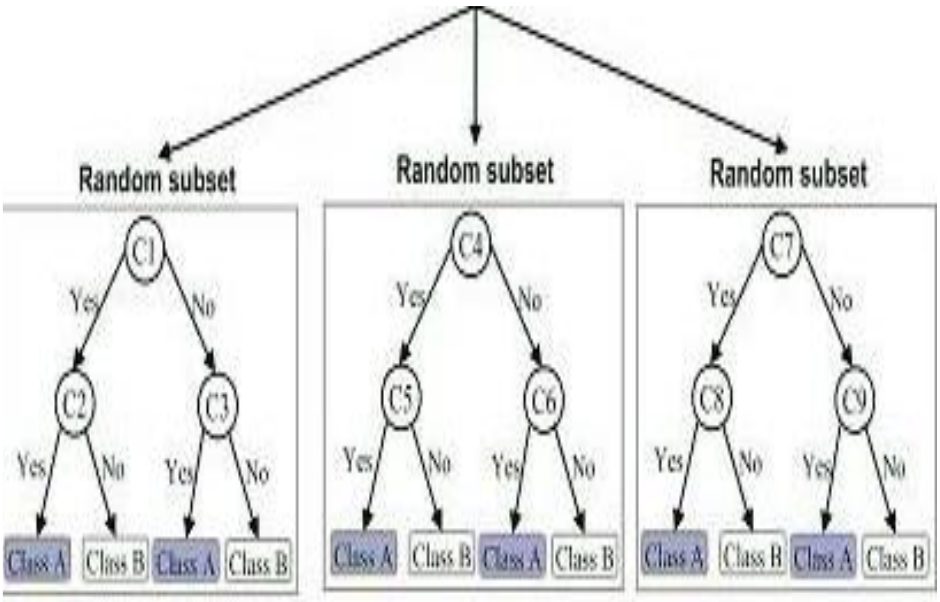


A simplified illustration of how the decision tree works

4 Random Forest (RF)

A random forest (RF) is an ensemble classifier and consisting of many DTs similar to the way a forest is a collection of many trees. DTs that are grown very deep often cause over fitting of the training data, resulting a high variation in classification outcome for a small change in the input data. They are very sensitive to their training data, which makes them error-prone to the test dataset. The different DTs of an RF are trained using the different parts of the training dataset.

Figure 2.4 shows an illustration of the RF algorithm which consists of three different decision trees. Each of those three decision trees was trained using a random subset of the training data. To classify a new sample, the input vector of that sample is required to pass down with each DT of the forest. Each DT then considers a different part of that input vector and gives a classification outcome. The forest then chooses the classification of having the most "votes" (for discrete classification outcome) or the average of all trees in the forest (for numeric classification outcome). Since the RF algorithm considers the outcomes from many different DTs, it can reduce the variance resulted from the consideration of a single DT for the same dataset.



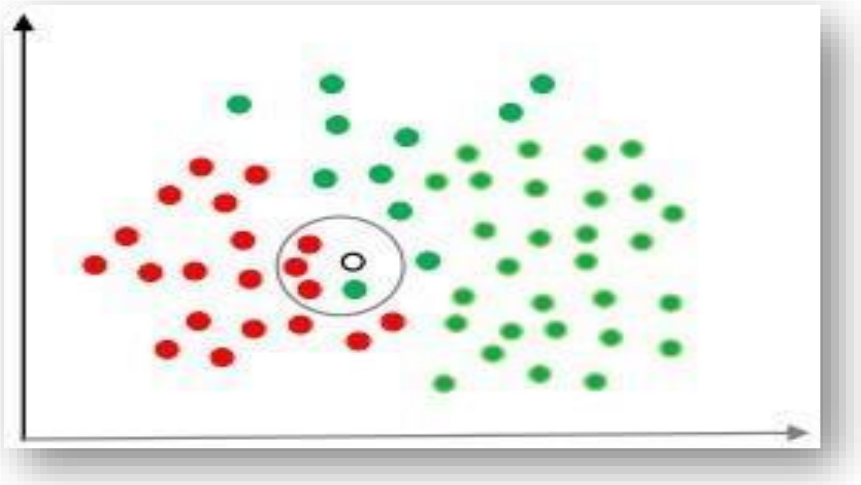
A simplified illustration of how the random forest works

5 Naïve Bayes (NB)

Naïve Bayes (NB) is a classification technique based on the Bayes‟ theorem. This theorem can describe the probability of an event based on the prior knowledge of conditions related to that event. This classifier assumes that a particular feature in a class is not directly related to any other feature although features for that class could have interdependence among themselves. By considering the task of classifying a new object (white circle) to either „green‟ class or „red‟ class, Figure 2.5 shows an illustration of the Naive Baiyes Algorithm. According to this figure, it is reasonable to believe that any new object is twice as likely to have „green‟ membership rather than „red‟ since there are twice as many „green‟ objects (40) as „red‟. In the Bayesian analysis, this belief is known as the prior probability. Therefore, the prior probabilities of „green‟ and „red‟ are 0.67 (40 ÷ 60) and 0.33

(20 ÷ 60), respectively. Now to classify the „white‟ object, we need to draw a circle around this object which encompasses several points (to be chosen prior) irrespective of their class labels. Four points (three „red‟ and one „green) were considered in Thus, the likelihood of „white‟ given „green‟ is 0.025 (1 ÷ 40) 14

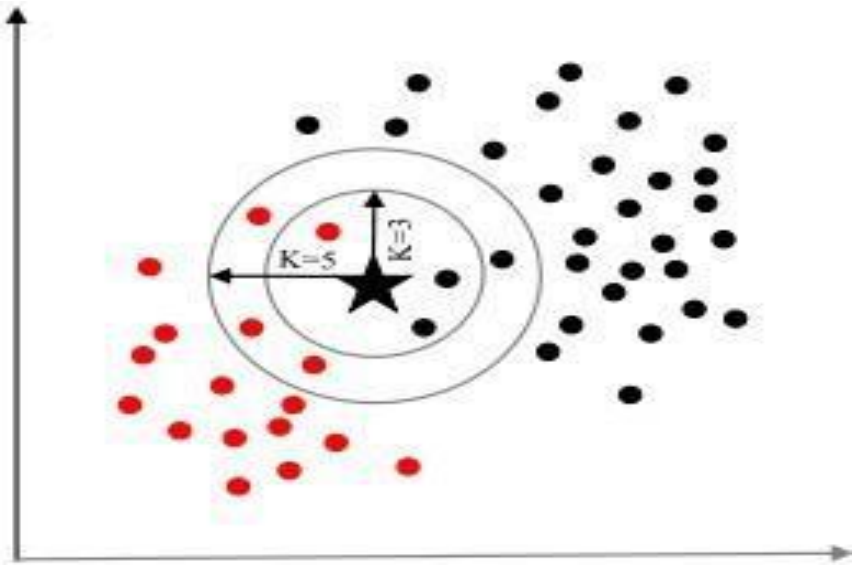
and the likelihood of „white‟ given „red‟ is 0.15 (3 ÷ 20). Although the prior probability indicates that the new „white‟ object is more likely to have „green‟ membership, the likelihood shows that it is more likely to be in the „red‟ class. In the Bayesian analysis, the final classifier is produced by combining both sources of information (i.e., prior probability and likelihood value). The „multiplication‟ function is used to combine these two types of information and the product is called the „posterior‟ probability. Finally, the posterior probability of „white‟ being „green‟ is 0.017 (0.67 × 0.025) and the posterior probability of „white‟ being „red‟ is 0.049 (0.33 × 0.15). Thus, the new „white‟ object should be classified as a member of the „red‟ class according to the NB technique.



6 K-Nearest Neighbor (KNN)

The K-nearest Neighbor (KNN) algorithm is one of the simplest and earliest classification algorithms. It can be thought a simpler version of an NB classifier. Unlike the NB tech- nique, the KNN algorithm does not require to consider probability values. The "K" is the

KNN algorithm is the number of nearest Neighbors considered to take "vote" from. The se- lection of different values for "K" can generate different classification results for the same sample object shows an illustration of the KNN algorithm. For K=3, the new object (star) is classified as "black"; however, it has been classified as "red" when K=5.

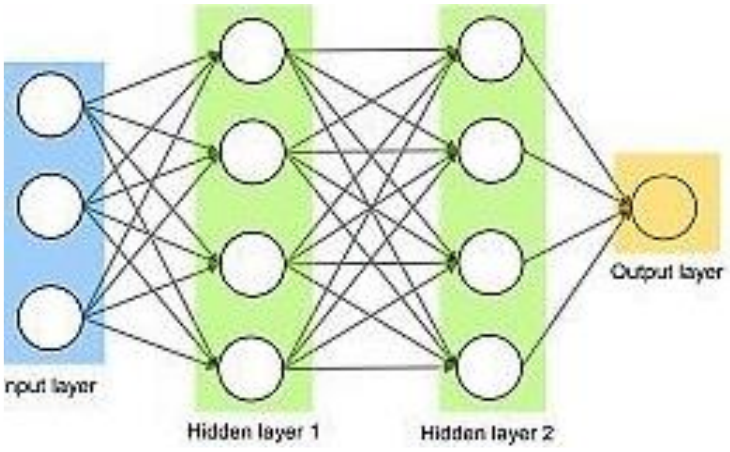


A simplified illustration of the K-nearest neighbor algorithm

7 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are a set of machine learning algorithms which are in- spired by the functioning of the neural networks of human brain. They were first proposed by McCulloch and Pitts and later popularized by the works of Rumelhart et al. in the 1980s. In the biological brain, neurons are connected to each other through multiple axon junctions forming a graph like architecture. These interconnections can be rewired (e.g., through neuroplasticity) that helps to adapt, process and store information. Figure 2.7 shows an illustration of artificial neural networks with two hidden layers. The arrows connect the output of nodes from one layer to the input of nodes of another layer. Likewise, ANN algorithms can be represented as an interconnected group of nodes. The output of one node goes as input to another node for subsequent processing according to the interconnection. Nodes are normally grouped into a matrix called layer depending on the transformation they perform. Apart 16

from the input and output layer, there can be one or more hidden layers in an ANN framework. Nodes and edges have weights that enable to adjust signal strengths of communication which can be amplified or weakened through re- peated training. Based on the training and subsequent adaption of the matrices, node and edge weights, ANNs can make a prediction for the test data.



7: An illustration of the artificial neural network structure with two hidden layers .

2.2.2 Unsupervised Machine Learning Algorithms

In unsupervised learning, only input data is provided to the model the use of labeled datasets. Unsupervised learning algorithms do not use labeled input and output data. An example of unsupervised learning is clustering. In contrast to supervised learning, unsu- pervised learning methods are suitable when the output variables (i.e. the labels) are not provided. Some examples of unsupervised learning algorithms include K-Means Cluster- ing, Principal Component Analysis and Hierarchical Clustering.

1 K-Mean Clustering: K mean is clustering algorithm that provides the partition of data in the form of small clusters. Algorithm is used to find out the similarity between different data points. Data points exactly consist of at least one cluster that is most suitable for the evaluation of big dataset.

2 C-Mean CLUSTERING: Clusters are identified on the similarity basis. Cluster that consists of similar data point belongs to one single family. In C mean algorithm each data point belongs to one single cluster. It is mostly used in medical images segmentation and disease prediction

3 Hierarchical Algorithm: Hierarchical algorithm mostly provides the evaluation of raw data in the form of matrix. Each cluster is separated from other clusters in the form of hierarchy. Every single cluster consists of similar data points. Probabilistic model is used to measure the distance between each cluster

4 Gaussian Mixture Algorithm: It is most popular technique of unsupervised learning. It is known as soft clustering technique which is used to compute the probability of different types of clustered data. The implementation of this algorithm is based on expectation maximization.

2.2.3 SemiSupervised Machine Learning Algorithms

Semi-supervised machine learning is a combination of supervised and unsupervised ma- chine learning methods. With more common supervised machine learning methods, you train a machine learning algorithm on a "labeled" dataset in which each record includes the outcome information. Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data). Semi supervised

learning is used in speech analysis. Since labeling of audio files is a very intensive task, Semi-Supervised learning is a very natural approach to solve this problem. Internet Content Classification: Labeling each webpage is an impractical and unfeasible process and thus uses Semi-Supervised learning algorithms.

2.2.4 Deep Learning Algorithms

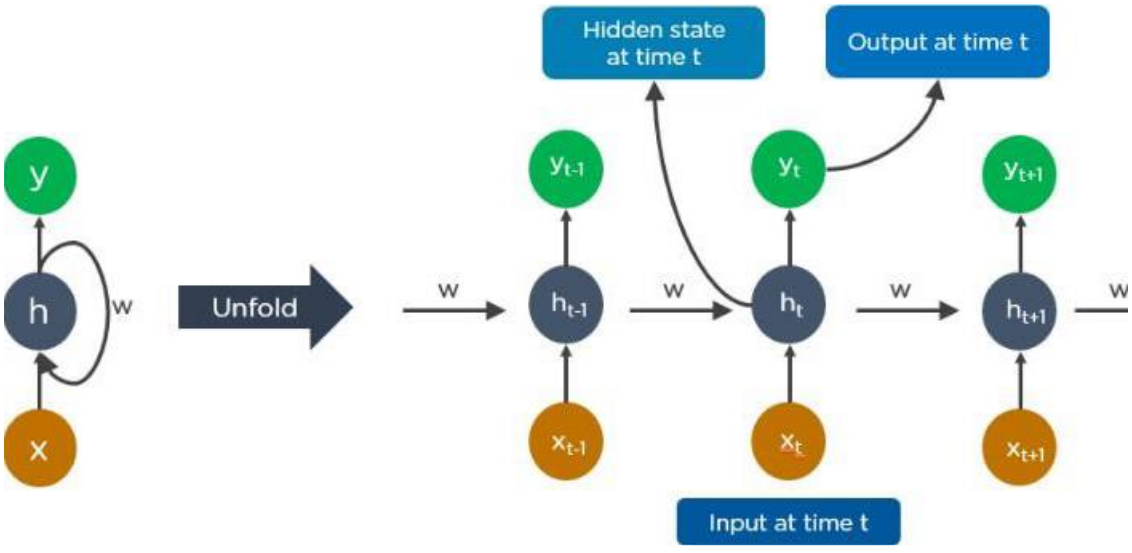
Deep learning has gained massive popularity in scientific computing, and its algorithms are widely used by industries that solve complex problems. All deep learning algorithms use different types of neural networks to perform specific tasks. Here is the list of top 10 most popular deep learning algorithms. [6]

1 Convolutional Neural Networks (CNNs): CNN‟s, also known as ConvNets, consist of multiple layers and are mainly used for image processing and object detection. Yann LeCun developed the first CNN in 1988 when it was called LeNet. It was used for recognizing characters like ZIP codes and digits. CNN‟s are widely used to identify satellite images, process medical images, forecast time series, and detect anomalies. [6]

2 Long Short-Term Memory Networks (LSTMs): LSTMs are a type of Recurrent Neural Network (RNN) that can learn and memorize long- term dependencies. Recalling past information for long periods is the default behavior.

LSTMs retain information over time. They are useful in time-series prediction because they remember previous inputs. LSTMs have a chain-like structure where four interacting layers communicate in a unique way. Besides time-series predictions, LSTMs are typically used for speech recognition, music composition, and pharmaceutical development. [6]

3 Recurrent Neural Networks (RNNs): An unfolded RNN is illustrated in Figure 2.8. RNNs have connections that form directed cycles, which allow the outputs from the LSTM to be fed as inputs to the current phase. The output from the LSTM becomes an input to the current phase and can memorize previous inputs due to its internal memory. RNNs are commonly used for image captioning, time-series analysis, natural-language processing, handwriting recognition and machine translation.



An illustration of recurrent neural networks

4 Generative Adversarial Networks (GANs): GANs are generative deep learning algorithms that create new data instances that resemble the training data. GAN has two components: a generator, which learns to generate fake data, and a discriminator, which learns from that false information. The usage of GANs has increased over a period of time. They can be used to improve astronomical images and simulate gravitational lensing for dark-matter research. Video game developers use GANs to upscale low-resolution, 2D textures in old video games by recreating them in 4K or higher resolutions via image training. GANs help generate realistic images and cartoon characters, create photographs of human faces, and render 3D objects. **5 Radial Basis Function Networks (RBFNs):** RBFNs are special types of feed forward neural networks that use radial basis functions as activation functions. They have an input layer, a hidden layer, and an output layer and are mostly used for classification, regression, and time-series prediction. **6 Multilayer Perceptrons (MLPs):** MLPs are an excellent place to start learning about deep learning technology. MLPs belong to the class of feed forward 20

neural networks with multiple layers of perceptrons that have activation functions. MLPs consist of an input layer and an output layer that are fully connected. They have the same number of input and output layers but may have multiple hidden layers and can be used to build speech-recognition, image-recognition, and machine-translation software. **7 Self-Organizing Maps (SOMs):** Professor Teuvo Kohonen invented SOMs, which enable data visualization to reduce the

dimensions of data through self-organizing artificial neural networks. Data visualization attempts to solve the problem that humans cannot easily visualize high- dimensional data. SOMs are created to help users understand this high-dimensional information. **8 Deep Belief Network (DBNS):** DBNs are generative models that consist of multiple layers of stochastic, latent variables. The latent variables have binary values and are often called hidden units. [6] DBNs are a stack of Boltzmann Machines with connections between the layers, and each RBM layer communicates with both the previous and subsequent layers. Deep Belief Networks (DBNs) are used for image-recognition, video-recognition, and motion-capture data.

9 Restricted Boltzmann Machines (RBMs): Developed by Geoffrey Hinton, RBMs are stochastic neural networks that can learn from a probability distribution over a set of inputs. This deep learning algorithm is used for dimensionality reduction, classification, regres- sion, collaborative filtering, feature learning, and topic modeling. RBMs constitute the building blocks of DBNs. RBMs consist of two layers: Visible units and Hidden units Each visible unit is connected to all hidden units. RBMs have a bias unit that is connected to all the visible units and the hidden units, and they have no output nodes. [6]

10 Auto encoders: Auto encoders are a specific type of feed forward neural network in which the input and output are identical. Geoffrey Hinton designed auto encoders in the

1980s to solve unsupervised learning problems. They are trained neural networks that replicate the data from the input layer to the output layer. Auto encoders are used for purposes such as pharmaceutical discovery, popularity prediction, and image processing. [6]

2.3 Review of Previous Works on Machine Learning for General Diseases Prediction

Extensive work was carried out in the field of Artificial Intelligence, especially Machine Learning, to detect common diseases. Dahiwade et al. [7] proposed a ML based system that predicts common diseases. The symptoms dataset was imported from the UCI ML depository, where it contained symptoms of many common diseases. The system used CNN and KNN as classification techniques to achieve multiple diseases prediction. Moreover, the proposed solution was supplemented with more information that concerned the living habits of the tested patient, which proved to be helpful in understanding the level of risk attached to the predicted disease. Dahiwade et al. compared the results between KNN and CNN algorithm in terms of processing time and accuracy. The accuracy and processing time of CNN were 84.5% and 11.1 seconds, respectively.

In light of this study, the findings of Chen et al. [8] also agreed that CNN outperformed typical supervised algorithms such as KNN, NB, and DT. The authors concluded that the proposed model scored higher in terms of accuracy, which is explained by the capability of the model to detect complex nonlinear relationships in the feature space. Moreover, CNN detects

features with high importance that renders better description of the disease, which enables it to accurately predict diseases with high complexity. This conclusion is well sup- ported and backed with empirical observations and statistical arguments. Nonetheless, the presented models lacked details, for instance, neural 22

networks parameters such as network size, architecture type, learning rate and back propagation algorithm, etc. In addition, the analysis of the performances is only evaluated in terms of accuracy, which debunks the validity of the presented findings. Moreover, the authors did not take into consideration the bias problem that is faced by the tested algorithms. In illustration, the incorporation of more feature variables could immensely ameliorate the performance metrics of under- performed algorithms. Uddin et al [5] compared the various supervised ML techniques. In their study, extensive research efforts were made to identify those studies that applied more than one supervised machine learning algorithm on single disease prediction. Two databases (i.e., Scopus and PubMed) were searched for different types of search items. Thus, they selected 48 articles in total for the comparison among variants supervised machine learning algorithms for dis- ease prediction. They found that the Support Vector Machine (SVM) algorithm is applied most frequently (in 29 studies) followed by the Na¨ıve Bayes algorithm (in 23 studies). However, the Random Forest (RF) algorithm showed superior accuracy comparatively. Of the 17 studies where it was applied, RF showed the highest accuracy in 9 of them, i.e., 53%. This was

followed by SVM which topped in 41% of the studies it was considered. **2.4 Review of Previous Works on Machine Learning for Breast Cancer Prediction**

Sengar et al. [9] attempted to detect breast cancer using ML algorithms, namely RF, Bayesian Networks and SVM. The researchers obtained the Wisconsin original breast cancer dataset from the UCI repository and utilized it for comparing the learning models in terms of key parameters such as accuracy, recall, precision, and area of ROC graph. The classifiers were tested using K-fold validation method, where the chosen value of K is equal to 10. The simulation results have proved that SVM excelled in terms of 23

recall, accuracy, and precision. However, RF had a higher probability in the correct classification of the tumor, which was implied by the ROC graph. In contrast, Yao [10] experimented with various data mining methods including RF and SVM to determine the best suited algorithm for breast cancer prediction. Per results, the classification rate, sensitivity, and specificity of Random Forest algorithm were 96.27%, 96.78%, and 94.57%, respectively, while SVM scored an accuracy value of 95.85%, a sensitivity of 95.95%, and a specificity of 95.53%. Yao came to the conclusion that the RF algorithm performed better than SVM because the former provides better estimates of information gained in each feature attribute. Furthermore, RF is the most adequate at breast diseases classification, since it scales well for large datasets and prefaces lower chances of variance and data over fitting. The studies advantageously presented multiple

performance metrics that solidified the underlined argument. Nevertheless, the inclusion of the preprocessing stage to prepare raw data for training proved to be disadvantageous for ML models. According to Yao, omitting parts of data reduces the quality of images, and therefore the performance of the ML algorithm is hindered.

Noreen Fatima et al. [1] performed a comparative review or machine learning techniques and analyzed their accuracy across various journals. Her main focus is to comparatively analyze different existing Machine Learning and Data Mining techniques in order to find out the most appropriate method that will support the large dataset with good accuracy of prediction. She found out that machine learning techniques were used in 27 papers, ensemble techniques were used in 4 papers, and deep learning techniques were used in 8 papers. She concluded by saying that each technique is suitable under different conditions and on different type of dataset, after the comparative analysis

of these algorithms we came to know that machine learning 24

algorithm SVM is the most suitable algorithm for prediction of breast cancer. Different researchers have provided the analysis of prediction algorithms by using the dataset from Wisconsin Diagnostic Breast Cancer (WDBC), and the analysis shows that each time the accuracy of SVM algorithm is higher than the other machine learning algorithms.

Delen et al. [11] used artificial neural networks, decision trees and logistic regression to develop prediction models for

breast cancer survival by analyzing a large dataset, the SEER cancer incidence database. Two popular data mining algorithms (artificial neural networks and decision trees) were used, along with a most commonly used statistical method (logis- tic regression) to develop the prediction models using a large dataset (more than 200,000 cases). 10-fold cross-validation method was used to measure the unbiased estimate of the three prediction models for performance comparison purposes. The results indicated that the decision tree (C5) is the best predictor with 93.6% accuracy on the holdout sample (this prediction accuracy is better than any reported in the literature), artificial neural networks came out to be the second with 91.2% accuracy and the logistic regression models came out to be the worst of the three with 89.2% accuracy. The comparative study of multiple prediction models for breast cancer survivability using a large dataset along with a 10-fold cross-validation provided us with an insight into the relative prediction ability of different data mining methods. Using sensitivity analysis on neural network models provided us with the prioritized importance of the prognostic factors used in the study.

Lundin et al. [12] used ANN and logistic regression models to predict 5, 10, and 15- year breast cancer survival. They studied 951 breast cancer patients and used tumor size, axillary nodal status, histological type, mitotic count, nuclear pleomorphism, tubule formation, B.M.Gayathri et al. [22] explained due to the change of life styles of women and avoid breast. Predicting this disease manually is very difficult and it is a time-consuming processing. To detect breast cancer

disease machine learning concepts are used. In this research work performed a comparative study of Relevance vector machine (RVM) with some other machine learning concepts are used in for breast cancer prediction.

Dana Bazazeh et al. [23] says that breast cancer most wide spread fatal disease among women throughout the world. Machine Learning approaches are used to diagnosis the breast cancer in early stage. In this research work the authors compared most commonly used machine leaning concepts are Support Vector Machine concept, Random Forest technique and Bayesian Networks approach Wisconsin original breast cancer data set was used to implemented machine learning concepts.

Zhiqiong Wang et al. [24] used the convolutional neural network (CNN) deep features to detect cancer disease. Initial step the authors used mass detection method based upon CNN deep learning features and unsupervised machine learning clustering concept. Then build a various feature set likes morphological features, texture features, and density features. Finally, ELM classifier was designed to classify benign and malignant breast tumors method

Yawen Xiao et al. [25] says that breast cancer disease is common disease in female category of the people. In this research work demonstrated a new system embedded with deep learning concept based unsupervised feature extraction algorithm. The stacked auto-encoder concept was also used with a support vector machine technique 30

to predict breast cancer. The proposed method was tested by using Wisconsin Diagnostic Breast Cancer data set. The result displays that SAE-SVM method used to increase accuracy level to 98.25%

Junaid Ahmad Bhat et al. [26] developed a new tool used to detect the breast cancer disease in early stage. In this research work the authors was presented preliminary results of the project BCDM developed by using Matlab software. The algorithm was implemented using adaptive resonance approach. P. Malathi et al. [27] proposed a research on computer aided detection system for women breast cancer diagnosis from digital mammographic images. The research about work conveyed to create PC helped conclusion instruments that can help the radiologists in mak- ing precise understanding of the computerized mammograms. The methodologies which applied to structure various phases of CAD framework are abridge

2.5 Survey of Previous Works After research, a great number of articles and publications on breast cancer prediction using deep learning and machine learning were investigated to find out the accuracy and repetition of selected algorithms with best performance. Total number of papers we have found using keyword search were 43,900 that we have got from different platforms like ACM, IEEE, Research Gate and Science Direct. Our search query was focused on four keywords: machine learning, deep learning, data mining and breast cancer prognosis.

2.6 Partial Conclusion

Investigation of the relevant literature helps in sorting out different deep learning and ma- chine learning techniques that could be exploited in detecting breast cancer. After reviewing all techniques, their performance, their accuracy, the number of times they appeared in a journal, the optimum machine learning techniques that I selected for breast cancer detection was artificial neural network, more precisely, Convolutional Neural Networks (CNN) because it was used in more references than any other algorithm. For my proposed methodology, the CNN architecture and functioning which shall be further explained in Chapter 3.

MATERIALS AND METHODS

3.1 Introduction

This chapter carefully explains the various methods and processes taken to realize the project. It reveals the model used, the training of the model in the software, and several other parameters.

3.2 Project Methodology

The summary of the project methodology is explained in Figure 3.1. This project aims to assess whether a lump in a breast could be malignant (cancerous) or benign (non-cancerous).

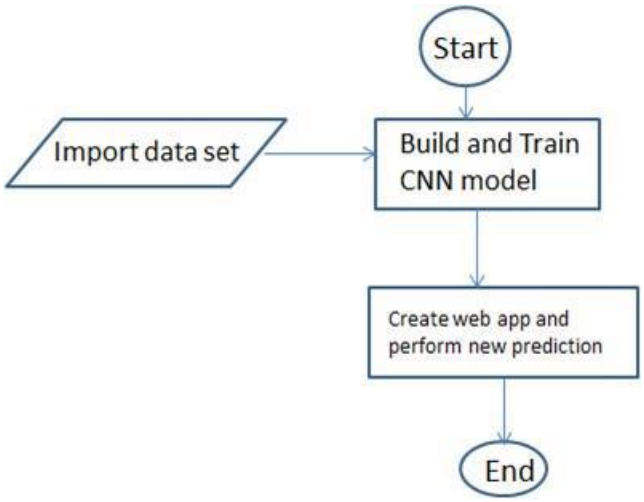


Figure 3. 1: Project Methodology Flowchart

For that, we use digitized histopathology images of fine-needle aspiration (FNA) biopsy using machine learning. First, the CNN model is built and trained in colab by importing the chosen data set to it. Then, once a high accuracy achieved, a web app is created in the front end to allow a new prediction to be made for any patient image data. Google Colab was chosen preferred to kaggle because it is very simple.

The Data Set

The data set for this project can be downloaded at kaggle.com/uciml/breast-cancer wisconsin- data. Dr. William H. Wolberg, from the University of Wisconsin Hospitals, Madison, obtained this breast cancer database [28]. Figure 3.2 shows the first five rows and columns of the data set. In this data set there are 30 input parameters more than 600 patient cases used. Target variables can only have two values

in a classification model: 0 (false) or 1 (true). Since this dataset doesn‟t contain image data, another dataset containing histopathological FNA biopsy images was also used to classify the instances into either benign or malignant, from the site kaggle.com/datasets/paultimothymooney/breasthistopathology-images. About 4000 images were used for the training.

Figure 3.

	mean radius	mean texture	mean perimeter	mean area	mean smoothness
472	14.92	14.93	96.45	686.9	0.08098
134	18.45	21.91	120.20	1075.0	0.09430
259	15.53	33.56	103.70	744.9	0.10630
32	17.02	23.98	112.80	899.3	0.11970
326	14.11	12.88	90.03	616.5	0.09309

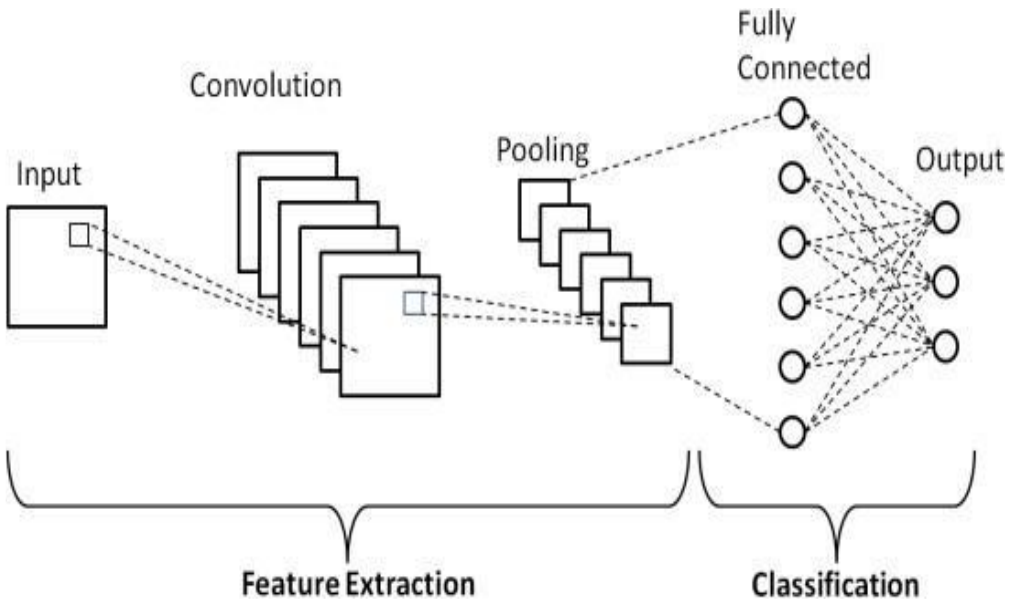
5 rows × 30 columns

3.2.2

Convolutional Neural Network Architecture The Convolutional Neural Network was built in Google Collaboratory, which is a free environment that provides users with free Graphics Processing Unit (GPU) and Tensor Processing Unit (TPU) runtimes for training their machine learning algorithms. Once in the Google Colab environment, a user will simply have to login to a Gmail account and create a new notebook in order to create a new neural network. CNN was chosen over other ANN algorithms because since digital images are a bunch of pixels with high values, it makes sense to use CNN to analyze them. CNN decreases their values, which is better for the training phase with less computational

power and less information loss The main reason why ReLu (Rectifying Linear Unit) is used in preference to other activation layer algorithms is because it is simple, fast, and empirically it seems to work well. Empirically, early papers observed that training a deep network with ReLu tended to converge much more quickly and reliably than training a deep network with sigmoid activation. CNNs are comprised of three types of layers. These are convolutional layers, pooling layers and fully-connected layers. When these layers are stacked, a CNN architecture has been formed [29]. Figure 3.2 shows the architecture. The CNN model takes as input the sequence of word embeddings, summarizes the sentence meaning by convolving the sliding window and pooling the saliency through the sentence, and yields the fixed-length distributed vector with other layers, such as dropout and

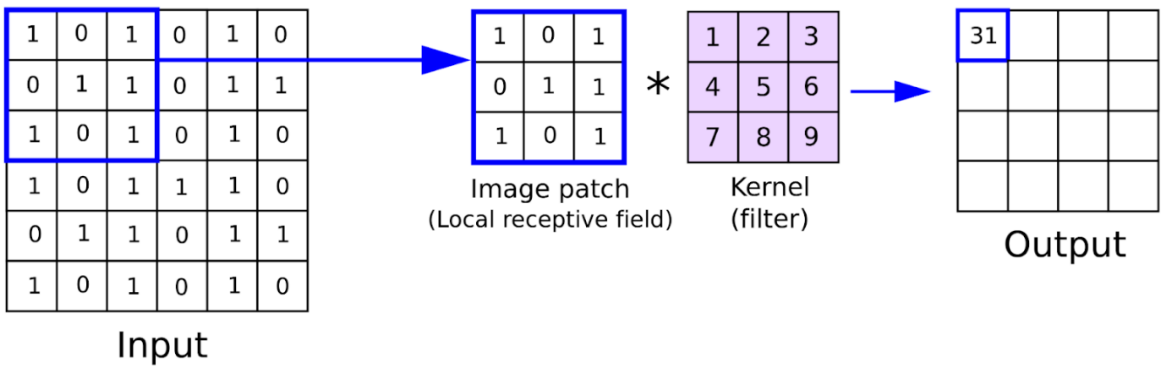
Fully concluded layers.

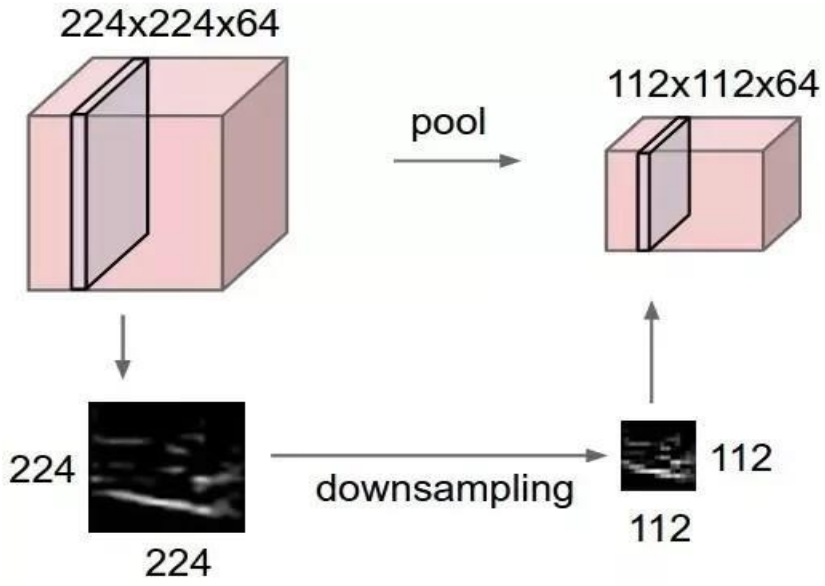


The CNN model was trained in Colab using Tensor Flow and Keras modules. The classification was done separating the data set into a training set and validation set. The training was set for ten epochs for the training and validation data sets. A resulting graph for the training and validation processing was made to highlight the accuracy and loss for both sets, which will be discussed in the next chapter.

Convolution Layer: Figure 3.4 shows the convolution operation. This is the first layer of the convolutional network that performs feature extraction by sliding the filter over the input image. The output or the convolved feature is the element-wise product of filters in the image and their sum for every sliding action. The output layer, also known as the feature map, corresponds to original images like curves, sharp edges, textures, etc.

In the case of networks with more convolutional layers, the initial layers are meant for extracting the generic features while the complex parts are removed as the network gets deeper.



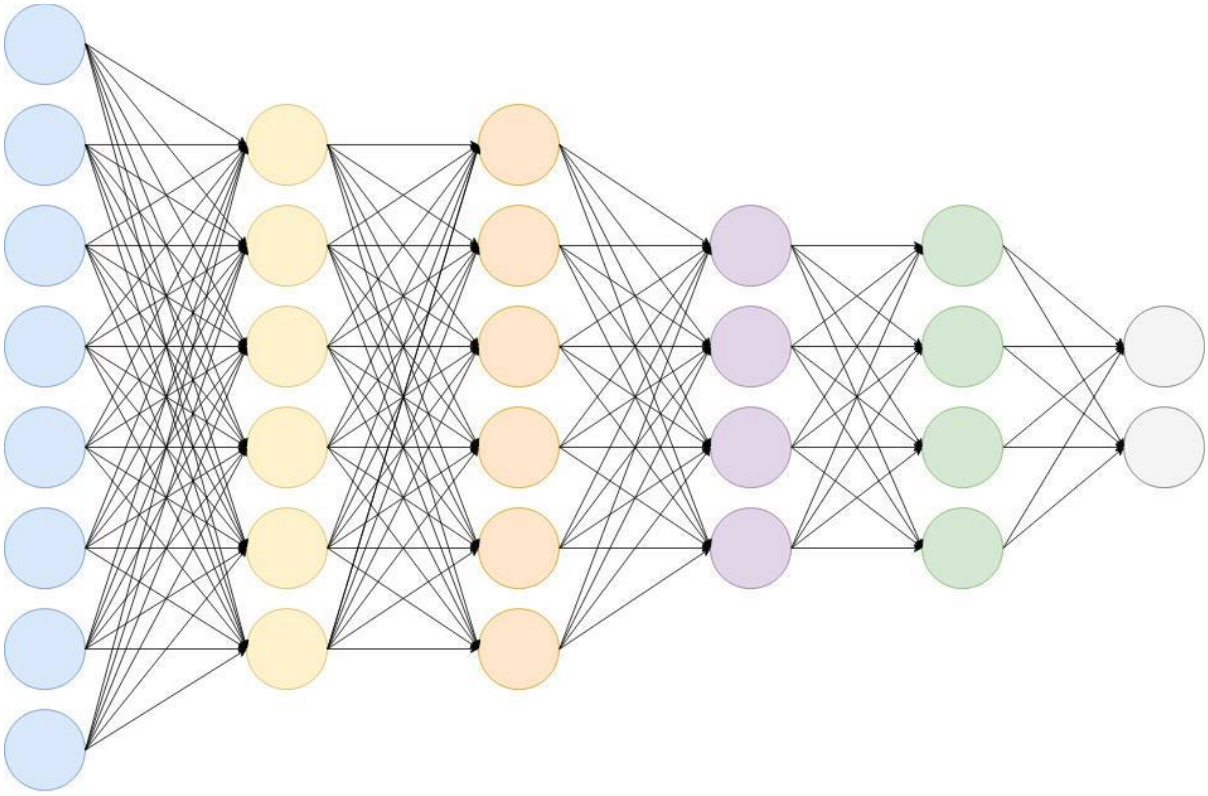


Pooling Layer Scheme

Pooling Layer: Figure 3.5 shows the functioning of the pooling layer. The primary purpose of this layer is to reduce the number of trainable parameters by decreasing the spatial size of the image, thereby reducing the computational cost. The image depth remains un- changed since pooling is done independently on each depth dimension. Max Pooling is the most common pooling method, where the most significant element is taken as input from the feature map. Max Pooling is then performed to give the output image with dimensions reduced to a great extent while retaining the essential information.

Fully Connected Layer: Figure 3.6 shows the functioning of the fully connected layer. The last few layers which determine the output are the fully connected layers. The output from the pooling layer is Flattened into a one-dimensional vector and then given as input to the fully connected layer. The

output layer has the same number of neurons as the number of categories we had in our problem for classification, thus associating features to a particular label.

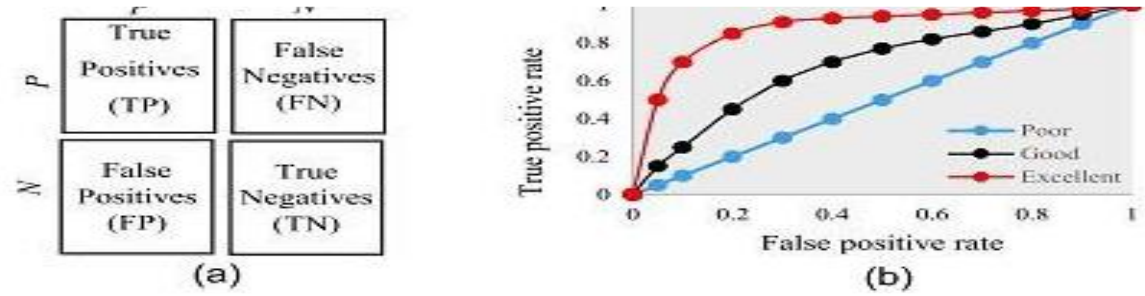


Fully Connected Layer scheme

After this process is known as forwarding propagation, the output so generated is compared to the actual production for error generation. The error is then back propagated to update the filters(weights) and bias values. Thus, one training is completed after this forwarding and backward propagation cycle.

Classifier Performance Index

The diagnostic ability of classifiers has usually been determined by the confusion matrix and the Receiver Operating Characteristic (ROC) curve. In the machine learning research domain, the confusion matrix is also known as error or contingency matrix. The basic framework of the confusion matrix has been provided in Figure 3.7 In this framework, true positives (TP) are the positive cases where the classifier correctly identified them. Similarly, true negatives (TN) are the negative cases where the classifier correctly identified them. False positives (FP) are the negative cases where the classifier incorrectly identified them as positive and the false negatives (FN) are the positive cases where the classifier incorrectly identified them as negative. The following measures, which are based on the confusion matrix, are commonly used to analyze the performance of classifiers, including those that are based on supervised machine learning algorithms. The acceptable ranges for the accuracy, precision, F1 score, sensitivity and specificity are from 90% to 100%, while the acceptable range for the false positive rate is from 0% to 10%.



a)Basic framework of confusion matrix b) ROC curve

Accuracy=TP+TN/TP+TN+FP+FN

Precision= TP/TP+FP

Sensitivity=true positive rate=TP/TP+FN

False positive rate=FP/FP+TN

ROC is one of the fundamental tools for diagnostic test evaluation and is created by plotting the true positive rate against the false positive rate at various threshold settings. The area under the ROC curve (AUC) is also commonly used to determine the predictability of a classifier. A higher AUC value represents the superiority of a classifier and vice versa. illustrates a presentation of three ROC curves based on an abstract dataset. The area under the blue ROC curve is half of the shaded rectangle. Thus, the AUC value for this blue ROC curve is 0.5. Due to the coverage of a larger area, the AUC value for the red ROC curve is higher than that of the black ROC curve. Hence, the classifier that produced the red ROC curve shows higher predictive accuracy compared with the other two classifiers that generated the blue and red ROC curves.

RESULTS AND DISCUSSIONS

4.1 Introduction

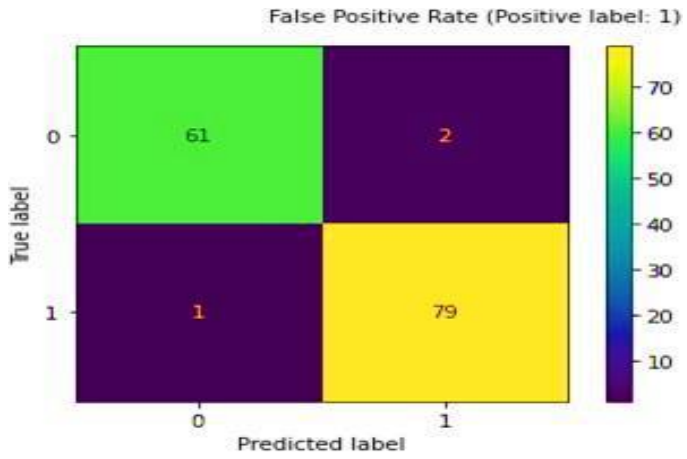
This chapter presents the results of training the CNN model, giving information about the ROC curve, the confusion matrix, the accuracy and other KPIs, and also, the application deployed.

4.2 Results from Google Colab

From the coding of the model which is found in Appendix A, Figure 4.1 shows the result of the ROC curve gotten after training the model. The output was a 0.98 AUC, indicating a strong classification.

Figure 4. 1: ROC curve performance

The next result obtained was the confusion matrix after training the dataset, which gave all the resulting KPI shown. Figure 4.2 shows the result of the confusion matrix, while Figure 4.3 shows a graph of the training and validation accuracy versus the training and validation



Confusion matrix performance

Accuracy= TP+TN/TP + TN+ FP + FN=140/143

F1 Score=2TP/2TP +FP +FN=122/125

Precision= TP/TP + FP=61/62

Sensitivity = True position rate =TP/TP + FN=61/63

False positive rate = FP/FP + TN=0.0125

From the results gotten, the accuracy of the prediction is 97.9%, implying a very good model trained. So here is a summary of the KPIs of the training.

Classification accuracy (ratio of instances correctly classified): 97.9%.

• **Error rate** (ratio of instances misclassified): 1.25%.

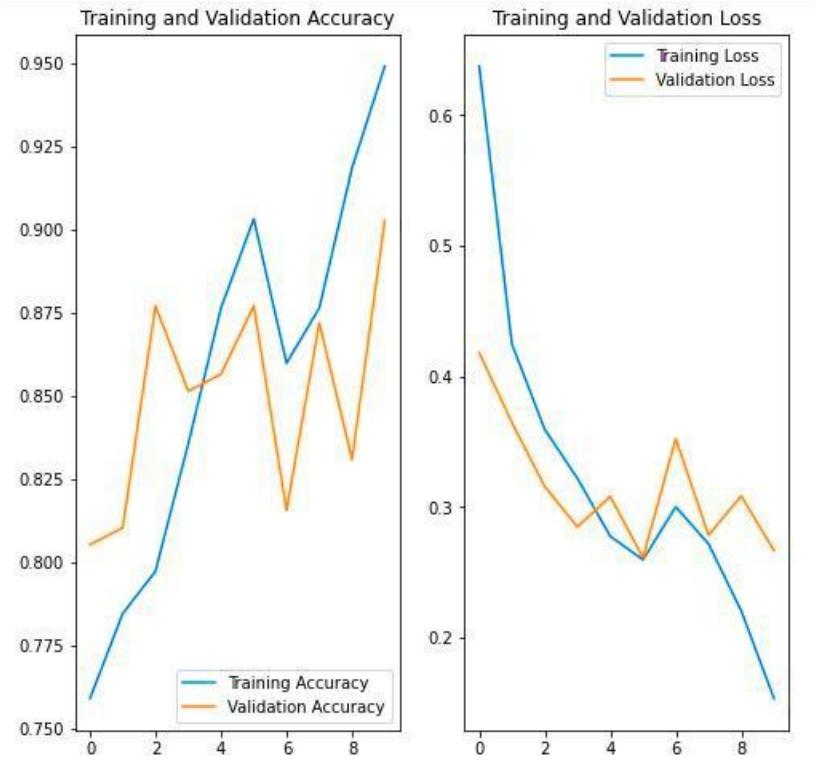
• **Specificity** (ratio of real negative which are predicted negative): 98.75%.

• **Sensitivity** (ratio of real positive which are predicted positive): 96.8%.

• **Precision** (ratio of real positive which are predicted positive): 98.4%.

A plot of the training and validation accuracies of the training and validation data sets were also gotten from the tensor flow prediction, together with a plot of their respective

losses, as a function of the epochs history. From Figure 4.3, the training and validation accuracy increases with an increase in the number of epochs, while the training and validation loss decreases with an increase in the number of epochs, indicating that the system becomes more accurate as more training time is given to it.



GENERAL CONCLUSION

5.1 Summary of Findings

In this dissertation, we proposed a simple and effective method for the classification of histopathology breast cancer images in case of a large training data. Following the training of the artificial neural network with the breast cancer dataset, we had the following results

- **Classification accuracy** (ratio of instances correctly classified): 97.9%.
- **Error rate** (ratio of instances misclassified): 1.25%.
- **Specificity** (ratio of real negative which are predicted negative): 98.75%.
- **Sensitivity** (ratio of real positive which are predicted positive): 96.8%.
- **Precision** (ratio of real positive which are predicted positive): 98.4%.

From the above values, we realize the training and classification of the model was accurately done, with ease and simplicity. Approximately 4,000 images were used for training, with 20% of this number used for validation. Therefore, the feature extraction and classification were done with more precision.

5.2 Implications to Existing Knowledge

The realization of this project means that there exist various ways through which breast cancer and other diseases could be easily detected using AI. Cancer detection in medical imaging is a field that can achieve many good results with deep learning technology. Re- viewed papers are summarized in Table 4.1. So far, the results very satisfactory, but the development of deep learning technology is very fast, the supply of researchable medical image data is getting bigger, and the research funds are getting

rich, so the future is bright. In the future, it will be easier and more accurate to diagnose not only medical images but also EHR and genetic information with the help of deep learning technology. The devel- opment of deep learning technologies is important for this, but the role of physicians who understand and use these technologies becomes increasingly important. [37]

5.3 Recommendations

This project can be widely used in the field of medicine, whereby diseases like breast cancer, heart disease, and other disease can be easily diagnosed for the good of everyone. This is a great contribution to Engineering and Technology, as the use of AI has been successfully applied to the good of the society. This study will be valuable to medical area as it will allow fast diagnostic of breast cancer even in areas without specialist. Moreover, it could be of high interest to patients in case they are to confirm their diagnosis.

5.4 Future Scope

This study had some limitations. The histopathology images were downsized to fit the available GPU. As more GPU memory becomes available, future studies will be able to train models using larger image sizes, or retain the original image resolution without the need for downsizing. Retaining the full resolution of the images will provide finer details of the KPIs and likely improve performance.

For future work, we intend to use and evaluate other CNN pretrained models for the features extraction stage, and extend the application usability to other types of cancer, such as colorectal, lung or prostate cancer.

Most papers published in the field of breast cancer detection and subtype classification 55

use machine learning techniques. However, deep learning models have not been heavily investigated in this domain. A thought for the future would be to present researchers with opportunities to use various deep learning mechanisms to predict patient status such as LSTM, GAN and RNN, as these types of research have not yet been conducted in the field.

We also intend to develop a mobile app for this solution in order to maximize the utility of the Project.

REFERENCES

[1] F. Noreen, L. Liu, H. Sha, and H. Ahmed, “Prediction of breast cancer, comparative review of machine learning techniques, and their analysis,” *IEEE Access*, vol. PP, pp. 1–1, 08 2020.

[2] A.cancer Society, “Breast cancer early detection and diagnosis.” https://www.cancer.org/cancer/breast-cancer/ screening-tests-and-early-detection/breast-mri-scans. html, 2022. Accessed: 2022-07-27.

[3] A. Victor,“10 uses of artificial intelligence in day to day life.” https://insights.daffodilsw.com/blog/ 10-uses-of-artificial-intelligence-in-day-to-day-life, 2021. Accessed: 2022-07-26.

[4] A. Dasgupta and A. Nath, “Classification of machine learning algorithms,” *Inter- national Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2763*, vol. 3, pp. 6–11, 03 2016.

[5] S. Uddin, A. Khan, M. Hossain, and M. A. Moni, “Comparing different supervised machine learning algorithms for disease prediction,” *BMC Medical Informatics and Decision Making*, vol. 19, 12 2019.

[6] A. Biswal, “Top 10 deep learning algorithms you should know in 2023.” https:

//www.Top10DeepLearningAlgorithmsYouShouldKnowin2022, 2022. Accessed: 2022-07-23.

[7] D. Dahiwade, G. Patle, and E. Meshram, “Designing disease prediction model using machine learning approach,” pp. 1211–1215, 03 2019.

[8] H. Chen, “An efficient diagnosis system for detection of parkinson‟s disease using fuzzy k-