Project Report

On

Quality of Life Prediction of Breat Cancer Patients Survival using XAI

Submitted to D Y Patil International University, Akurdi, Pune
in partial fulfilment of full-time degree

Master of Computer Applications

Submitted By:

Khushboo Gupta(20220804002)

Akanksha B. Kumbhar(20220804038)

Sharda S.Kamble(20220804047)

Under the Guidance of

Mr.Vaishnaw Kale

School of Computer Science, Engineering and Applications

D Y Patil International University, Akurdi,Pune, INDIA, 411044

[Session 2022-23]

This report on **(Quality of Life Prediction of Breat Cancer Patients Survival using XAI)** is submitted for the partial fulfillment of project,which is part of the Second Year Master of Computer Applications curriculum, under my supervision and guidance.

Mr.Vaishnaw Kale
(DYPIU Guide)

Mrs. Vaishali Kumar
(Project Coordinator)

Dr.Maheshwari Biradar
(HOD BCA & MCA)

Dr. Bahubali Shiragapur
Director
School of Computer Science Engineering & Applications
D Y Patil International University, Akurdi
Pune, 411044, Maharashtra, INDIA

DECLARATION

I, hereby declare that the following Project which is being presented in the Project entitled as **Quality of Life Prediction of Breat Cancer Patients Survival using XAI** is an authentic documentation of my own original work to the best of my knowledge. The following Project and its report in part or whole, has not been presented or submitted by me for any purpose in any other institute or organization. Any contribution made to my work, with whom i have worked at D Y Patil International University, Akurdi, Pune, is explicitly acknowledged in the report.

Khushboo Gupta(20220804002) Signature :
Akanksha B. Kumbhar(20220804038) Signature :
Sharda S. Kamble(20220804047) Signature :

ACKNOWLEDGEMENT

With due respect, we express our deep sense of gratitude to our respected guide Mr.Vaishnaw Kale, for his valuable help and guidance. We are thankful for the encouragement that he has given us in completing this Project successfully.

It is imperative for us to mention the fact that the report of project could not have been accomplished without the periodic suggestions and advice of our project supervisor Mrs.Vaishali Kumar.

We are also grateful to our respected, Dr. Bahubali Shiragapur(Director), Dr. Maheshwari Biradar (HOD BCA & MCA) and Hon'ble Vice Chancellor, DYPIU, Akurdi, Prof. Prabhat Ranjan for permitting us to utilize all the necessary facilities of the college.

We are also thankful to all the other faculty, staff members and laboratory attendants of our department for their kind cooperation and help. Last but certainly not the least; we would like to express our deep appreciation towards our family members and batch mates for providing support and encouragement.

Khushboo Gupta(20220804002)
Akanksha B. Kumbhar(20220804038)
Sharda S.Kamble(20220804047)

Abstract

Breast cancer remains a critical global health concern, necessitating advancements in predictive modeling to enhance patient care and outcomes. This project focuses on employing state-of-the-art machine learning techniques, specifically Explainable Artificial Intelligence (XAI), to predict the quality of life and survival probabilities for breast cancer patients. The integration of XAI aims to provide transparent and interpretable insights into the model's decision-making process, ensuring that healthcare professionals and patients can comprehend and trust the predictions.

The project leverages a comprehensive dataset, encompassing diverse clinical, genomic, and demographic features of breast cancer patients. Machine learning algorithms, including but not limited to deep neural networks, will be employed to extract patterns and relationships from the data. The use of XAI techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) will facilitate the understanding of the model's predictions at both the individual and population levels.

The anticipated outcomes of this project include a robust predictive model capable of assessing the quality of life and survival probabilities of breast cancer patients. Additionally, the incorporation of XAI will empower healthcare professionals to interpret the model's decisions, fostering trust and facilitating personalized patient care strategies. Ultimately, this research contributes to the ongoing efforts to enhance breast cancer treatment and patient outcomes, thereby improving the overall quality of life for individuals affected by this disease.

TABLE OF CONTENTS

DECLARATION i

ACKNOWLEDGEMENT ii

ABSTRACT iii

LIST OF FIGURES vi

LIST OF TABLES vii

1 INTRODUCTION 1

1.1 Introduction/Background 1
1.2 Problem Statement 1
1.3 Objectives 1

2 Literature Survey 3

2.1 Information about Breast cancer 3
2.2 Statistical Analysis of Breast Cancer in India 4
2.3 Treatment for the Breast Cancer 6
2.4 Quality Life of Breast Cancer Patients 7
2.5 AI , ML and DL for Breast Cancer 8
2.5.1 AI , ML And DL algorithms used for breast cancer survival prediction . 10
2.6 Gap Analysis 11

3 METHODOLOGY 14

3.1 Dataset with Explaination 14
3.2 Block Diagram with Explaination 17
3.3 Flow Chart with Explaination 18
3.4 UML Diagrams 19
3.4.1 Activity Diagram 20
3.4.2 Sequence Diagram 20
3.4.3 Use Case Diagram 21
3.5 Performance Parameters with formulas 21

4 PERFORMANCE ANALYSIS 23

4.1 Result Oriented Tables 23
4.2 Comparative Analysis 32
4.3 Graphs and Charts 33
4.3.1 Graph for Distribution of Age 33
4.3.2 Graph for Distribution of Patients's Vital Status 33

4.3.3 Pie Chart for Cancer Type 34
4.3.4 Graph for Overall Survival Status 34
4.3.5 Correlation between different features 35
4.3.6 Distribution of Actual and Predicted Output 35
4.4 Result with Discussion 36

5 CONCLUSION **37**
5.1 Conclusion 37
5.2 Advantages 37
5.3 Applicability 37
5.4 Scope 38

REFERENCES **39**

List of Figures

1 Breast Cancer in India 5
2 Quality of Life in breast cancer survivors 8
3 Block Diagram 17
4 Flow Chart 19
Sequence Diagram 20
5 Sequence Diagram 20
6 Use Case Diagram 21
Training Confusion Matrix 23
Testing Confusion Matrix 24
7 Explainability of Random Forest Classifier 25
8 Explainability of Random Survival Forest 27
9 Explainability of Just Neural Network 29
10 Explainability of Coherant Voting Network 31
11 Distribution of Age 33
12 Distribution of Patients's Vital Status 33
13 Pie Chart for Cancer Type 34
14 Overall Survival Status 34
15 Correlation between different features 35
16 Distribution of Actual and Predicted Output 35
17 Actual Output 36

List of Tables

2.1 Gap Analysis Table 11
3.1 Clinical attributes in the dataset 14
4.1 Training Classification Report for Random Forest Classifier 23
4.2 Testing Classification Report for Random Forest Classifier 24
4.3 Testing Classification Report for JNN 28
4.5 Testing Classification Report for CVN 30
4.7 Classification Report for (XGBoot) 32
4.8 Comparative analysis 32

1. INTRODUCTION

1.1. Introduction/Background

Breast cancer is a significant global health concern, impacting millions of women and their families each year. Early detection and accurate prognosis are crucial for improving the outcomes and quality of life for those diagnosed with breast cancer. In the pursuit of enhancing patient care and treatment strategies, the development of advanced predictive models for breast cancer prognosis has become a paramount research area. These predictive models not only help in estimating the survival and recurrence rates of breast cancer patients but also aid in tailoring individualized treatment plans, thereby increasing the chances of successful outcomes. The aim of this project is to predict breast cancer survival using machine learning models with clinical data and gene expression profiles. Using machine learning models on genetic data has the potential to improve our understanding of cancers and survival prediction .Leveraging the power of data analytics, machine learning, and medical knowledge, we will explore and harness the wealth of information available from clinical records, genetic data, and histopathological features to create a predictive tool that offers valuable insights into the future course of the disease for individual patients.

1.2. Problem Statement

Most of us know someone who struggled with breast cancer, or at least heard about the struggles facing patients who are fighting against breast cancer. Breast cancer is the most frequent cancer among women, impacting 2.1 million women each year. Breast cancer causes the greatest number of cancer-related deaths among women. In 2018 alone, it is estimated that 627,000 women died from breast cancer. The most important part of a process of clinical decision-making in patients with cancers in general is the accurate estimation of prognosis and survival duration. Breast cancer patients with the same stage of disease and the same clinical characteristics can have different treatment responses and overall survival, but why? Cancers are associated with genetic abnormalities. Gene expression measures the level of gene activity in a tissue and gives information about its complex activities. Comparing the genes expressed in normal and diseased tissue can bring better insights about the cancer prognosis and outcomes. Using machine learning techniques on genetic data has the potentials of giving the correct estimation of survival time and can prevent unnecessary surgical and treatment procedures.

1.3. Objectives

The purpose of the ”Life Prediction for Breast Cancer Patients” project is to develop a predictive model that can accurately estimate the life expectancy or survival time of individuals diagnosed with breast cancer. This project aims to provide healthcare professionals and patients with a

valuable tool to make informed decisions regarding treatment options, care planning, and to better understand the prognosis of breast cancer patients.

2. Literature Survey

2.1. Information about Breast cancer

Breast cancer is a disease in which abnormal breast cells grow out of control and form tumours. If left unchecked, the tumours can spread throughout the body and become fatal.
Breast cancer cells begin inside the milk ducts and/or the milk-producing lobules of the breast. The earliest form (in situ) is not life-threatening. Cancer cells can spread into nearby breast tissue (invasion). This creates tumours that cause lumps or thickening.
After skin cancer, breast cancer is the most common cancer diagnosed in women in the United States. Breast cancer can occur in both men and women, but it's far more common in women.

Types of Breast Cancer are as follows:

1. Angiosarcoma
2. Ductal carcinoma in situ (DCIS)
3. Inflammatory breast cancer
4. Invasive lobular carcinoma
5. Lobular carcinoma in situ (LCIS)
6. Male breast cancer
7. Paget's disease of the breast
8. Recurrent breast cancer

Signs and symptoms:

Breast cancer can have combinations of symptoms, especially when it is more advanced. Most people will not experience any symptoms when the cancer is still early.

Symptoms of breast cancer can include:

1. a breast lump or thickening, often without pain
2. change in size, shape or appearance of the breast
3. dimpling, redness, pitting or other changes in the skin
4. change in nipple appearance or the skin surrounding the nipple (areola)
5. abnormal or bloody fluid from the nipple.
6. People with an abnormal breast lump should seek medical care, even if the lump does not hurt.

Most breast lumps are not cancer. Breast lumps that are cancerous are more likely to be successfully treated when they are small and have not spread to nearby lymph nodes. Breast cancers may spread to other areas of the body and trigger other symptoms. Often, the most common first detectable site of spread is to the lymph nodes under the arm although it is possible to have cancer-bearing lymph nodes that cannot be felt. Over time, cancerous cells may spread to other organs including the lungs, liver, brain and bones. Once they reach these

sites, new cancer-related symptoms such as bone pain or headaches may appear.

2.2. Statistical Analysis of Breast Cancer in India

Percentage of Breast Cancer in India:

With being the most common type of cancer in women, breast cancer accounts for 14% of cancers in Indian women. It is reported that with every four minutes, an Indian woman is diagnosed with breast cancer. Breast cancer is on the rise, both in rural and urban India. A 2018 report of Breast Cancer statistics recorded 1,62,468 new registered cases and 87,090 reported deaths.

Cancer survival becomes more difficult in higher stages of its growth, and more than 50% of Indian women suffer from stage 3 and 4 of breast cancer. Post cancer survival for women with breast cancer was reported 60% for Indian women, as compared to 80% in the U.S.

Women can self-diagnose their condition and know of the presence of lumps or masses that suggest cancerous outgrowths. The very reason for a low breast cancer survival rate of women in India accounts from its lack of awareness and poor early screening and diagnosis rates.

Which State in India has the highest Cancer Rate?

With the latest study reports, India's highest cancer rate is listed in the state of Kerala. Other states with high cancer rates in India include Mizoram, Haryana, Delhi and Karnataka. Mizoram accounted for the highest cancer death rates in the country, followed by Kerala and Haryana.

As the most common cancer type in Indian women, women in their early thirties till fifties are at considerable risk to develop breast cancer, and the incidence risk increases till its peak by the time they reach 50-64 years of age. One in twenty-eight Indian women is likely to develop breast cancer during her lifetime. It is more (1 in 22) for urban women than the rural group (1 in 60). A report stated that cancer caused 5% of the total disability-adjusted life years(DALYs) in the Indian population in 2016. Palliative care has proved helpful for patients to alleviate them of the disability from cancer.

The Rise of Breast Cancer in India:

Breast cancer is the most common form of cancer in the country, having overtaken cervical cancer. In cities like Mumbai, Delhi, Bengaluru, Bhopal, Kolkata, Chennai, Ahmedabad, breast cancer accounts for 25% to 32% of all female cancers, more than 1/4th of all female cancers.

It's also more common in the younger age group. Almost 50% of all cases are in the age group of 25-50. And more than 70% of the cases present in the advanced stage had poor survival and high mortality.

The numbers are staggering and constantly rising. The Indian Council for Medical Research recently published a report which stated that in 2016 the total number of new cancer cases is expected to be about 14.5 lakhs. This figure will likely increase to 17.3 lakhs in 2020.

The survival rates of breast cancer in India are low because the detection takes place late. The only way to change these numbers is by increasing awareness. Breast cancer is a treatable disease and chances of survival are higher if it's detected in time. The only way to do so is by being aware of how it can be detected and early diagnosis can be done.
This includes leading a healthy lifestyle, being aware of family medical histories – so that if you know you're genetically inclined towards it, you can take preventive drugs or undergo preventive surgery. The simplest way to breast cancer prevention is by being able to do a self-breast examination. Women should be doing this on a regular basis after they turn 30.

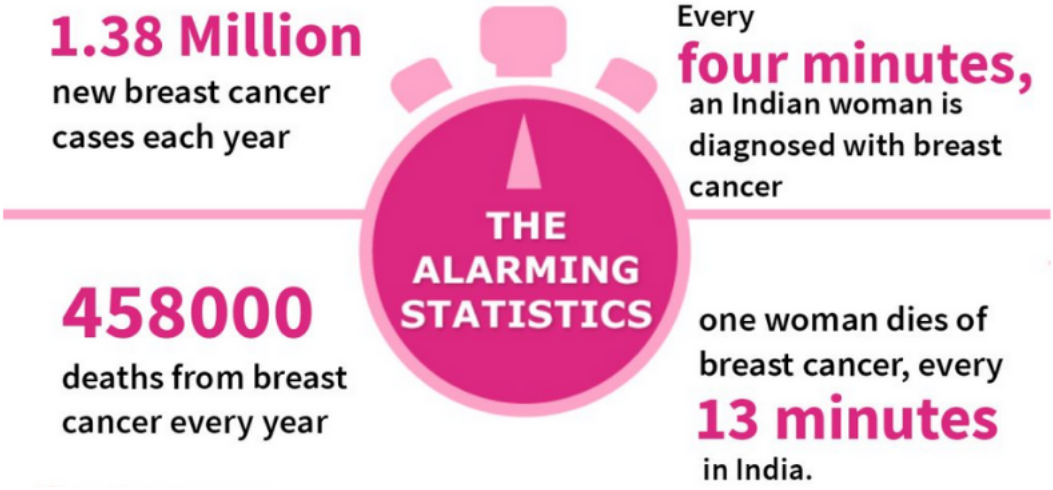

Fig. 1: Breast Cancer in India

Who is at risk?
Female gender is the strongest breast cancer risk factor. Approximately 0.5–1% of breast cancers occur in men. The treatment of breast cancer in men follows the same principles of management as for women.
Certain factors increase the risk of breast cancer including increasing age, obesity, harmful use of alcohol, family history of breast cancer, history of radiation exposure, reproductive history (such as age that menstrual periods began and age at first pregnancy), tobacco use and postmenopausal hormone therapy. Approximately half of breast cancers develop in women who have no identifiable breast cancer risk factor other than gender (female) and age (over 40 years).
Family history of breast cancer increases the risk of breast cancer, but most women diagnosed with breast cancer do not have a known family history of the disease. Lack of a known family history does not necessarily mean that a woman is at reduced risk.
Certain inherited high penetrance gene mutations greatly increase breast cancer risk, the most dominant being mutations in the genes BRCA1, BRCA2 and PALB-2. Women found to have

mutations in these major genes may consider risk reduction strategies such as surgical removal of both breasts.

Global impact

Age-standardized breast cancer mortality in high-income countries dropped by 40% between the 1980s and 2020. Countries that have succeeded in reducing breast cancer mortality have been able to achieve an annual breast cancer mortality reduction of 2–4% per year.

The strategies for improving breast cancer outcomes depend on fundamental health system strengthening to deliver the treatments that are already known to work. These are also important for the management of other cancers and other non-malignant noncommunicable diseases (NCDs). For example, having reliable referral pathways from primary care facilities to district hospitals to dedicated cancer centres.

The establishment of reliable referral pathways from primary care facilities to district hospitals to dedicated cancer centres is the same approach as is required for the management of cervical cancer, lung cancer, colorectal cancer and prostate cancer. To that end, breast cancer is an "index" disease whereby pathways are created that can be followed for the management of other diseases.

2.3. Treatment for the Breast Cancer

Treatment for breast cancer depends on the subtype of cancer and how much it has spread outside of the breast to lymph nodes (stages II or III) or to other parts of the body (stage IV).

Doctors combine treatments to minimize the chances of the cancer coming back (recurrence). These include:

1.surgery to remove the breast tumour

2.radiation therapy to reduce recurrence risk in the breast and surrounding tissues

3.medications to kill cancer cells and prevent spread, including hormonal therapies, chemotherapy or targeted biological therapies.

Treatments for breast cancer are more effective and are better tolerated when started early and taken to completion. Surgery may remove just the cancerous tissue (called a lumpectomy) or the whole breast (mastectomy). Surgery may also remove lymph nodes to assess the cancer's ability to spread.

Radiation therapy treats residual microscopic cancers left behind in the breast tissue and/or lymph nodes and minimizes the chances of cancer recurring on the chest wall.

Advanced cancers can erode through the skin to cause open sores (ulceration) but are not necessarily painful. Women with breast wounds that do not heal should seek medical care to have a biopsy performed.

Medicines to treat breast cancers are selected based on the biological properties of the cancer as determined by special tests (tumour marker determination). The great majority of drugs

used for breast cancer are already on the WHO Essential Medicines List (EML).
Lymph nodes are removed at the time of cancer surgery for invasive cancers. Complete removal of the lymph node bed under the arm (complete axillary dissection) in the past was thought to be necessary to prevent the spread of cancer. A smaller lymph node procedure called "sentinel node biopsy" is now preferred as it has fewer complications.
Medical treatments for breast cancers, which may be given before ("neoadjuvant") or after ("adjuvant") surgery, is based on the biological subtyping of the cancers. Cancer that express the estrogen receptor (ER) and/or progesterone receptor (PR) are likely to respond to endocrine (hormone) therapies such as tamoxifen or aromatase inhibitors. These medicines are taken orally for 5–10 years and reduce the chance of recurrence of these "hormone-positive" cancers by nearly half. Endocrine therapies can cause symptoms of menopause but are generally well tolerated.
Cancers that do not express ER or PR are "hormone receptor negative" and need to be treated with chemotherapy unless the cancer is very small. The chemotherapy regimens available today are very effective in reducing the chances of cancer spread or recurrence and are generally given as outpatient therapy. Chemotherapy for breast cancer generally does not require hospital admission in the absence of complications.

Breast cancers may independently overexpress a molecule called the HER-2/neu oncogene. These "HER-2 positive" cancers are amenable to treatment with targeted biological agents such as trastuzumab. These biological agents are very effective but also very expensive, because they are antibodies rather than chemicals. When targeted biological therapies are given, they are combined with chemotherapy to make them effective at killing cancer cells.
Radiotherapy plays a very important role in treating breast cancer. With early-stage breast cancers, radiation can prevent a woman having to undergo a mastectomy. With later stage cancers, radiotherapy can reduce cancer recurrence risk even when a mastectomy has been performed. For advanced stage of breast cancer, in some circumstances, radiation therapy may reduce the likelihood of dying of the disease.

The effectiveness of breast cancer therapies depends on the full course of treatment. Partial treatment is less likely to lead to a positive outcome.

2.4. Quality Life of Breast Cancer Patients

Breast Cancer survival rate in India has a 66%.Last knowledge update in January 2022, the survival rates for breast cancer in India vary based on factors such as the stage at diagnosis, access to healthcare, and treatment options. It's important to note that cancer statistics can change over time due to advancements in treatment, changes in healthcare infrastructure, and other factors.

As a reference, the National Cancer Registry Program (NCRP) in India periodically releases cancer statistics, including survival rates. However, the most recent data may not be available, and I recommend checking with authoritative sources such as the Indian Council of Medical Research (ICMR), the World Health Organization (WHO), or reputable cancer research organizations for the latest and most accurate information.

Additionally, specific survival rates can vary for different stages of breast cancer (Stage I, II, III, IV) and other factors like age and hormone receptor status. Collaborative efforts between healthcare organizations, government agencies, and non-governmental organizations are crucial in addressing breast cancer and improving survival rates through early detection, accessible treatment, and supportive care.

The average 5-year survival rate for all people with breast cancer is 89%. The 10-year rate is 83%, and the 15-year rate is 78%. If the cancer is located only in the breast (Stage I), the 5-year survival rate is 99%. More than 70% of breast cancers are diagnosed at an Early Stage.

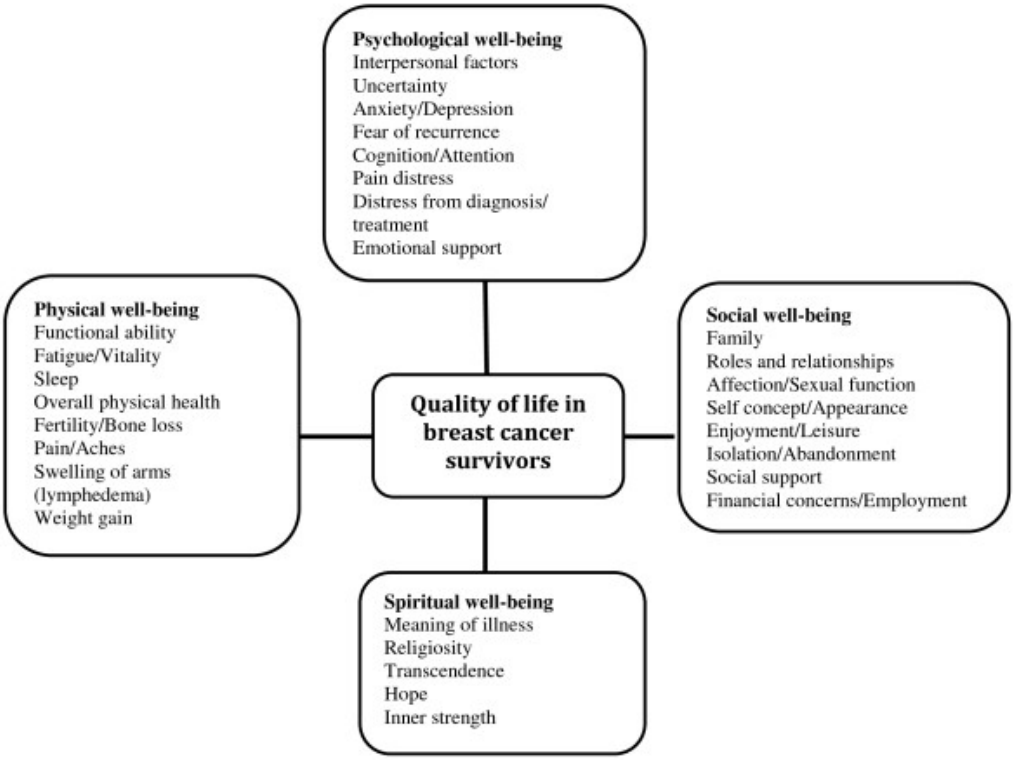

Fig. 2: Quality of Life in breast cancer survivors

2.5. AI , ML and DL for Breast Cancer

Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have shown promising potential in various applications related to breast cancer detection, diagnosis, and treatment. Here are some ways these technologies are being utilized in the context of breast cancer:

1. Early Detection and Diagnosis:

Mammogram Analysis: AI algorithms can analyze mammograms to identify potential signs of breast cancer. ML models can learn patterns and anomalies in mammogram images, aiding in early detection.

Computer-Aided Diagnosis (CAD): ML algorithms assist radiologists in interpreting medical images by highlighting suspicious areas, reducing false negatives, and improving overall diagnostic accuracy.

2.Predictive Analytics:

Risk Prediction Models: ML models can assess individual patient data, including genetic information, lifestyle factors, and family history, to predict the risk of developing breast cancer. This information can be valuable for personalized screening and prevention strategies.

3. Treatment Planning:

Personalized Medicine: ML and AI are used to analyze genomic data, helping in the identification of specific biomarkers and genetic mutations. This information enables the development of targeted therapies for individual patients, optimizing treatment effectiveness.

4. Pathology and Histology:

Digital Pathology: AI can analyze pathology slides and histological images, aiding pathologists in diagnosing breast cancer more accurately and efficiently. DL models can identify patterns and characteristics in tissue samples that may not be easily noticeable by human observers.

5.Prognosis and Outcome Prediction:

Survival Analysis: ML models can analyze patient data to predict the likely outcome of breast cancer treatment, helping clinicians make informed decisions about patient care and follow-up.

6. Research and Drug Discovery:

Data Mining: AI and ML algorithms can sift through vast amounts of scientific literature, clinical trial data, and patient records to identify potential targets for drug development and treatment innovation.

7. Patient Management:

Health Records Analysis: AI tools can analyze electronic health records to assist healthcare providers in managing patient care, ensuring timely follow-ups, and enhancing communication among the healthcare team.

It's important to note that while AI, ML, and DL technologies offer significant promise in improving breast cancer-related processes, they are not meant to replace healthcare professionals. Instead, they serve as powerful tools to assist medical experts in making more accurate and timely decisions. Ongoing research and collaborations between AI specialists,

healthcare professionals, and researchers are crucial for the continued advancement of these technologies in the field of breast cancer.

2.5.1. AI , ML And DL algorithms used for breast cancer survival prediction

Predicting breast cancer survival involves analyzing various factors to estimate the likelihood of a patient's survival over a certain period. Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) algorithms can play a crucial role in this task. Here are some algorithms commonly used for breast cancer survival prediction:

Logistic Regression:
Logistic regression is a simple and interpretable algorithm that can be used for binary outcomes, such as survival or non-survival.

Decision Trees:
Decision trees can be employed to create a model that represents decisions and their possible consequences. They are interpretable and can handle both categorical and numerical features.

Random Forest:
Random Forest like Random Forest Classifier and Random Survival Forest (RSF)is an ensemble learning method that combines multiple decision trees to improve accuracy and robustness. It can handle a large number of features and is less prone to overfitting.

Support Vector Machines (SVM):
SVM is a powerful classification algorithm that can be used for survival prediction. It works well in high-dimensional spaces and is effective in capturing complex relationships between features.

K-Nearest Neighbors (KNN):
KNN is a non-parametric algorithm that can be used for both classification and regression tasks. It calculates the survival probability based on the majority class of its k-nearest neighbors.

Naive Bayes:
Naive Bayes is a probabilistic algorithm that assumes independence between features. It is particularly useful when dealing with a large number of features.

Neural Networks (Deep Learning):
Deep Learning techniques, including neural networks like Just Neural Network(JNN) and Coherent Voting Network(CVN), can be applied to breast cancer survival prediction. Deep neural networks can automatically learn intricate patterns from data but may require a large amount of labeled data and computational resources.

Gradient Boosting Machines:
Gradient Boosting algorithms like XGBoost or LightGBM can be effective for survival prediction. They build a series of weak learners to create a strong predictive model.

Cox Proportional Hazards Model:
This is a survival analysis model commonly used in medical research. It considers the hazard function to model the survival time of patients based on various features.

Long Short-Term Memory (LSTM) Networks:
In the context of sequential data (time series data in medical records), LSTMs, which are a type of recurrent neural network (RNN), can be used to capture temporal dependencies and patterns in patient data.

2.6. Gap Analysis

Table 2.1: Gap Analysis Table

Author(s)	**Research paper**	**Methods used**	**Important Parameters**	**Gap Analysis**
Mohammadia et al.	[1]Diagnostic Accuracy of Machine Learning Models on Predicting Chemo-Brain in Breast Cancer Survivors Previously Treated with Chemotherapy:A Meta Analysis	Machine Learning	Sensitivity, Specificity,AUC	ML models demonstrate high accuracy in predicting chemo-brain, but the performance of different algorithms varies.
Khan et al.	[2] Machine Learning Based Comparative Analysis for Breast Cancer Prediction	Machine Learning	Random Forest, Logistic Regression, Decision Tree, K-Nearest Neighbor	Logistic regression offers the best results with 98 percent accuracy. Data preprocessing steps and feature selection need further optimization.

Abu Shawarib et al.	[3] Breast Cancer Diagnosis and Survival Prediction Using JNN	Artificial Neural Network (ANN)	ANN with JNN environment	Achieved 99.57 percent accuracy in breast cancer prediction, which is higher than previous methods.
Moncada-Torres et al.	[4] Explainable machine Machine earning can outperform Cox regression predictions and provide insights in breast cancer survival	Machine Learning (Random Survival Forests, Survival Support Vector Machines, Extreme Gradient Boosting)	Harrell's concordance index, SHAP values	ML-based models can perform at least as well as CPH regression, with XGB showing even better performance. ML models need to be interpretable and explainable for adoption in clinical settings.
Li et al.	[5] Predicting breast cancer 5-year survival using machine learning: A systematic review	Machine Learning	Various ML methods (Decision Trees, Artificial Neural Networks, Support Vector Machines, Ensemble Learning)	ML models do not consistently outperform traditional statistical methods for breast cancer survival prediction. More standardization and validation are required.
Nuutinen et al.	[6]Impact of Machine Learning Assistance on the Quality of Life Prediction for Breast Cancer Patients	Machine Learning, Clinical Decision Support System (CDSS)	AUROC, Average Accuracy (ACC)	Clinicians' performance in predicting patients' QoL improves with machine learning predictions, indicating the potential benefit of CDSS with ML in healthcare.

We conducted an in-depth literature review to understand existing research and the state of the art in Quality of Life Prediction of Breat Cancer Patients Survival using XAI. Many research paper had the similar gap of the integrating the techniques and methodologies in this process. By integrating the latest methodologies, we also overcomes the gap of less accuracy.

3. METHODOLOGY

3.1. Dataset with Explaination

The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset is a valuable resource in breast cancer research, particularly for studies involving molecular subtypes and survival prediction.

Link of Dataset:
`https://drive.google.com/drive/folders/1vABhM9a9d6J8H2-sNbArz3XdUotFQPrw?usp=drive_link`

Table 3.1: Clinical attributes in the dataset

Name	Type	Description
patient_id	object	Patient ID
age_at_diagnosis	float	Age of the patient at diagnosis time
type_of_breast_surgery	object	Breast cancer surgery type: 1- MASTECTOMY, which refers to a surgery to remove all breast tissue from a breast as a way to treat or prevent breast cancer.2-BREAST CONSERVING, which refers to a urgery where only the part of the breast that has cancer is removed
cancer_type	object	Breast cancer types: 1- Breast Cancer or 2- Breast Sarcoma
cancer_type_detailed	object	Detailed Breast cancer types: 1- Breast Invasive Ductal Carcinoma 2- Breast Mixed Ductal and Lobular Carcinoma 3- Breast Invasive Lobular Carcinoma 4- Breast Invasive Mixed Mucinous Carcinoma 5- Metaplastic Breast Cancer
cellularity	object	Cancer cellularity post chemotherapy, which refers to the amount of tumor cells in the specimen and their arrangement into clusters
chemotherapy	int	Whether or not the patient had chemotherapy as a treatment (yes/no)

pam50+_claudin-low_subtype	object	Pam 50: is a tumor profiling test that helps show whether some estrogen receptor-positive (ER-positive), HER2-negative breast cancers are likely to metastasize (when breast cancer spreads to other organs). The claudin-low breast cancer subtype is defined by gene expression characteristics, most prominently: Low expression of cell–cell adhesion genes, high expression of epithelial–mesenchymal transition (EMT) genes, and stem cell-like/less differentiated gene expression patterns
cohort	float	Cohort is a group of subjects who share a defining characteristic (It takes a value from 1 to 5)
er_status_measured_by_ihc	float	To assess if estrogen receptors are expressed on cancer cells by using immune-histochemistry (a dye used in pathology that targets specific antigen, if it is there, it will give a color, it is not there, the tissue on the slide will be colored) (positive/negative)
er_status	object	Cancer cells are positive or negative for estrogen receptors
neoplasm_histologic_grade	int	Determined by pathology by looking the nature of the cells, do they look aggressive or not (It takes a value from 1 to 3)
her2_status_measured_by_snp6	object	To assess if the cancer positive for HER2 or not by using advance molecular techniques (Type of next generation sequencing)
her2_status	object	Whether the cancer is positive or negative for HER2

tumor_other_histologic_subtype	object	Type of the cancer based on microscopic examination of the cancer tissue (It takes a value of 'Ductal/NST', 'Mixed', 'Lobular', 'Tubular/ cribriform', 'Mucinous', 'Medullary', 'Other', 'Metaplastic')
hormone_therapy	int	Whether or not the patient had hormonal as a treatment (yes/no)
inferred_menopausal_state	object	Whether the patient is is post menopausal or not (post/pre)
integrative_cluster	object	Molecular subtype of the cancer based on some gene expression (It takes a value from '4ER+', '3', '9', '7', '4ER-', '5', '8', '10', '1', '2', '6')
primary_tumor_laterality	object	Whether it is involving the right breast or the left breast
lymph_nodes_examined textunderscore positive	float	To take samples of the lymph node during the surgery and see if there were involved by the cancer
mutation_count	float	Number of gene that has relevant mutations
nottingham_prognostic_index	float	It is used to determine prognosis following surgery for breast cancer. Its value is calculated using three pathological criteria: the size of the tumour; the number of involved lymph nodes; and the grade of the tumour.
oncotree_code	object	The OncoTree is an open-source ontology that was developed at Memorial Sloan Kettering Cancer Center (MSK) for standardizing cancer type diagnosis from a clinical perspective by assigning each diagnosis a unique OncoTree code.
overall_survival_months	float	Duration from the time of the intervention to death
overall_survival	object	Target variable wether the patient is alive of dead.

pr_status	object	Cancer cells are positive or negative for progesterone receptors
radio_therapy	int	Whether or not the patient had radio as a treatment (yes/no)
3-gene_classifier_subtype	object	Three Gene classifier subtype It takes a value from 'ER-/HER2-', 'ER+/HER2- High Prolif', nan, 'ER+/HER2- Low Prolif','HER2+'
tumor_size	float	Tumor size measured by imaging techniques
tumor_stage	float	Stage of the cancer based on the involvement of surrounding structures, lymph nodes and distant spread
death_from_cancer	int	Wheather the patient's death was due to cancer or not (yes/no)

3.2. Block Diagram with Explaination

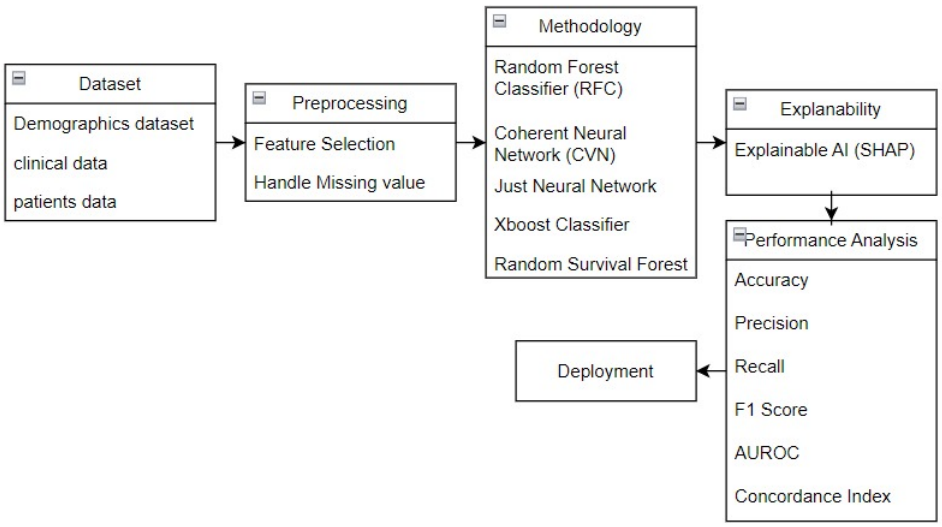

Fig. 3: Block Diagram

Here is a more detailed explanation of each step:

Dataset: The dataset is the most important part of any machine learning project. The quality and quantity of the data will have a direct impact on the performance of the model. The dataset should be representative of the real-world data that the model will be used on.

Preprocessing: Preprocessing is the process of cleaning and preparing the data for machine learning. This may involve removing outliers, handling missing values, and converting the data into a consistent format. Preprocessing is an important step because it can help to improve the performance and accuracy of the machine learning model.

Machine learning models: There are many different types of machine learning models, each with its own strengths and weaknesses. For this pipeline, we are using a neural network. Neural networks are a type of machine learning model that can be used to solve a wide range of problems, including classification, regression, and object detection.

Explanability: Explanability is the ability to understand how a machine learning model makes predictions. This is important for several reasons. First, it can help to identify any potential biases in the model. Second, it can help to improve the trust of users in the model. Finally, it can help to improve the performance

3.3. Flow Chart with Explaination

The steps are as follows:

Data collection: This step involves collecting data from relevant sources, such as databases, surveys, and experiments. The data should be representative of the problem that the model is being developed to solve.

Data preprocessing: This step involves cleaning and preparing the data for modeling. This may include tasks such as removing outliers, handling missing values, and converting data to a consistent format.

Feature selection: This step involves identifying the subset of features in the data that are most relevant to the target variable. This can help to improve the performance of the model and reduce overfitting.

Model building: This step involves selecting a machine learning algorithm and training a model on the prepared data. The algorithm should be appropriate for the type of data and the target variable.

Model training: This step involves feeding the prepared data to the selected algorithm and training the model to learn the relationships between the features and the target variable. **Model evaluation:** This step involves evaluating the performance of the trained model on a held-out test set. This helps to ensure that the model generalizes well to new data.

Explainability (XAI): This step involves interpreting the model to understand how it makes predictions. This can be helpful for identifying potential biases in the model and for building trust with users.

Model deployment: This step involves deploying the trained model to production so that it can be used to make predictions on new data.

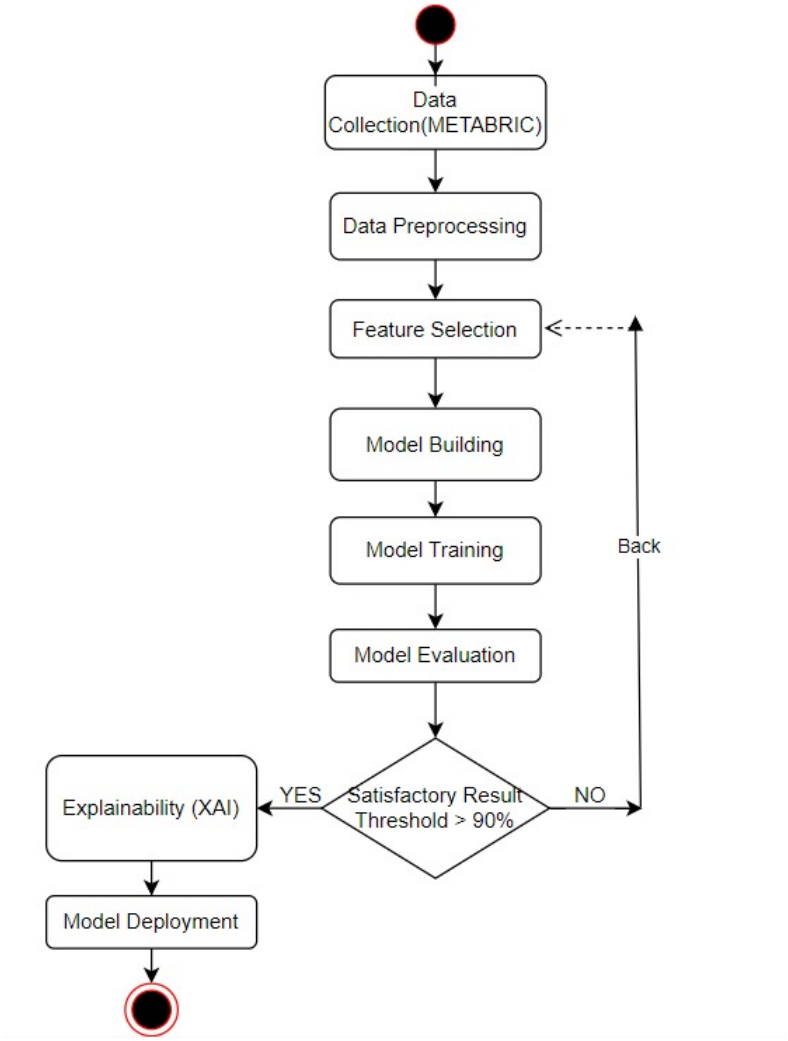

Fig. 4: Flow Chart

3.4. UML Diagrams

The Unified Modeling Language (UML) is a standard language for specifying, visualizing, constructing, and documenting the artifacts of software systems, as well as business modeling and other system-models. The UML offers a set of graphical notations to create a variety of diagrams that describe the structure, behavior, and other aspects of a system. These diagrams can be used to communicate ideas, design software systems, and document existing systems.

3.4.1. Activity Diagram

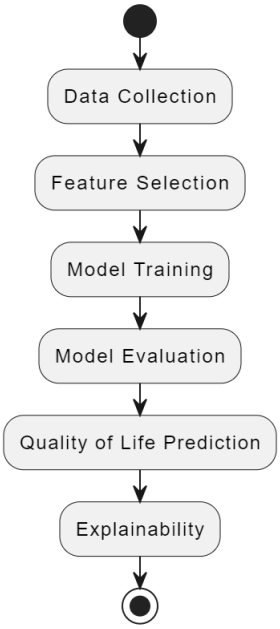

Fig. : Sequence Diagram

3.4.2. Sequence Diagram

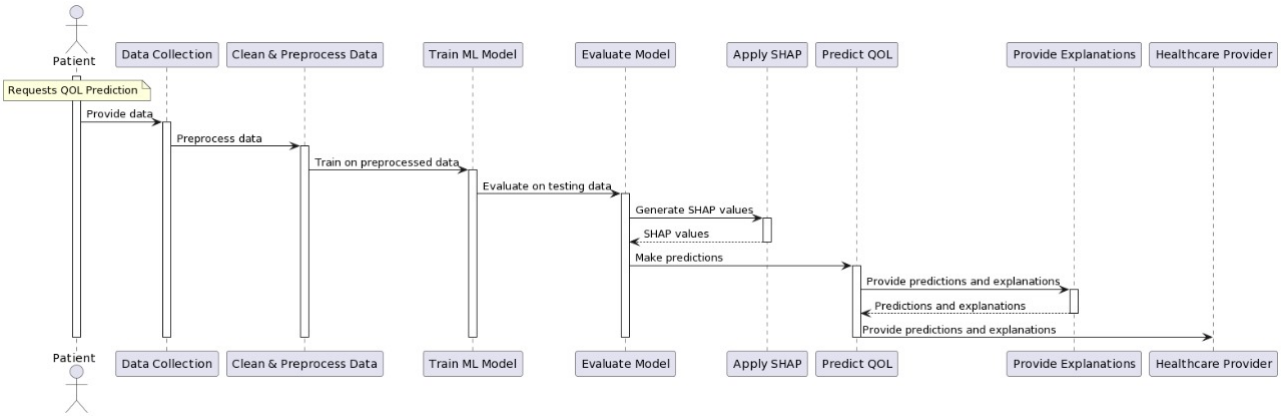

Fig. 5: Sequence Diagram

3.4.3. Use Case Diagram

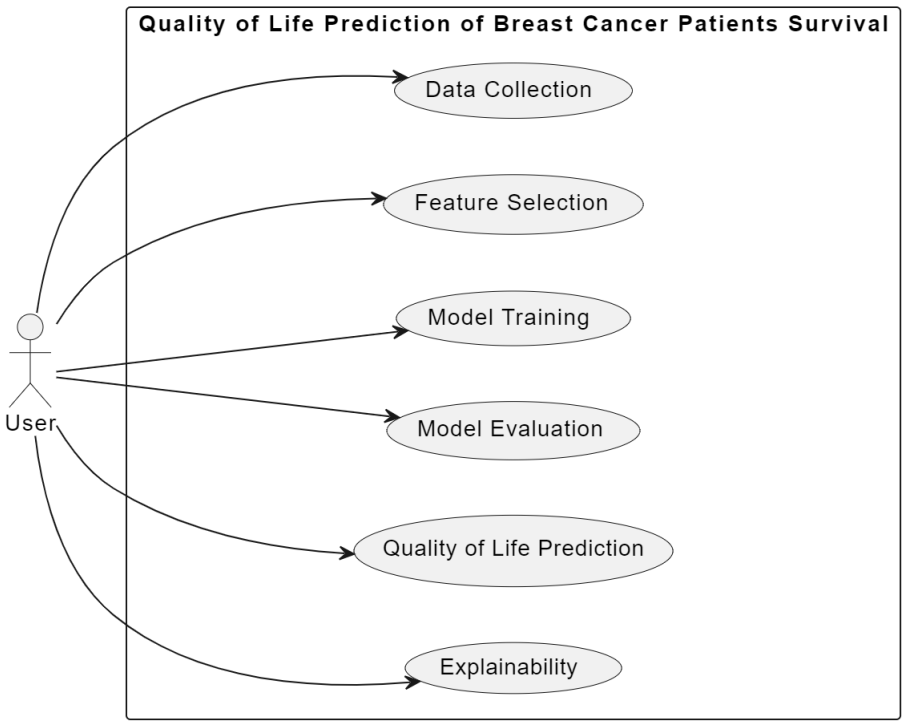

Fig. 6: Use Case Diagram

3.5. Performance Parameters with formulas

1.Accuracy:

Accuracy is a measure of the overall correctness of the model. It is the ratio of correctly predicted instances to the total instances. Mathematically, it is calculated as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

2. Precision:

Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of the accuracy of the positive predictions. Mathematically, it is calculated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

3. Recall (Sensitivity):

Recall, also known as sensitivity or true positive rate, is the ratio of correctly predicted positive observations to the all observations in actual class. It measures the ability of the model to

capture all the relevant instances. Mathematically, it is calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

4. F1 Score:

The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall. Mathematically, it is calculated as:

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

5.Area Under the Receiver Operating Characteristic (AUROC) Curve:

The AUROC is a graphical representation of the trade-off between true positive rate (sensitivity) and false positive rate (1 - specificity) at various thresholds. It provides an overall measure of the model's ability to discriminate between classes. A higher AUROC indicates better performance.

6.Confusion Matrix:

A confusion matrix is a table that summarizes the performance of a classification algorithm. It includes counts of true positives, true negatives, false positives, and false negatives. It is particularly useful for understanding where the model is making errors.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive	False Negative
Actual Negative	False Positive	True Negative

Where:

• **True Positive (TP)** Is when predicted values are positive and turns out to be true. For instance, the number of cases correctly identified that breast cancer patients survive.

• **False positive (FP)** is when values predicted as positive and turns out to be false. For instance, the number of cases incorrectly identified that breast cancer patients will survive.

• **False Negative (FN)** is when values predicted as negative and turns out to be false. This is the number of cases incorrectly identified that breast cancer patients will not survive.

• **True Negative (TN)** is when values predicted as negative and turns out to be negative, the number of cases correctly identified that breast cancer patients will not survive.

4. PERFORMANCE ANALYSIS

4.1. Result Oriented Tables

1. Random Forest Classifier

Table 4.1: Training Classification Report for Random Forest Classifier

	Precison	**Recall**	**F1 Score**	**Support**
Died of Disease	1.00	1.00	1.00	523
Died of Other Causes	1.00	1.00	1.00	805
Living	1.00	1.00	1.00	679
accuracy	-	-	1.00	2007
macro avg	1.00	1.00	1.00	2007
weighted avg	1.00	1.00	1.00	2007

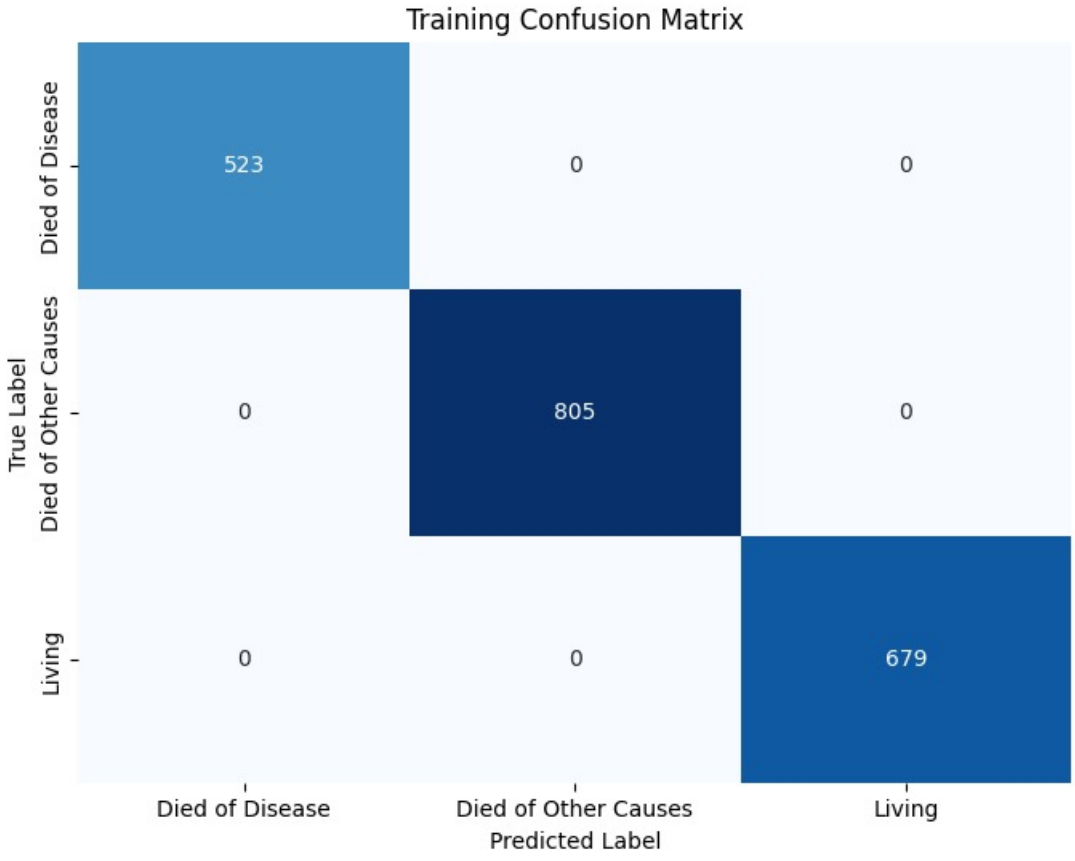

Fig. : Training Confusion Matrix

Training Accuracy:1.0

Table 4.2: Testing Classification Report for Random Forest Classifier

	Precison	**Recall**	**F1 Score**	**Support**
Died of Disease	0.91	1.00	0.95	124
Died of Other Causes	1.00	0.94	0.97	218
Living	1.00	1.00	1.00	160
accuracy	-	-	0.98	502
macro avg	0.97	0.98	0.98	502
weighted avg	0.98	0.98	0.98	502

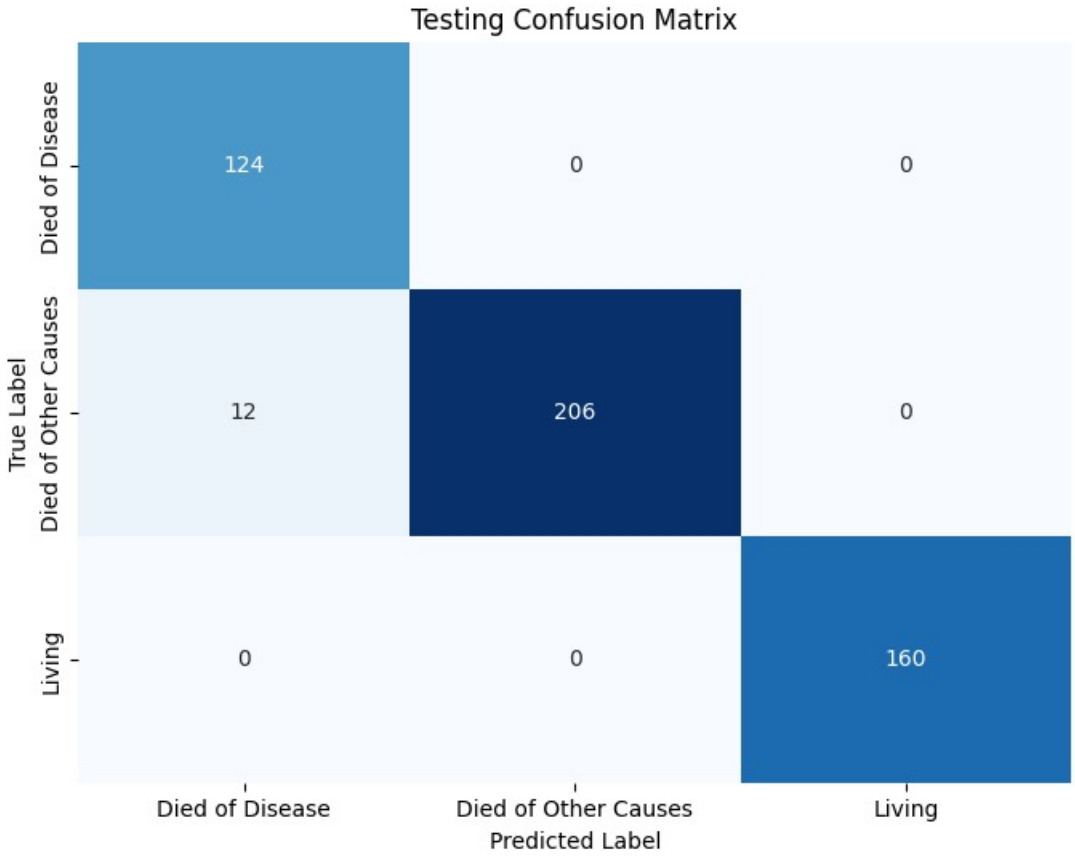

Fig. : Testing Confusion Matrix

Testing Accuracy:0.9760956175298805

A random forest classifier was trained on the 548 patients in the training dataset to classify patients as having either ”low QoL” or ”high QoL” after six months. Performance metrics such as AUROC, recall, and precision were used to evaluate the model’s performance on the test dataset.

Output:

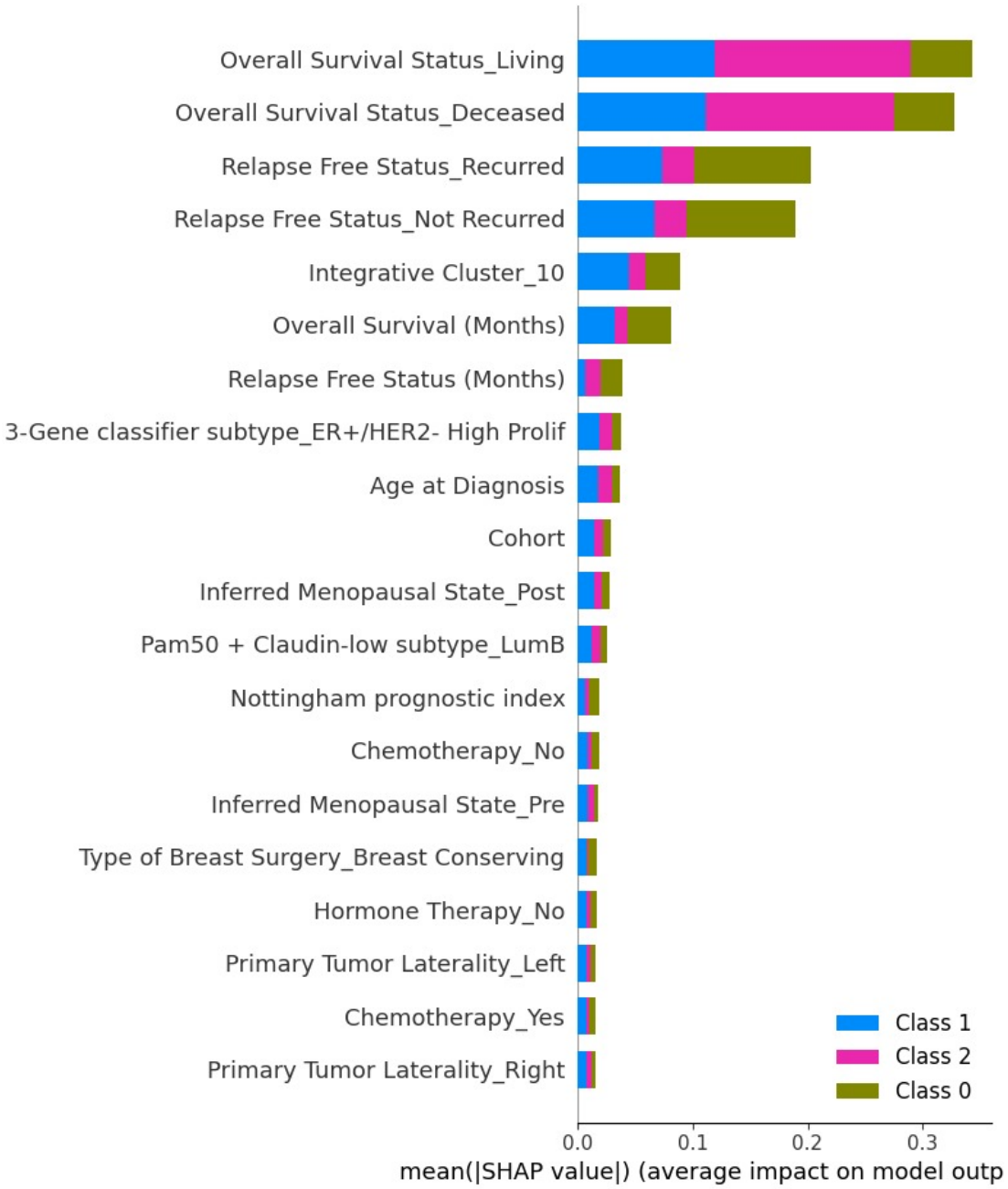

Fig. 7: Explainability of Random Forest Classifier

2. Random Survival Forest

Concordance Index: (0.9454243730092877, 35322, 2039, 0, 2)

The concordance index is a measure of how well the model's predicted probabilities align with the actual outcomes. It ranges from 0 to 1, where 1 indicates perfect concordance (the model ranks higher survival probabilities for observed longer survival times) and 0.5 indicates random chance.

The output provided seems to include the following information:
Concordance Index (C-index):**0.9454243730092877**
Number of Concordant Pairs: **35322**
Number of Discordant Pairs: **2039**
Number of Ties in Survival Times: **0**
Number of Ties in Predicted Probabilities:**2**

where:
Number of Concordant Pairs: This is the count of pairs where the model correctly predicted the order of survival times. Higher values are desirable.
Number of Discordant Pairs: This is the count of pairs where the model incorrectly predicted the order of survival times. Lower values are desirable.
Number of Ties in Survival Times: Ties occur when there are identical survival times for different individuals. In this case, there are 0 ties in observed survival times.
Number of Ties in Predicted Probabilities: Ties in predicted probabilities occur when the model assigns the same survival probability to multiple individuals. In this case, there are 2 ties in predicted probabilities.

Random Survival Forest (RSF) is a machine learning algorithm particularly suited for predicting survival outcomes in the presence of right-censored data, a common challenge in medical studies. It has gained significant popularity in breast cancer survival prediction due to its advantages

Output:

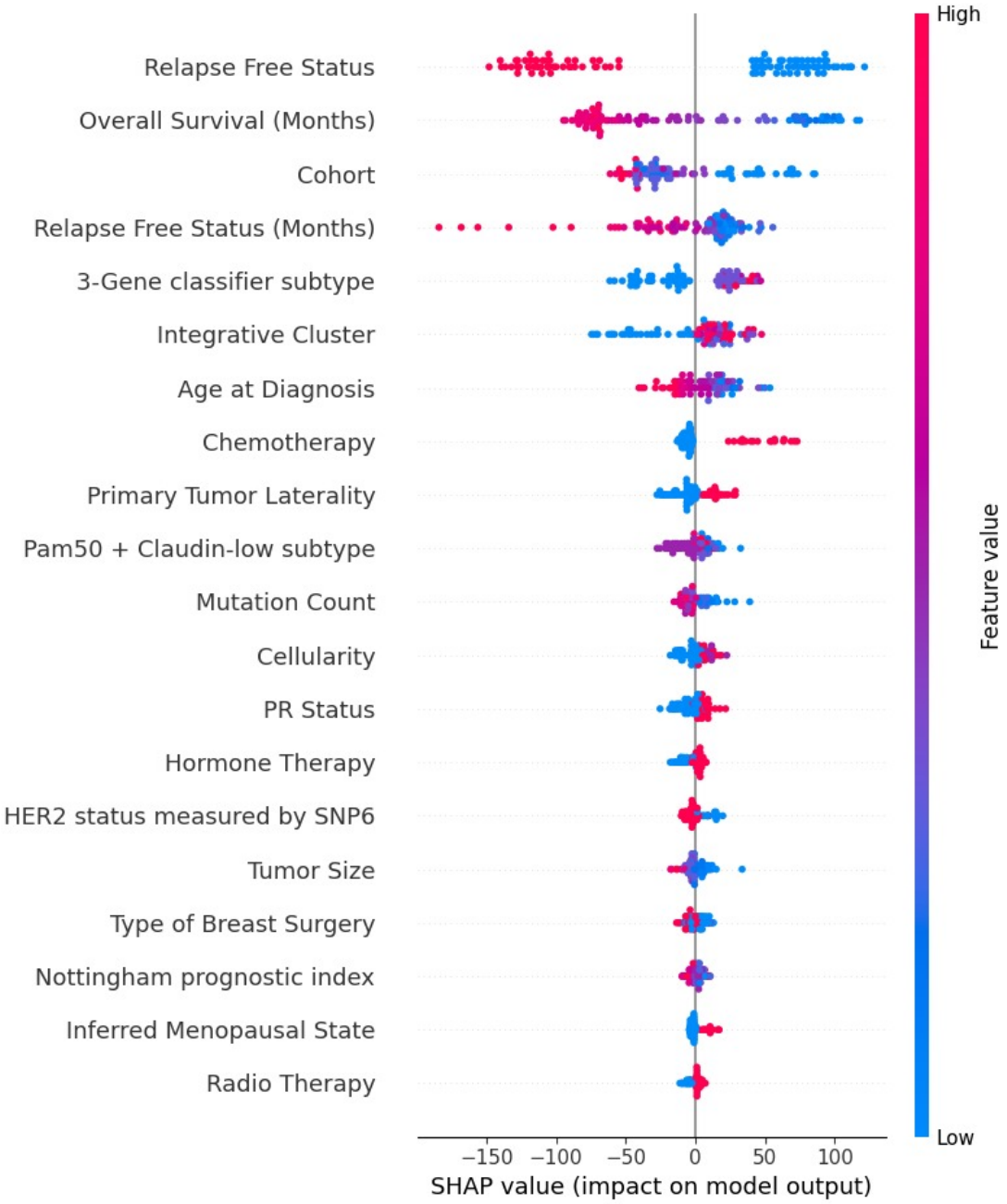

Fig. 8: Explainability of Random Survival Forest

3.Just Neural Network(JNN)

Table 4.3: Testing Classification Report for JNN

Epoch	**Loss**	**Accuracy**	**val_loss**	**val_accuracy**
1/10	0.6597	0.6413	0.4989	0.7610
2/10	0.5227	0.7404	0.4392	0.7928
3/10	0.4726	0.7718	0.4132	0.8048
4/10	0.4439	0.7942	0.3956	0.8207
5/10	0.4143	0.8092	0.3843	0.8267
6/10	0.3969	0.8132	0.3748	0.8287
7/10	0.3854	0.8161	0.3688	0.8307
8/10	0.3770	0.8326	0.3649	0.8247
9/10	0.3713	0.8256	0.3616	0.8247
10/10	0.3692	0.8301	0.3594	0.8307

AUROC	**Accuracy**	**Precison**	**Recall**	**F1 Score**
0.904495614	0.830677290	0.724550898	0.75625	0.740061162

A type of neural network called ”Just Neural Network” (JNN) to diagnose breast cancer and predict the survival of patients based on datasets.

Output:

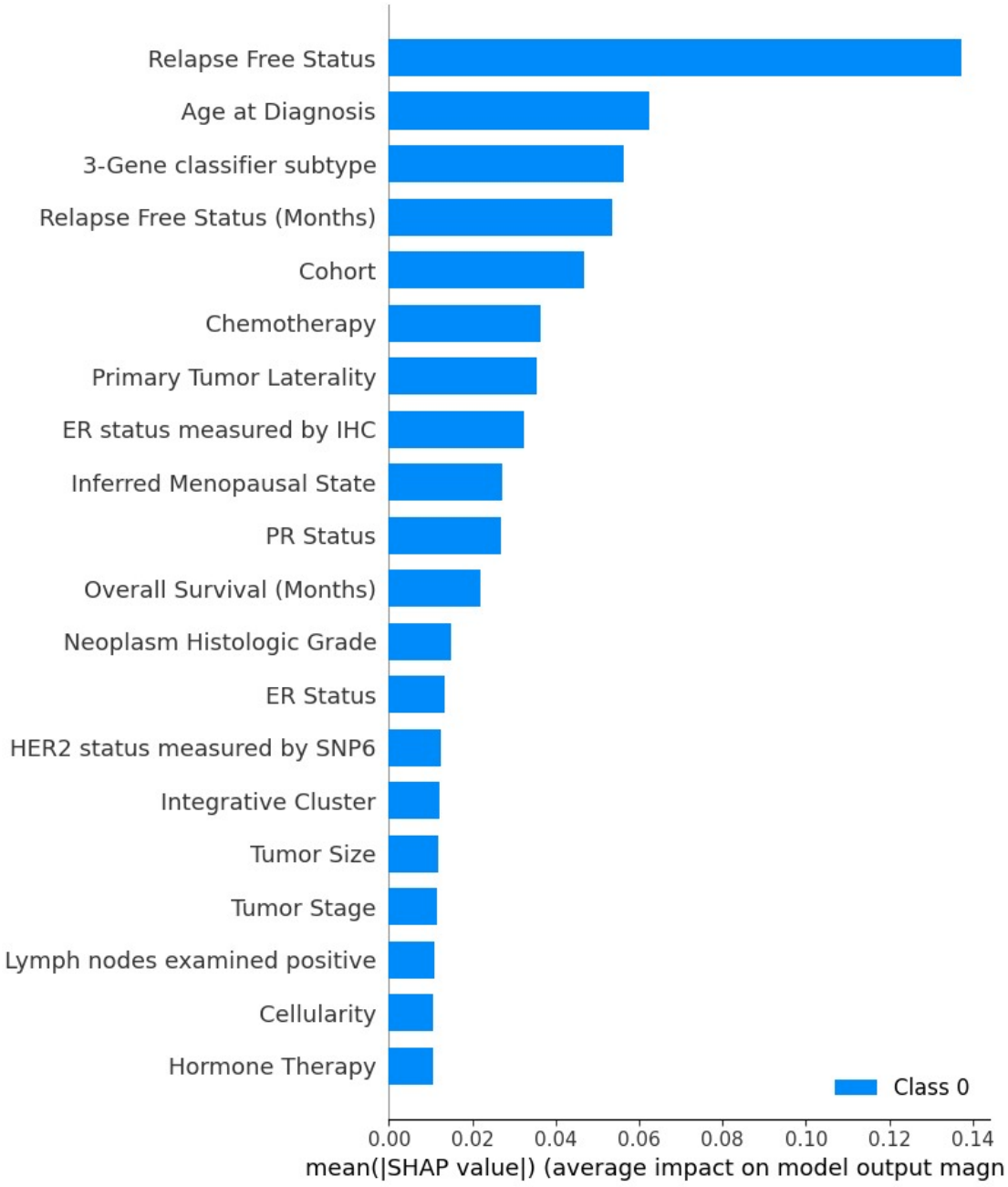

Fig. 9: Explainability of Just Neural Network

4.Coherant Voting Network(CVN)

Table 4.5: Testing Classification Report for CVN

Epoch	**Loss**	**Accuracy**	**val_loss**	**val_accuracy**
Epoch 1/10	0.6232	0.6542	0.4503	0.7869
Epoch 2/10	0.4951	0.7549	0.4017	0.7968
Epoch 3/10	0.4416	0.7922	0.3826	0.8068
Epoch 4/10	0.4222	0.7947	0.3705	0.8088
Epoch 5/10	0.4078	0.8107	0.3652	0.8227
Epoch 6/10	0.3904	0.8142	0.3578	0.8207
Epoch 7/10	0.3708	0.8256	0.3549	0.8267
Epoch 8/10	0.3588	0.8351	0.3499	0.8287
Epoch 9/10	0.3540	0.8291	0.3458	0.8307
Epoch 10/10	0.3483	0.8445	0.3468	0.8307

AUROC	**Accuracy**	**Precison**	**Recall**	**F1 Score**
0.91379751	0.83067729	0.7094972067	0.79375	0.74926253

The CVN is a mathematical model used to make predictions about patients' survival based on the expression of certain genes. Methodology using CVNs to predict breast cancer survival based on gene expressions and various patient features.

Output:

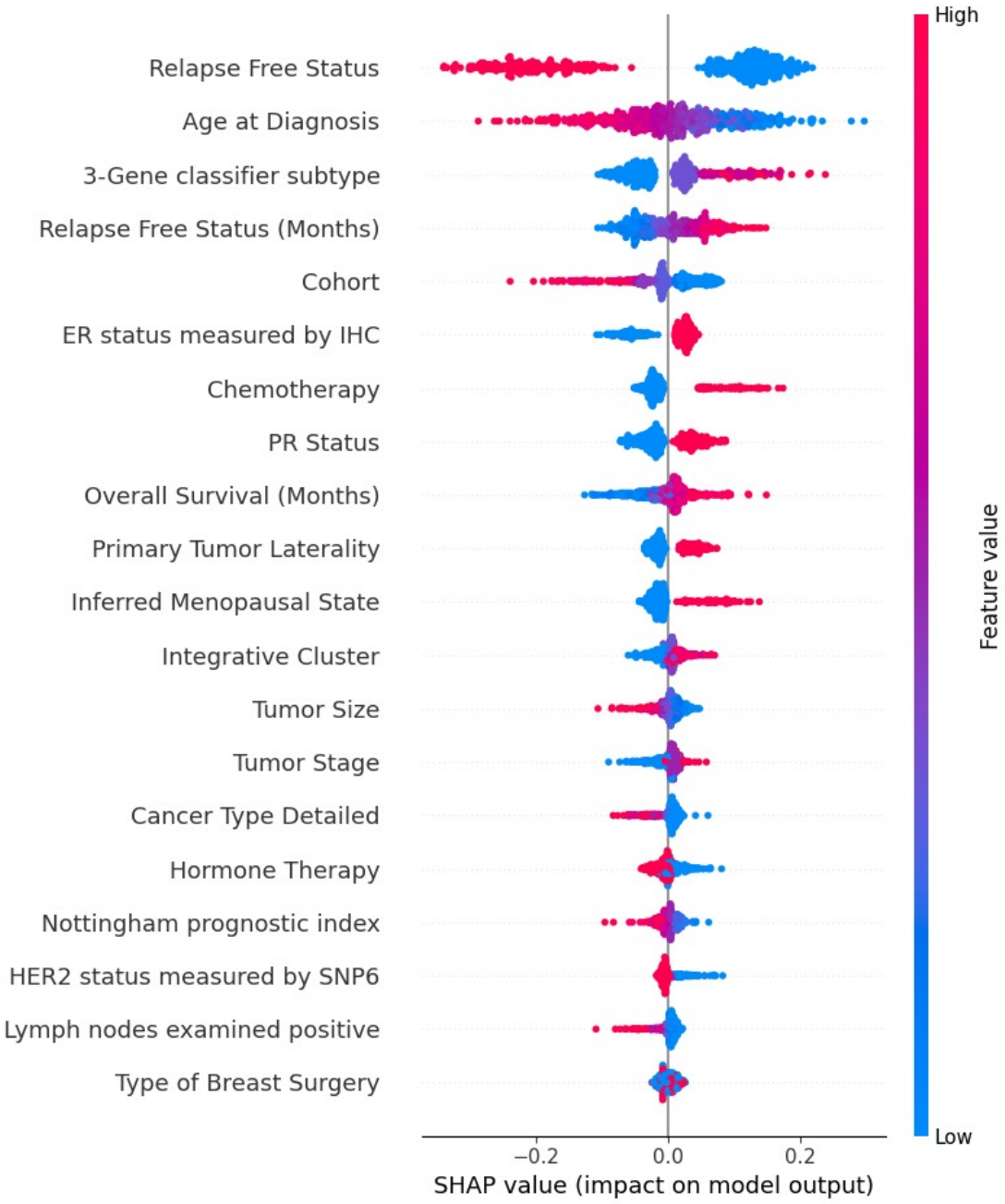

Fig. 10: Explainability of Coherent Voting Network

5. Extreme Gradient Boosting (XGBoot)

Table 4.7: Classification Report for (XGBoot)

	Precison	**Recall**	**F1 Score**	**Support**
Died of Disease	0.94	0.97	0.95	85
Died of Other Causes	0.95	0.91	0.93	58
Living	1.00	1.00	1.00	128
accuracy	-	-	0.97	273
macro avg	0.96	0.96	0.96	273
weighted avg	0.97	0.97	0.97	273

Confusion Matrix:

$$\begin{bmatrix} 84 & 3 & 0 \\ 5 & 53 & 0 \\ 0 & 0 & 128 \end{bmatrix}$$

Accuracy:0.97

4.2. Comparative Analysis

Table 4.8: Comparative analysis

Algorithms	**Precison**	**Recall**	**F1 Score**	**Accuracy**
Random Forest Classifier	0.97	0.98	0.98	0.976
XGBoot	0.96	0.96	0.96	0.970
Just Neural Network	0.72	0.75	0.74	0.830
Coherant Voting Network	0.70	0.79	0.74	0.836

4.3. Graphs and Charts

4.3.1. Graph for Distribution of Age

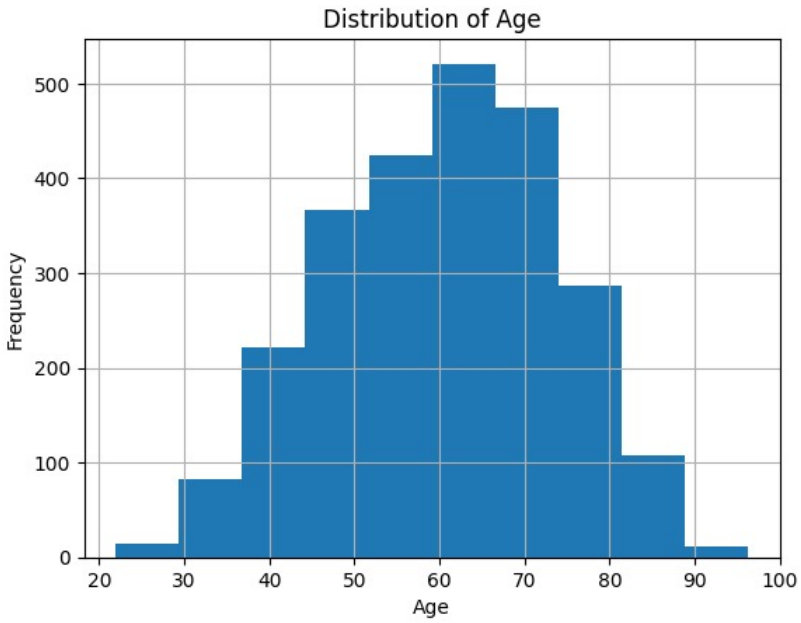

Fig. 11: Distribution of Age

4.3.2. Graph for Distribution of Patients's Vital Status

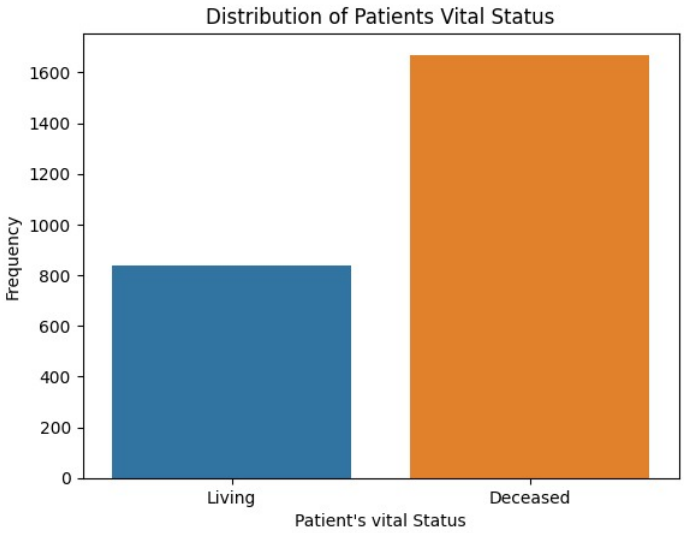

Fig. 12: Distribution of Patients's Vital Status

4.3.3. Pie Chart for Cancer Type

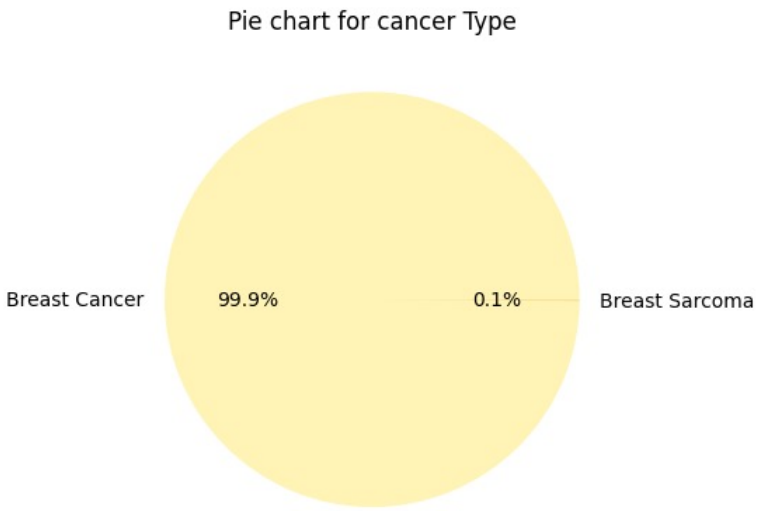

Fig. 13: Pie Chart for Cancer Type

4.3.4. Graph for Overall Survival Status

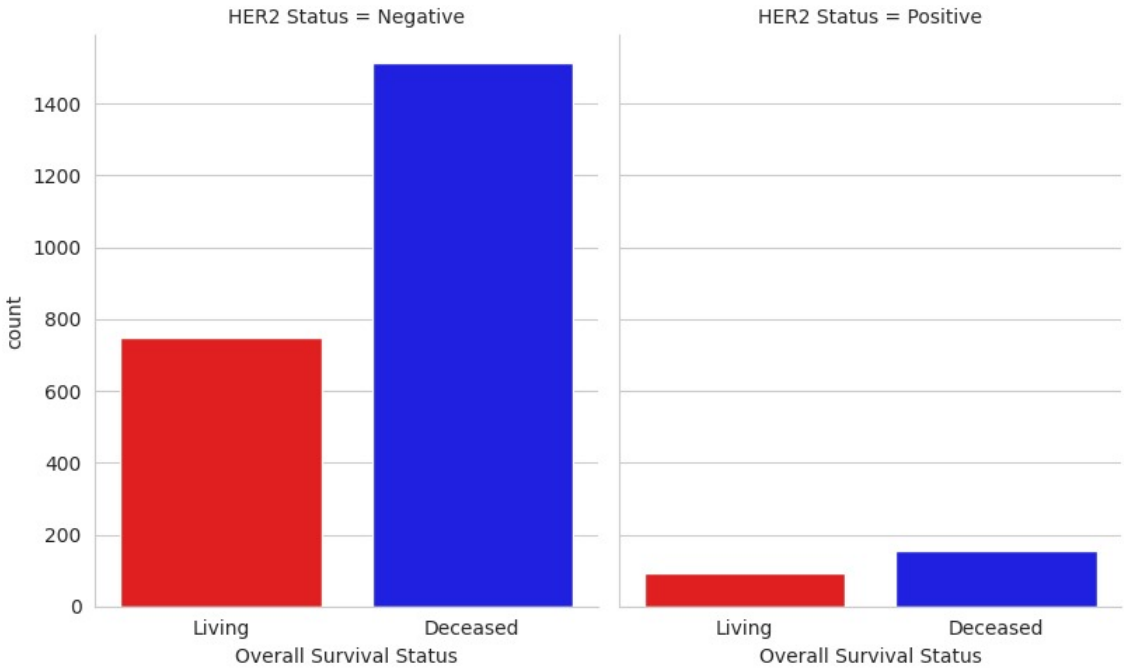

Fig. 14: Overall Survival Status

4.3.5. Correlation between different features

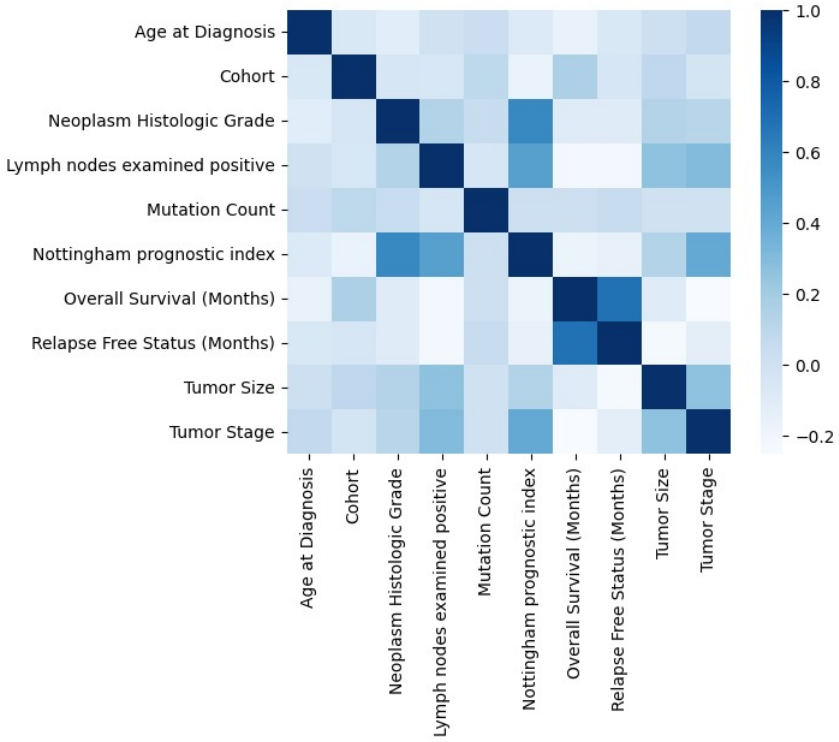

Fig. 15: Correlation between different features

4.3.6. Distribution of Actual and Predicted Output

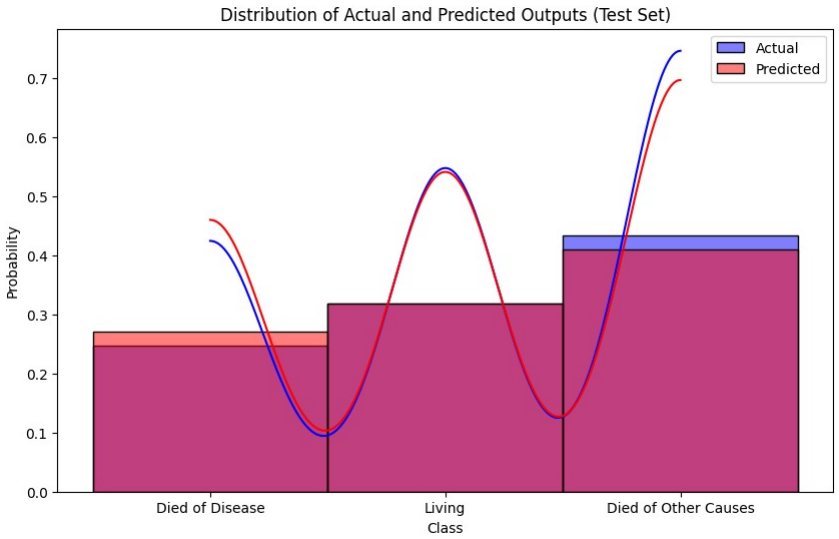

Fig. 16: Distribution of Actual and Predicted Output

4.4. Result with Discussion

Following is the result of our project.Prediction of survival of patient using some features in the data.

```
# Create an example individual data point with random values
individual_data_random = pd.DataFrame({
    'Age at Diagnosis': [45],
    'Type of Breast Surgery': ['Mastectomy'],
    'Cancer Type Detailed': ['Breast Invasive Ductal Carcinoma'],
    'Chemotherapy': ['Yes'],
    'Pam50 + Claudin-low subtype': ['LumA'],
    'ER Status': ['Negative'],
    'HER2 status measured by SNP6': ['Negative'],
    'Hormone Therapy': ['Yes'],
    'Integrative Cluster':  [2],
    'Primary Tumor Laterality': ['Right'],
    'Mutation Count': [4],
    'Nottingham prognostic index': [6],
    'Oncotree Code': ['MDLC'],
    'Overall Survival (Months)': [30],
    'Overall Survival Status': ['Living'],
    'Relapse Free Status (Months)': [112.4],
    'Relapse Free Status': ['Not Recurred'],
    '3-Gene classifier subtype': ['ER+/HER2-'],
    'Tumor Size': [24],
    'Tumor Stage':  [2],
}, index=[0])

# Match columns and handle missing columns
for column in X_train.columns:
    if column not in individual_data_random.columns:
        individual_data_random[column] = 0

# Ensure the columns are in the same order
individual_data_random = individual_data_random[X_train.columns]

# Preprocess the individual data point
individual_data_random_preprocessed = pd.get_dummies(individual_data_random)

# Ensure that the columns of individual_data_random_preprocessed match the columns used during training
missing_columns = set(X_train.columns) - set(individual_data_random_preprocessed.columns)
for column in missing_columns:
    individual_data_random_preprocessed[column] = 0

# Predict the outcome for the random individual data point
individual_prediction_random = rf_classifier.predict(individual_data_random_preprocessed)

# Print or use the prediction
print("Predicted Outcome for Random Individual Data:", individual_prediction_random)
```

```
Predicted Outcome for Random Individual Data: ['Living']
```

Fig. 17: Actual Output

To implement our model,we created a google form to collect the information of patient and to predict its survival status.

Google Form Link:
`https://docs.google.com/forms/d/1CVIOh4AJmXWtOuFGwFJg3J28GPqI-QE39HSisPyiOGO/edit`

5. CONCLUSION

5.1. Conclusion

In conclusion,Breast cancer is a complex and devastating disease, demands innovative approaches to enhance diagnosis, survival prediction, and quality of life assessment. In this project, we amalgamate methodologies from diverse research papers to create a comprehensive framework for breast cancer care. Leveraging insights from these studies, we strive to improve patient outcomes by integrating advanced technology and predictive models.

5.2. Advantages

1.Clinical Decision Support:

Explainable AI can assist healthcare professionals in making informed decisions about treatment plans and interventions for breast cancer patients. Understanding the model's reasoning helps clinicians tailor treatments to individual patient needs.

2.Identification of Relevant Features:

Explainability allows for the identification of key features or variables that significantly contribute to the quality of life and survival predictions. This information can guide healthcare professionals in focusing on critical factors during patient care.

3.Customization of Treatment Plans:

With a clear understanding of the model's decision factors, healthcare providers can customize treatment plans based on individual patient characteristics. This promotes the concept of personalized medicine, optimizing the chances of positive outcomes.

Enhanced Survivorship Programs:

Predictive models can assist in the development of survivorship programs that address the unique needs of breast cancer survivors, including physical, emotional, and social aspects. This comprehensive approach contributes to better long-term outcomes and improved quality of life

It's important to note that the implementation of these predictive models should be accompanied by ethical considerations, clear communication, and collaboration between patients, healthcare providers, and data scientists to ensure the responsible and ethical use of patient data.

5.3. Applicability

The "Life Prediction for Breast Cancer Patients" project is designed to benefit a wide range of applicability like

Clinical Decision Support:
Healthcare professionals can use the predictive model to assist in treatment planning and discussions with breast cancer patients, providing a more accurate prognosis.

Patient Education:
Breast cancer patients and their families can gain a better understanding of the expected trajectory of the disease, potentially reducing anxiety and aiding in informed decision-making.

Research:
The project can contribute to the field of oncology by providing a new tool for studying the factors that influence the survival of breast cancer patients.

Healthcare Policy:
The project's findings may inform healthcare policy decisions related to breast cancer care and resource allocation.

Pharmaceutical Development: Pharmaceutical companies can potentially use the insights generated by the project to develop more personalized treatment strategies for breast cancer.

5.4. Scope

The "Life Prediction for Breast Cancer Patients" project aims to develop a comprehensive and accurate predictive model to estimate the life expectancy and survival rates of individuals diagnosed with breast cancer. This project encompasses a range of activities, data analysis, and the creation of predictive tools to assist medical professionals and patients in making informed decisions regarding treatment and care.

References

[1] A. Turcu-Stiolica, M. Bogdan, E. A. Dumitrescu, D. L. Zob, V. Gheorman, M. Aldea, V. C. Dinescu, M.-S. Subtirelu, D.-L. Stanculeanu, D. Sur *et al.*, "Diagnostic accuracy of machine-learning models on predicting chemo-brain in breast cancer survivors previously treated with chemotherapy: A meta-analysis," *International Journal of Environmental Research and Public Health*, vol. 19, no. 24, p. 16832, 2022.

[2] M. Monirujjaman Khan, S. Islam, S. Sarkar, F. I. Ayaz, M. M. Kabir, T. Tazin, A. A. Albraikan, F. A. Almalki *et al.*, "Machine learning based comparative analysis for breast cancer prediction," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[3] M. Z. A. Shawarib, A. E. A. Latif, B. E. E.-D. Al-Zatmah, and S. S. Abu-Naser, "Breast cancer diagnosis and survival prediction using jnn," 2020.

[4] A. Moncada-Torres, M. C. van Maaren, M. P. Hendriks, S. Siesling, and G. Geleijnse, "Explainable machine learning can outperform cox regression predictions and provide insights in breast cancer survival," *Scientific reports*, vol. 11, no. 1, p. 6968, 2021.

[5] J. Li, Z. Zhou, J. Dong, Y. Fu, Y. Li, Z. Luan, and X. Peng, "Predicting breast cancer 5-year survival using machine learning: A systematic review," *PloS one*, vol. 16, no. 4, p. e0250370, 2021.

[6] M. Nuutinen, S. Korhonen, A.-M. Hiltunen, I. Haavisto, P. Poikonen-Saksela, J. Mattson, H. Kondylakis, K. Mazzocco, R. Pat-Horenczyk, B. Sousa *et al.*, "Impact of machine learning assistance on the quality of life prediction for breast cancer patients." SCITEPRESS Science And Technology Publications, 2021.