BVRIT HYDERABAD College of Engineering for Women

Department of Information Technology

BREAST CANCER PROGNOSIS USING MACHINE LEARNING

Under the guidance of

Guide Name: Mr. B.Srinivasulu

Designation: Assistant Professor

Team- 09

R. Parimala (19WH1A1268)

P. Aditi Kiran(19WH1A1277)

Ch. Lakshmi Durga(19WH1A1295)

Raveena Yadlapalli (19WH1A1296)

Agenda

- Introduction
- Existing System
- Problem statement
- Literature Survey
- Proposed system
- Tools and Technology
- Feasibility study
- Societal impact

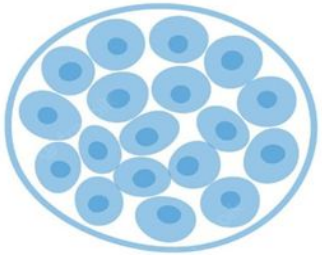
- Architecture
- Modules
- UML Diagram
- Implementation
- Report
- Project timeline
- References

Introduction

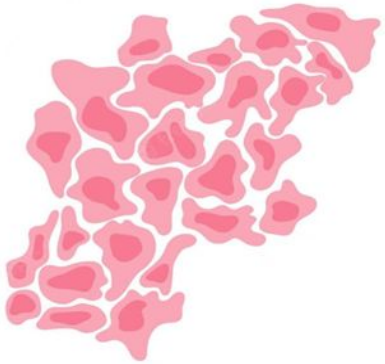
- Basically, there are two types of tumors found in breasts. They are malignant and benign.

  Types of breast cancer:
  - **> Malignant**
  - **>Benign**

- The main objective of this project to build the model for predicting cancer using various machine learning techniques and predicting the chances of reoccurrence of cancer.



Benign tumor
- Non-cancerous
- Capsulated
- Non-invasive
- Slow growing
- Do not metastasize (spread) to other parts of the body
- Cells are normal



Maligant tumor
- Cancerous
- Non-capsulated
- Fast growing
- Metastasize (spread) to other parts of the body
- Cells have large, dark nuclei, may have abnormal shape

VISHNU
UNIVERSAL LEARNING

Existing system

There are 2 existing systems that are already being used and they are

- Breast Cancer Risk Assessment Tool (BCRAT).
- Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm (BOADICEA) model.

VISHNU
UNIVERSAL LEARNING

Problem Statement

- Computer Science & Engineering is used in Bioinformatics and Biomedical to diagnose and prognoses disease Cancer. This can be further directed to a field called Machine Learning where various techniques are available to predict the cancer on the basis of collected standard data sets.
- We need to apply some classifiers of Machine Learning Techniques to signify the cancer in a human.

VISHNU
UNIVERSAL LEARNING

Literature Survey

S.No	Author	Title	Published year	Observation
1	Krishna Mridha	Early Prediction of Breast Cancer by using Artificial Neural Network and Machine Learning Techniques	2021	The paper focuses on numerous models that are applied to the dataset taken. In terms of accuracy, cross-validation, sensitivity, and specificity gained, each of the algorithm was calculated and compared.Random Forest gives the best accuracy .
2	Prateek P.Sengar, Mihir J.Gaikwad, Ashlesha S.Nagdive	Comparative study of Machine Learning Algorithms for Breast Cancer Prediction.	2020	2 machine learning algorithms is proposed to compare namely Logistic Regression and Decision Tree algorithm on the Wisconsin (Diagnostic) Data Set and use the algorithm with the best accuracy for predicting Breast Cancer.
3	Mamatha Sai Yarabaria, Lakshmi Kavya Ravi, A. Sivasangari	Breast Cancer Prediction via Machine Learning.	2019	predicting whether a person is suffering with breast cancer or not is done with the help of the trained data.

VISHNU
UNIVERSAL LEARNING

Proposed System

- The objective of this project is to predict and diagnose breast cancer, using machine learning algorithms, and find out the most effective result based on the performance of each classifier/algorithm in terms of confusion matrix, precision, accuracy, and sensitivity. Our project also Identify the recurrence of breast cancer.

VISHNU
UNIVERSAL LEARNING

Tools and Technology

Environment	Specification
Hardware	System Hard disk Ram - 4GB Processor - intel core i5 7th generation Hard drive - 64 GB
Software	Operating System - Windows 10 Jupyter Notebook Google colab Ms-word Python

VISHNU
UNIVERSAL LEARNING

Feasibility Study

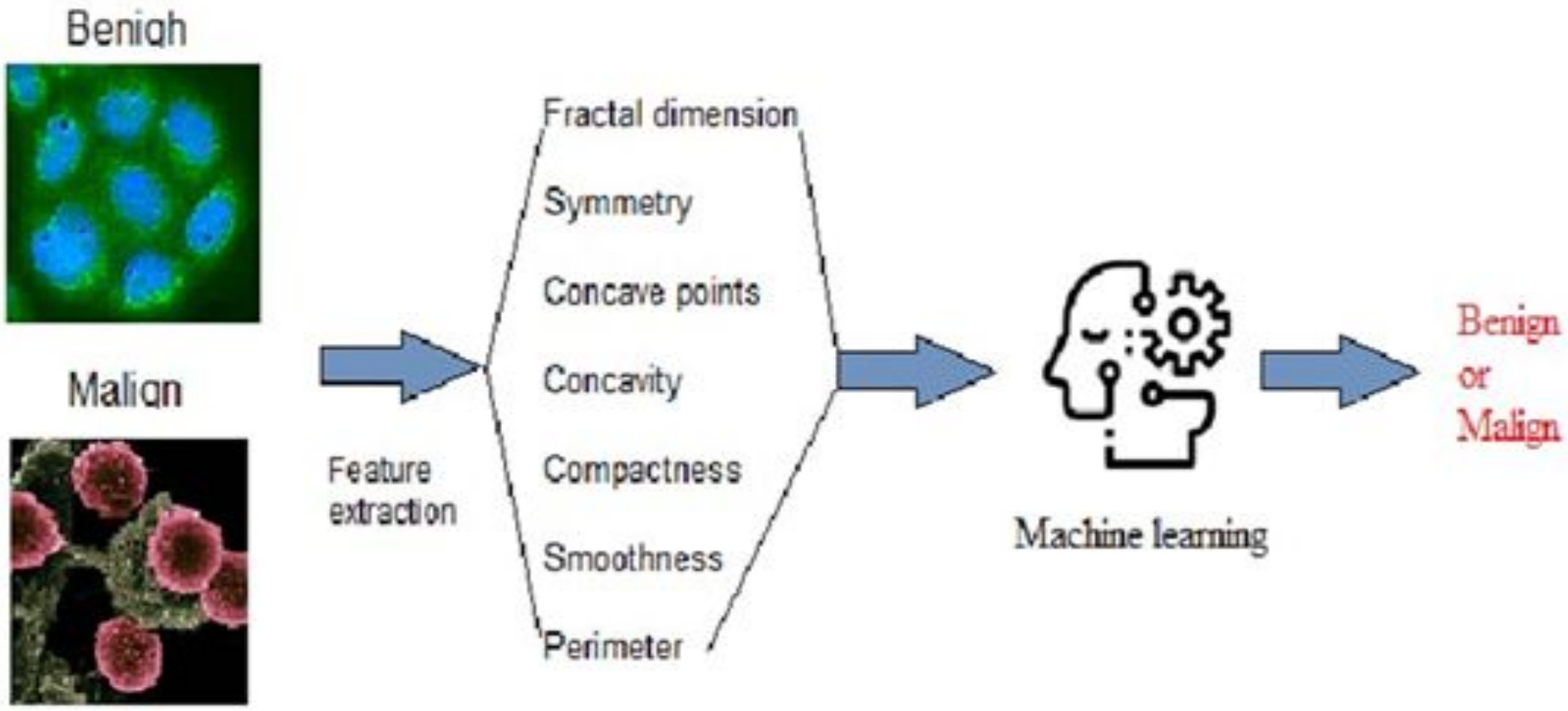
- The system utilizes Machine Learning Algorithms like Logistic Regression,Decision Tree,Naive Bayes,K Nearest Neighbors,Support Vector Machine,Random Forest.
- Tools required are Google Colab,Jupyter Notebook.

VISHNU
UNIVERSAL LEARNING

Societal Impact

- Breast cancer is considered to be one of the significant causes of death in women. So, early detection of breast cancer plays an essential role to save woman's life.
- Machine Learning models are getting better than pathologists at accurately predicting the development of cancer.
- We are intending how to parametrize our classification techniques to achieve high accuracy. We are looking into many datasets and how further Machine Learning algorithms can be used to characterize Breast Cancer. We want to reduce the error rates with maximum accuracy.

Architecture

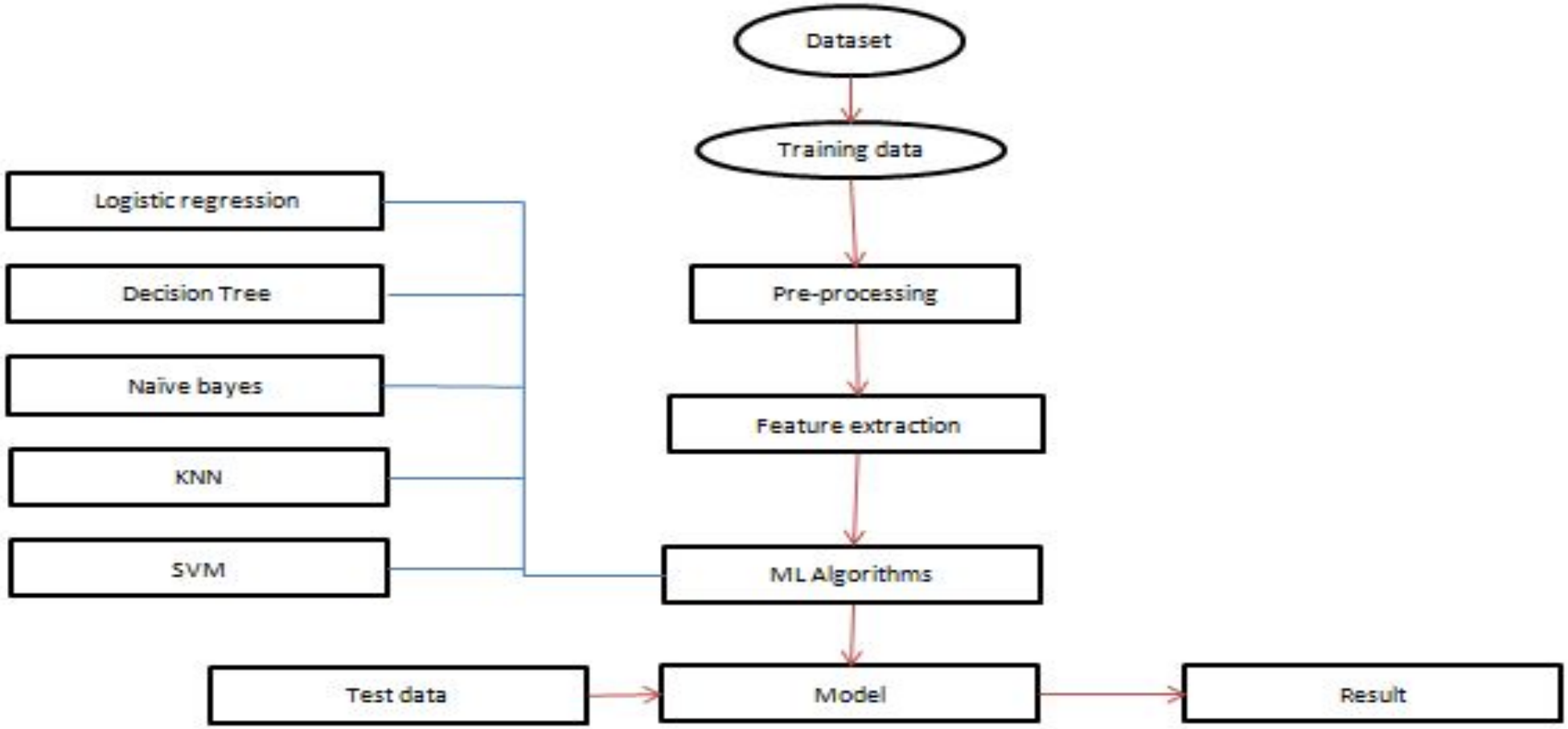


VISHNU
UNIVERSAL LEARNING

Modules

1. Importing Data
2. Data Pre-processing
3. Training and Testing data
4. Prediction

UML Diagram


Dataset
Training data
Logistic regression
Decision Tree
Naïve bayes
KNN
SVM
Pre-processing
Feature extraction
ML Algorithms
Test data
Model
Result

VISHNU
UNIVERSAL LEARNING

Partial Implementation

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
%matplotlib inline
```

```
dataset = pd.read_csv('data.csv')
```

```
X = dataset.iloc[:, 1:31].values
Y = dataset.iloc[:, 31].values
```

```
dataset.head()
```

```
print("Cancer data set dimensions : {}".format(dataset.shape))
```

```
Cancer data set dimensions : (569, 33)
```

```
diagnosis_unique = dataset.diagnosis.unique()
print(diagnosis_unique)
```

```
['M' 'B']
```

```
benign,malignant=dataset["diagnosis"].value_counts()
```

```
print('Number of cells labeled Benign: ', benign)
print('Number of cells labeled Malignant : ', malignant)
```

```
Number of cells labeled Benign:  357
Number of cells labeled Malignant :  212
```

```
print('% of cells labeled Benign', round(benign / len(dataset) * 100, 2), '%')
print('% of cells labeled Malignant', round(malignant / len(dataset) * 100, 2), '%')
```

```
% of cells labeled Benign 62.74 %
% of cells labeled Malignant 37.26 %
```

```
sns.set_style('darkgrid')
plt.figure(figsize=(15, 5))
plt.xlabel("Diagnosis")
plt.subplot(1, 2, 2)
plt.title("Count of Diagnosis")
sns.countplot(x='diagnosis', data=dataset);
```

VISHNU
UNIVERSAL LEARNING

Partial Implementation

```
#Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.fit_transform(X_test)
```

```
def model_building(model, X_train, X_test, y_train, y_test):
    """

    Model Fitting, Prediction And Other stuff
    return ('score', 'accuracy_score', 'predictions' )
    """

    model.fit(X_train, y_train)
    score = model.score(X_train, y_train)
    predictions = model.predict(X_test)
    accuracy = accuracy_score(predictions, y_test)
    return (score, accuracy, predictions)
```

```
models_list = {
    "LogisticRegression" :  LogisticRegression(random_state=0),
    "K-NearestNeighbor" :  KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2),
    "SVC" :  SVC(kernel = 'rbf', random_state = 2,C = 2),
    "SVM" :  SVC(kernel = 'linear', random_state = 0),
    "NaiveBayes" :  GaussianNB(),
    "DecisionTreeClassifier" :  DecisionTreeClassifier(criterion='entropy', random_state=0),
    "RandomForestClassifier" :  RandomForestClassifier(n_estimators=10, criterion='entropy', random_state=0),
}
```

```
import random
a = random.random()
meanR=round(random.uniform(0,30),3)
meanT=round(random.uniform(0,50),3)
meanS=round(random.uniform(0,1),5)
meanC=round(random.uniform(0,1),5)
meanSy=round(random.uniform(0,1),4)
meanF=round(random.uniform(0,1),5)
seR=round(random.uniform(0,2),4)
seT=round(random.uniform(0,3),4)
seS=round(random.uniform(0,1),6)
seC=round(random.uniform(0,1),6)
seSy=round(random.uniform(0,1),6)
seF=round(random.uniform(0,1),6)
input_1=[]
lst =  map(lambda x : x[1], filter(lambda x : x[0].startswith('mean'), globals().items()))
for i in lst:
    input_1.append(i)
lst1 =  map(lambda x : x[1], filter(lambda x : x[0].startswith('se'), globals().items()))
for i in lst1:
    input_1.append(i)
input_array = np.asarray(input_1)
input_reshaped = input_array.reshape(1,-1)
predict = model.predict(input_reshaped)
if (predict[0] < 0.5):
  print("breast cancer is malignant")
else:
   print("breast cancer is benign")
```

VISHNU
UNIVERSAL LEARNING

Result

	model_name	score	accuracy_score	accuracy_percentage
0	LogisticRegression	0.934673	0.959064	95.91%
1	K-NearestNeighbor	0.937186	0.964912	96.49%
2	SVC	0.962312	0.964912	96.49%
3	SVM	0.932161	0.959064	95.91%
4	NaiveBayes	0.912060	0.929825	92.98%
5	DecisionTreeClassifier	1.000000	0.883041	88.30%
6	RandomForestClassifier	0.994975	0.918129	91.81%

```
df_pred.sort_values('score', ascending=False)
```

	model_name	score	accuracy_score	accuracy_percentage
5	DecisionTreeClassifier	1.000000	0.883041	88.30%
6	RandomForestClassifier	0.994975	0.918129	91.81%
2	SVC	0.962312	0.964912	96.49%
1	K-NearestNeighbor	0.937186	0.964912	96.49%
0	LogisticRegression	0.934673	0.959064	95.91%
3	SVM	0.932161	0.959064	95.91%
4	NaiveBayes	0.912060	0.929825	92.98%

```
import random
a = random.random()
meanR=round(random.uniform(0,30),3)
meanT=round(random.uniform(0,50),3)
meanS=round(random.uniform(0,1),5)
meanC=round(random.uniform(0,1),5)
meanSy=round(random.uniform(0,1),4)
meanF=round(random.uniform(0,1),5)
seR=round(random.uniform(0,2),4)
seT=round(random.uniform(0,3),4)
seS=round(random.uniform(0,1),6)
seC=round(random.uniform(0,1),6)
seSy=round(random.uniform(0,1),6)
seF=round(random.uniform(0,1),6)
input_1=[]
lst =  map(lambda x : x[1], filter(lambda x : x[0].startswith('mean'), globals().items()))
for i in lst:
    input_1.append(i)
lst1 =  map(lambda x : x[1], filter(lambda x : x[0].startswith('se'), globals().items()))
for i in lst1:
    input_1.append(i)
input_array = np.asarray(input_1)
input_reshaped = input_array.reshape(1,-1)
predict = model.predict(input_reshaped)
if (predict[0] < 0.5):
  print("breast cancer is malignant")
else:
   print("breast cancer is benign")
```

```
breast cancer is benign
```

VISHNU
UNIVERSAL LEARNING

Project Timeline

Dates(To-From)	Duration	Tasks
06.09.2022-22.10.2022	4 weeks	Domain Selection References Specifications Requirements
23.10.2022-12.11.2022	3 weeks	Literature Survey
13.11.2022-18.12.2022	5 weeks	Data Pre-processing & Algorithms
19.12.2022-10.01.2023	3 weeks	Full Implementation

VISHNU
UNIVERSAL LEARNING

References

- Usman Naseem,Junaid Rashid, ”An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers”12 May 2022
- L Yang, B Fu, Y Li, Y Liu, W Huang, S Feng et al., "Prediction model of the response to neoadjuvant chemotherapy in cancers by a Naive Bayes algorithm", *Computer methods and programs in biomedicine*, vol. 192, pp. 105458, 2020.
- B. Akbugday, "Classification of Breast Cancer Data Using Machine Learning Algorithms," 2019 Medical Technologies Congress (TIPTEKNO), Izmir, Turkey, 2019, pp. 1-4.

VISHNU
UNIVERSAL LEARNING

THANK YOU