BENIGN AND MALIGNANT BREAST TUMOR CLASSIFICATION

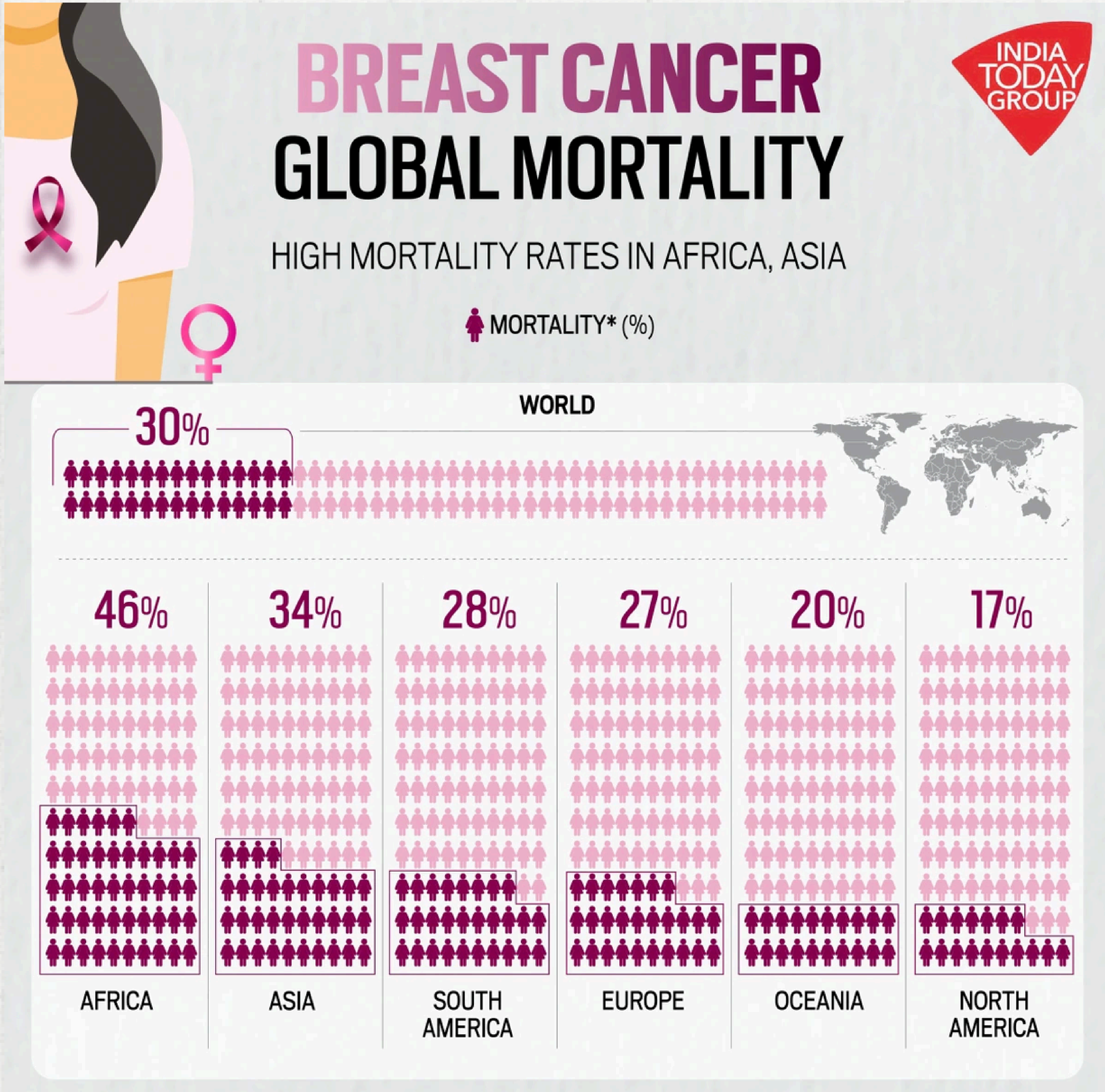
Project Presentation

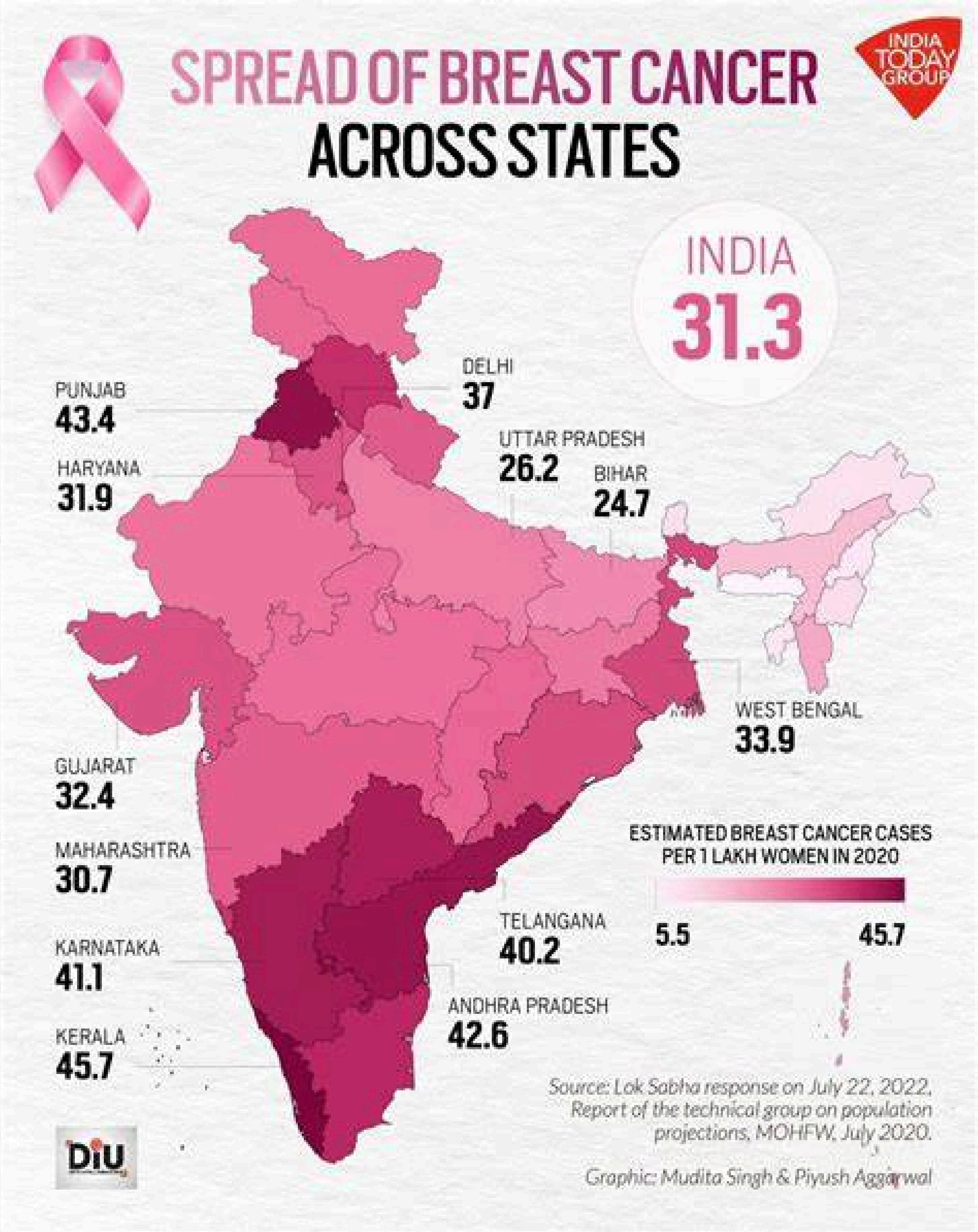
Presented by HARSH (20) and ROHAN (47)

Breast Cancer

Malignant tumors are defined as +ve

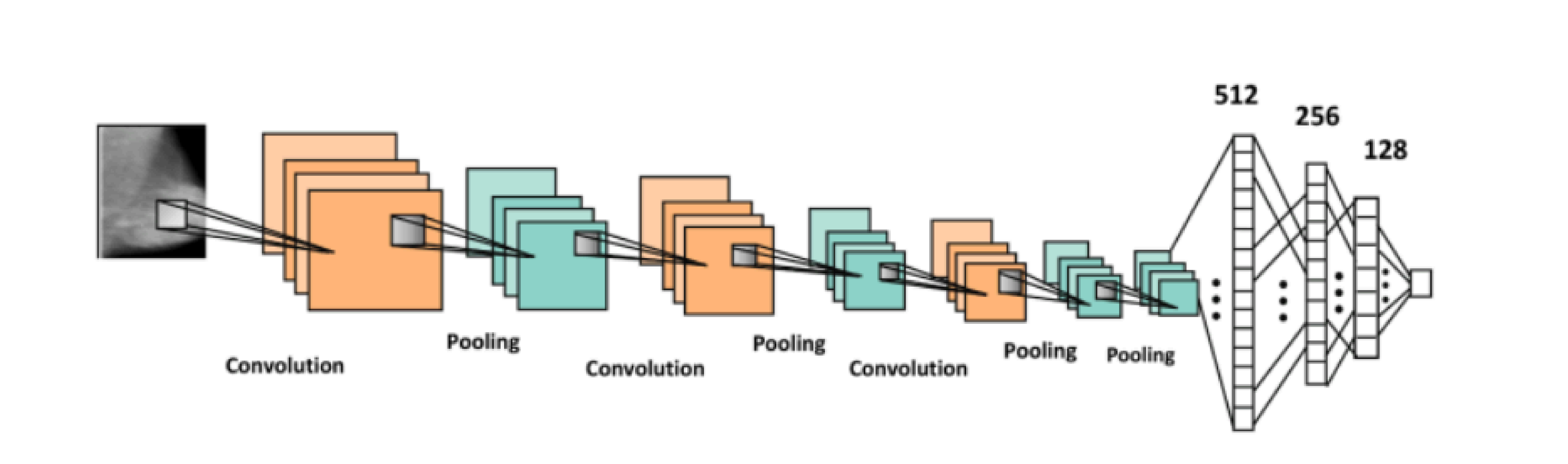
Benign tumors are defined as -ve


BREAST CANCER
GLOBAL MORTALITY
HIGH MORTALITY RATES IN AFRICA, ASIA
INDIA TODAY GROUP
MORTALITY* (%)
WORLD
30%
46%
34%
28%
27%
20%
17%
AFRICA
ASIA
SOUTH AMERICA
EUROPE
OCEANIA
NORTH AMERICA


SPREAD OF BREAST CANCER
ACROSS STATES
INDIA TODAY GROUP
INDIA
31.3
PUNJAB
43.4
HARYANA
31.9
DELHI
37
UTTAR PRADESH
26.2
BIHAR
24.7
WEST BENGAL
33.9
GUJARAT
32.4
MAHARASHTRA
30.7
KARNATAKA
41.1
KERALA
45.7
TELANGANA
40.2
ANDHRA PRADESH
42.6
ESTIMATED BREAST CANCER CASES PER 1 LAKH WOMEN IN 2020
5.5
45.7
Source: Lok Sabha response on July 22, 2022, Report of the technical group on population projections, MOHFW, July 2020.
Graphic: Mudita Singh & Piyush Aggarwal
DIU

Literature Review

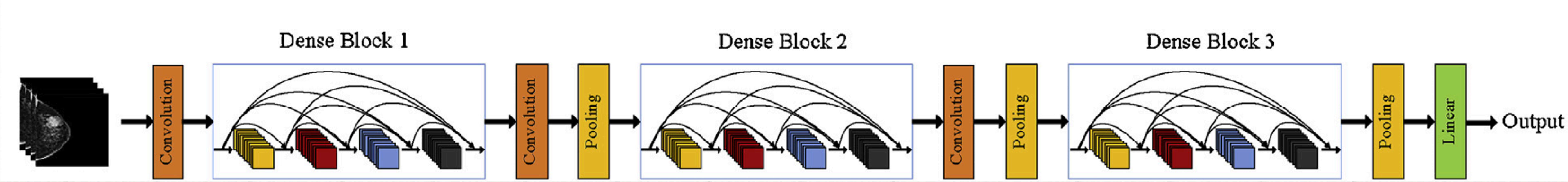
Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks.

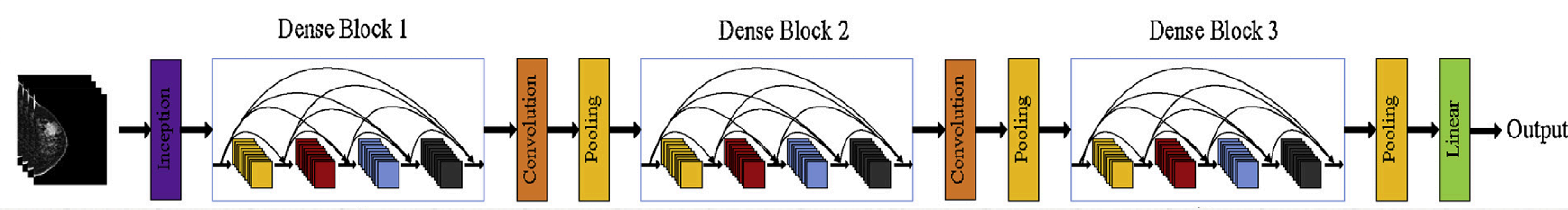


The average results of different datasets using 5-fold cross-validation.

Dataset	Tp	Fn	Tn	Fp	sensitivity	specificity	F1 score	Test acc.	AUC
Annotated INbreast	56	2	55	2	96.55	96.49	96.55	96.52	0.98
Annotated MIAS	51	1	50	4	98	92.6	95.33	95.3	0.974
INbreast	55	3	52	5	94.83	91.23	93.22	93.04	0.946
MIAS	51	4	48	3	92.72	94.12	93.58	93.39	0.945
DDSM	156	13	135	15	92.31	90	91.76	91.2	0.924

Benign and malignant classification of mammogram images based on deep learning

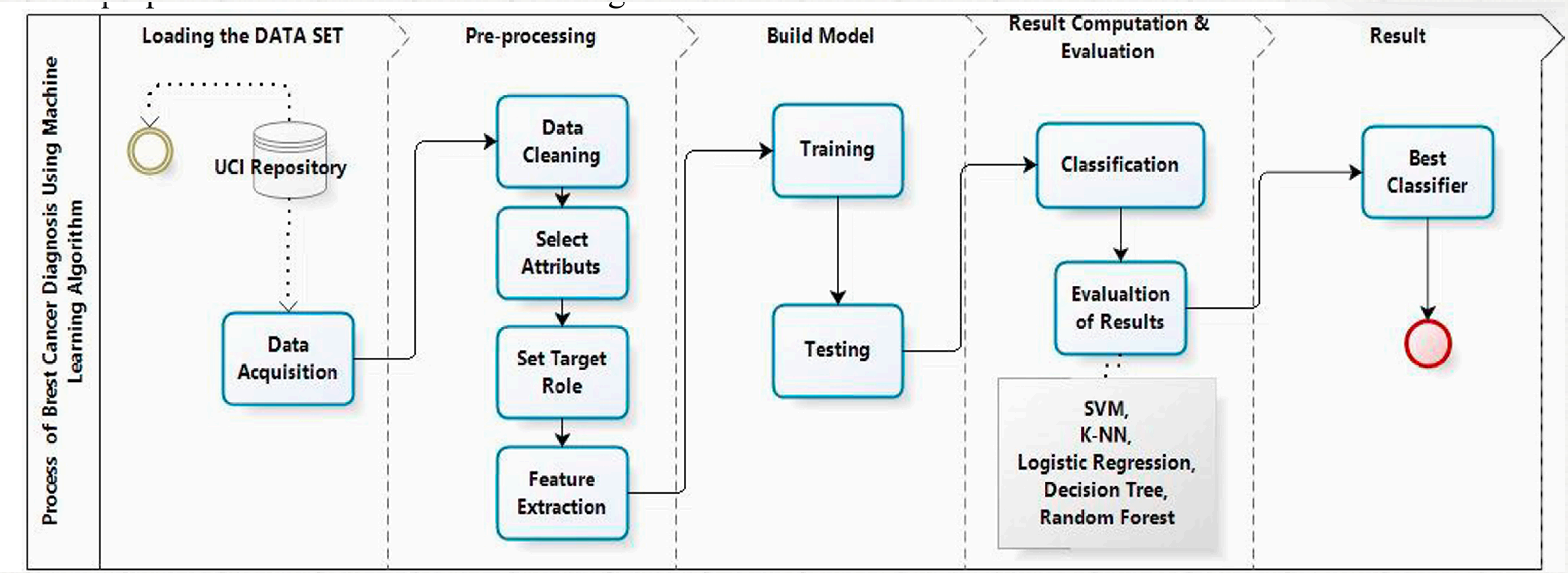




The performance table of breast cancer benign and malignant classification(the 10-fold cross-validation average results).

Neural Networks	Performance		
	Acc(%)	Sen(%)	Spec(%)
AlexNet	92.70	93.60	91.78
VGGNet	92.78	93.58	92.42
GooLeNet	93.54	93.90	93.17
DenseNet	93.87	94.59	93.90
DenseNet-II	94.55	95.60	95.36

Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis



Algorithms	Precision	Sensitivity	F-Measure	Class
SVM	0.98	0.94	0.96	Benign
	0.97	0.99	0.98	Malignant
Random Forests	0.96	0.94	0.95	Benign
	0.97	0.98	0.97	Malignant
Logistic Regression	0.98	0.91	0.94	Benign
	0.95	0.99	0.97	Malignant
Decision Tree	0.94	0.92	0.93	Benign
	0.96	0.97	0.96	Malignant
K-NN	0.92	0.91	0.91	Benign
	0.95	0.96	0.95	Malignant

Implementation

Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach

Methodology

01
Input Dataset

02
Model Building

03
Model Training and Test

04
Results

Wisconsin (WDBC)

Dataset Used

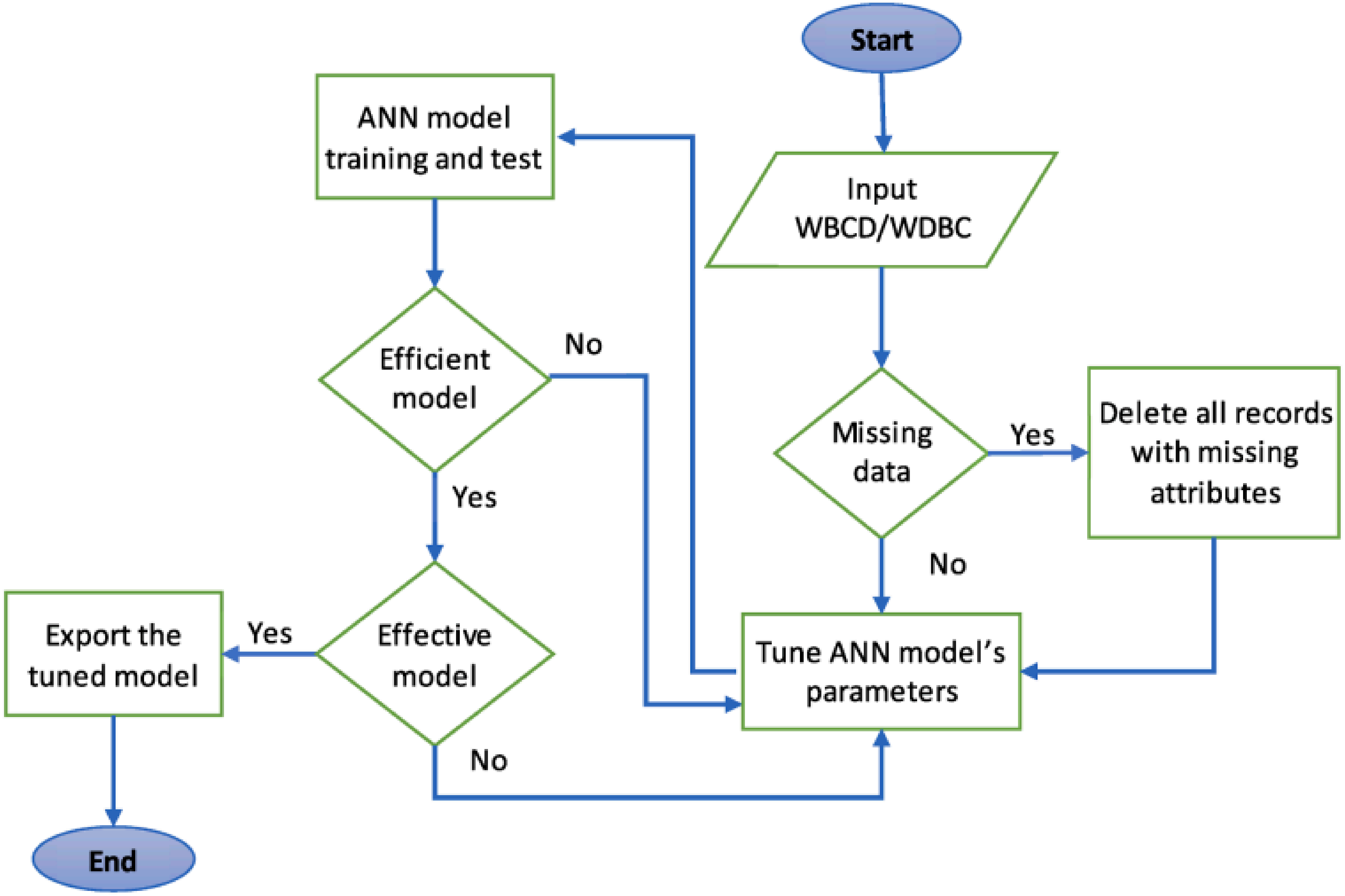
Contains the extracted features of the acquired images after the image pre-processing, image segmentation, and feature extraction steps have been conducted

Instances	Features	Benign	Malignant
569	30	212	357

```
Data columns (total 33 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   id                       569 non-null    int64  
 1   diagnosis                569 non-null    object 
 2   radius_mean              569 non-null    float64
 3   texture_mean             569 non-null    float64
 4   perimeter_mean           569 non-null    float64
 5   area_mean                569 non-null    float64
 6   smoothness_mean          569 non-null    float64
 7   compactness_mean         569 non-null    float64
 8   concavity_mean           569 non-null    float64
 9   concave points_mean      569 non-null    float64
 10  symmetry_mean            569 non-null    float64
 11  fractal_dimension_mean   569 non-null    float64
 12  radius_se                569 non-null    float64
 13  texture_se               569 non-null    float64
 14  perimeter_se             569 non-null    float64
 15  area_se                  569 non-null    float64
 16  smoothness_se            569 non-null    float64
 17  compactness_se           569 non-null    float64
 18  concavity_se             569 non-null    float64
 19  concave points_se        569 non-null    float64
 20  symmetry_se              569 non-null    float64
 21  fractal_dimension_se     569 non-null    float64
 22  radius_worst             569 non-null    float64
 23  texture_worst            569 non-null    float64
 24  perimeter_worst          569 non-null    float64
 25  area_worst               569 non-null    float64
 26  smoothness_worst         569 non-null    float64
 27  compactness_worst        569 non-null    float64
 28  concavity_worst          569 non-null    float64
 29  concave points_worst     569 non-null    float64
 30  symmetry_worst           569 non-null    float64
 31  fractal_dimension_worst  569 non-null    float64
 32  Unnamed: 32              0 non-null      float64
dtypes: float64(31), int64(1), object(1)
memory usage: 146.8+ KB
```

Wisconsin (WDBC)

Flowchart



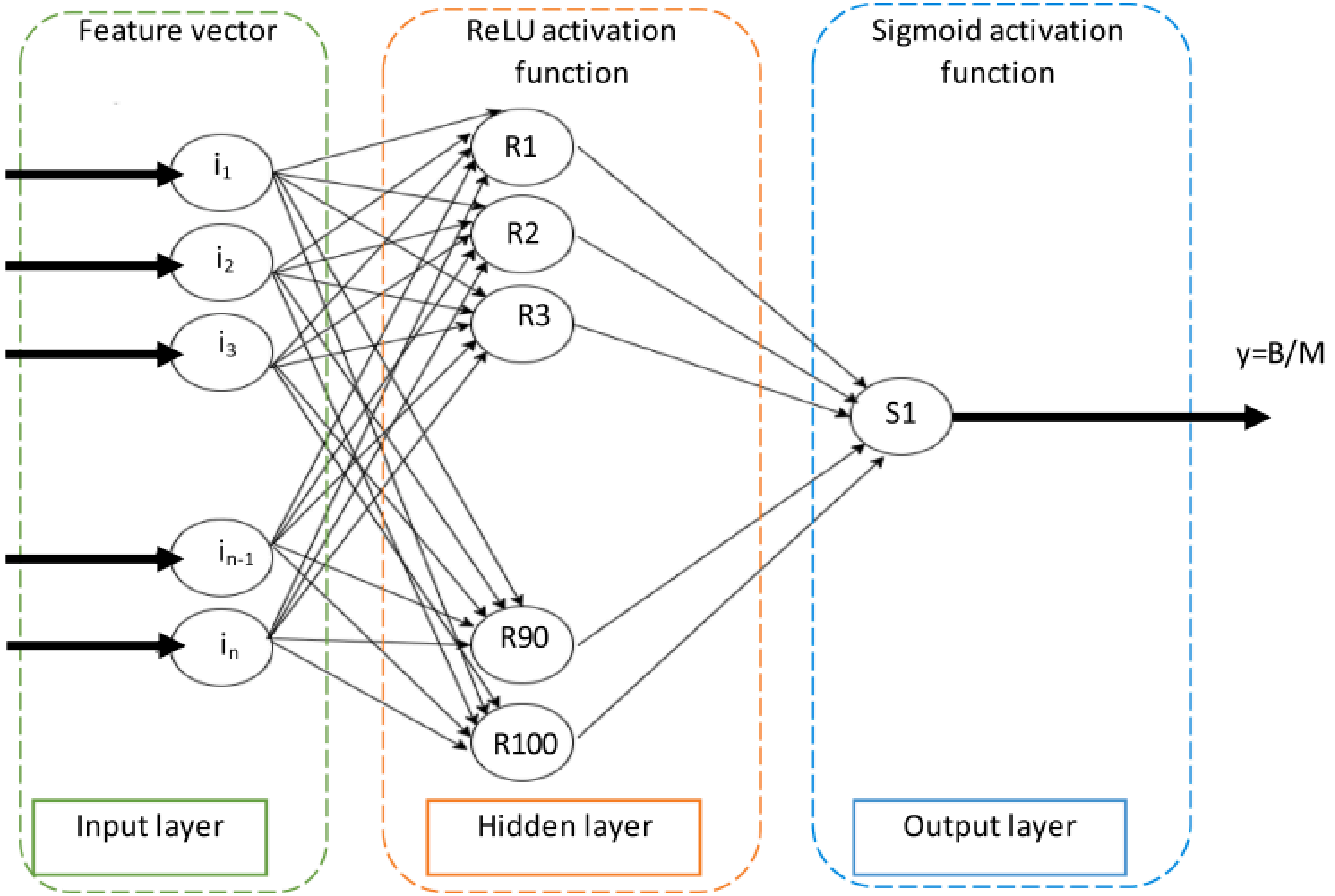
Start
Input WBCD/WDBC
Missing data
Yes
Delete all records with missing attributes
No
Tune ANN model's parameters
ANN model training and test
Efficient model
No
Yes
Effective model
Yes
No
Export the tuned model
End

Preprocessing

- Irrelevant features are removed
- Feature scaling
- Binary encoding of the target variable
- Oversampling of the minor class

Model

A Shallow ANN with one hidden layer having 100 neurons



Workings

Minimize the loss function by recalibrating the weights of the hidden layer nodes using the back-propagation algorithm until the convergence criteria are met.

Once the loss function is close to zero, the model becomes fully trained and the validation/testing dataset should be used to evaluate its performance.

01. Weights were randomly initialized

02. Calculation of the 100 neurons

03. Calculation of the output of the output layer neuron

Initialize the weights W (w_1, w_2, w_3, ….w_n) and bias b randomly.

Calculate the input to the R^{th} neuron in the hidden layer using Eq. (1):

$$R^{th}\ \text{input} = \sum_{i=1}^{n} w_{iR} i_i + b \quad (1)$$

Calculate the R^{th} neuron output using Eqs. (2) and (3):

$$R^{th}\ \text{output} = \text{ReLU}\left(\sum_{i=1}^{n} w_{iR} i_i + b\right) \quad (2)$$

$$\text{ReLU}(x) = \max(0, x) \quad (3)$$

Calculate the output of the output layer neuron S1 using Eqs. (4) and (5):

$$S1 = \sigma\left(\sum_{i=1}^{100} R_i \text{output}\right) \quad (4)$$

$$\sigma(x) = 1/(1 + e^{-x}) \quad (5)$$

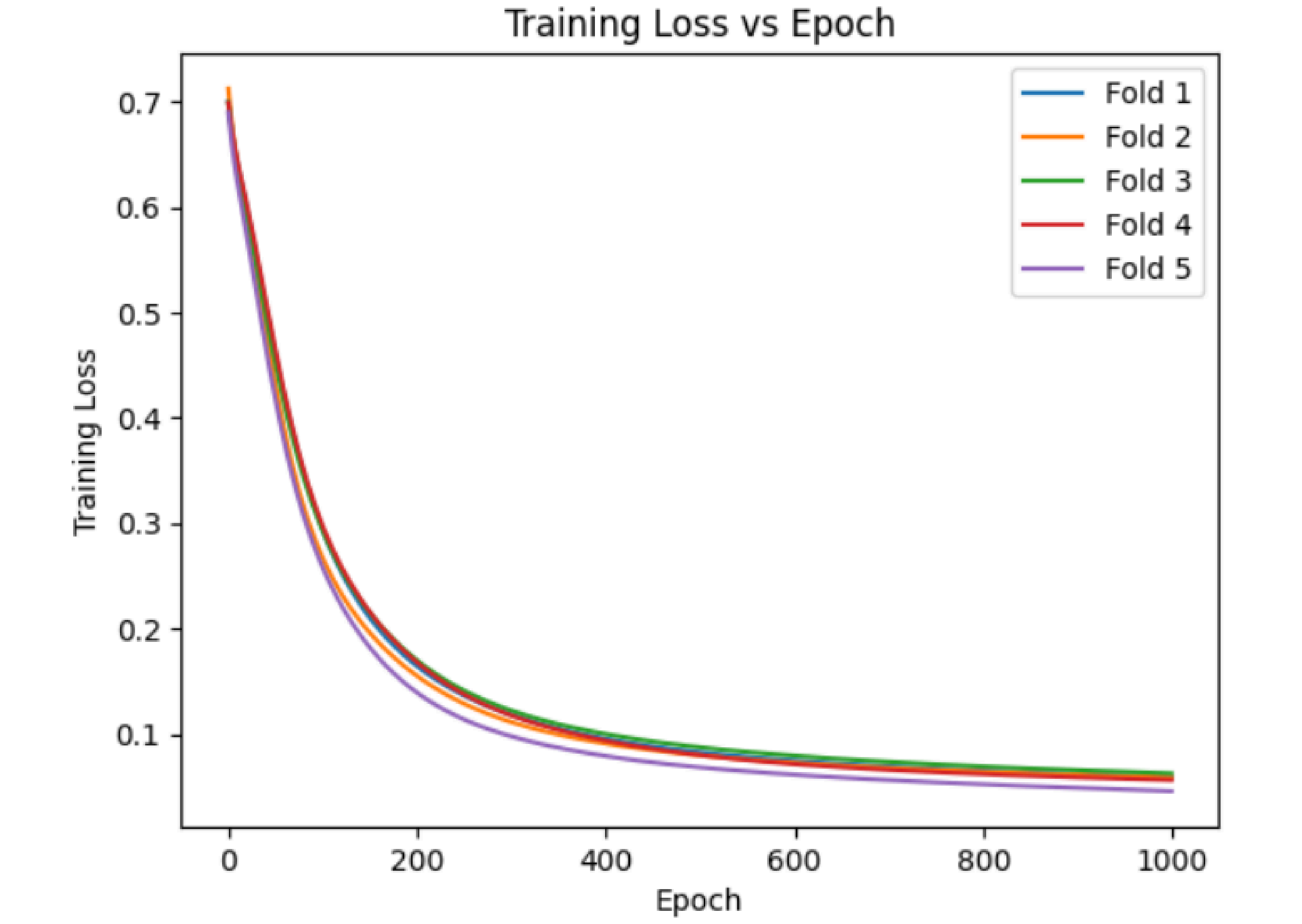
Techniques Used

- 5-Fold cross-validation
- Adam Optimizer
- Batch gradient
- Loss = Binary Cross Entropy

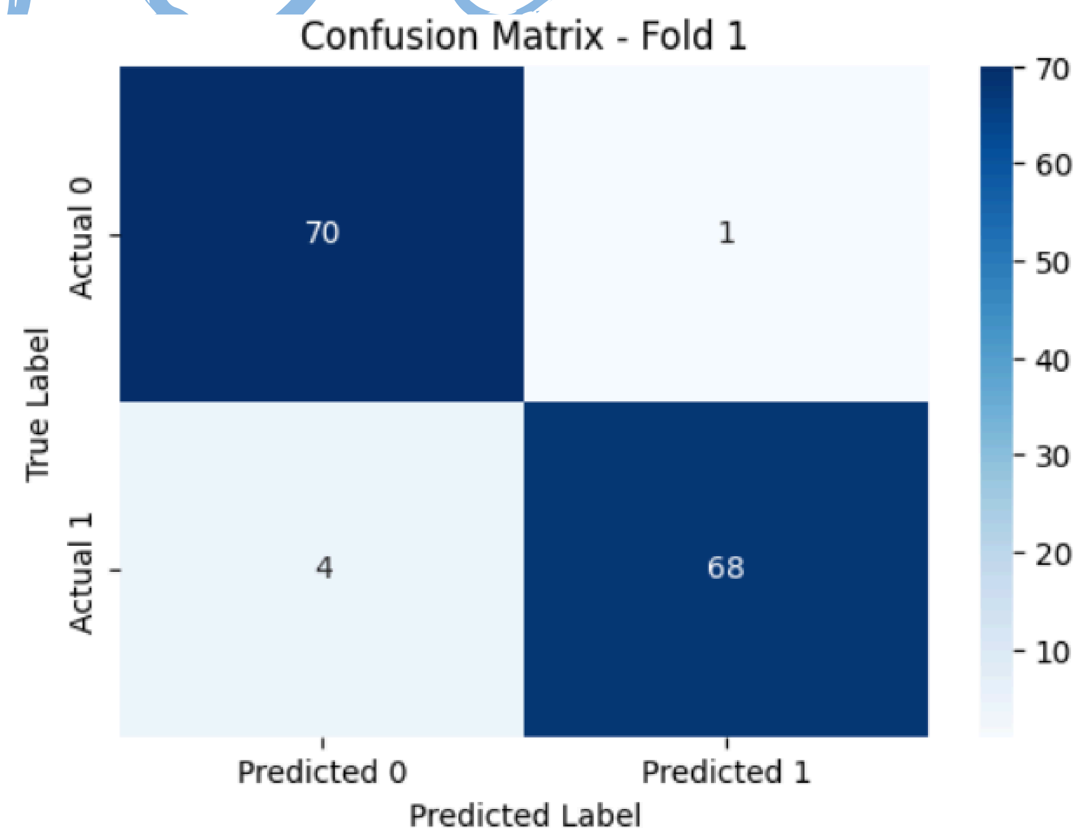
$$loss = y \log(S1) + (1 - y)\log(1 - S1)$$

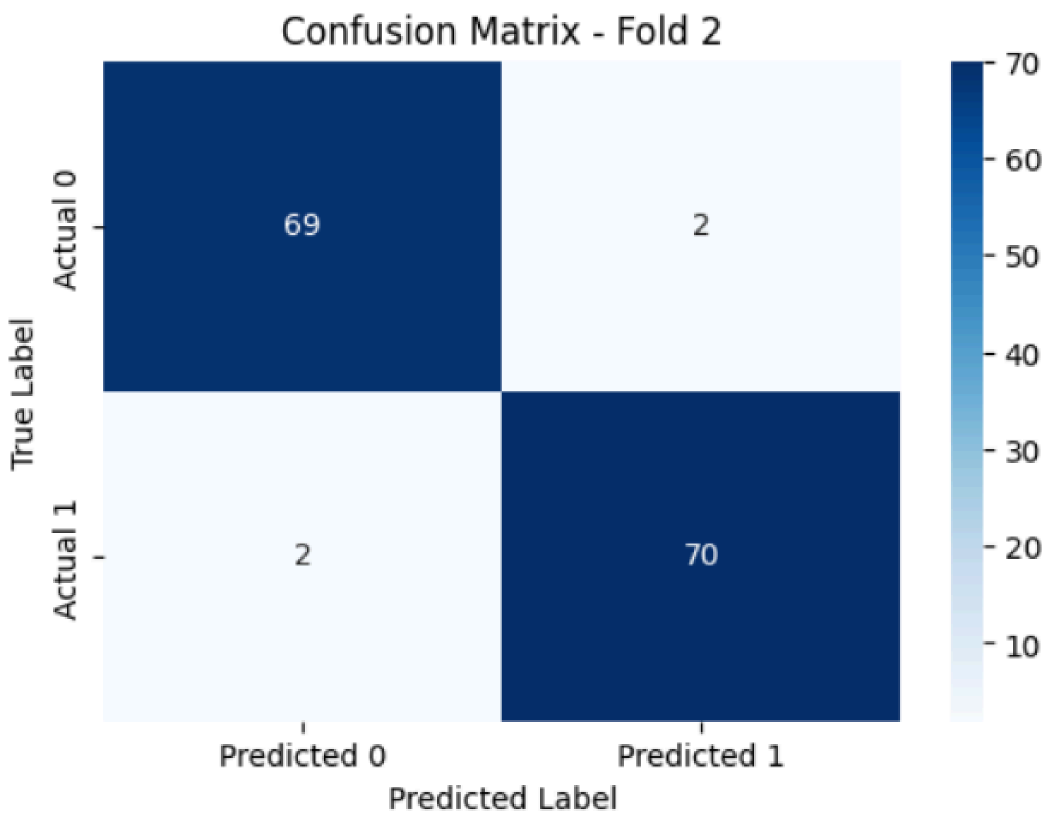
Result after 1000 epochs

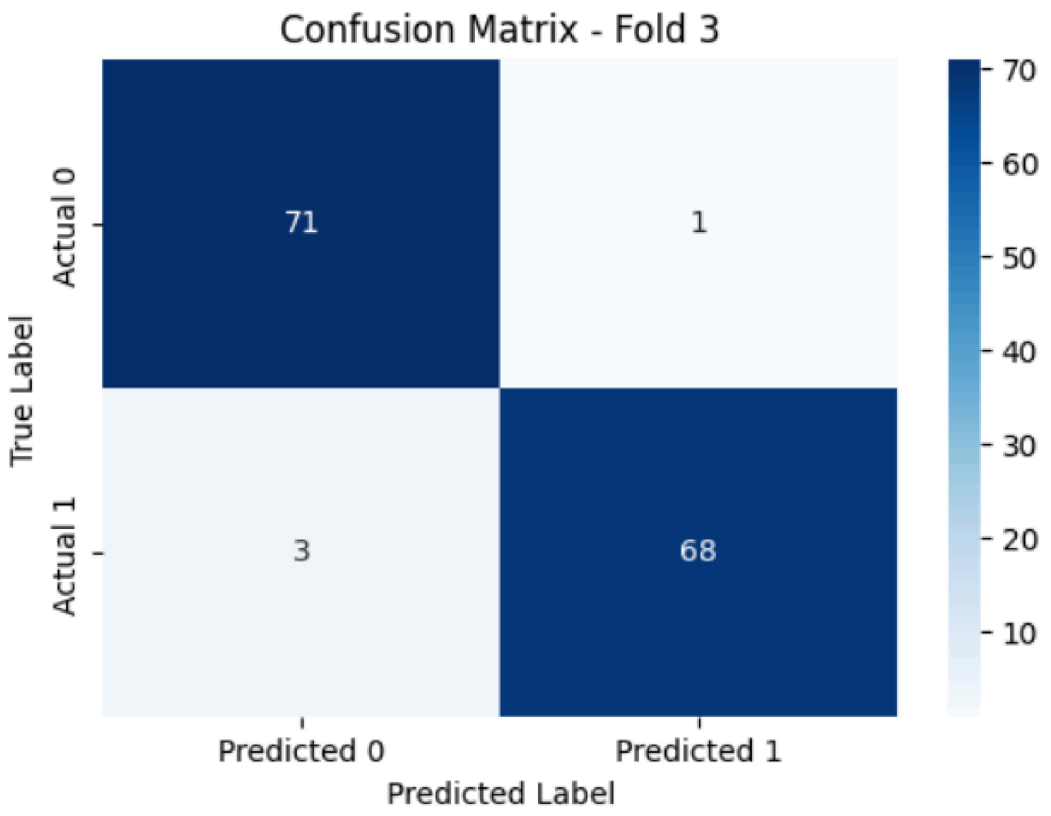
Mean Accuracy: 0.9663547719885747
Mean Sensitivity: 0.9691705790297339
Mean Specificity: 0.9634976525821596

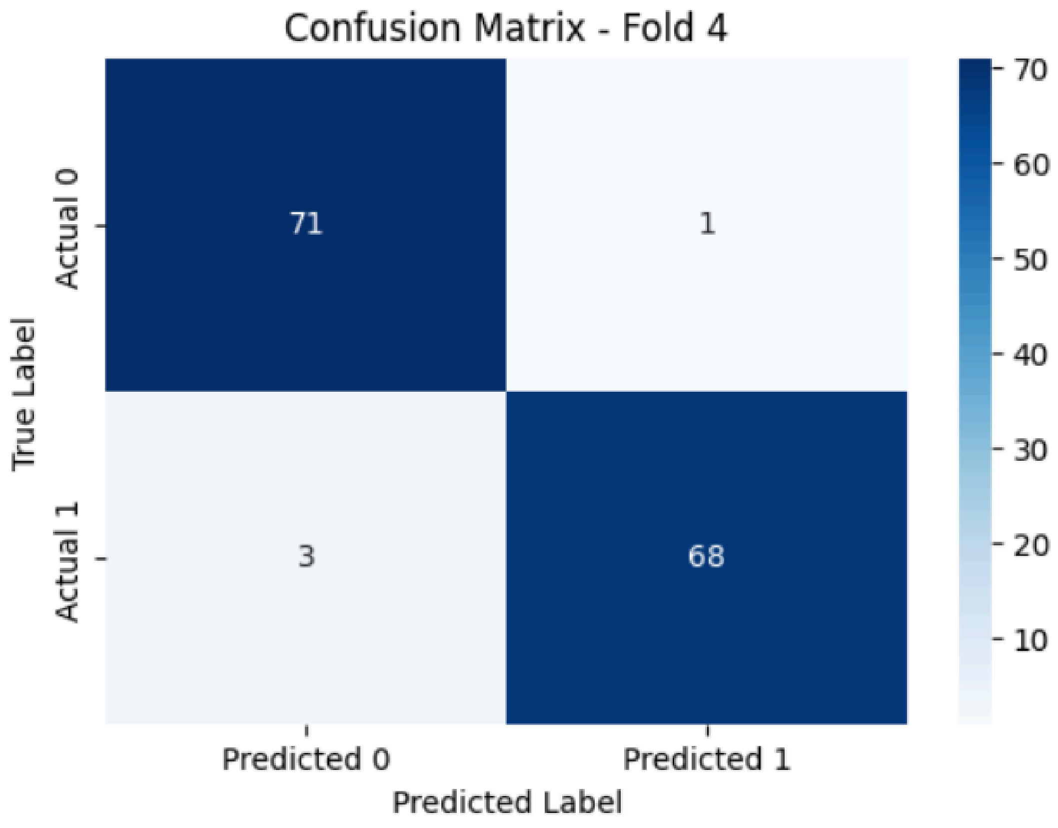


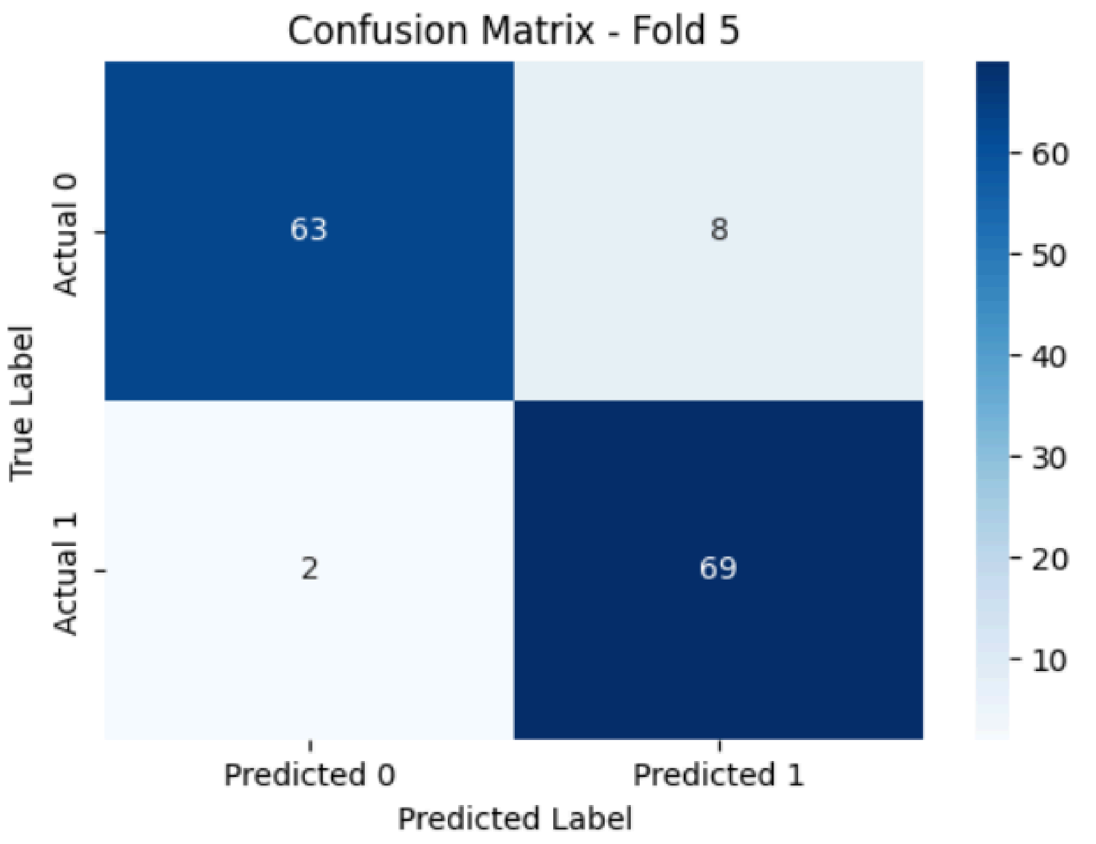
CONFUSION MATRIX











Refrences

Project Report	CLICK HERE
Paper Summaries	CLICK HERE
Notebook	CLICK HERE

Refrences

@article{LI2019347,
title = {Benign and malignant classification of mammogram images based on deep learning},
journal = {Biomedical Signal Processing and Control},
volume = {51},
pages = {347-354},
year = {2019},
issn = {1746-8094},
doi = {https://doi.org/10.1016/j.bspc.2019.02.017},
url = {https://www.sciencedirect.com/science/article/pii/S1746809419300618},
author = {Hua Li and Shasha Zhuang and Deng-ao Li and Jumin Zhao and Yanyun Ma},
keywords = {Breast cancer, Mammogram images, Deep learning, Inception structure, DenseNet-II neural network model}
}

@article{ELHOUBY2021102954,
title = {Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks},
journal = {Biomedical Signal Processing and Control},
volume = {70},
pages = {102954},
year = {2021},
issn = {1746-8094},
doi = {https://doi.org/10.1016/j.bspc.2021.102954},
url = {https://www.sciencedirect.com/science/article/pii/S1746809421005516},
author = {Enas M.F. {El Houby} and Nisreen I.R. Yassin},
keywords = {Breast cancer, Classification, Computer-aided diagnosis, Convolution neural network, Deep learning, Mammogram}

}

Refrences


```
@article{NAJI2021487,
title = {Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis},
journal = {Procedia Computer Science},
volume = {191},
pages = {487-492},
year = {2021},
note = {The 18th International Conference on Mobile Systems and Pervasive Computing (MobiSPC), The 16th International Conference on Future Networks and Communications (FNC), The 11th International Conference on Sustainable Energy Information Technology},
issn = {1877-0509},
doi = {https://doi.org/10.1016/j.procs.2021.07.062},
url = {https://www.sciencedirect.com/science/article/pii/S1877050921014629},
author = {Mohammed Amine Naji and Sanaa El Filali and Kawtar Aarika and EL Habib Benlahmar and Rachida Ait Abdelouhahid and Olivier Debauche},
keywords = {Breast Cancer, Prediction, Diagnostic, SVM, Random Forest, Logistic regression, C4.5, k-NN, Accuracy, Precision}
}

@article{ALSHAYEJI2022103141,
title = {Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach},
journal = {Biomedical Signal Processing and Control},
volume = {71},
pages = {103141},
year = {2022},
issn = {1746-8094},
doi = {https://doi.org/10.1016/j.bspc.2021.103141},
url = {https://www.sciencedirect.com/science/article/pii/S1746809421007382},
author = {Mohammad H. Alshayeji and Hanem Ellethy and Sa'ed Abed and Renu Gupta}
}
```

Thank you!

HARSH (20) and ROHAN (47)