BREAST CANCER DIAGNOSIS USING MACHINE LEARNING AND DEEP LEARNING ALGORITHM

Project Report submitted by

Elton Arthur Pereira	***4NM20CS068***
Pranav P Shetty	***4NM20CS133***
Ryan D'Souza	***4NM20CS148***
Shashank K T	***4NM20CS164***

Under the Guidance of

Ms.Jayapadmini Kanchan

Assistant Professor
Dept. of CSE, NMAMIT, NITTE

In partial fulfillment of the requirements for the

award of the Degree of

Bachelor of Engineering in Computer Science and Engineering

from

Visvesvaraya Technological University, Belagavi

Department of Computer Science Engineering
NMAM Institute of Technology,

Nitte- 574110

(An Autonomous Institution affiliated to VTU, Belagavi)

APRIL 2024

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the project work entitled

““Breast Cancer Diagnosis Using Machine Learning and Deep Learning Algorithm”

is a bonafide work carried out by

Elton Arthur Pereira (4NM20CS068) , Pranav P Shetty(4NM20CS133),

Ryan D’Souza(4NM20CS148) and Shashank K T (4NM20CS164).

in partial fulfilment of the requirements for the award of

Bachelor of Engineering Degree in Computer Science and Engineering

prescribed by ***Visvesvaraya Technological University, Belagavi*** *during*

the year ***2023-2024****.It is certified that all corrections/suggestions*

indicated for Internal Assessment have beenincorporated in the report

deposited in the departmental library.

The project report has been approved as it satisfies the academic requirements in respect of

the project work prescribed for the Bachelor of Engineering Degree.

Signature of the Guide	***Signature of the HOD***	***Signature of the Principal***

Semester End Viva Voce Examination

Name of the Examiners	Signature with Date
1.	
2.	

ACKNOWLEDGEMENT

The satisfaction that accompanies the completion of any task would be incomplete without the mention of all the people, without whom this endeavour would have been a difficult one to achieve. Their constant blessings, encouragement, guidance and suggestions have been a constant source of inspiration

First and foremost, my gratitude to my project guide**, Ms. Jayapadmini Kanchan** for her constant guidance throughout the course of this project and for the valuable suggestions.

I also take this opportunity to express a deep sense of gratitude to the **Project Coordinators** for their valuable guidance and support.

I acknowledge the support and valuable inputs given by, **Dr. Jyothi Shetty** the Head of the Department, Computer Science and Engineering, NMAMIT, Nitte.

My sincere thanks to our beloved principal, **Dr. Niranjan N Chiplunkar** for permitting us to carry out this project at our college and providing us with all needed facilities.

Finally, thanks to staff members of the Department of Computer Science and Engineering and our friends for their honest opinions and suggestions throughout the course of our project .

Elton Arthur Pereira(4NM20CS068)
Pranav P Shetty(4NM20CS133)
Ryan D’Souza (4NM20CS148)
Shashank K T (4NM20CS164)

ABSTRACT

This abstract introduces an innovative approach for breast cancer diagnosis through the application of machine learning (ML) and deep learning (DL) algorithms. Specifically, K-Nearest Neighbors (KNN), Random Forest (RF),Support Vector Machine (SVM) and Artificial Neural Network (ANN) methodologies are employed to enhance diagnostic accuracy. Leveraging a rich dataset comprising diverse breast cancer samples, the study aims to effectively differentiate between malignant and benign cases. Additionally, deep learning techniques are integrated to further augment diagnostic capabilities. The research methodology involves pre-processing the dataset to extract relevant features crucial for classification. Subsequently, the models are trained using the KNN, RF, and SVM algorithms, allowing for comprehensive exploration of various ML techniques. Performance evaluation metrics including accuracy, precision, recall, and F1-score are employed to assess the efficacy of each algorithm. Results from extensive experimentation showcase promising outcomes, indicating the potential of the proposed approach in facilitating precise and timely breast cancer diagnosis. The fusion of traditional ML algorithms with cutting-edge deep learning methodologies presents a holistic framework for improving diagnostic accuracy. The study's findings underscore the importance of leveraging advanced computational techniques in healthcare, particularly in critical areas like cancer diagnosis, ultimately contributing to enhanced patient care and outcomes.

CONTENT

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Breast cancer is one of the most prevalent forms of cancer worldwide, posing a significant threat to women's health. Timely and accurate diagnosis is crucial for effective treatment planning and improving patient outcomes. In recent years, the integration of machine learning (ML) and deep learning (DL) algorithms has emerged as a promising approach to enhance the accuracy and efficiency of breast cancer diagnosis.

This study aims to investigate the potential of ML and DL algorithms in breast cancer diagnosis, specifically focusing on methodologies such as K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) . By leveraging these algorithms, the research seeks to develop a robust diagnostic framework capable of accurately classifying breast cancer cases.

The research utilizes a comprehensive dataset comprising various breast cancer samples, encompassing diverse clinical and pathological features. Through feature extraction and selection techniques, relevant information is extracted from the dataset to train and evaluate the ML and DL models.

By integrating traditional ML algorithms with more complex DL methodologies, this study endeavors to leverage the complementary strengths of each approach to improve diagnostic accuracy. The findings of this research have the potential to significantly impact clinical practice by providing clinicians with advanced tools for more precise and timely breast cancer diagnosis, ultimately leading to improved patient care and outcomes.

Overall, machine learning and deep learning techniques offer powerful tools for analysing breast cancer data, improving diagnostic accuracy, predicting patient outcomes, and advancing our understanding of the disease.

CHAPTER 2

LITERATURE SURVEY

A literature survey in the context of breast cancer diagnosis using machine learning (ML) and deep learning (DL) algorithms, along with image processing, involves an exploration of existing research and studies related to the application of these technologies in medical imaging for breast cancer detection.

In [2], Zeyad Q. Habeeb *et al.* proposed a model that Challenges in defining features on mammograms limit accuracy. Deep learning identifies crucial features, but research is constrained by small datasets. Innovative methods, including MRI-CT fusion, show promise. Pre-trained CNNs, especially InceptionV3, outperform in breast cancer detection. datasets have been used to train and test the proposed framework have been described, and the architectures of the CNN have been presented(CBIS-DDSM dataset ,coco and INbreast dataset).

In [3], V. Durga Prasad Jasti *et al.* proposed a model in which Image preprocessing using geometric mean filter, Feature extraction using AlexNet, Feature selection using relief algorithm, Classification using LS-SVM . The presented studies employ diverse methodologies for enhancing breast cancer detection and analysis. Data mining processes are utilized to improve statistical analysis from cancer log data, offering potential benefits to future research.
Deep learning, ensemble methods, and convolutional neural networks contribute to accurate diagnosis and risk prediction.

In [4], Prannoy Giri et al. proposed Mammography pre-processing, segmentation, and feature extraction are followed by edge detection, texture analysis, and classification using a Neural Network, addressing challenges in training complexity. Detecting macrocalcification in dense breast tissue is challenging due to white pixel similarity. Machine learning aids cancer diagnosis, enhancing traditional methods, especially in pattern recognition.

In [5], Shubhangi A. Joshi et al. proposed a model where bibliometric analysis focuses on breast cancer detection using machine learning techniques. It emphasizes automated diagnostic techniques to aid pathologists. The focus is on machine learning, especially deep learning, to automate the process. A key dataset, BreaKHis, and various techniques like transfer learning and convolutional neural networks are discussed. The bibliometric analysis reveals global research trends, collaborations, and top contributors in the field, emphasizing the significance of automated tools in breast cancer diagnosis.

In [6], Sushovan Chaudhury *et al.* proposed a model where it employs image processing and machine learning for breast cancer detection, using mammogram images, CLAHE for enhancement, and classifiers (fuzzy SVM, Bayesian, random forest) for categorization. explores image enhancement, segmentation, and classification techniques in mammography. Adaptive histogram equalization, difference pictures, and region growth enhance images. Various segmentation methods, including adaptive thresholding and -means clustering, improve lesion detection. Image classification involves SVM, perceptron, and Bayesian classifier for accurate diagnosis in breast cancer detection.

In [7], V Harvind Viswanath *et al.* proposed a model where it aims to classify mammogram images into normal, benign, or malignant conditions. It uses preprocessing, feature extraction, dimensionality reduction, feature selection, and three classifiers (SVM, K-NN, RF). automated mammogram classification system using SVM, K-NN, and RF. RF achieves the highest accuracy, emphasizing the significance of image pre-processing for improved results. Future work aims for a fully automatic system to aid radiologists in mammographic interpretation.

In [8], Neela A G *et al* . machine learning (ML) in biomedical applications, emphasizing supervised and unsupervised learning. It explores neural network structures, learning types, and tasks like regression and clustering. Challenges include data quality and preprocessing for effective ML implementation. explores breast cancer detection through data mining, image processing, and machine learning techniques. It compares algorithms for classifying benign and malignant cancers, focusing on error estimation and model selection. The goal is early detection to prevent cancer spread.

In [9], Sri Hari Nallamala *et al.* Logistic Regression predicts binary outcomes based on independent variables, using a derived equation. SVM is a supervised algorithm for classification, finding a hyperplane to separate classes in n-dimensional space. K-Nearest Neighbour (KNN) is versatile for classification and regression, evaluated based on interpretability, computation time, and predictive power.

This exertion is projected an ensemble voting ML technique for analysis breast cancer. And we see in diagram that the suggested strategy has acquired the 98.50% precision. For this work, we took 16 features only into the consideration of breast cancer analysis.

In [10], Monika Tiwari *et al.* employs machine learning and deep learning algorithms for breast cancer diagnosis, utilizing the Wisconsin Breast Cancer Dataset. It details pre-processing steps and the application of Logistic Regression, KNN, SVM, and deep learning techniques like CNN and ANN to enhance prediction accuracy. explores breast cancer prediction using machine learning (SVM, Random Forest) and deep learning (CNN, ANN) on the Wisconsin dataset. SVM and Random Forest achieve 96.5% accuracy, while CNN and ANN outperform with 97.3% and 99.3%, highlighting deep learning's superior efficiency.

In [11] , Pradeeba.R et al . This paper proposes an efficient breast cancer detection system using FCM segmentation on digital mammograms. The architecture involves image acquisition, pre-processing for intensity improvement and noise removal using a median filter, and segmentation through FCM to identify early-stage tumors accurately, reducing the risk associated with breast cancer.

i: evaluates machine learning algorithms for breast cancer detection from fine needle aspirate digitized images. Deep learning shows promising results, but challenges include

CHAPTER 3

PROBLEM DEFINITION

To harness machine learning techniques for breast cancer diagnosis, striving to enhance accuracy and early detection while ensuring interpretability, ultimately improving patient outcomes and healthcare efficiency.

OBJECTIVE

To implement an automatic breast cancer detection based on statistical dataset, and a pre-trained model.
To develop an accurate, efficient, and reliable system that enhances the early detection and diagnosis of breast cancer.
To improve patient outcomes and assist healthcare professionals in making more informed and timely decisions

CHAPTER 4

SYSTEM DESIGN

The system design for breast cancer diagnosis integrating machine learning (ML) and deep learning (DL) algorithms, specifically K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) necessitates a meticulous approach to data preprocessing, model implementation, and evaluation.

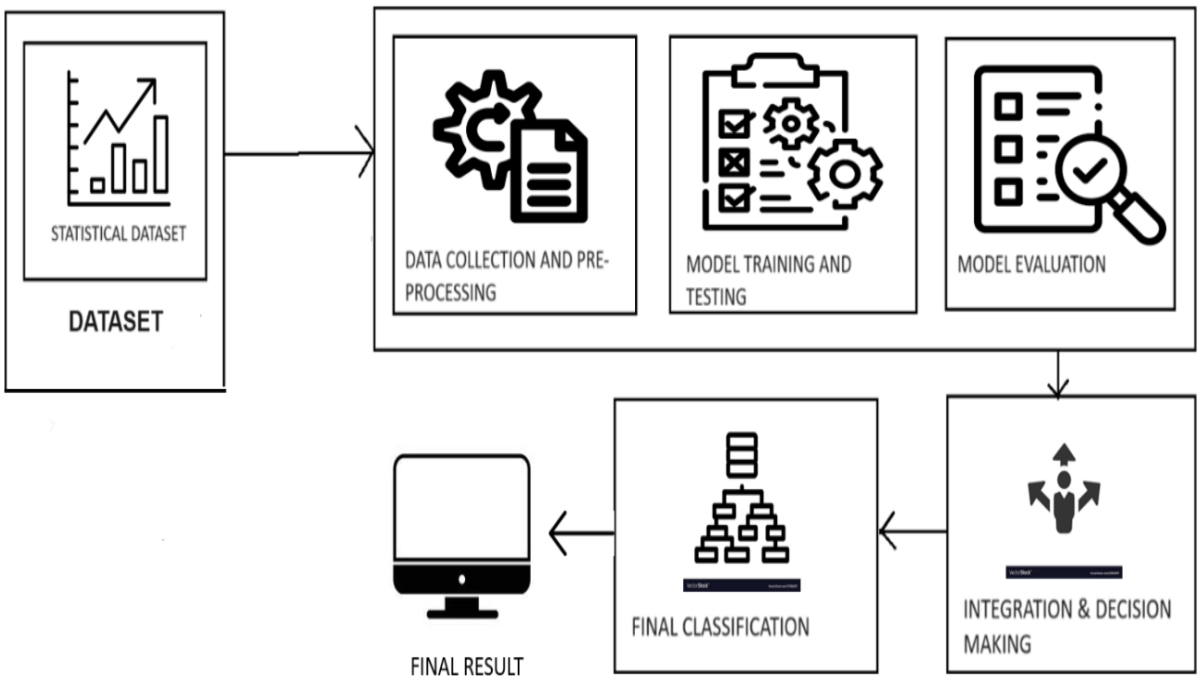


Fig.1-Architecture Design Diagram

1. Data Collection and Preprocessing:
    - Component**:** Data Collection and Preprocessing Module
    - Description**:** Gathers mammogram datasets, preprocesses raw images, and ensures data quality for input to the ensemble.

2. Individual Machine Learning Models:
    - Component: ML Model 1, ML Model 2, ..., ML Model N
    - Description: Implements diverse machine learning algorithms
3. Individual Deep Learning Models:
    - Component: DL Model 1, DL Model 2, ..., DL Model N
    - Description: Implements various deep learning architectures to capture intricate patterns and features in mammogram images.
4. Ensemble Module:
    - Component: Ensemble Model
    - Description: Combines predictions from individual machine learning and deep learning models to produce a final, more accurate prediction. Common ensemble methods include:
        - Voting: Combines predictions through voting mechanisms (e.g., majority voting).
        - Averaging: Takes the average of predicted probabilities or class labels.
        - Stacking: Trains a meta-model to combine predictions from individual models.
5. Integration Layer:
    - Component: Integration Module
    - Description: Manages the flow of predictions between individual models and the ensemble module.
6. Output Module:
    - Component: Diagnostic Report Module
    - Description: Generates detailed diagnostic reports based on the final prediction from the ensemble model.
7. User Interface:
    - Component: UI Module
    - Description: Provides an interface for healthcare professionals to input data, visualize results, and interpret diagnostic reports.
8. Database Management:
    - Component: Database Module

- Description**:** Manages a secure database to store patient data, model parameters, and diagnostic results.

9. Security and Compliance:
    - Component**:** Security Module
    - Description**:** Ensures HIPAA compliance, implements encryption, and enforces secure access controls.
10. Scalability and Performance:
    - Component**:** Optimization Module
    - Description**:** Optimizes models, algorithms, and ensemble techniques for better performance and scalability.

USE CASE DIAGRAM

A use case diagram for breast cancer diagnosis using machine learning and deep learning algorithms would illustrate the various actors, such as users and systems, and the interactions between them.

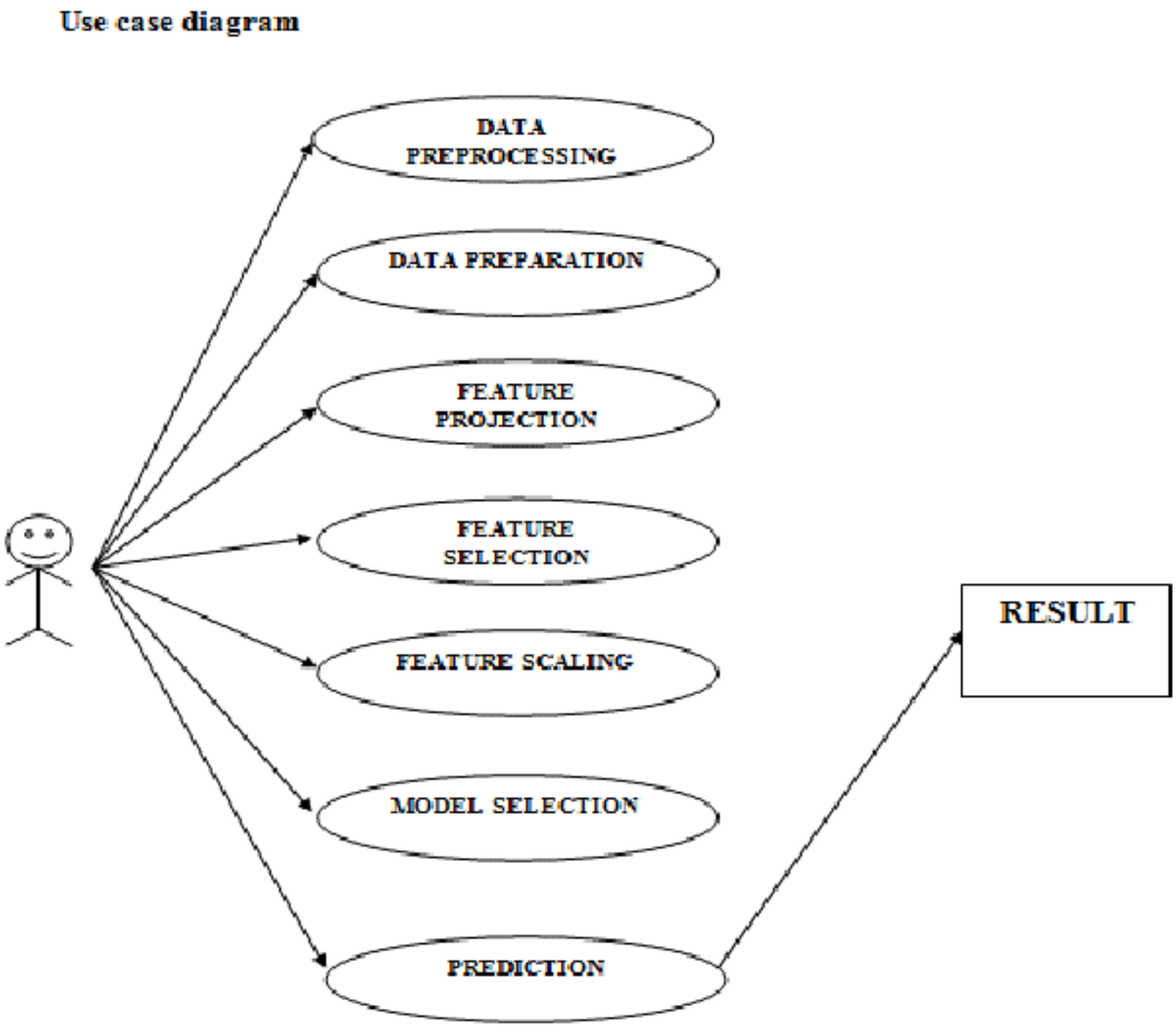


Fig.2- Use case diagram

SEQUENCE DIAGRAM

In the sequence diagram for breast cancer diagnosis, the process begins with data collection, including medical history and imaging. The data is pre-processed for analysis. Machine learning and deep learning algorithms are then trained and tested.

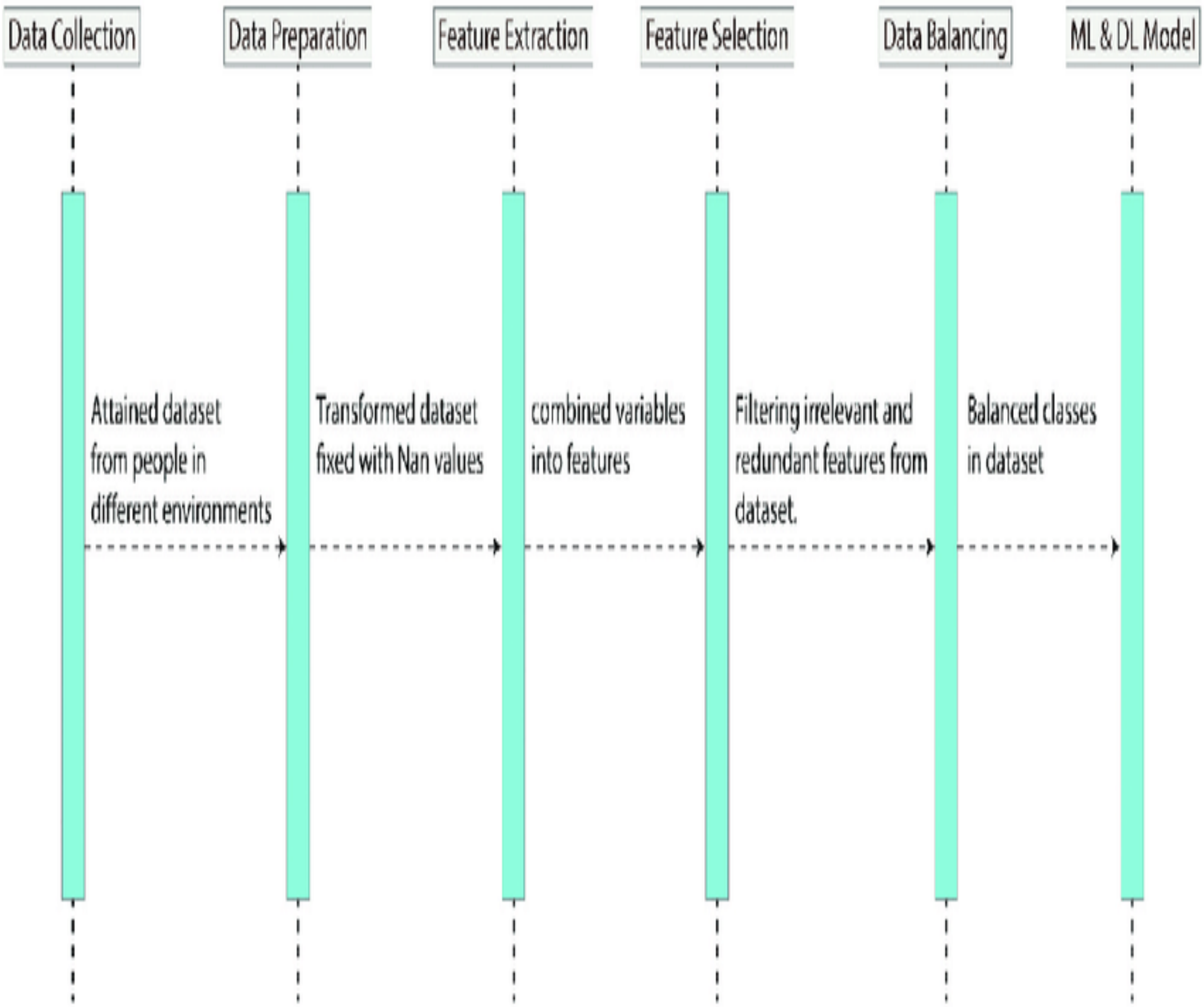


Fig.3- Sequence diagram

CHAPTER 5

SYSTEM REQUIREMENTS SPECIFICATION

1. Hardware Requirements:

- Processor: Ryzen 5 4800H
- GPU: GTX 1650ti 4GB 8GB VRAM.
- Memory: RAM (16GB)

2. Software Requirements:

- Deep Learning Frameworks.
- Library : Scikit-Learn
- Programming Language: Python 3.7 .
- IDE- Jupyter notebook

CHAPTER 6

IMPLEMENTATION

In breast cancer analysis using machine learning and deep learning techniques, you'll typically use a combination of libraries for data manipulation, visualization, model building, and evaluation. Here are some commonly used libraries:

IMPORT LIBRARIES:

```
#Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud
plt.style.use('ggplot')
sns.set_style('whitegrid')
from statsmodels.graphics.gofplots import qqplot
from scipy.stats import shapiro


from sklearn.compose import ColumnTransformer
from sklearn.preprocessing import StandardScaler, MinMaxScaler
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import cross_validate


from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.neighbors import KNeighborsClassifier


from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score
from sklearn.metrics import roc_auc_score, roc_curve
from sklearn.metrics import precision_recall_curve, average_precision_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.model_selection import train_test_split
```

Fig.4- Code for importing libraries

DATASET:

One commonly used dataset for breast cancer analysis using machine learning and deep learning techniques is the Breast Cancer Wisconsin (Diagnostic) dataset. This dataset is publicly available and widely used for research and educational purposes.

	Age	T Stage	N Stage	6th Stage	differentiate	Grade	A Stage	Tumor Size	Estrogen Status	Progesterone Status	Regional Node Examined	Regional Node Positive	Survival Months	Status
0	68	T1	N1	IIA	Poorly differentiated	3	Regional	4	Positive	Positive	24	1	60	Alive
1	50	T2	N2	IIIA	Moderately differentiated	2	Regional	35	Positive	Positive	14	5	62	Alive
2	58	T3	N3	IIIC	Moderately differentiated	2	Regional	63	Positive	Positive	14	7	75	Alive
3	58	T1	N1	IIA	Poorly differentiated	3	Regional	18	Positive	Positive	2	1	84	Alive
4	47	T2	N1	IIB	Poorly differentiated	3	Regional	41	Positive	Positive	3	1	50	Alive

Fig.5- Dataset

DATASPLIT AND FEATURE SCALING:

In breast cancer analysis using machine learning and deep learning techniques, it's common to split the data into training and testing sets for model training and evaluation. Additionally, feature scaling is often applied to ensure that features have similar scales, which can improve the performance of many machine learning algorithms.

```
#Data Split and Feature Scaling
X= SS.iloc[:,1 :-1].values
y = SS.iloc[:, -1].values
```

```
cat_cols.remove('Status')
cat_cols
```

```
['T_Stage',
 'N_Stage',
 '6th_Stage',
 'differentiate',
 'Grade',
 'A_Stage',
 'Estrogen_Status',
 'Progesterone_Status']
```

```
X= (X - np.min(X))/(np.max(X)- np.min(X))
```

```
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.10,random_state=1)
```

```
skf = StratifiedKFold(n_splits = 5,
                      shuffle = True,
                      random_state = 123)
```

Fig.6- Code for datasplit and feature scaling

ML ALGORITHM:

For breast cancer analysis using machine learning algorithms with a statistical dataset, you can follow a similar approach as mentioned before. Here's how you can apply machine learning algorithms to a statistical dataset:

```python
#ML algorithms
def plot_metrics(metrics_train:dict, metrics_test:dict, metrics:list)->None:


    df_metrics = pd.DataFrame(index = metrics,
                              data = {'Train':[metrics_train[metric] for metric in metrics],
                                      'Test':[metrics_test[metric] for metric in metrics]})


    n = len(df_metrics.index)
    x = np.arange(n)


    width = 0.25

    fig,ax = plt.subplots(figsize=(6,4))

    rects1 = ax.bar(x-width, df_metrics.Train, width=width, label='Train',linewidth=1.6,edgecolor='black',color='blue')

    rects2 = ax.bar(x, df_metrics.Test, width=width, label='Test',linewidth=1.6, edgecolor='black', color = 'red')

    ax.set_title('Metrics',fontsize=12, fontweight='bold')
    ax.set_ylabel('Score',fontsize=10, fontweight='bold')
    ax.spines['top'].set_visible(False)
    ax.spines['right'].set_visible(False)
    ax.set_xticks(x-0.13)
    ax.set_xticklabels(df_metrics.index, fontsize=10, fontweight='bold')
    ax.legend()

    def autolabel(rects):

        for rect in rects:
            height = rect.get_height()
            ax.annotate('{}'.format(height),
                        xy=(rect.get_x() + rect.get_width() / 2, height-0.005),
                        xytext=(0, 3),  # 3 points vertical offset
                        textcoords="offset points",
                        ha='center', va='bottom', size = 7, weight = 'bold')

    autolabel(rects1)
    autolabel(rects2)
    fig.tight_layout()
    fig.show()
```

Fig.7- Code for ML Algorithm

COMPARE MODELS:

To compare machine learning and deep learning models for breast cancer analysis, we'll consider a few commonly used algorithms from both domains. We'll evaluate their performance using accuracy as the metric. Here's a comparison between kNN, Random Forest, Support Vector Machine (SVM), and a basic Artificial Neural Network (ANN) for breast cancer analysis:

```
#Compare Models
# We create a dictionary that stores the names of the models and their respective metrics.
models_name = {'random_forest':metrics_test_rf,
               'Knn' :metrics_test_knn,
               'SVM':metrics_test_svm}

df_train_metrics = pd.DataFrame.from_dict(models_name,orient='index')
# We order from highest to lowest by the f1 score metric, which is the one we chose because we have unbalanced classes.
df_train_metrics = df_train_metrics.sort_values('recall', ascending = False)

# we visualize the training metrics
fig,ax = plt.subplots(figsize=(9,4.5))
sns.heatmap(df_train_metrics, annot=True, cmap = 'coolwarm', annot_kws = {'fontweight':'bold'},fmt = '.3f', ax = ax)
ax.xaxis.tick_top()
ax.set_ylabel('Models', fontsize = 11, fontweight = 'bold', color = 'blue')
ax.set_title('Metrics train', fontsize = 11, fontweight = 'bold', color = 'blue')
fig.show()
```

Fig.8- Code for comparing the models.

FIND THE BEST MODEL BASED ON ACCURACY:

To find the best model based on accuracy for breast cancer analysis using machine learning and deep learning, you can train multiple models and evaluate their performance using accuracy as the metric.

```
#finds the best model based on accuracy
s=[]
s.append(cv_results_svm['test_accuracy'][4])
s.append(cv_results_rf['test_accuracy'][4])
s.append(cv_results_knn['test_accuracy'][4])
print(s)

print(max(s))

for i in range (0,3):
    if max(s) == s[i]:
        if i == 0:
            print("SVM is the best model")
        elif i == 1:
            print("Random forest is the best model")
        elif i == 2:
            print("KNN is the best model")
```

Fig.9- Code for finding the best model based on accuracy

DATASET

Source: Kaggle

ATTRIBUTES: Age, T-Stage, N-Stage, 6^{th} Stage, Differentiate, Grade, A-Stage, Tumor Size, Estrogen Status, Progesterone Status,
Regional Node Examined, Regional Node Positive, Survival Months, Status

1. Data Loading and Preprocessing:

- Import necessary libraries.
- Check and clean column names.
- Handle missing values.
- Analyze and visualize data distributions, correlations, and relationships among variables.

2. Data Splitting and Feature Scaling:

- Split the data into training and testing sets.
- Scale numerical features to a uniform range (0 to 1).

3. Model Building and Evaluation:

- Use three machine learning algorithms: Support Vector Machine (SVM), Random Forest, k-Nearest Neighbors (KNN) and Artificial Neural Network (ANN).
- Use stratified k-fold cross-validation to evaluate the models. Visualize and compare model performance using various metrics like accuracy, precision, recall, and F1-score.

4, Model comparison:

- Compare the test accuracies of Random Forest, SVM, KNN and ANN.
- Identify which model has the highest accuracy on the test data.
- Visualize and compare the test accuracies of the three models.

5. Choosing the best model:

- Make predictions using these models on the test data.
- Calculate the test accuracies for each model.Compare the test accuracies of all models to determine the best-performing model

CHAPTER 7

RESULT AND ANALYSIS

In the domain of breast cancer diagnosis using machine learning techniques, all three models (SVM, KNN, Random Forest and ANN) demonstrated commendable performance)

Support Vector Machine (SVM)

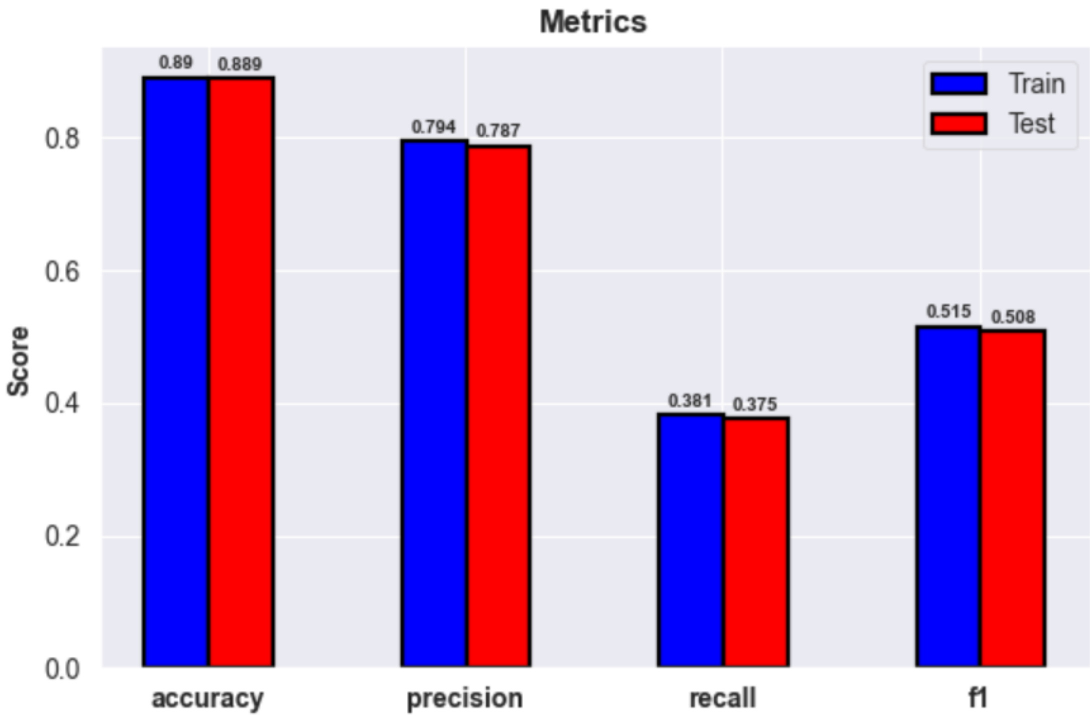


Fig.10- Metrics of SVM

Test Accuracy: 88.9%

Performance Analysis: SVM demonstrated a robust performance with an accuracy of 88.9%. SVM is known for its ability to handle high-dimensional data effectively by finding the optimal hyperplane that best separates the classes in the feature space.

k-Nearest Neighbors (kNN)

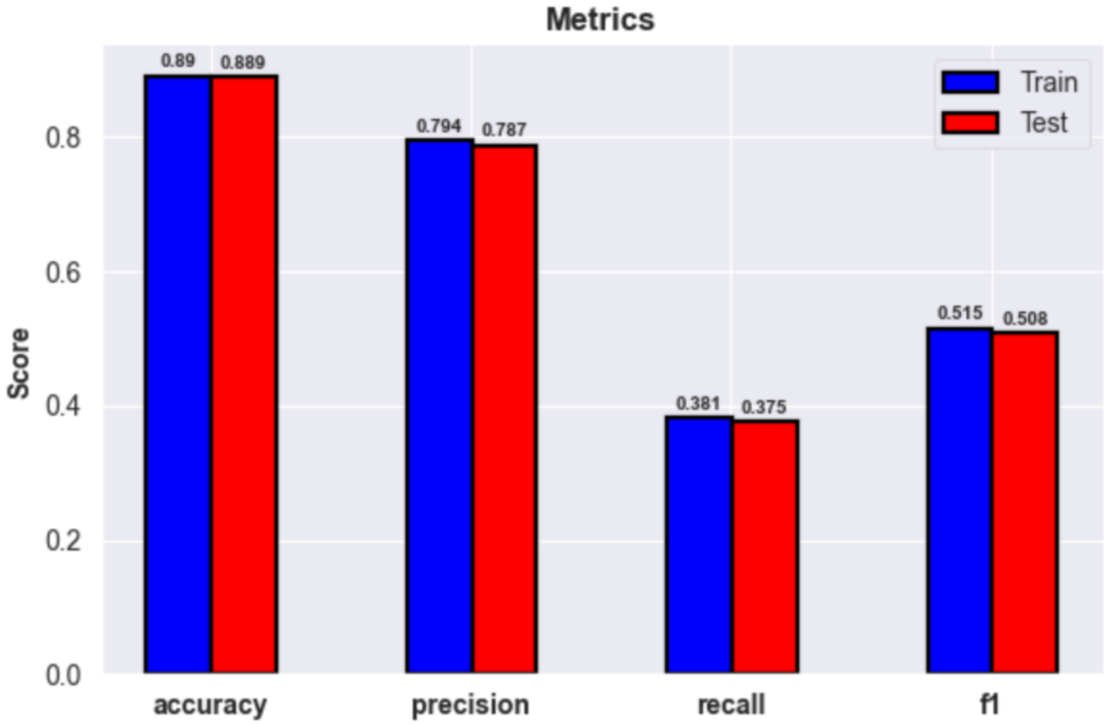


Fig.11- Metrics of kNN

Test Accuracy: 88.8%

Performance Analysis: kNN performed closely to SVM with a test accuracy of 88.8%. kNN is a non-parametric, lazy learning algorithm that classifies a data point based on how its neighbors are classified.

Random Forest

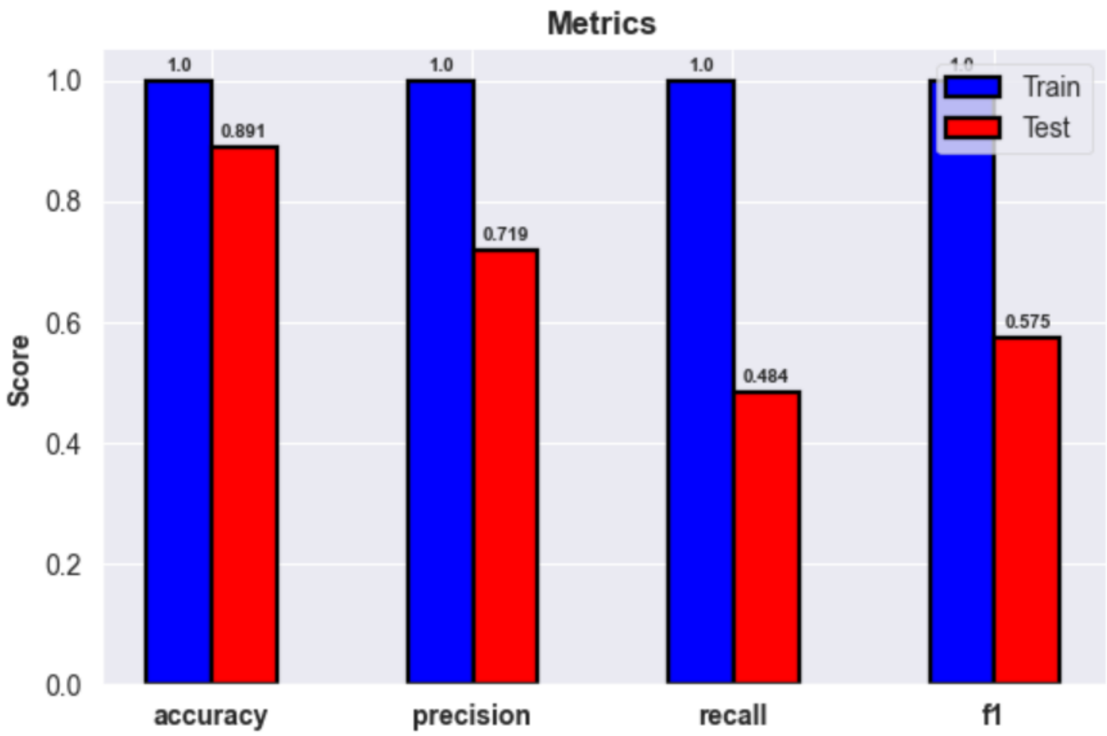
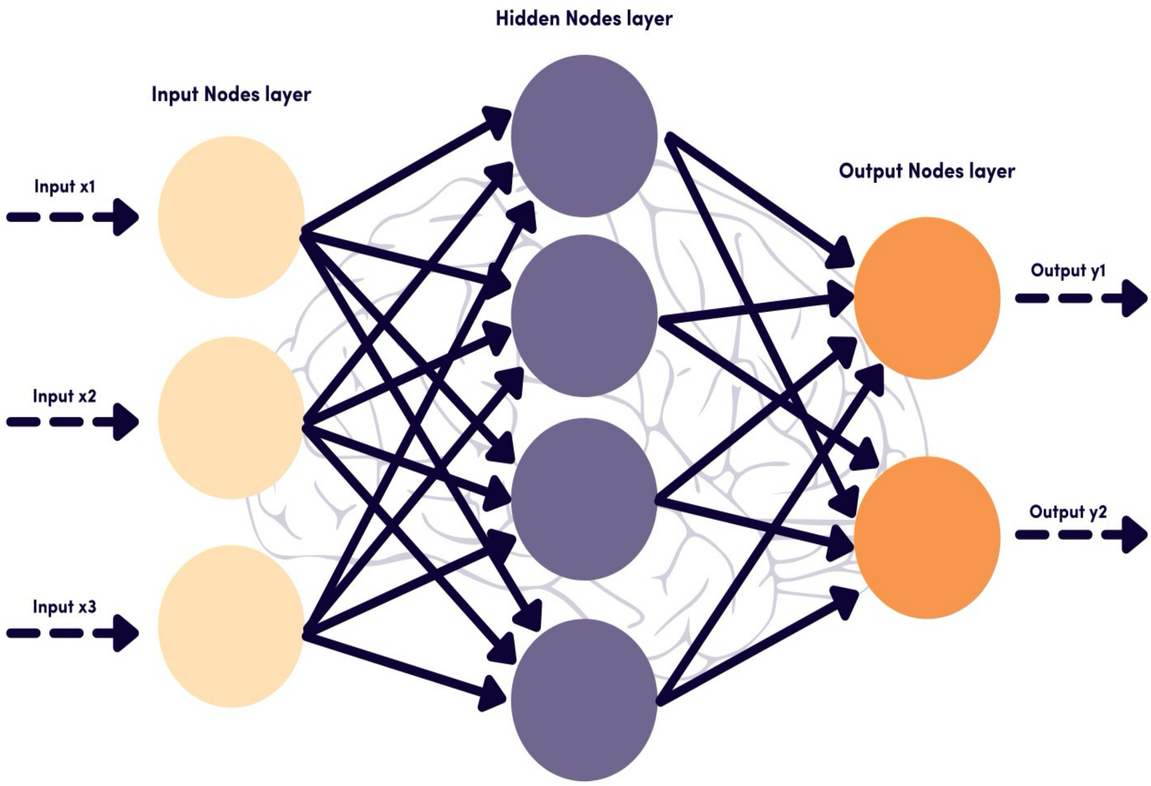


Fig.12- Metrics of Random Forest

Test Accuracy: 89%

Performance Analysis: Random Forest emerged as the top-performing model with an accuracy of 89%. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions, which often leads to improved accuracy and robustness.

Artificial Neural Network



Test Accuracy: 86.3%

Performance Analysis: Artificial Neural Network emerged with an accuracy of 86.3%. It is a brain-inspired network which processes data through interconnected nodes, learning patterns for tasks like classification and prediction in fields such as image recognition and natural language processing.Therefore accuracy could be increased with further addition to the data set.

CHAPTER 8

CONCLUSION

To develop an automatic breast cancer detection system using machine learning, we evaluated four models: Random Forest, SVM, KNN and ANN. The Random Forest model demonstrated the highest accuracy at 89%, followed closely by SVM at 88.9%, KNN at 88.8% and ANN at 86.3%. These models successfully addressed our objectives of enhancing early detection and diagnosis, aiding healthcare professionals in making timely and informed decisions. While Random Forest proved most effective, all models contributed to a reliable and efficient system, supporting improved patient outcomes in breast cancer diagnosis.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] N. Howlader, A. M. Noone, M. Krapcho et al., *SEER Cancer Statistics Review*, Natl. Cancer Inst, Bethesda, MD, 2015.

[3] T. C. Lewis, V. J. Pizzitola, M. E. Giurescu et al., "Contrast-enhanced digital mammography: a single-institution experience of the first 208 cases," *Breast Journal*, vol. 23, no. 1, pp. 67–76, 2017.

[4] A. A. Tabl, A. Alkhateeb, W. ElMaraghy, L. Rueda, and A. Ngom, "A machine learning approach for identifying gene biomarkers guiding the treatment of breast cancer," *Frontiers in Genetics*, vol. 10, p. 256, 2019.

[5] B. Ehteshami Bejnordi, M. Veta, P. Johannes van Diest et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *JAMA*, vol. 318, no. 22, pp. 2199–2210, 2017.

[6] M. P. Sampat, A. C. Bovik, and G. J. Whitman, "A model-based framework for the detection of spiculated masses on mammographya)," Medical Physics, vol. 35, no. 5, pp. 2110–2123, 2008. doi: 10.1118/1.289008

[7] N. H. Eltonsy, G. D. Tourassi, A. S. Elmaghraby, et al., "A concentric morphology model for the detection of masses in mammography," IEEE Transactions on Medical Imaging, vol. 26, no. 6, pp. 880–889, 2007.

[8] H. Jung, B. Kim, I. Lee, et al., "Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network," PLoS One, pp. 1–16, 2018. doi: 10.1016/j.acra.2011.09.014

[9] R. Agarwal, O. Diaz, X. Lladó, et al., "Automatic mass detection in mammograms using deep convolutional neural networks," J. Med. Imaging, vol. 6, no. 3, p. 1, 2019. doi: 10.1117/1.jmi.6.3.031409

[10] N. Dhungel, G. Carneiro, and A. P. Bradley, “A deep learning approach for the analysis of masses in mammograms with minimal user intervention,” Med. Image Anal., 2017. doi: 10.1016/j.media.2017.01.009