MALLA REDDY UNIVERSITY
(Telangana State Private Universities Act No. 13 of 2020 &
G.O.Ms. No. 14, Higher Education (UE) Department)

Maisammaguda, Kompally,
Medchal - Malkajgiri District
Hyderabad - 500100,
Telangana State.
www.mallareddyuniversity.

Department of Computer Science & Engineering

Mammary Carcinoma

Batch no. 39

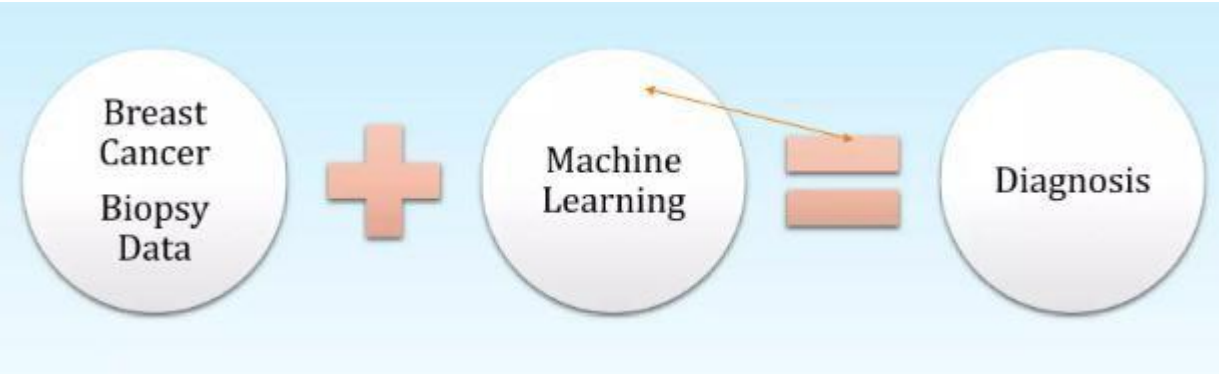
1. 2211CS010363 (M. Bhargav)
2. 2211CS010459 (P. Charan Sai)
3. 2211CS010407 (N. Uday Kumar)
4. 2211CS010514 (S. Harsha Vardhan)

Guide
Mr. G. Raju

Table of Contents

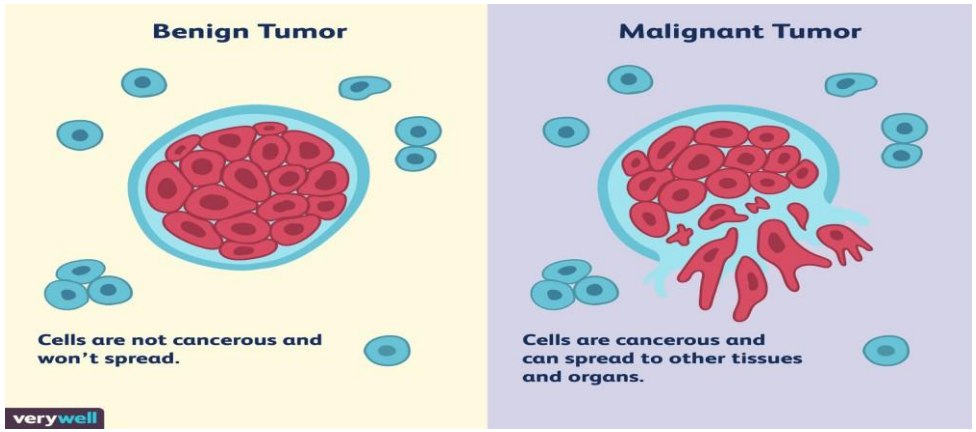
INTRODUCTION

This project focuses on analyzing breast cancer datasets to identify key patterns, correlations, and trends. By leveraging data analytics techniques such as statistical analysis, visualization, and predictive modeling, we aim to extract meaningful insights from patient records, biopsy results, and tumor characteristics.



LITERATURE SURVEY

Breast cancer is one of the most prevalent cancers among women worldwide. It occurs due to the uncontrolled growth of cells in breast tissue and can be classified as **benign** (non-cancerous) or **malignant** (cancerous).The number of new breast cancer patients is expected to increase by 70% in the next 20 years. Therefore, early and accurate diagnosis plays a pivotal role in improving prognosis and increasing the survival rate of cancer patients from 30 to 50%



LIMITATIONS OF EXISTING ALGORITHMS

- **Incomplete or Missing Data**: Patient records may have missing values for important attributes like tumor size, hormone receptor status, etc.
- **Poor User Interface**: Doctors and healthcare professionals may find existing systems difficult to use.
- **Slow Data Processing**: Existing systems may not process large datasets efficiently, causing delays in diagnosis.
- **Overfitting**: AI/ML models may perform well on training data but fail in real-world clinical applications.

ABSTRACT

Breast cancer is one of the most common forms of cancer among women, and early detection plays a critical role in improving survival rates. This project focuses on leveraging data analytics to enhance understanding, diagnosis, and treatment of breast cancer. Utilizing a dataset comprising patient demographic information, diagnostic features, and medical outcomes, we apply machine learning techniques and statistical analyses to identify patterns and key indicators associated with breast cancer.

The project involves preprocessing and cleaning the dataset, feature selection to identify significant variables, and building predictive models to classify malignant and benign cases with high accuracy. Advanced visualization techniques are used to explore the relationships between clinical features such as tumor size, texture, and shape.

PROPOSED METHODOLOGY

1. Problem Definition
2. Data Collection
3. Data Preprocessing & Cleaning
4. Exploratory Data Analysis (EDA)
5. Model Development
6. Deployment & Interpretation
7. Conclusion & Future Work

HARDWARE REQUIREMENTS

- **Processor (CPU):** Intel Core i3/i5 or AMD Ryzen 3/5 (Quad-core or higher)
- **RAM:** 8GB – 16GB
- **Storage:** 256GB – 512GB SSD (HDD not recommended)
- **GPU:** Not required for basic statistical analysis or small ML models
- **OS:** Windows 10/11, macOS
- **Internet:** Required for cloud-based analytics (Google Colab)

SOFTWARE REQUIREMENTS

- **Operating System:** Windows 10/11,
- **Programming Languages:** Python
- **Development Environment:** Jupyter Notebook, VS Code

Libraries & Frameworks:

- **Data Handling & Processing:** Pandas, NumPy
- **Data Visualization:** Matplotlib, Seaborn
- **Machine Learning & Model Building:** Scikit-learn (Logistic Regression, SVM, Decision Tree, Random Forest)
- **Data Preprocessing & Feature Selection:** Scikit-learn
- **Deployment**: Streamlit

CONCLUSION

Breast cancer data analytics plays a crucial role in early detection, diagnosis, and treatment planning. By leveraging data-driven insights, healthcare professionals can improve patient outcomes through accurate risk assessment, predictive modeling, and personalized treatment strategies.

This project has demonstrated the power of machine learning and statistical analysis in identifying key patterns and risk factors associated with breast cancer. The findings highlight the importance of early screening, feature selection, and model optimization to enhance predictive accuracy.

REFERENCES

Datasets

- Wisconsin Breast Cancer Dataset (WBCD)

  https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

- Breast Cancer Histopathological Image Dataset

  https://www.kaggle.com/datasets/paultimothymooney/breast-histopathology-images

Research Papers & Studies

- Breast Cancer Prediction Using Fine Needle Aspration

  https://pmc.ncbi.nlm.nih.gov/articles/PMC9913345/

- Explainable AI in Breast Cancer Diagnosis

https://pmc.ncbi.nlm.nih.gov/articles/PMC10840064/#:~:text=Recently%2C%20explainable%20artificial%20intelligence%20(XAI,and%20ultrasound

THANK YOU