

A  
PROJECT REPORT  
ON

## **CANCER DIAGNOSIS**

*Submitted in partial fulfilment of the  
requirements for the award of the degree  
of*  
**BACHELOR OF TECHNOLOGY**

*by*

**PRANJALI KOTHARI  
(EN. NO. - 2018554)**

*To*

**Mr. Aniruddha Prabhu**

Associate Professor

Department Of Computer Science & Engineering



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
GRAPHIC ERA HILL UNIVERSITY, DEHRADUN  
DEHRADUN – 248002 (INDIA)**

**July 2023**

# DECLARATION

I, **Pranjali Kothari**, student of **B. Tech CS VI Semester**, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, declare that the technical project work entitled “Cancer Diagnosis” has been carried out by me and submitting partial fulfilment of the course requirements for the award of degree in Bachelor of Technology of **Graphic Era Hill University, Dehradun** during the academic year **2022-2023**. The matter embodied in this synopsis has not been submitted to any other university or institution for the award of any other degree or diploma.

**Place: Dehradun**

**Date: 12 July 2023**

# ACKNOWLEDGEMENT

Here by I am submitting the project report on “Cancer Diagnosis”, as per the scheme of Graphic Era Hill University, Dehradun.

I express the deepest sense of gratitude to Mr. Aniruddha Prabhu, Asst. Professor, Department of Computer Science and Engineering for the invaluable guidance extended at every stage and in every possible way.

I am very much thankful to all the faculty members of the Department of Computer Science and Engineering, friends and my parents for their constant encouragement, support and help throughout the period of project conduction.

**Pranjali Kothari**  
2018554

# INDEX

1. Introduction
2. Data Exploration
3. Dataset Description
4. Data Preprocessing
5. Model Training
6. User Interface Development
7. Prediction
8. Results and Evaluation
9. Conclusion
10. Future Scope

# Introduction

The aim of this project was to develop a predictive model to classify cancer cells as either benign or malignant based on their individual characteristics. The dataset consists of 570 cancer cells and 30 features. The prediction model was built using Python, Tkinter for the graphical user interface, and the Random Forest Classifier model from the scikit-learn library.

# Data Exploration

The dataset contains information about various characteristics of cancer cells, such as radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. The target variable is the diagnosis, which can take two values: 'B' for benign cancer and 'M' for malignant cancer. The dataset consists of 285 instances of each type of cancer, providing a balanced representation for effective model training and evaluation.

# Dataset Description

The dataset consists of 30 features that provide various characteristics of cancer cells. These features are used to predict whether the cancer cells are benign or malignant. Here is a brief description of the columns in the dataset:

1. **diagnosis:** Target variable indicating the diagnosis of the cancer cells. It takes two values: 'B' for benign cancer and 'M' for malignant cancer.
2. **radius\_mean, texture\_mean, perimeter\_mean, area\_mean:** Mean values of the radius, texture, perimeter, and area of the cancer cells, respectively.
3. **smoothness\_mean, compactness\_mean, concavity\_mean, concave\_points\_mean:** Mean values of the smoothness, compactness, concavity, and concave points of the cancer cells, respectively.
4. **symmetry\_mean, fractal\_dimension\_mean:** Mean values of the symmetry and fractal dimension of the cancer cells, respectively.
5. **radius\_se, texture\_se, perimeter\_se, area\_se:** Standard errors of the radius, texture, perimeter, and area of the cancer cells, respectively.
6. **smoothness\_se, compactness\_se, concavity\_se, concave\_points\_se:** Standard errors of the smoothness, compactness, concavity, and concave points of the cancer cells, respectively.

7. symmetry\_se, fractal\_dimension\_se: Standard errors of the symmetry and fractal dimension of the cancer cells, respectively.
8. radius\_worst, texture\_worst, perimeter\_worst, area\_worst: Worst or largest values of the radius, texture, perimeter, and area of the cancer cells, respectively.
9. smoothness\_worst, compactness\_worst, concavity\_worst, concave\_points\_worst: Worst or largest values of the smoothness, compactness, concavity, and concave points of the cancer cells, respectively.
10. symmetry\_worst, fractal\_dimension\_worst: Worst or largest values of the symmetry and fractal dimension of the cancer cells, respectively.



# Data Preprocessing

Before training the model, data preprocessing steps were performed. This included splitting the dataset into features (X) and the target variable (y), and further splitting them into training and testing sets using the `train_test_split` function. The missing values were handled by dropping columns with missing data.

# Model Training

The RandomForestClassifier model was chosen for its ability to handle high-dimensional datasets and provide accurate predictions. The model was trained using the training data (X\_train and y\_train). Random Forest works by creating an ensemble of decision trees and aggregating their predictions to make the final classification. The model was fitted to the training data using the fit() function.

# User Interface Development

Tkinter, a Python library for creating GUI applications, was used to develop the user interface. The main window of the application was created with a title and dimensions. UI elements such as labels, entry fields, and buttons were added to allow users to input the value for the "radius mean" feature and trigger the prediction process.

# Prediction

The `predict()` function was implemented to read the input value from the UI, prepare the input data for prediction, and make predictions using the trained Random Forest Classifier model. The predicted cancer type ('B' or 'M') was displayed in the prediction label.

# Results and Evaluation

The developed application allows users to input the radius mean of a cancer cell and predict whether it is benign or malignant. The accuracy of the model can be evaluated using various metrics such as accuracy, precision, recall, and F1 score. Additionally, techniques like cross-validation and hyperparameter tuning can be applied to optimize the model's performance.

# Conclusion

The project successfully achieved the goal of creating a predictive model to classify cancer cells as benign or malignant based on their individual characteristics. The user-friendly GUI allows users to easily interact with the model and obtain predictions. Further improvements can be made by incorporating additional features, exploring different models, and enhancing the user interface.

# Future Scope

The developed application can serve as a starting point for further research and development in cancer diagnosis. It can be extended to handle larger datasets, incorporate more advanced machine learning models, and integrate additional features for enhanced prediction accuracy. The user interface can also be improved to provide more interactive and informative visualizations.