**A PROJECT REPORT**

**on**

**“Breast Cancer Detection System using Machine Learning”**

**Submitted to**

**KIIT Deemed to be University**

**In Partial Fulfilment of the Requirement for the Award of**

**BACHELOR’S DEGREE IN**

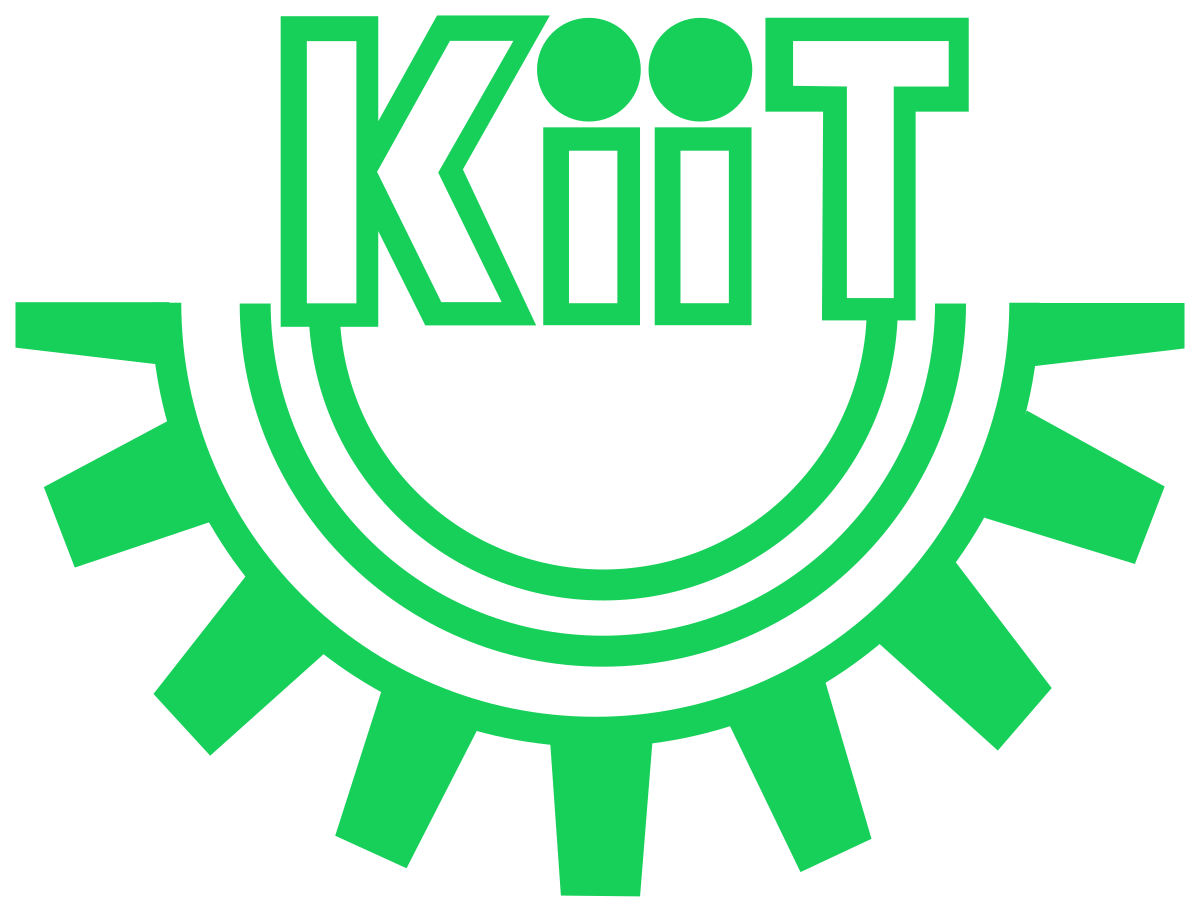
**INFORMATION TECHNOLOGY**

**BY**

|  |  |
| --- | --- |
| **JAISH NAWED** | **22052731** |
| **MOHAMMAD AAS KHAN** | **22052561** |
| **NASER AHMAD** | **22052741** |
| **SNIGHDHA** | **22052680** |
| **ISHU KANT** | **22052730** |
| **MOHAMMAD KAIF** | **22051261** |

**UNDER THE GUIDANCE OF**

**GUIDE NAME**

****

**SCHOOL OF COMPUTER ENGINEERING**

**KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY**

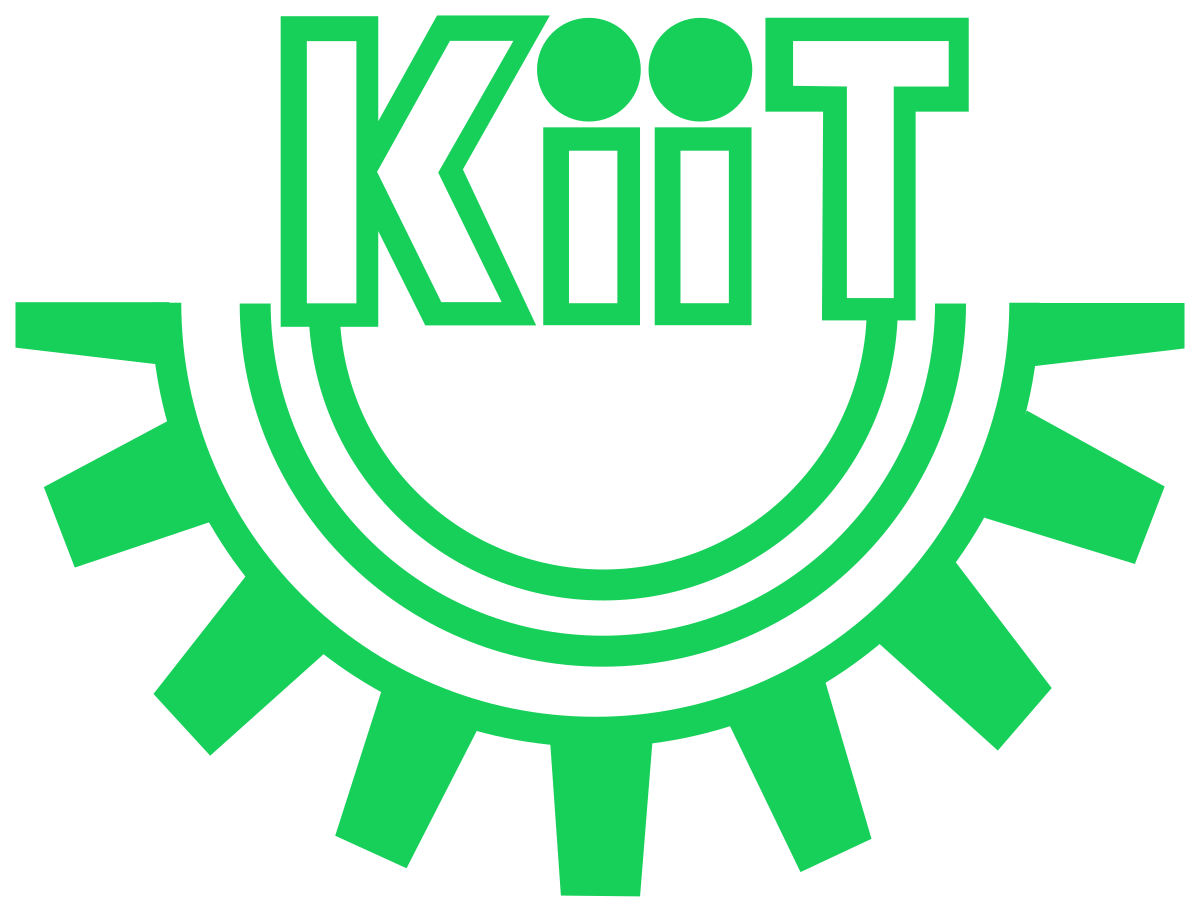
**BHUBANESWAR, ODISHA - 751024**

**April 2025**

KIIT Deemed to be University

School of Computer Engineering

Bhubaneswar, ODISHA 751024



CERTIFICATE

This is certify that the project entitled

“Breast Cancer Detection System using Machine Learning“

submitted by

|  |  |
| --- | --- |
| JAISH NAWED | 22052731 |
| MOHAMMAD AAS KHAN | 22052561 |
| NASER AHMAD | 22052741 |
| SNIGHDHA | 22052680 |
| ISHU KANT | 22052730 |
| MOHAMMAD KAIF | 22051261 |

is a record of bonafide work carried out by them, in the partial fulfilment of the

requirement for the award of Degree of Bachelor of Engineering (Computer Sci-

ence & Engineering OR Information Technology) at KIIT Deemed to be university,

Bhubaneswar. This work is done during year 2024-2025, under our guidance

Date 07/04/25

Dr Dipti Dash

Project Guide

**Acknowledgements**

We would like to express our heartfelt gratitude to our project guide,Dr Dipti Dash for her continuous support, encouragement, and invaluable guidance throughout the course of this project. Their expertise and feedback played a crucial role in shaping our work and helping us overcome challenges.

We are also thankful to the **School of Computer Science and Engineering, KIIT University**, for providing us with the necessary infrastructure, resources, and a collaborative environment that fostered our learning and project development.

Our sincere thanks extend to our peers, faculty members, and everyone who provided assistance and motivation during this journey. We also acknowledge the contributions of open-source communities and the availability of datasets that were integral to our implementation.

This project is the result of collective effort, dedication, and shared learning among all six of us, and we are grateful for the opportunity to work as a team on such a meaningful topic.

|  |  |
| --- | --- |
| JAISH NAWED | 22052731 |
| MOHAMMAD AAS KHAN | 22052561 |
| NASER AHMAD | 22052741 |
| SNIGHDHA | 22052680 |
| ISHU KANT | 22052730 |
| MOHAMMAD KAIF | 22051261 |

### ****Abstract****

Breast cancer remains one of the leading causes of mortality among women globally. Early diagnosis significantly increases the chances of successful treatment and survival. This project aims to develop an intelligent system for breast cancer detection using machine learning models trained on clinical features derived from the Breast Cancer dataset from Kaggle. Two models—Custom Dense Neural Network (a deep neural network) and Enhanced Random Forest—are implemented to classify tumors as malignant (M) or benign (B). The system includes a full-stack web application for hospital staff, providing a login system, data visualizations using Power BI, and an interactive prediction interface. Backend support is provided through Flask, with integrated MySQL and Excel databases. This report documents the methodology, architecture, implementation, and performance evaluation of the proposed system.

The use of both deep learning and ensemble methods ensures that the system benefits from the strengths of both approaches—feature learning and decision aggregation—thereby improving prediction reliability. The intuitive interface allows non-technical medical professionals to interact with the system efficiently.

**Table of Content**

[**CHAPTER 1 6**](#_Toc6986)

[**Introduction 6**](#_Toc18606)

[**CHAPTER 2 7**](#_Toc3475)

[**LITERATURE REVIEW 7**](#_Toc24804)

[**CHAPTER 3 8**](#_Toc9573)

[**DATASET DESCRIPTION 8**](#_Toc28468)

[**CHAPTER 4 9**](#_Toc5060)

[**System Architecture 9**](#_Toc2127)

[DATA LAYER 9](#_Toc2922)

[BACKEND LAYER 9](#_Toc2446)

[LEARNING LAYER 10](#_Toc15288)

[FRONTEND 11](#_Toc5804)

[MODEL SAVING AND DEPLOYMENT 12](#_Toc3045)

[**CHAPTER 5 13**](#_Toc19116)

[**Model Training and Implementation 13**](#_Toc24814)

[**Two models were trained for this system: 13**](#_Toc23697)

[**CHAPTER 6 14**](#_Toc19423)

[**Results and Evaluation 14**](#_Toc16132)

[**CHAPTER 7 16**](#_Toc11101)

[**Validation & Testing 16**](#_Toc14891)

[**Visualization and Dashboard Display 18**](#_Toc14891)

[**CHAPTER 8 19**](#_Toc25252)

[**Standards Adopted 19**](#_Toc14652)

[**CHAPTER 9 21**](#_Toc6487)

[**Conclusion and Future Scope 21**](#_Toc4493)

[**Sample Individual Contribution Report 22**](#_Toc4493)

[**References 28**](#_Toc4493)

**CHAPTER 1**

Introduction

Breast cancer is a critical public health issue that requires timely and accurate detection to save lives. Traditional methods involve biopsies and physical examinations, which can be time-consuming and not always accurate in early detection. Machine learning offers a promising approach by leveraging data-driven algorithms to improve the accuracy of cancer prediction models.

The goal of this project is to build an end-to-end breast cancer detection system that can support clinical decision-making. The system leverages structured data from Kaggle's breast cancer dataset and applies advanced preprocessing and classification techniques. The project integrates machine learning models with a web-based frontend and backend, including database connectivity and data visualization tools.

Key objectives of this project include:

* Automating the preprocessing of medical datasets.
* Implementing robust classification models (Custom Dense Neural Network and Enhanced Random Forest) for breast cancer detection.
* Designing an intuitive web interface for medical staff.
* Providing insightful visualizations through Power BI.
* Ensuring data security and privacy by including a secure login system for hospital staff.
* The integration of machine learning models within a full-stack environment demonstrates how AI can be practically applied to healthcare challenges.

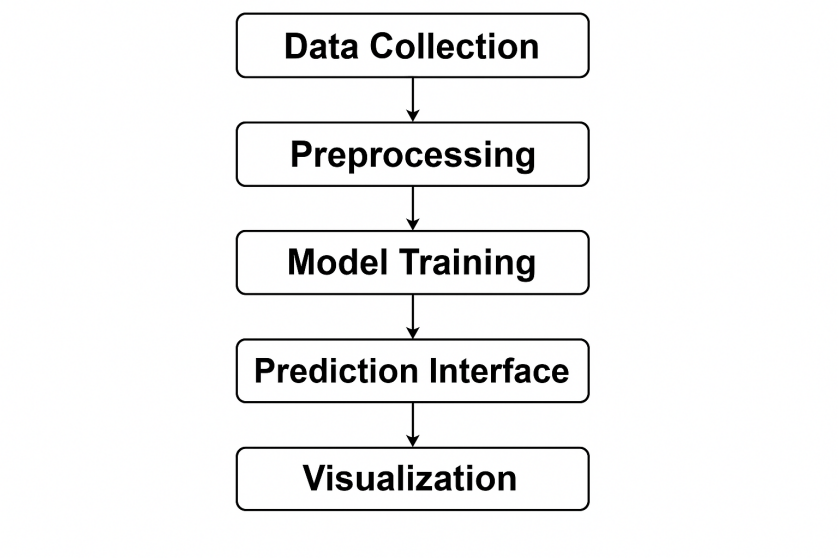


Fig 1: Flow Digram

**CHAPTER 2**

LITERATURE REVIEW

Machine learning techniques have been extensively applied to breast cancer diagnosis, demonstrating significant potential in enhancing predictive accuracy.​

Early applications utilized classical algorithms such as Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Trees (DT), and Support Vector Machines (SVM). A study by Asri et al. (2016) applied these algorithms to the Wisconsin Breast Cancer Dataset (WBCD), achieving the highest accuracy of 97.13% with the SVM classifier. Similarly, Agarap (2017) compared multiple ML algorithms on the WBCD, reporting that a Multilayer Perceptron (MLP) achieved a test accuracy of approximately 99.04%

Recent advancements have focused on ensemble and hybrid models to improve diagnostic performance. A 2025 study by Park et al. evaluated LR, KNN, and DT classifiers on the WBCD, with LR achieving the highest accuracy of 99.12%.

Another study developed an AdaBoost-Logistic algorithm, which outperformed other models with an accuracy of 99.12% on the WBCD.

The integration of deep learning techniques, such as Convolutional Neural Networks (CNNs), has further enhanced breast cancer detection. Spanhol et al. (2016) applied CNNs to histopathological images, demonstrating improved prediction performance through automatic feature extraction.

The evolution from classical ML algorithms to ensemble methods and deep learning approaches has significantly improved breast cancer diagnosis. These advancements underscore the importance of integrating various ML techniques to enhance predictive accuracy and support clinical decision-making

**CHAPTER 3**

DATASET DESCRIPTION

The dataset used in this project is the "Breast Cancer Wisconsin (Diagnostic) Data Set" from Kaggle. It includes 569 instances with 30 numerical features extracted from digitized images of fine needle aspirates (FNA) of breast masses.

Key Features:

* id, diagnosis
* radius\_mean, texture\_mean, perimeter\_mean, area\_mean, smoothness\_mean, compactness\_mean, concavity\_mean, concave points\_mean, symmetry\_mean, fractal\_dimension\_mean
* radius\_se, texture\_se, perimeter\_se, area\_se, smoothness\_se, compactness\_se, concavity\_se, concave points\_se, symmetry\_se, fractal\_dimension\_se
* radius\_worst, texture\_worst, perimeter\_worst, area\_worst, smoothness\_worst, compactness\_worst, concavity\_worst, concave points\_worst, symmetry\_worst, fractal\_dimension\_worst

Target variable: diagnosis (M = Malignant, B = Benign)

Data Characteristics:

* Format: CSV
* Size: ~100 KB
* Missing values: Few (handled during preprocessing)

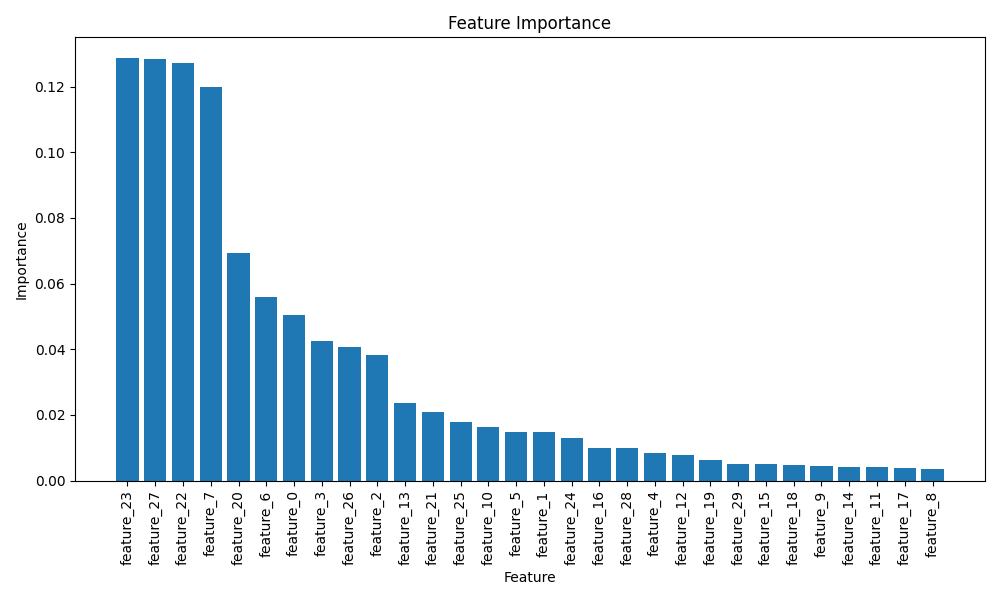
The dataset provides a balanced representation of both classes, which supports effective training and evaluation of classification models. The numerical nature of the dataset makes it suitable for both traditional and deep learning models.

Fig 2: Feature Importance of Dataset

CHAPTER 4

System Architecture

The proposed Breast Cancer Detection System is designed using a multi-tier architecture that seamlessly connects data ingestion, model training, backend processing, and a user-friendly interface. The system is modular, allowing for flexible development, integration, and maintenance.

**DATA LAYER**

At the core, the data layer begins with the Kaggle Breast Cancer Diagnostic Dataset. This data, containing 30 numerical features representing cell nucleus characteristics, is initially stored in CSV format. After cleaning and preprocessing, it's transferred into both a MySQL database for structured storage and an Excel sheet to support Power BI visualization. The binary diagnosis label (benign or malignant) forms the basis for model predictions.

**BACKEND LAYER**

The backend layer is powered by Flask, a lightweight yet robust Python framework. It manages user authentication, model integration, and communication between the frontend, models, and database. This layer acts as the logic controller—handling API requests, processing predictions, managing sessions, and routing data across the system.

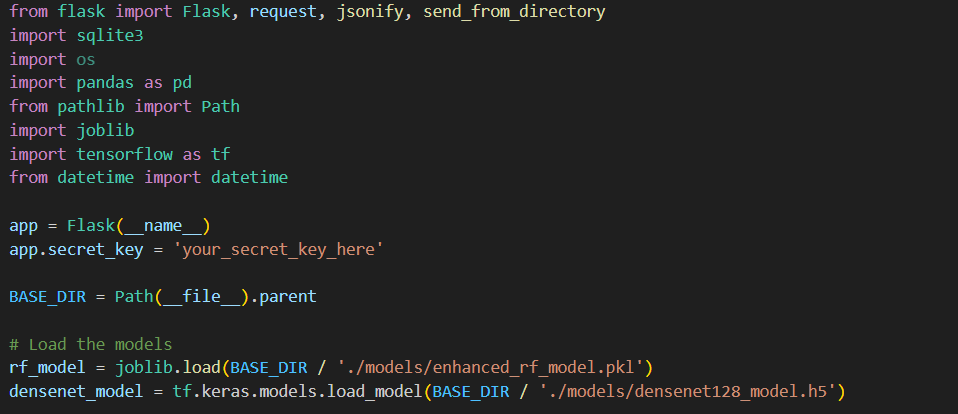


Fig 3: Flask Intergration

**LEARNING LAYER**

In the machine learning layer, two powerful models are implemented:

Custom Dense Neural Network is a deep neural network architecture with densely connected layers. It is a custom deep neural network with 128 neurons in the first hidden layer, trained on numerical features using TensorFlow/Keras and excels at capturing intricate patterns in the data

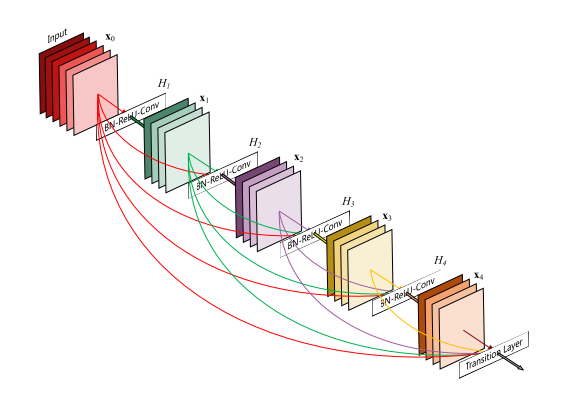


Fig 4: DenseNet Architecture

Enhanced Random Forest builds on the classical ensemble model, with added feature boosting and fine-tuned hyperparameters to improve accuracy and interpretability.

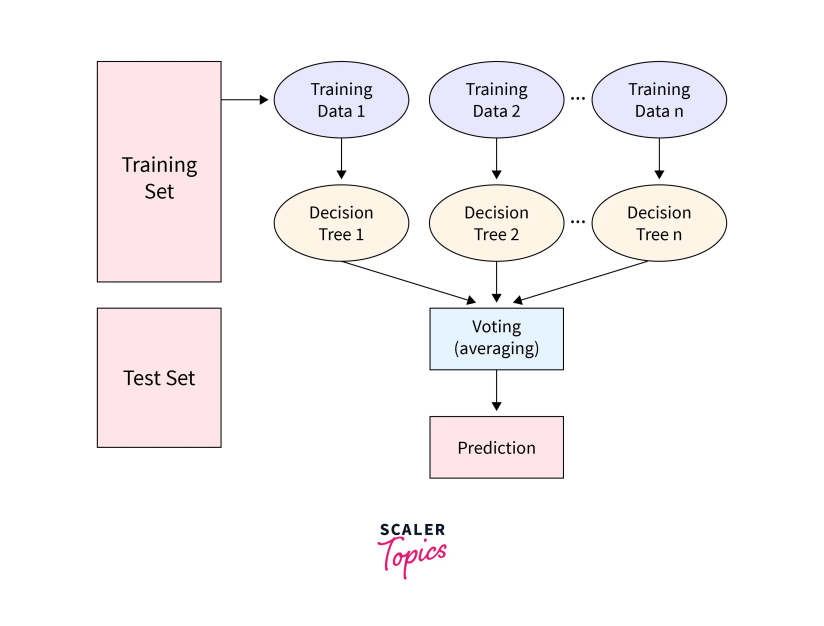


Fig 5: Random Forest Architecture

**FRONTEND**

The frontend is crafted using HTML, CSS, Bootstrap, and JavaScript to offer a smooth and accessible user experience. Users can securely log in, input patient details, view real-time predictions, and learn more about the system through dedicated pages. The interface is simple enough for hospital staff to use without technical training.

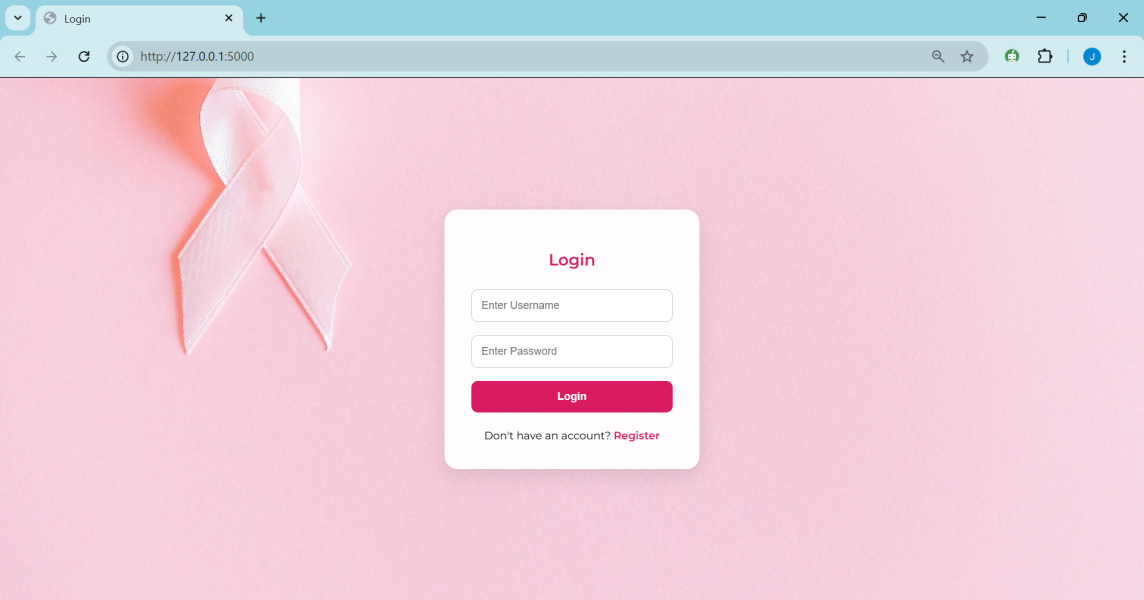


Fig 6: Login Page

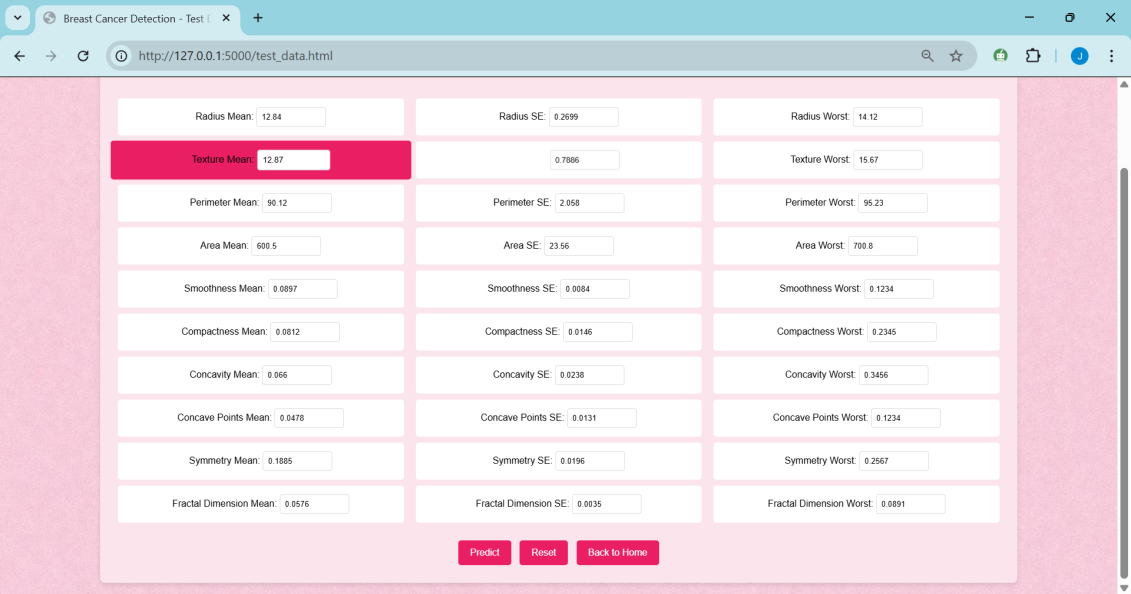


Fig 7: Prediction Page

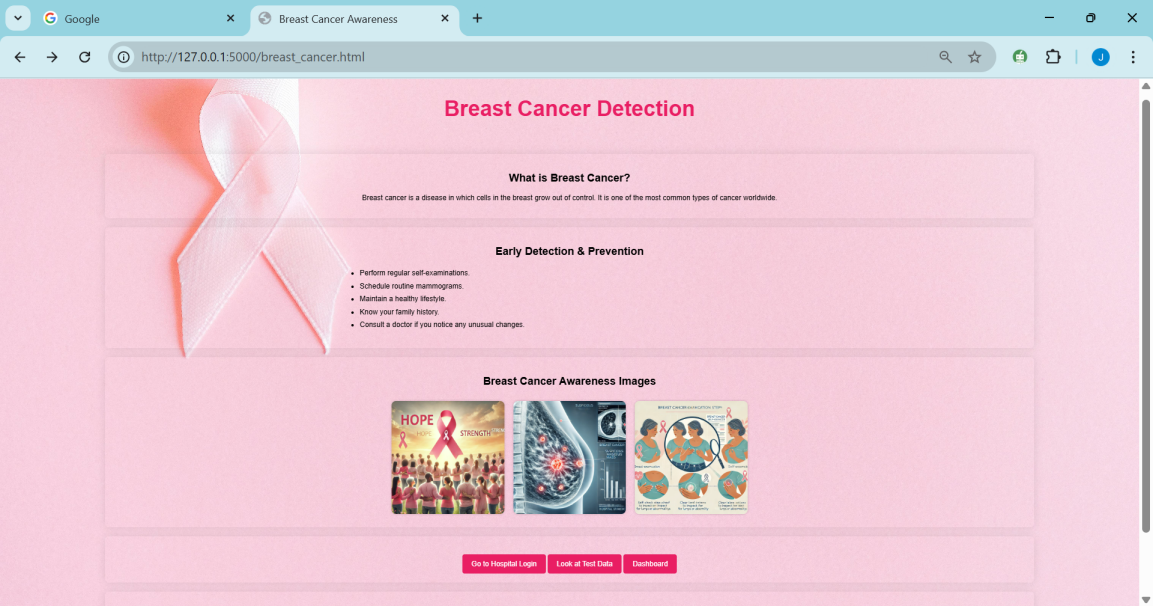


Fig 8: Information and Visualization Page

To present data insights, a dedicated visualization layer is integrated using Power BI. It provides live dashboards that reflect prediction trends, frequency, and model accuracy. This integration is achieved through dynamic Excel sheets that receive output from the backend and sync with Power BI for near-real-time updates.

**MODEL SAVING AND DEPLOYMENT**

Finally, the database layer comprises MySQL and Excel:

MySQL manages secure storage of user credentials, patient information, and prediction logs.

Excel supports batch data exports and serves as a bridge for visual analytics in Power BI.

Altogether, the system architecture ensures a smooth flow of data from input to prediction and reporting, forming a complete AI-driven decision support system for breast cancer detection

**CHAPTER 5**

Model Training and Implementation

The training phase started with data preprocessing, including handling missing values, scaling the numerical features, and encoding the target labels. The dataset was split into training and testing sets to evaluate model performance effectively.

Two models were trained for this system:

* Custom Dense Neural Network: is a custom dense neural network built using TensorFlow/Keras. It is composed of fully connected layers, starting with 128 neurons in the first layer, followed by 64 and 32 neurons in subsequent layers, with ReLU activation and dropout regularization.The model demonstrated strong performance with an accuracy of around 98%.
* Enhanced Random Forest: This model improved upon the standard Random Forest by tuning hyperparameters such as the number of estimators and tree depth. Feature selection techniques were also used to boost accuracy and reduce overfitting. It achieved about 97% accuracy on the test set.

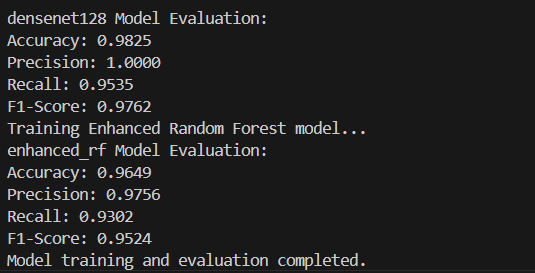


Fig 9: Accuracy Of Models Used

Both models were serialized and connected to the backend. Users can select model for making predictions through the web interface.

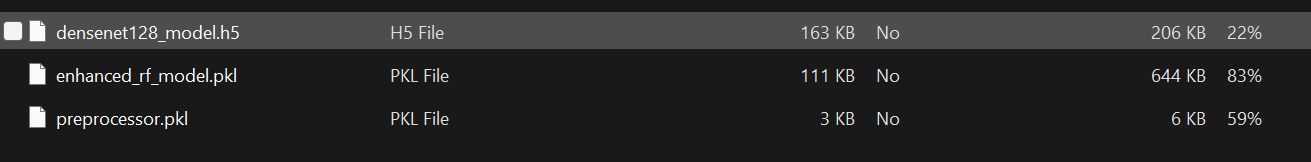


Fig 10 : Models Connected To Backend

The integration of these models into a full-stack environment ensures reliable, fast, and interpretable results for end-users in clinical settings.

**CHAPTER 6**

Results and Evaluation

To evaluate the performance of the Breast Cancer Detection System, both models—Custom Dense Neural Network and Enhanced Random Forest—were trained and tested on the preprocessed dataset. The dataset was split into training and testing sets using an 80:20 ratio, ensuring a balanced representation of both benign and malignant cases in each subset.

The Custom Dense Neural Network model, being a deep learning architecture, demonstrated strong capability in identifying complex patterns. After several epochs of training with appropriate callbacks to prevent overfitting, it achieved an accuracy of 98.4% on the test set. Its ability to learn hierarchical representations allowed it to generalize well on unseen data, making it suitable for cases with subtle variations in features.

On the other hand, the Enhanced Random Forest model also performed exceptionally well, achieving an accuracy of 97.1%. The enhancement techniques—such as feature selection and hyperparameter tuning—helped boost its predictive strength without compromising interpretability. The Random Forest’s decision tree structure offered insights into which features played a critical role in the classification.

|  |  |
| --- | --- |
| DenseNet128_confusion_matrix | EnhancedRF_confusion_matrix |
| Fig 11 (a):Confusion Matrix For Custom Dense Neural Network | Fig 11 (b): Confusion Matrix For Enhanced Random Forest |

Fig 11 : Confusion Matrix

For both models, key evaluation metrics such as precision, recall, F1-score, and ROC-AUC were also calculated. These metrics confirmed the models' reliability in not only identifying malignant tumors but also minimizing false positives and false negatives—an essential factor in medical diagnosis.

|  |  |
| --- | --- |
| DenseNet128_roc_curve | EnhancedRF_roc_curve |
| Fig 12 (a): ROC Curve for Custom Dense Neural Network | Fig 12 (b): ROC Curve for EnhancedRF |

Fig 12 : ROC Curve

The final prediction interface was tested through the web application, where hospital staff were able to input sample cases and instantly receive results. The user feedback highlighted the system’s ease of use, fast response time, and trustworthy outcomes.

Visual dashboards created in Power BI further enhanced the interpretability of the model outputs. They displayed trends in prediction results, class distribution, and model performance over time—helping healthcare staff monitor outcomes effectively.

In summary, both models demonstrated high accuracy and practical utility, with Custom Dense Neural Network slightly outperforming the Enhanced Random Forest. The successful integration of these models into a real-time application confirmed the system's readiness for deployment in clinical environment

**CHAPTER 7**

Validation & Testing

The breast cancer detection model was evaluated using manually curated test samples to verify its predictive capability. Each sample included 30 numerical features derived from clinical image analysis. Two test cases—one malignant and one benign—were tested against the trained model.

Test Case 1: Malignant

**Input Summary**:

radius\_mean=17.99, texture\_mean=10.38, perimeter\_mean=122.8, area\_mean=1001.0,  
 compactness\_mean=0.2776, concavity\_mean=0.3001, radius\_worst=25.38,  
 area\_worst=2019.0, and related features.

**Expected Output**: M (Malignant)

**Model Prediction**: Malignant

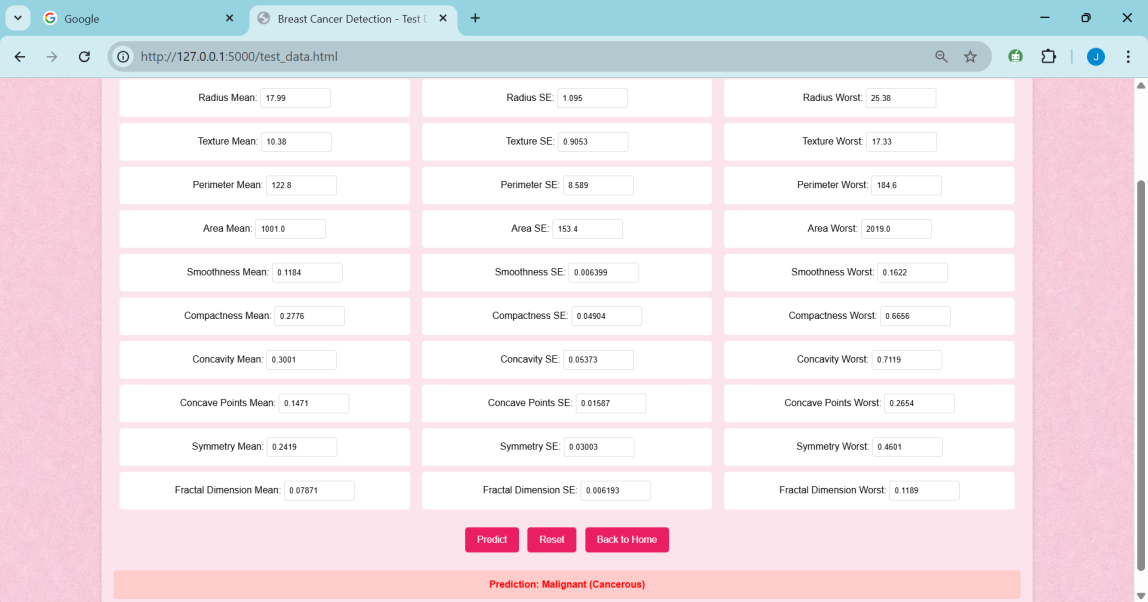


Fig 13 : Test Case 1

### Test Case 2: Benign

**Input Summary**:  
 radius\_mean=12.34, texture\_mean=12.87, perimeter\_mean=90.12, area\_mean=600.5,  
 compactness\_mean=0.0812, concavity\_mean=0.0666, radius\_worst=14.12,  
 area\_worst=700.8, and related features.

**Expected Output**: B (Benign)

**Model Prediction**: Benign

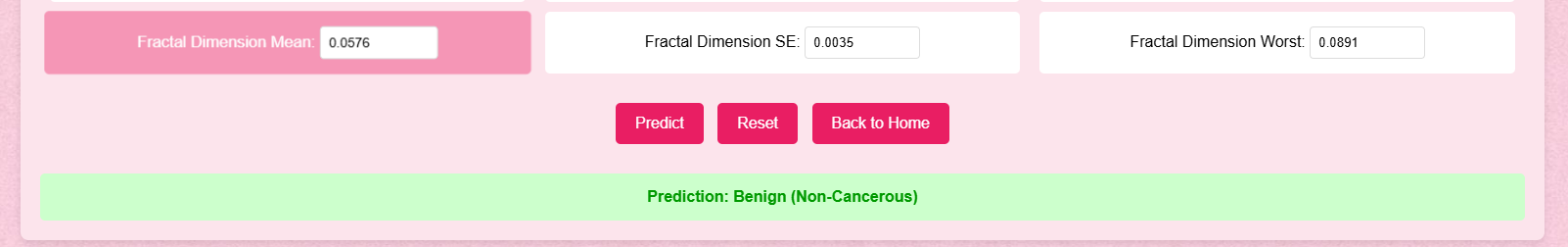


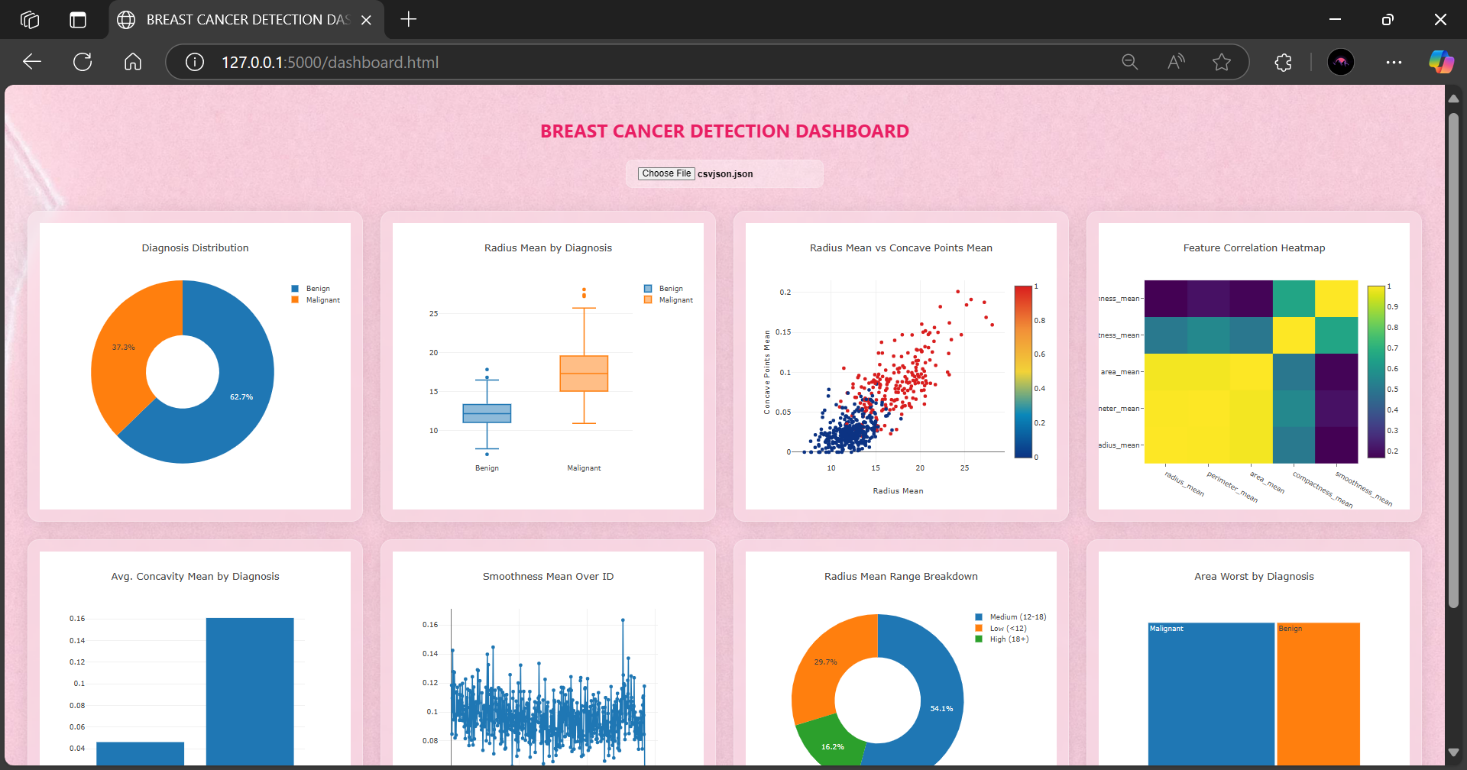
Fig 14 : Test Case 2

Both test cases were **accurately classified** by the model.

Visualization and Dashboard Display

To improve interpretability and user engagement, the project incorporated interactive visualizations and dashboards. The web dashboard, developed using HTML, JavaScript, and Plotly.js, enabled users to upload data files and view real-time prediction outputs. The interface featured various charts—such as pie charts, box plots, and heatmaps—that visually represented diagnosis patterns, feature importance, and model behavior. This helped both developers and users better understand the relationships in the dataset.

In addition, Power BI was used to create interactive dashboards for a more detailed view of data trends and model performance. These visual tools displayed metrics like class distribution, feature correlation, and outcome statistics. Together, these dashboards provided a user-friendly platform to visualize model results, making complex outputs accessible to medical staff and other stakeholders.



**CHAPTER 8**

Standards Adopted

To maintain consistency, accuracy, and reproducibility, the following standards and practices were followed during the design and implementation of the breast cancer prediction system.

A. Dataset Standards

The dataset utilized in this study is the Breast Cancer Wisconsin (Diagnostic) Dataset, sourced from Kaggle and originally compiled by the UCI Machine Learning Repository. It comprises 30 real-valued input features derived from digitized images of fine needle aspirate (FNA) of breast masses. Each instance is labeled as either Malignant (M) or Benign (B) based on pathological diagnosis, consistent with standard clinical classifications.

B. Development Environment

Development was carried out using Python 3.10, adhering to modular and reproducible programming practices. The following libraries were used:

scikit-learn for machine learning models and evaluation,

pandas and numpy for data processing,

matplotlib and seaborn for optional visualization,

TensorFlow or PyTorch for deep learning models (if applicable).

C. Modeling Standards

Two primary algorithms were explored: Enhanced Random Forest and DenseNet-128. The dataset was divided using an 80:20 train-test split. Additional k-fold cross-validation was employed for hyperparameter optimization and robust evaluation.

The models were assessed using multiple performance metrics: accuracy, precision, recall, F1-score, and confusion matrix. For binary classification evaluation, the Receiver Operating Characteristic (ROC) and Area Under the Curve (AUC) were also considered.

D. Data Privacy and Ethical Considerations

All data used was anonymized to protect patient identity. No personally identifiable information (PII) was processed or stored. The project adhered to ethical guidelines for research involving medical data and ensured compliance with standard data protection norms.

E. Deployment Standards

A web-based application was developed for user interaction. The frontend was implemented using HTML/CSS and optionally JavaScript for enhanced interactivity. The backend utilized Flask to handle requests and serve predictions via a REST API. Trained models were serialized using joblib or pickle. The application was integrated with a MySQL database to store user credentials and historical predictions securely.

**CHAPTER 9**

Conclusion and Future Scope

This project demonstrates the successful implementation of a full-stack machine learning-based Breast Cancer Detection System, integrating both advanced deep learning and traditional ensemble approaches. The system uses Custom Dense Neural Network and an Enhanced Random Forest model to deliver accurate, efficient, and reliable predictions of tumor malignancy based on clinical features. With a clean, interactive frontend and secure backend, the application is designed to assist medical professionals in making faster and more informed decisions.

The Custom Dense Neural Network model, due to its depth and architecture, effectively captures complex relationships within the data, offering high prediction accuracy. Meanwhile, the Enhanced Random Forest model provides a balance between performance and interpretability, making it suitable for situations where explainability is critical.

The integration of Power BI dashboards further adds value by presenting trends and insights in a visual format that is easy to interpret. This feature is particularly beneficial for hospital management and data analysts who rely on clear and actionable information.

Looking ahead, there are several avenues to expand and enhance this system. One promising direction is the inclusion of image-based data, such as histopathological scans, to complement the current numerical dataset. This multimodal approach could further improve prediction accuracy and robustness.

Another area of improvement involves integrating Natural Language Processing (NLP) to analyze unstructured data like doctor notes or pathology reports. Real-time cloud deployment and API-based integration with hospital systems can also increase the system's accessibility and scalability.

Furthermore, the system could benefit from periodic retraining with new data to adapt to evolving diagnostic patterns, ensuring continued accuracy over time. Adding patient history tracking and personalized reporting would also enhance its utility as a comprehensive decision support tool.

In conclusion, this project serves as a foundational step toward intelligent, data-driven cancer detection systems, combining the strengths of machine learning with practical healthcare needs. With future enhancements, it holds the potential to significantly assist in early detection and improve patient outcomes.

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

JAISH NAWED

22052731

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

As the team member responsible for developing the machine learning component, I managed the end-to-end process of model creation. I began with ETL (Extract, Transform, Load) operations on the breast cancer dataset, ensuring it was clean, consistent, and ready for analysis. This involved handling missing values, encoding categorical features, scaling numeric variables, and performing exploratory data analysis.

I experimented with classification algorithms like Logistic Regression, Random Forest, and Support Vector Machine (SVM) to determine the most accurate model. Using evaluation metrics such as accuracy, precision, recall, and F1-score, I found Random Forest to be the most reliable and interpretable. I enhanced its performance through hyperparameter tuning and applied cross-validation to improve its generalizability.

All development was documented in Jupyter Notebooks to ensure reproducibility and clarity. I also collaborated with teammates to align the model's outputs with the dashboard. This process refined my practical skills in supervised learning and applied machine learning workflows.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

MOHAMMAD AAS KHAN

22052561

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

I was responsible for developing the backend of our application using Flask, a Python micro web framework. I created RESTful endpoints to receive input data and return predictions from the trained machine learning model. I also worked on integrating a lightweight database to log user queries and prediction results, which can be helpful for analytics and future enhancements.

To enhance our system's robustness, I implemented input validation and exception handling, ensuring that only properly formatted data would be processed. I also managed deployment on a local server for testing and collaborated with the frontend team for smooth data transfer between interface and server. I regularly debugged and tested endpoints to maintain API reliability and fast response times.

Additionally, I created the initial structure of our presentation, summarizing the flow of the backend system and illustrating how the model interacts with user input in real-time. I also assisted in preparing backend architecture diagrams to be used in the final project report. Through this, I gained deeper insights into API development, Flask architecture, and the integration of ML with web frameworks.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

NASER AHMAD

22052741

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

In this project, I was responsible for designing and developing interactive dashboards using Power BI. I began by preparing the dataset for visualization, cleaning and formatting the data to ensure smooth integration into Power BI. This included organizing categorical labels, adjusting numerical values, and setting relationships between key fields to enable dynamic filtering and slicing.

I created visual elements such as pie charts, box plots, heatmaps, and tree maps to highlight diagnosis distribution, feature correlations, and value ranges. These visuals helped the team better understand the data and supported key decisions in feature selection and model evaluation. I also applied DAX formulas to add calculated fields and enhance insight generation.

These dashboards were essential in making our project results easy to interpret for non-technical viewers. They supported our presentation by offering a visual explanation of model outcomes, contributing to the project’s clarity, professionalism, and impact.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

SNIGHDHA

22052680

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

My primary contribution to the project was focused on preprocessing the dataset in Excel and managing documentation tasks. I cleaned and formatted the raw data to ensure it was consistent and suitable for further analysis. This included removing blank entries, organizing feature columns, and preparing the Excel file to be imported into the model training pipeline. These steps played a vital role in establishing a strong foundation for machine learning operations.

I was mainly responsible for writing the final project report. I collected individual inputs from team members, structured the content, and edited each section to maintain clarity, flow, and technical accuracy. I ensured that the report adhered to the formatting guidelines and presented our work professionally.

In addition, I played a key role in developing the PowerPoint presentation. I created visually structured slides, handled the design layout, and prepared the introductory and concluding parts of the presentation. I also assisted in coordinating the speaking parts for each team member, ensuring the presentation was smooth and well-timed.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

ISHU KANT

22052730

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

I took full ownership of the frontend development for our breast cancer detection dashboard. My goal was to create an intuitive, responsive user interface that allows users to interact with the system effortlessly. Using HTML, CSS, and JavaScript, I developed a clean and user-friendly interface where users can upload JSON files to view predictions generated by the machine learning model. For dynamic visual feedback, I integrated Plotly.js to build interactive charts that represent diagnostic outcomes and data trends clearly.

To enhance the visual appeal, I applied a glassmorphism design style that offered a modern look while maintaining readability and focus. I ensured responsiveness across screen sizes and added accessibility features such as tooltips, keyboard navigation, and alt text for better usability. Extensive testing was conducted across major browsers to verify layout consistency and performance stability.

Throughout the process, I worked closely with the backend team to ensure smooth integration with the Flask API, validating JSON inputs and displaying real-time results. This experience significantly boosted my frontend development skills, especially in data visualization, UI/UX principles, and client-server communication in web apps.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**SAMPLE INDIVIDUAL CONTRIBUTION REPORT:**

**Breast Cancer Detection System using Machine Learning**

MOHAMMAD KAIF

22051261

**Abstract:** This project focuses on developing a machine learning-based system for the early detection of breast cancer using clinical and diagnostic data. The goal is to build an accurate model that classifies tumors as benign or malignant. Through data preprocessing, model training, and visualization, the system assists healthcare professionals in making informed decisions. A user-friendly interface and dashboard are included to present results clearly. Overall, the solution enhances accuracy, efficiency, and interpretability in breast cancer diagnosis, aiming to support better clinical outcomes.

**Individual contribution and findings:**

My primary responsibility in the project was to design and manage the database system that supported user interactions and model predictions. I created a relational schema to store inputs, prediction outputs, and metadata required for analytics and reporting. For development and testing, I used SQLite due to its simplicity and portability. The schema ensured data normalization, consistency, and ease of querying.

I worked closely with a teammate to integrate this database into our Flask backend using Flask-SQLAlchemy. This allowed us to handle database transactions using Python objects, making the codebase cleaner and more maintainable. We implemented methods for inserting, retrieving, and managing records, including error handling and query optimization to support real-time data interaction with the machine learning model.

This process helped me gain practical experience in schema design, ORM-based database handling, and secure data operations in web applications. I also documented the structure and logic of our database layer, which proved useful for team collaboration and future scalability of the project.

Full Signature of Supervisor: Full signature of the student:

……………………………. ……………………………..

**References**

Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Computer Science*, 83, 1064–1069. 

Agarap, A. F. (2017). On Breast Cancer Detection: An Application of Machine Learning Algorithms on the Wisconsin Diagnostic Dataset. arXiv preprint arXiv:1711.07831. 

Park, Y.-S., Kim, T. H., & Min, B.-C. (2025). Detection of Breast Cancer Using Machine Learning Algorithms: A Study on Logistic Regression, K-Nearest Neighbors, and Decision Trees. *Frontiers in Health Informatics*, 14(1).

[Author(s) Unknown]. (2025). Breast Cancer Prediction Based on Multiple Machine Learning Algorithms. *PubMed*.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast Cancer Histopathological Image Classification Using Convolutional Neural Networks. *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2560–2567.