

**Karnataka State Council for Science and Technology**

**Bengaluru-560012**

**Visvesvaraya Technological University**

**Jnana Sangama, Belagavi - 590018**



**PROJECT REPORT**

**On**

**“RANK-BASED ENSEMBLE OF CNN MODELS FOR BREAST CANCER  
DETECTION IN HISTOPATHOLOGY IMAGES”**

This Project is sponsored by KSCST 46<sup>th</sup> Series of Student Project Programme  
[Project Reference Number: 46S\_BE\_1051]

Submitted in partial fulfillment of the requirement for the award of the Degree of

**BACHELOR OF ENGINEERING**

**In**

**COMPUTER SCIENCE & ENGINEERING**

**By**

**ANUJ KUMAR**

**4MT19CS026**

**POOJA P POOJARY**

**4MT19CS105**

**PRA THEEKSHA P SHETTY**

**4MT19CS111**

**PRA THEEKSHA SHETTY**

**4MT19CS112**

Under the Guidance of

**Dr. PRADEEP B S**

Professor



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

(Accredited by NBA)

**MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING**

*Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution*

*(A Unit of Rajalaxmi Education Trust®, Mangalore - 575001)*

*Affiliated to VTU, Belagavi, Approved by AICTE, New Delhi*

*Badaga Mijar, Moodabidri-574225, Karnataka*

**2022-23**

# MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution  
(A Unit of Rajalaxmi Education Trust<sup>®</sup>, Mangalore - 575001)  
Affiliated to VTU, Belugavi, Approved by AICTE, New Delhi  
Badaga Mijar, Moodabidri-574225, Karnataka



## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Accredited by NBA)

### CERTIFICATE

This is to certify that the project work entitled “**RANK-BASED ENSEMBLE OF CNN MODELS FOR BREAST CANCER DETECTION IN HISTOPATHOLOGY IMAGES**” is a bonafide work carried out by **Anuj Kumar (4MT19CS026)**, **Pooja P Poojary (4MT19CS105)**, **Pratheeksha P Shetty (4MT19CS111)**, **Pratheeksha Shetty (4MT19CS112)** in partial fulfillment for the award of degree of **Bachelor of Engineering in Computer Science & Engineering** of the **Visvesvaraya Technological University, Belagavi** during the year **2022–23**. It is certified that all corrections and suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering degree.

**Dr. Pradeep B S**

**Project Guide**

**Dr. Ravinarayana B**

**Head of the Department**

Head of the Dept. of Comp. Sci. & Engg.  
Mangalore Institute of Technology & Engineering  
Badaga Mijar, MOODABIDRI - 574 225

**Dr. Prashanth C M**

**Principal**

Mangalore Institute of Technology & Engineering  
Badaga Mijar, MOODABIDRI - 574 225

#### External Viva

Name of the Examiners

Signature with Date

1)

2)

# MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution

(A Unit of Rajalaxmi Education Trust)

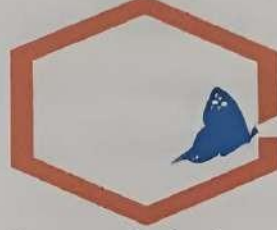
Affiliated to Visvesvaraya Technological University, Belagavi

Badaga Mijar, Moodabidri – 574 225 , Karnataka

## DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Accredited by NBA)

MITE

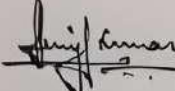



Invent Solutions

## DECLARATION


We, **Anuj Kumar**(4MT19CS026), **Pooja P Poojary**(4MT19CS105), **Pratheeksha P Shetty**(4MT19CS111) and **Pratheeksha Shetty**(4MT19CS112) students of 8<sup>th</sup> semester BE in Computer Science & Engineering, **Mangalore Institute of Technology and Engineering, Moodabidri**, hereby declare that the project work entitled “**RANK-BASED ENSEMBLE OF CNN MODELS FOR BREAST CANCER DETECTION IN HISTOPATHOLOGY IMAGES**”, submitted to the **Visvesvaraya Technological University, Belagavi** during the academic year **2022-23**, is a record of an original work done by us under the guidance of **Dr. Pradeep B S**, Professor, Department of Computer Science & Engineering , Mangalore Institute of Technology and Engineering Moodabidri. This project work is submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science & Engineering. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree.

Date:

 Anuj Kumar

 Pooja P Poojary

 Pratheeksha P Shetty

 Pratheeksha Shetty

Place: Moodabidri



# Karnataka State Council for Science and Technology

(An autonomous organisation under the Dept. of Science & Technology, Govt. of Karnataka)

Indian Institute of Science Campus, Bengaluru – 560 012

Telephone: 080-23341652, 23348848, 23348849, 23348840

Email: office.kscst@iisc.ac.in, office@kscst.org.in Website: www.kscst.iisc.ernet.in, www.kscst.org.in

**Dr. U T Vijay**  
Executive Secretary

24th April, 2023

Ref: 7.1.01/SPP/33

To,  
The Principal,  
Mangalore Institute of Technology and Engineering,  
Badaga Mijar, Moodabidri,  
Mangaluru - 574 225.

Dear Sir/Madam,

Sub : Sanction of Student Project - 46th Series: Year 2022-2023

**Project Proposal Reference No. : 46S\_BE\_1051**

Ref : Project Proposal entitled **RANK BASED ENSEMBLE OF CNN MODELS FOR BREAST CANCER DETECTION IN HISTOPATHOLOGY IMAGES**

We are pleased to inform that your student project proposal referred above, has been approved by the Council under "Student Project Programme - 46th Series". The project details are as below:

<b>Student(s)</b>	Mr. ANUJ KUMAR	<b>Department</b>	COMPUTER SCIENCE AND ENGINEERING
	Ms. POOJA P POOJARY		
	Ms. PRATHEEKSHA P SHETTY		
	Ms. PRATHEEKSHA SHETTY		
<b>Guide(s)</b>	Dr. PRADEEP B S	<b>Sanctioned Amount (in Rs.)</b>	4,000.00

### Instructions:

- The project should be performed based on the objectives of the proposal submitted.
- Any changes in the project title, objectives or students team is liable for rejection of the project and your institution shall return the sanctioned funds to KSCST.
- Please quote your project reference number printed above in all your future correspondences.
- After completing the project, 2 to 3 page write-up (synopsis) needs to be uploaded on to the following Google Forms link <https://forms.gle/nWTaJjvrwzp3Wmvt6>. The synopsis should include following:
  - Project Reference Number
  - Title of the project
  - Name of the College & Department
  - Name of the students & Guide(s)
  - Keywords
  - Introduction / background (with specific reference to the project, work done earlier, etc) - about 20 lines
  - Objectives (about 10 lines)

**46S\_BE\_1051**

- 8) Methodology ( about 20 lines on materials, methods, details of work carried out, including drawings, diagrams etc)
  - 9) Results and Conclusions (about 20 lines with specific reference to work carried out)
  - 10) Scope for future work (about 20 lines).
- e) In case of incometed projects, the sanctioned amount shall be returned to KSCST.
  - f) The sanctioned amount will be transferred by NEFT to the bank account provided by the College/Institute.
  - g) The sponsored projects evaluation will be held in the Nodal Centre/Online Mode and the details of the same will be intimated shortly by email / Website announcement.
  - h) After completion of the project, soft copy of the project report duly signed by the Principal, the HoD, Guide(s) and studetn(s) shall be uploaded in the following Google Forms Link <https://forms.gle/YWz56TrGg7fnSQgc7>. The report should be prepared in the format prescribed by the university.

Please visit our website for further announcements / information and for any clarifications please email to [spp@kscst.org.in](mailto:spp@kscst.org.in)

Thanking you and with best regards,

Yours sincerely,



(U T Vijay)

Copy to:

- 1) The HoD  
COMPUTER SCIENCE AND ENGINEERING  
MANGALORE INSTITUTE OF TECHNOLOGY AND ENGINEERING, MOODBIDRI
- 2) Dr. PRADEEP B S  
COMPUTER SCIENCE AND ENGINEERING  
MANGALORE INSTITUTE OF TECHNOLOGY AND ENGINEERING, MOODBIDRI
- 3) THE ACCOUNTS OFFICER  
KSCST, BENGALURU

## ABSTRACT

Breast cancer is one of the most common types of cancer in women worldwide. Breast cancer can be invasive or non-invasive. Invasive breast cancer can spread into surrounding tissues also in distant organs whereas non-invasive (in-situ) breast cancer spread surrounding the milk ducts or lobules of the mammary glands in the breast. However early detection followed by the diagnosis can not only reduce the treatment cost but also increase the chance of survival. One of the most effective methods for the detection of tumor malignancies is breast histopathology image analysis. However, manual image analysis for breast histopathology is subjective, time-consuming, and prone to human error. Convolutional neural network (CNN) based models have produced encouraging results for the classification of breast histopathology imaging. In this work, we will combine the results of three transfer learning experiments to create a model utilizing a novel rank-based ensemble method rather than depending on a single CNN model, namely GoogleNet, VGG16, and MobileNetV3 Small. Here, it is suggested that the 2-class classification problem for breast histopathology pictures can be handled using the Gompertz function. This project has been developed using technologies such as deep learning and ensemble method for breast cancer image classification.

**Keywords:** Breast Cancer, Histopathology Image, Deep Learning, Ensemble Method, Gompertz Function.

## ACKNOWLEDGEMENTS

The satisfaction and the successful completion of this project would be incomplete without the mention of the people who made it possible, whose constant guidance encouragement crowned our efforts with success.

This project is made under the guidance of **Dr. Pradeep B S, Professor**, in the Department of Computer Science and Engineering. We would like to express my sincere gratitude to our guide for all the helping hand and guidance in this project.

We would like to thank our project coordinator **Mr. Annappa Swamy D R, Associate Professor** in the Department of Computer Science and Engineering, for his cordial support, valuable information and guidance, which helped us in completing this project through the various stages.

We would like to express appreciation to **Dr. Ravinarayana B, Associate Professor and Head**, Department of Computer Science and Engineering, for his support and guidance.

We would like to thank our Principal **Dr. Prashanth C M**, for encouraging us and giving us an opportunity to accomplish the project.

We are thankful to **KSCST** for selecting and providing the funding for the project.

We also thank our management who helped us directly or indirectly in the completion of this project.

Our special thanks to faculty members and others for their constant help and support.

Above all, we extend our sincere gratitude to our parents and friends for their constant encouragement with moral support.

**Anuj Kumar**  
**Pooja P Poojary**  
**Pratheeksha P Shetty**  
**Pratheeksha Shetty**

# TABLE OF CONTENTS

<b>Contents</b>	<b>Page No.</b>
<b>ABSTRACT</b>	i
<b>ACKNOWLEDGEMENT</b>	ii
<b>TABLE OF CONTENTS</b>	iii
<b>LIST OF FIGURES</b>	v
<b>LIST OF TABLES</b>	vi

<b>Chapter No.</b>	<b>TITLE</b>	
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
1.1	Problem Statement	1
1.2	Objectives	2
1.3	Features	2
1.4	Scope	3
1.5	Organization of the Report	3
<b>2</b>	<b>LITERATURE SURVEY</b>	<b>5</b>
2.1	Existing System	5
2.2	Related Work	6
2.3	Limitations of Existing System	8
2.4	Proposed System	8
<b>3</b>	<b>SYSTEM REQUIREMENT SPECIFICATION</b>	<b>10</b>
3.1	Product Functionality	10
3.2	Specific Requirements	11
	3.2.1 Hardware Requirements	11
	3.2.2 Software Requirements	11
	3.2.3 Functional Requirements	13
	3.2.4 Non-functional Requirements	13
<b>4</b>	<b>GANTT CHART</b>	<b>14</b>



<b>5</b>	<b>SYSTEM DESIGN</b>	16
5.1	Architectural Diagram	16
5.2	Activity Diagram	17
5.3	Data flow Diagram	19
<b>6</b>	<b>IMPLEMENTATION AND TESTING</b>	20
6.1	Algorithm and Pseudo Code	20
6.1.1	Code Snippet for implementing base html page	20
6.1.2	Code Snippet for implementing index html page	21
6.1.3	Code Snippet for implementing main.js Script	22
6.1.4	Code Snippet for implementing function in app.py	24
6.2	Test Levels	25
6.2.1	Unit Testing	25
6.2.2	Integration Testing	25
6.2.3	System Testing	25
6.2.4	Acceptance Testing	25
6.3	Test Cases	26
<b>7</b>	<b>RESULTS AND DISCUSSION</b>	28
<b>8</b>	<b>CONCLUSION</b>	35
8.1	Scope for future work	35
	<b>REFERENCES</b>	36

## LIST OF FIGURES

<b>Figure No.</b>	<b>TITLE</b>	<b>Page No.</b>
5.1.1	Architectural Diagram of Proposed System	16
5.2.1	Activity Diagram of Proposed System	18
5.3.1	Data Flow Diagram of Proposed System	19
7.1	Confusion matrices generated by the CNN models on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively	30
7.2	Confusion matrices generated by the proposed ensemble model on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively	30
7.3	ROC curves obtained by the proposed ensemble model from the four sets of magnification i.e., a) 40X, b) 100X, c) 200X and d) 400X	31
7.4	Screenshot of Home Page(Part-1)	32
7.5	Screenshot of Home Page(Part-2)	32
7.6	Screenshot of before selecting a image	33
7.7	Screenshot of after selecting a image	33
7.8	Screenshot of result page(Part-1)	34
7.9	Screenshot of result page(Part-2)	34

## LIST OF TABLES

<b>Table No.</b>	<b>TITLE</b>	<b>Page No.</b>
4.1	Gantt chart of planning and scheduling of project	14
4.2	Gantt chart	15
6.3.1	Test Cases for Project	27
7.1	Accuracy generated by the CNN models on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively	28

**Chapter – 1**  
**INTRODUCTION**

## CHAPTER 1

### INTRODUCTION

Breast cancer is a significant global health concern, being the most prevalent type of cancer among women worldwide. According to the World Health Organization (WHO), breast cancer ranks second in terms of incidence, just after lung cancer, and fifth in terms of mortality among female cancer types (Organization et al., 2018). The development of cancer occurs when healthy cells in the breast undergo abnormal changes and uncontrollable multiplication, resulting in the formation of a mass or cluster of cells known as a tumor. Breast cancer can be invasive or non-invasive. Invasive breast cancer can spread into surrounding tissues also in distant organs whereas non-invasive (in-situ) breast cancer spread surrounding the milk ducts or lobules of the mammary glands in the breast. However early detection followed by the diagnosis can not only reduce the treatment cost but also increase the chance of survival. False results can occur with mammograms and CT scans, and patients are exposed a little amount of ionizing radiation during the process. Pathologists can histologically assess the microscopic organelles and structures of the breast tissue using breast cancer biopsies. The creation of histology slides is done to observe and evaluate breast tissue under a microscope in order to determine whether the breast tumor as (i) healthy or normal breast tissue, (ii) benign lesions, (iii) in-situ carcinoma, or (iv) invasive cancer. Even among pathologists with years of experience, the manual examination process is lengthy and prone to inter-observer variations. To address this challenge, several researchers have recently explored the use of ensemble learning-based methods and deep learning approaches in various domains, including breast cancer diagnosis.

#### 1.1 Problem Statement

Histopathology images are commonly used to diagnose breast cancer, but accurate interpretation of these images can be challenging. While deep learning models have shown promising results in detecting breast cancer, their accuracy can be limited by factors such as model selection and training data.

To address this challenge, we propose a novel Rank-based ensemble method for Breast Cancer Detection using histopathology images by combining outcomes of three transfer learning CNN models. This approach aims to reduce the dependency on a single CNN model and improve the accuracy and performance of breast cancer detection.

## 1.2 Objectives

The objective is to develop a novel rank-based ensemble method for breast cancer detection using histopathology images. This will be achieved by combining the outcomes of three transfer learning CNN models, namely GoogleNet, VGG16, and MobileNetV3 Small. The ensemble model will be designed using the Gompertz function for solving a 2-class classification problem of breast histopathology images. The project aims to produce better classification results for different levels of magnification, including 40X, 100X, 200X, and 400X, which will provide a more accurate and robust tool for breast cancer detection.

## 1.3 Features

**1. Rank-based ensemble:** Rather than relying on a single CNN model to make predictions, this approach involves using an ensemble of models and ranking their outputs to make a final prediction. This can help to reduce the impact of any individual model's biases or errors, resulting in more accurate predictions overall.

**2. CNN architecture:** CNN models are commonly used in image analysis tasks due to their ability to learn features directly from the raw image data. The specific architecture of the CNN models used in this approach may vary, but typically involves several layers of convolution and pooling operations, followed by one or more fully connected layers to perform classification.

**3. Image pre-processing:** Before being input to the CNN models, the breast tissue images may undergo various pre-processing steps to enhance image quality and remove noise. This could include steps such as color normalization, contrast enhancement, and image resizing.

**4. Data augmentation:** To improve the generalization ability of the CNN models, data augmentation techniques may be used to generate additional training images by applying transformations such as rotation, flipping, and zooming to the original images.

**5. Transfer learning:** Rather than training the CNN models from scratch on the breast tissue images, transfer learning techniques may be used to leverage pre-trained CNN models that have been trained on large datasets of general images (such as ImageNet). This can help to reduce training time and improve the accuracy of the final model.

**6. Feature extraction:** In addition to using the CNN models for classification, they can also be used to extract features from the breast tissue images. These features can then be fed into other machine learning models for tasks such as clustering or visualization.

## 1.4 Scope

The scope of this approach is significant, as breast cancer is one of the most common cancers among women, and early detection is critical for successful treatment. Histopathology images provide a wealth of information that can aid in the diagnosis of breast cancer, but interpreting these images requires extensive expertise and can be time-consuming. The use of deep learning-based approaches such as rank-based ensembles of CNN models can help improve the accuracy and speed of breast cancer diagnosis by automating the process of analyzing histopathology images.

## 1.5 Organization of the Report

Chapter 1: Introduction deals with the organization of the report, objectives of the proposed model, and a brief introduction of our proposed model.

Chapter 2: Literature Survey focuses on the initial study conducted before the development of the project. It includes observations and details about the existing system and its disadvantages.

Chapter 3: System Requirement Specification provides information about the product functionality, specific requirements and resource requirements.

Chapter 4: Gantt Chart deals with the use of a Gantt chart, which is a type of bar chart developed by Henry Gantt. It is used to illustrate a project schedule.

Chapter 5: System Design deals with the formal description of the system through the Architectural diagram, the flow of activities through the Activity diagram, and the graphical representation of data flow in the system through the Data Flow Diagram.

Chapter 6: Implementation and Testing deals with the steps involved in creating the project work, which are explained using code snippets for reader convenience. It also covers various test cases and test levels, including unit testing, system testing, integration testing, and acceptance testing.

Chapter 7: Results and Discussion focuses on the graphical user interface of the project, showcasing the application's output, and highlighting performance comparisons.

Chapter 8: Conclusion deals with providing a comprehensive summary of the project and discussing potential ways for future enhancements to expand its scope.



**Chapter – 2**  
**LITERATURE SURVEY**

## CHAPTER 2

### LITERATURE SURVEY

A literature survey shows the various analysis and research made in the field of interest and results already published, taking into account the various parameters of the project and the extent of the project. It includes researches made by various analysts-their methodology and the conclusion they have arrived at. It is the most important part of the report as it gives the direction in the area of research.

#### 2.1 Existing System

**Mammography:** Mammography is a common screening method for finding breast cancer. Images of the breast tissue are produced using low-dose X-rays, and these images can then be examined for indications of malignancy.

**Magnetic Resonance Imaging (MRI):** Another imaging method that can be used to find breast cancer is magnetic resonance imaging (MRI). To produce finely detailed images of the breast tissue, it uses a magnetic field and radio waves.

**Ultrasound:** Breast cancer can also be identified via ultrasound imaging. Images of the breast tissue are produced using high-frequency sound waves, and these images can then be examined for indications of malignancy.

**Biopsy:** During a biopsy, a small sample of breast tissue is removed and examined under a microscope for indications of malignancy. Numerous methods, such as needle biopsy, core biopsy, and surgical biopsy, can be used to accomplish this.

**Computer-Aided Detection (CAD):** CAD systems analyse mammography images and identify potential breast cancer indicators by using computer algorithms. This can help radiologists find possible tumors that could go undetected during visual inspection.

**Artificial intelligence (AI):** Deep learning models and other AI systems have demonstrated promising results in the early identification of breast cancer. These systems train models that can precisely identify images as normal or malignant using vast datasets of histopathology or mammography images.

## 2.2 Related Work

**DeepBreastNet by Fatih Demir [1]:** In this paper, a deep learning model called DeepBreastNet was proposed for automated breast cancer detection from histopathological images. The model is based on a convolutional neural network (CNN) architecture and was trained using a sizable dataset of mammography pictures. The algorithm divides mammograms into two categories: normal and abnormal, and has demonstrated good results in a number of investigations. In particular, the model achieved an accuracy on par with that of skilled radiologists, and it has the potential to increase the diagnosis of breast cancer while decreasing false-positive and false-negative outcomes. Overall, DeepBreastNet is a promising advancement in the field of breast cancer detection and has the potential to have a big impact on millions of women's lives around the world. Along with being extremely accurate, DeepBreastNet is also made to be scalable and adaptable to various environments and demographics. The model can be trained on data from diverse geographic regions, ethnicities, and age groups to ensure that it can be utilized in a number of scenarios. DeepBreastNet may be utilized to increase detection rates and save lives in a variety of communities around the world thanks to its scalability and adaptability, making it a potent weapon in the fight against breast cancer.

**Gamma function-based ensemble of CNN models for breast cancer detection in histopathology images by Samriddha Majumdar et al. [2],** the authors propose a new ensemble learning approach for breast cancer detection in histopathology images. The authors note that while deep learning techniques have shown promise in detecting breast cancer, there is still room for improvement. One area for improvement is in the use of ensembles of models, which can improve performance by combining the predictions of multiple models. The authors propose an ensemble approach based on the gamma function, which is a well-known mathematical function used in statistics and probability theory. The proposed approach involves training multiple convolutional neural network (CNN) models on different subsets of the dataset, and then combining their predictions using the gamma function. The authors show that this approach can improve performance compared to using a single CNN model. To evaluate their approach, the authors conduct experiments on two publicly available datasets of breast cancer histopathology images. They compare their ensemble approach to several other state-of-the-art methods, including single CNN models and other ensemble approaches. The results show that the proposed ensemble approach achieves better

performance than the other methods on both datasets. The authors also perform ablation studies to investigate the effect of different factors on performance, such as the number of models in the ensemble and the size of the training dataset. The results show that the proposed approach is robust to changes in these factors, and that increasing the number of models in the ensemble can improve performance up to a certain point.

**Pei-Wen Huang et al. [3] proposed a deep learning-based approach for detecting breast cancer in histopathological images.** The authors developed a novel framework that can detect breast cancer in cross-staining histopathology images, which is considered a challenging task. Their framework is based on a deep convolutional neural network (CNN) and features a multi-scale architecture that captures both local and global information from the images. The authors trained and tested their model on a dataset of 2,578 images from 88 patients, achieving an overall accuracy of 86.6%. The authors compared their proposed approach to several state-of-the-art methods and found that their model outperformed them in terms of accuracy. They also conducted a comprehensive analysis of their model's performance in various scenarios, including different staining types, tissue regions, and magnifications. Their results showed that their model is robust and can work well in different conditions. Furthermore, the authors also conducted a sensitivity analysis to determine the effect of different hyperparameters on their model's performance. They found that the choice of network architecture and the number of training epochs had a significant impact on the model's accuracy. The authors also compared their model to the interpretation of a pathologist and found that their model achieved a comparable accuracy.

**Hameed et al. [4] developed an ensemble of deep learning models for the classification of breast cancer histopathology images.** The authors focused on utilizing deep learning models due to their effectiveness in classifying histopathology images with high accuracy. The ensemble of models developed by the authors combined multiple deep learning models, including Convolutional Neural Networks (CNN), to achieve higher accuracy rates. The authors used the publicly available BreakHis dataset to train and test their ensemble of models. The dataset consisted of 7,909 breast biopsy images, which were classified into four categories: normal, benign, in situ carcinoma, and invasive carcinoma. The authors preprocessed the dataset by resizing and normalizing the images to improve the efficiency of the training process. The authors compared their model's performance with other state-of-the-art models and found that their ensemble of models outperformed other models in terms of

accuracy, sensitivity, and specificity. The authors identified some limitations of their research, including the limited size of the dataset, the lack of diversity in the dataset, and the potential bias in the dataset. The authors suggested that future research should focus on addressing these limitations and exploring the use of larger and more diverse datasets for developing deep learning models for breast cancer detection.

### 2.3 Limitations of Existing System

**False positives and false negatives:** False positive and negative results can sometimes result from mammography, MRI, and ultrasound imaging, necessitating unnecessary biopsies or failing to detect cancer.

**Limited accessibility:** Not all patients may have access to certain imaging procedures, such as MRI.

**Dependence on operator skill:** The proficiency of the operator performing imaging and biopsy procedures may have an impact on the accuracy of the process.

**Limited sensitivity and specificity:** Certain kinds of breast cancer may be harder to detect with some imaging modalities.

**High price:** Imaging and biopsy procedures can be costly, making some patients unable to afford them.

### 2.4 Proposed System

We have proposed an ensemble of deep learning models for the detection of breast cancer in histopathology images. At first, we obtain the decision score for an image from three standard CNN models: GoogleNet, MobileNetV3 Small, VGG16, and the next objective is the fusion of decision scores using the Gompertz Function function employing a fuzzy rank based ensemble for making the final decision. Since ensembling is a potent tool for incorporating the discriminating properties of all the contributing models, a novel ensemble technique has been developed to fuse the decision scores of the aforementioned models. Fuzzy fusion outperforms standard ensemble approaches because it uses adaptive priority depending on the confidence scores of the classifiers for each sample to be predicted. The Gompertz function has exponential growth before reaching an asymptote, which is helpful for ensembling the CNN models' decision scores because a classifier's predicted decision

score rarely approaches zero. Deep learning models can be used to automatically identify and categorize the presence of malignant cells in breast tissue in the context of breast histopathology image analysis. For the early diagnosis and treatment of breast cancer, one of the major causes of cancer-related deaths in women globally, the accuracy of such models is essential. Deep learning models can perform better in breast histopathology image processing when ensemble learning is used. For instance, many deep neural networks may be trained using various hyper parameters or on various subsets of the training data. An ensemble model, which can attain more accuracy and resilience than any of the individual models, can then be created by combining the predictions of these separate models.

**Chapter – 3**

**SYSTEM REQUIREMENT  
SPECIFICATION**

## CHAPTER 3

### SYSTEM REQUIREMENT SPECIFICATION

A system analysis is a description of a software system to be developed. It lays out functional and non-functional requirements, and may include a set of use cases that describe user interactions that the software must provide.

#### 3.1 Product Functionality

Many studies have employed deep learning to find and categorize breast cancer in histopathology images. The histopathology photos give a thorough view of the breast tissue and make it possible to identify malignant cells, but they need to be meticulously manually interpreted by qualified pathologists. Pathologists can use deep learning algorithms to precisely identify and categorize malignant cells, which can help them with their diagnosis and treatment choices. Deep learning techniques commonly use a convolutional neural network (CNN) architecture when applied to breast histopathology pictures. The CNN processes the images via multiple layers of convolutional and pooling procedures in order to extract features from the images. The classification task is subsequently carried out by these features being sent into fully connected layers. Deep learning techniques commonly use a convolutional neural network (CNN) architecture when applied to breast histopathology pictures. The CNN processes the images via multiple layers of convolutional and pooling procedures in order to extract features from the images. The classification task is subsequently carried out by these features being sent into fully connected layers. Nevertheless, there are some difficulties with applying deep learning to breast histopathology images. With changes in staining, tissue preparation, and picture quality, the data is frequently highly varied. Due to this, it may be challenging to complete classification tasks with a high degree of accuracy. Deep learning models can also be tricky to interpret, making it difficult to comprehend how they generate their predictions.

A machine learning technique called ensemble learning combines different models to increase the reliability and accuracy of predictions. The accuracy of breast cancer classification tasks has been enhanced using ensemble approaches in the setting of breast histopathology images. In order to reach a judgement, this process requires training numerous models on the same dataset and integrating their predictions. In breast histopathology images,



a variety of ensemble methods, such as bagging, boosting, and stacking, can be applied. The process of bagging entails training numerous models on various subsets of the data and aggregating their predictions via a voting mechanism. Boosting involves training models one after the other, each one concentrating on the mistakes made by the preceding models. Using a meta-learner to combine the predictions of numerous models with various architectures or characteristics is known as stacking. It has been demonstrated that using ensemble approaches increases the precision of breast cancer classification tasks in histopathology images. Ensemble approaches can get beyond the weaknesses of individual models and produce more reliable and accurate forecasts by combining the strengths of various models. Additionally, ensemble methods can offer a level of prediction uncertainty that can help direct further research and decision-making. Ensemble approaches do, however, have significant drawbacks. They can be computationally costly and demand a significant amount of practise data. Furthermore, ensemble methods can be challenging to interpret, making it difficult to comprehend how they generate their predictions. However, ensemble methods continue to be a valuable tool for enhancing the reliability and accuracy of tasks involving breast cancer classification in histopathology images.

## **3.2 Specific Requirements**

### **3.2.1 Hardware Requirements**

Hardware requirements refer to the physical parts of a computer and related devices. Internal hardware devices include motherboards, hard drives and RAM External hardware devices include monitors, keyboards, mouse, printers, and scanner.

- Processor: Intel i5/i7, 1.8GHz machine or above.
- Hard Drive: 240GB or above.
- Memory (RAM): 8GB RAM or more.
- Monitor: 64 bit.
- Graphics card: NVIDIA RTX 30 series or above.

### **3.2.2 Software Requirements**

Software requirement is a field within Software Engineering that deals with establishing the Deeds of stakeholders that are to be solved by the software. The software requirements of our

project given below:

## **1. HTML**

HTML, also known as Hyper Text Markup The most used mark-up language for web pages is language. By indicating structural semantics for text elements like headers, paragraphs, lists, etc. as well as for links, quotes, and other objects, it offers a way to generate structured texts. It enables the embedding of objects and pictures and can be applied to the development of interactive forms. Within the text of the web page, it is written as HTML elements made up of "tags" enclosed in angle brackets. It can load or include programmes written in languages like JavaScript that modify how HTML processors, like Web browsers, parse HTML, as well as Cascading Style Sheets (CSS), which specify how text and other content should look and be organized.

## **2. CSS**

The goal of CSS, the most recent development of the Cascading Style Sheets language, is to expand CSS. It works in conjunction with HTML to provide content structure, and CSS3 is used to format that material. It works in conjunction with HTML to provide content structure, and CSS3 is used to format that material. Font characteristics, colour schemes, text alignments, graphics, background pictures, tables, and other elements are all within its control. Additional features offered by this tool include absolute, fixed, and relative positioning of different elements. Due to its significance in presenting more alternatives throughout the design process, CSS3 is strongly advised to assist in the development of highly dynamic online pages.

## **3. JAVASCRIPT**

A dynamic computer programming language is called JavaScript. Its implementations enable client-side script to interact with users and create dynamic pages, and it is most frequently used as a component of web pages. It is an object-oriented programming language that may be interpreted. JavaScript is a computer language that is largely utilised by Web browsers to give users an engaging and dynamic experience. The majority of the features and programmes that make the Internet so crucial to modern life are written in JavaScript in one way or another. The most popular application for JavaScript is as a client-side scripting language. This indicates that HTML pages contain JavaScript code.

## 4. PYTHON

Python is renowned for having simple syntax, which makes it a great language for new programmers to learn. Additionally, it has a sizable collection of libraries and frameworks that offer strong tools for data analysis, machine learning, and scientific computing, including NumPy, Pandas, Scikit-learn, TensorFlow, and PyTorch. Python is an interpreted language, thus it doesn't require compilation before use, making it simple to test and debug programmes. Additionally, it supports procedural, object-oriented, and functional programming paradigms, making it a flexible language with a wide range of applications.

### 3.2.3 Functional Requirements

Functional requirements describe the system function in detail, its input and output. Functional requirements may explicitly state what the system should do. These requirements depend on the type of software being developed, the expected users of the software. It gives the overall details of how the proposed system works.

- Input: Users should be able to upload medical(histopathology) images to the system.
- System requirement: The system should be Operating system Windows 7 or above and browser Chrome v8 or Edge.
- Pre-processing: The system includes pre-processing features that clean and normalize the uploaded images to ensure they are suitable for analysis.
- Image Analysis: The system will have a robust image analysis module that uses machine learning algorithms to accurately detect and classify breast cancer cells.
- Output: The system should generate comprehensive diagnostic reports for each image uploaded, including information on the type of breast cancer.

### 3.2.4 Non-Functional Requirements

Non-functional requirements are those that have no direct bearing on the particular function that the system provides. Instead of describing particular behaviour, they define the standards that can be applied to evaluate how a system works. They may relate to emergent system characteristics like dependability, response time, and store occupancy. Non-functional requirements are brought about by user needs due to financial restrictions, organisational rules, and the requirement for software interoperability.

**Chapter – 4**  
**GANTT CHART**

## CHAPTER 4

### GANTT CHART

A Gantt chart is a type of bar chart, developed by Henry Gantt that illustrates a project schedule. Gantt charts illustrate the start and finish of the terminal elements and summary elements of the project. Terminal elements and summary elements comprise the work breakdown structure of the project.

The following is the Gantt chart of the project “Rank-based Ensemble of CNN models for breast cancer detection in histopathology images”.

Number	Task	Start	End	Duration(d ays)
1	Synopsis	10-Oct-2022	25-Oct-2022	16
2	Presentation on idea	27-Oct-2022	27-Oct-2022	1
3	Software Requirement Specification	07-Nov-2022	18-Nov-2022	12
4	System Design	10-Dec-2022	24-Dec-2022	15
5	Implementation	6-Feb-2023	27-Apr-2023	80
6	Presentation on work progress	28-Apr-2023	28-Apr-2023	1
7	Testing	29-Apr-2023	02-May-2023	5
8	Result and Report	03-May-2023	05-May-2023	3

Table 4.1: Gantt chart of planning and scheduling of project

ACTIVITY/ MONTH	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL
SYNOPSIS										
PRESENTATION ON IDEA										
SRS										
DESIGN										
IMPLEMENTATI ON										
TESTING										
REPORT										

Table 4.2: Gantt chart

**Chapter – 5**  
**SYSTEM DESIGN**

## CHAPTER 5

# SYSTEM DESIGN

### 5.1 Architectural Diagram

An architecture description is a formal description of a system, organized in a way that supports reasoning about the structural properties of the system. It defines the system components or building blocks and provides a plan from which products can be procured, and system developed, that will work together to implement the overall system.

Architectural diagram helps in:

- They help with comprehension: architectural diagrams help convey complex information in a single image like there is a saying a picture is worth a thousand words.
- They improve communication and collaboration: when we are working on anything that involves multiple people, there is always a risk of miscommunication between them .so it is important to standardize information, which is where the architectural diagrams are very useful.

The architectural design gives the description about how the overall system is designed. It is specified by identifying the components defining the control and data flow between them. The arrows indicate the connection and the rectangular boxes represent the functional units. The Figure 5.1.1 shows the Architectural Design Diagram of this project.

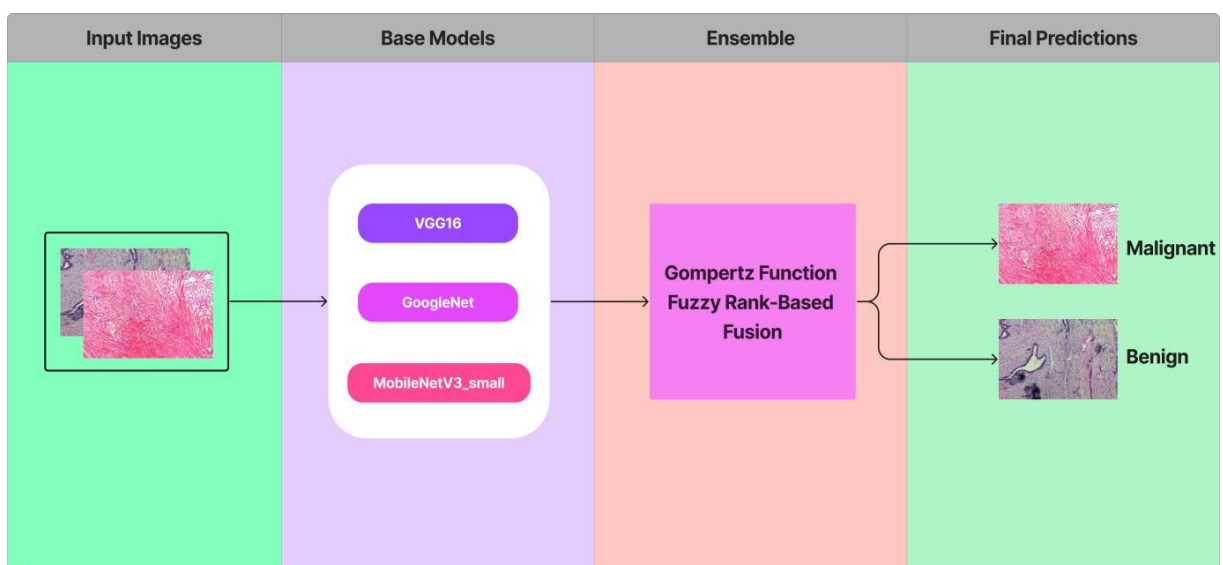


Figure 5.1.1: Architectural Diagram of Proposed System



The first step in the proposed framework is to preprocess the images by resizing them to a smaller size (224 x 224 pixels) and normalizing the pixel values. The images are then fed into three standard CNN models - GoogleNet, VGG16, and MobileNetV3 Small - which have been pre-trained on the ImageNet dataset. The models are used to extract features from the images, and the resulting feature vectors are fed into separate fully connected layers that generate decision scores for the two classes - benign and malignant.

These decision scores are then passed through a fuzzy rank function to assign ranks based on the decision scores of the above models. These ranks are then used to assign weights to the decision scores. The Gompertz function is used to assign these weights as it is a non-linear activation function that has an exponential growth phase followed by saturation to an asymptote, making it useful for ensembling the decision scores of the CNN models. The weights generated by the function and the decision scores are ensembled to get the final ensemble value. Finally, the ensemble value generated is used to determine whether the input image belongs to the "Benign" or "Malignant" class.

The BreakHis dataset is used for evaluating the performance of this framework. The BreakHis dataset is a publicly available dataset that contains 7,909 breast histopathology images. The dataset is divided into two groups based on the type of breast tissue: benign (2,480 images) and malignant (5,429 images). Each image in the dataset is of size 700 x 460 pixels and has been stained with Hematoxylin and Eosin (H&E) to highlight the morphological features of the breast tissue.

## 5.2 Activity Diagram

Activity diagram is an important diagram in UML to describe the dynamic aspects of the system. Activity diagram is basically a flowchart to represent the flow from one activity to another activity. The activity can be described as an operation of the system.

The components used in this are as follows:

Rounded Rectangle: It indicates the process.

Arrow: It indicates a transition line.

Rhombus: It indicates the decision.

Bars: It represents the start or end of concurrent activities.

Solid Circle: It represents the initial state of workflow.

Encircled Black Circle: It represents the final state of workflow.

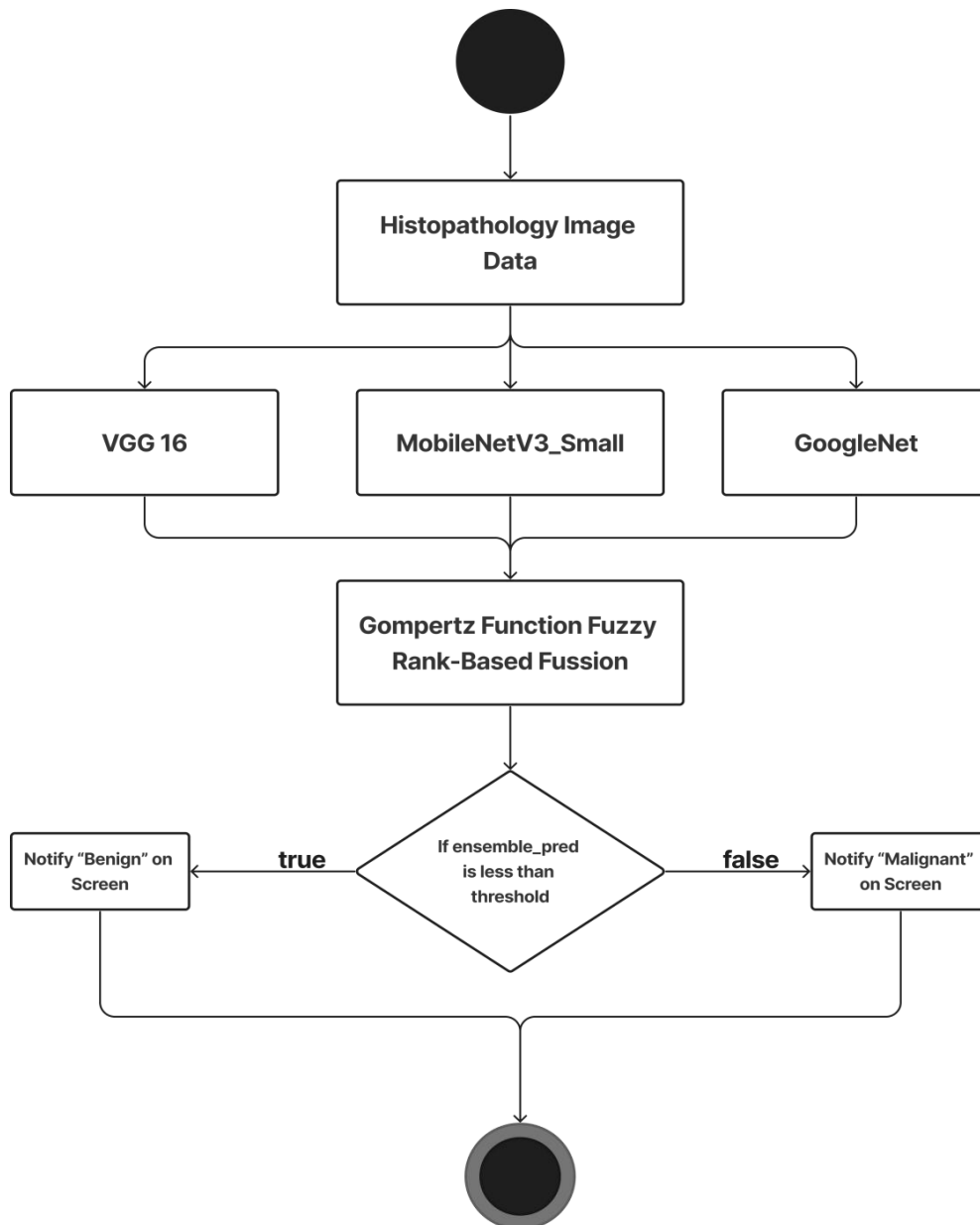


Figure 5.2.1: Activity Diagram of Proposed System

When an activity diagram is activated, control starts at the solid circle and proceeds. Initially pre-processed histopathology image data is sent to 3 models VGG16, MobileNetV3\_Small and GoogLeNet, then the decision scores of these models are sent to the ensemble function. It checks for a threshold, based on threshold values and ensemble prediction value it will notify the user whether the given histopathology image is “benign” or “Malignant” on the monitor. A bull’s eye shows the termination point. When control reaches a bull’s eye, the overall activity is complete and execution of the activity diagram ends.

### 5.3 Data Flow Diagram

A data flow diagram (DFD) is a graphical representation of the flow of data through an information system. It is a structured analysis and design technique that models how information moves between different entities within a system. The main purpose of a DFD is to depict the system's components, processes, and data flows, as well as the inputs and outputs associated with each component. It is a useful tool for understanding, analyzing, and designing information systems. The Figure 5.3.1 shows the Data Flow Diagram of Proposed System.

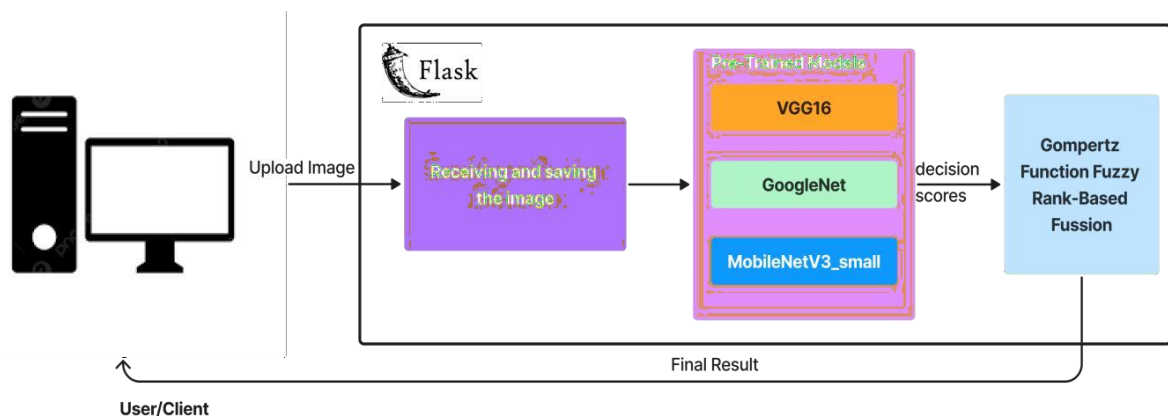


Figure 5.3.1: Data Flow Diagram of Proposed System

- User/Client: The user uploads a histopathology image to the system.
- Image Upload: The uploaded image is received by the system, and a copy of the image is stored for future reference.
- Pre-Trained Model Section: The uploaded image is then passed to the pre-trained model section, which contains three pre-trained CNN models: GoogleNet, VGG16, and MobileNetV3 Small. Each of the pre-trained models generates a decision score for the input image.
- Ensemble Function: The decision scores generated by the pre-trained models are then passed through an ensemble function. It finally determine whether the input image belongs to the Benign or Malignant class. The final prediction result is sent back to the user as output.

**Chapter – 6**

**IMPLEMENTATION AND  
TESTING**

## CHAPTER 6

### IMPLEMENTATION AND TESTING

**Implementation** is the realization of an application, or execution of a plan, idea, model, design, specification, algorithm or policy. Implementation refers to conversion of a new system design to an operation. An implementation plan is made before starting the actual implementation of the system. The new system may be totally new, replacing an existing manual or automated system or it may be a major modification to an existing system.

**Testing** is an important phase in the development life cycle of the product. During the testing, the program to be tested was executed with a set of test cases and the output of the program for the test cases was evaluated to determine whether the program is performing as expected. Errors were found and corrected by using the following testing steps and correction was recorded for future references. Thus, a series of testing was performed on the system before it was ready for implementation. An important point is that software testing should be distinguished from the separate discipline of Software Quality Assurance (SQA), which encompasses all business process areas, not just testing.

#### 6.1 Algorithm and Pseudo Code

##### 6.1.1 Code Snippet for implementing base html page

```
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <meta http-equiv="X-UA-Compatible" content="ie=edge">
  <title>Brest Cancer Detection Web App</title>
  <link href="https://cdn.bootcss.com/bootstrap/4.0.0/css/bootstrap.min.css" rel="stylesheet">
  <script src="https://cdn.bootcss.com/popper.js/1.12.9/umd/popper.min.js"></script>
  <script src="https://cdn.bootcss.com/jquery/3.3.1/jquery.min.js"></script>
  <script src="https://cdn.bootcss.com/bootstrap/4.0.0/js/bootstrap.min.js"></script>
  <link href="{{ url_for('static', filename='css/main.css') }}" rel="stylesheet">
</head>
<body>
  <nav class="navbar navbar-dark bg-dark">
    <div class="container">
      <a class="navbar-brand" href="#">Brest Cancer Detection Web App</a>
    </div>
```

```

</nav>
<div class="container">
  <div id="content" style="margin-top:2em">{% block content %}{% endblock %}</div>
</div>
</body>
<footer>
  <script src="{ { url_for('static', filename='js/main.js') } }" type="text/javascript"></script>
</footer>
</html>

```

The above code snippet is an HTML page for a breast cancer detection web application. It includes dependencies for Bootstrap, jQuery, and Popper.js, along with custom CSS and JavaScript files. This code is used to create page structure consists of a navigation bar, a content container, and a footer. It provides a foundation for building the user interface of the web application.

### 6.1.2 Code Snippet for implementing index html page

```

{% extends "base.html" %} {% block content %}
<body>
  <div class="title"><h1>Breast Cancer Early Detection</h1></div>
  <div class="text parbase">
    <p>Breast cancer(Content) .. </p>
  </div>
  <div style="display: flex; justify-content: center;">
    <div class="card">
      <h3>Can Breast Cancer Be Found Early?</h3>
      <p>Breast cancer is sometimes found after symptoms appear, but many women with breast cancer have no symptoms. This is why regular breast cancer screening is so important. Learn more.</p>
      <a href="link1">Read more</a>
    </div>
    <div class="card">
      <h3>Imaging Tests to Find Breast Cancer</h3>
      <p>Different tests can be used to look for and diagnose breast cancer. If your doctor finds an area of concern on a screening test (a mammogram), or if you have symptoms that could mean breast cancer, you will need more tests to know for sure if it's cancer. </p>
      <a href="link2">Read more</a>
    </div>
    <div class="card">
      <h3>How is Breast Cancer Treated?</h3>
      <p>The treatment for breast cancer depends on the stage of the disease, the type of cancer, and other factors.</p>
      <a href="link3">Read more</a>
    </div>
  </div>

```

```

</div><br>
<div class="text parbase">
  <p>Content...</p>
  <p>Our web application..... presence of breast cancer.</p>
</body>
<h2></h2>
<div>
  <form id="upload-file" method="post" enctype="multipart/form-data">
    <label for="imageUpload" class="upload-label">
      Select Image
    </label>
    <input type="file" name="file" id="imageUpload" accept=".png, .jpg, .jpeg">
  </form>
  <div class="image-section" style="display:none;">
    <div class="img-preview">
      <div id="imagePreview">
      </div>
    </div>
    <div>
      <button type="button" class="btn btn-primary btn-lg " id="btn-predict">Predict!</button>
    </div>
  </div>
  <div class="loader" style="display:none;margin-left: 50%;"></div>
  <h3 id="result" style="margin-left: 42%;">
    <span> </span>
  </h3>
</div>{% endblock %}

```

The above code snippet represents the content section of a web page for a breast cancer detection web application. It includes information about breast cancer and its early detection. The code also includes a form for image upload, a preview section, a prediction button, and a result section.

### 6.1.3 Code Snippet for implementing main.js Script

```

$(document).ready(function () {
  $('.image-section').hide();
  $('.loader').hide();
  $('#result').hide();
  function readURL(input) {
    if (input.files && input.files[0]) {
      var reader = new FileReader();
      reader.onload = function (e) {

```

```
reader.onload = function (e) {
    $('#imagePreview').css('background-image', 'url(' + e.target.result + ')');
    $('#imagePreview').hide();
    $('#imagePreview').fadeIn(650);
}
reader.readAsDataURL(input.files[0]);
}}
$("#imageUpload").change(function () {
    $('.image-section').show();
    $('#btn-predict').show();
    $('#result').text("");
    $('#result').hide();
    readURL(this);
});
$('#btn-predict').click(function () {
    var form_data = new FormData($('#upload-file')[0]);
    $(this).hide();
    $('.loader').show();
    $.ajax({
        type: 'POST',
        url: '/predict',
        data: form_data,
        contentType: false,
        cache: false,
        processData: false,
        async: true,
        success: function (data) {
            $('.loader').hide();
            $('#result').fadeIn(600);
            $('#result').text(' Result: ' + data);
            console.log('Success!');
        },
    });
});
});
```

The above code snippet is from the main.js file, which contains JavaScript code that performs various functions for our web application. It handles features such as uploading an image and displaying a preview, initiating the prediction process, showing a loading animation while waiting for the prediction result, making an API call to the /predict endpoint, and getting and displaying the prediction result. This file plays a crucial role in enabling user interaction and providing real-time feedback during the breast cancer detection process.



### 6.1.4 Code Snippet for implementing function in app.py

```

model1 = load_model("C:/Users/anujk/Final_year/model/model_v1.weights.best.hdf5")# load
VGG16 model here
model2 = load_model("C:/Users/anujk/Final_year/model/netv3small-4.weights.best.hdf5") # load
MobileNetV3_Small model here
model3 = load_model("C:/Users/anujk/Final_year/model/googlenet_vtrue.weights.best.hdf5") #
load GoogleNet model here

# Define the fuzzy rank function
def fuzzy_rank(scores):
    order = np.argsort(-scores)
    rank = np.empty(len(scores), int)
    rank[order] = np.arange(len(scores))
    return rank

# Define the Gompertz function
def gompertz(x, a, b, c):

    return a * np.exp(-b * np.exp(-c * x))
def model_predict(img_path, model1, model2, model3):
    img = ... # Preprocessing the image takes place
    pred1 = model1.predict(img)[:, 1]
    pred2 = model2.predict(img)[:, 1]
    pred3 = model3.predict(img)[:, 1]
    ..
    ranked_models = fuzzy_rank(scores)
    weights = []
    for i in range(len(ranked_models)):
        rank = i + 1
        weight = gompertz(rank, a=1, b=1, c=0.5)
        weights.append(weight)
    weights = weights / np.sum(weights)
    ensemble_pred = (weights[0] * pred1) + (weights[1] * pred2) + (weights[2] * pred3)
    ...
    return ensemble_label

```

The given code snippet demonstrates a breast cancer detection model implemented in Python. It combines pre-trained models (VGG16, MobileNetV3\_Small, and GoogleNet) to make predictions. The code includes functions for fuzzy ranking and the Gompertz function to assign weights to the predictions. The `model_predict` function processes an input image, obtains predictions from each model, combines them using the assigned weights, and returns the predicted label.

## 6.2 Test Level

Testing is part of Verification and Validation. Testing plays a very critical role for quality assurance and for ensuring the reliability of the software.

The objective of testing can be stated in the following ways.

- A successful test is one that uncovers as-yet-undiscovered bugs.
- A better test case has high probability of finding un-noticed bugs.
- A pessimistic approach of running the software with the intent of finding errors.

Testing can be performed in various levels like unit test, integration test and system test.

### 6.2.1 Unit Testing

Unit testing tests the individual components to ensure that they operate correctly. Each component is tested independently, without other system component. This system was tested with the set of proper test data for each module and the results were checked with the expected output. Unit testing focuses on verification effort on the smallest unit of the software design module.

### 6.2.2 Integration Testing

Integration testing is another aspect of testing that is generally done in order to uncover errors associated with the flow of data across interfaces. The unit-tested modules are grouped together and tested in small segment, which makes it easier to isolate and correct errors. This approach is continued until we have integrated all modules to form the system as a whole.

### 6.2.3 System Testing

System testing tests a completely integrated system to verify that it meets its requirements. The proposed system is used to detect the given histopathology image is Benign or Malignant and notify the User.

### 6.2.4 Acceptance Testing

After the system tests for its accuracy, efficiency, and reliability in detecting cancer in histopathology images. The test cases will be designed to cover all possible scenarios and

edge cases to ensure that the model performs well in real-world scenarios. The test results will be evaluated against the predetermined acceptance criteria to determine whether the model meets the specified requirements and is fit for deployment.

### 6.3 Test Cases

A test case is a software testing document, which consists of events, action, input, output, expected result and actual result. Technically a test case includes test description, procedure, expected result and remarks. Test cases should be based primarily on the software requirements and developed to verify correct functionality and to establish conditions that reveal potential errors.

Individual PASS/FAIL criteria are written for each test case. All the tests need to get a PASS result for proper working of an application.

Test No.	Test Case	Expected Output	Actual Output	Status
1.	Test three models on Break-His dataset for different levels of magnification	Accuracy should be greater than 50% for all models	Accuracy was less than 50% for all models	<b>Fail</b>
2.	Optimize models and tune hyperparameters	Accuracy should be greater than 50% for all models	Accuracy improved to more than 60% for all models	<b>Pass</b>
3.	Test Gompertz function with different parameter values	Different parameter values should result in different prediction results	Different parameter values result in different prediction results	<b>Pass</b>
4.	Upload image with correct format	Save a copy of the uploaded image in local file and pass it to the model to generate classification result	Image not saved in local storage, image path is empty and model does not receive any input image	<b>Fail</b>

<b>5.</b>	Fixed the code to use os module to save and get the path of saved images	Save a copy of the uploaded image in local file and pass it to the model to generate classification result	Image was successfully saved in local store but Model was not correctly pre-processing images, resulting in incorrect classification	<b>Fail</b>
<b>6.</b>	Optimized the image pre-processing code for the model	Image should be successfully saved in local store and passed to model, it has to correctly pre-processing images, resulting in correct classification	Image was successfully saved in local store and passed to model. Model was not correctly pre-processing images, and correct classification result was obtained	<b>Pass</b>

Table 6.3.1: Test Cases for Project

**Chapter – 7**  
**RESULTS AND DISCUSSION**

**CHAPTER 7****RESULTS AND DISCUSSION**

In this deep learning project, we trained three different models, namely VGG16, GoogleNet, and MobileNetV3, for the task of histopathology image detection. The training process involved using a dataset of histopathological images and implementing the respective architectures of each model.

The training results showed promising performance for all three models. The accuracy of each model where evaluated. The results showed in Table 7.1.

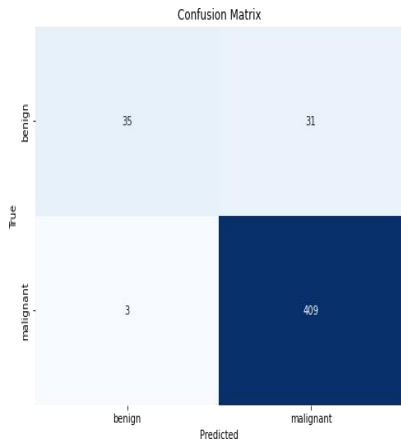
Magnification	VGG16 Accuracy	MobileNetV3 Accuracy	GoogleNet Accuracy
40X	0.928	0.769	0.688
100X	0.898	0.752	0.746
200X	0.827	0.667	0.737
400X	0.872	0.776	0.848

Table 7.1: Accuracy generated by the CNN models on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively

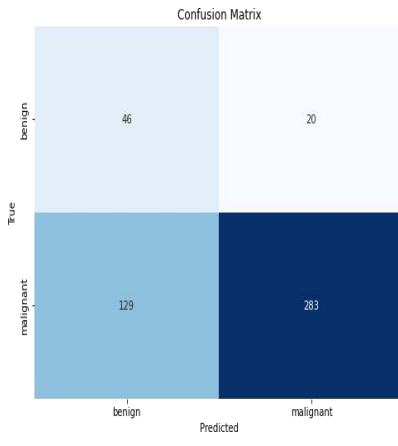
However, it is important to note that the actual performance of these models may vary depending on the specific dataset and evaluation metrics used.

We extracted confusion matrix for three different models, VGG16, GoogleNet, and MobileNetV3Small. The models were tested using histopathology images of varying magnification levels, including 40x, 100x, 200x, and 400x. In addition to the individual model results, the final confusion matrix for the ensembled model is also displayed. This shows the classification accuracy for each magnification level. The ensembled model shows better classification results than any of the individual models, across all magnification levels. This demonstrates the effectiveness of the ensemble approach for improving the accuracy of cancer detection in histopathology images. The results showed in below Figure 7.1 and Figure 7.2.

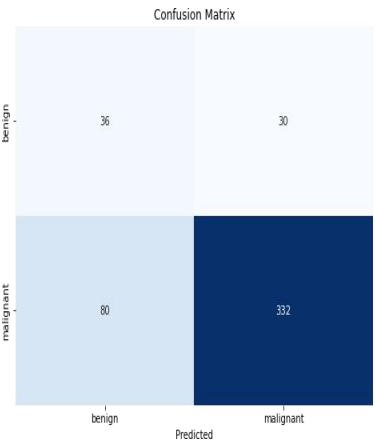
**40X:**



**Vgg16**

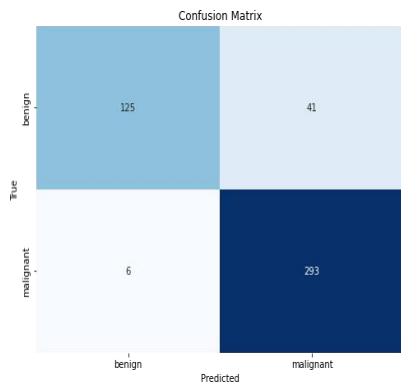


**GoogleNet  
( a )**

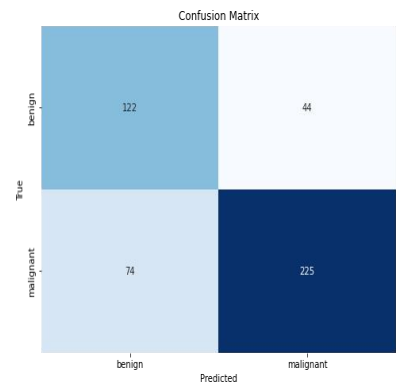


**MobileNetV3\_small**

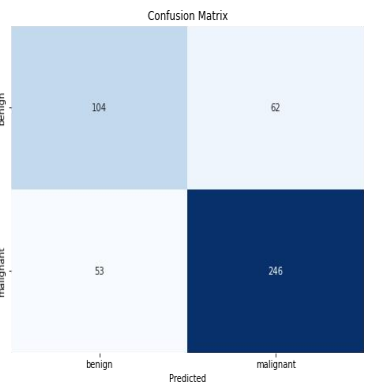
**100X:**



**Vgg16**

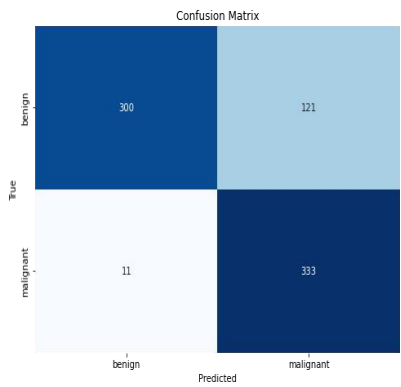


**GoogleNet  
( b )**

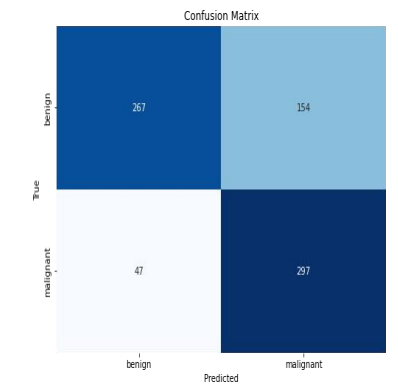


**MobileNetV3\_small**

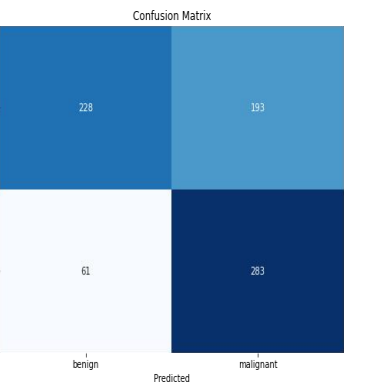
**200X:**



**Vgg16**



**GoogleNet  
( c )**



**MobileNetV3\_small**

**400X:**

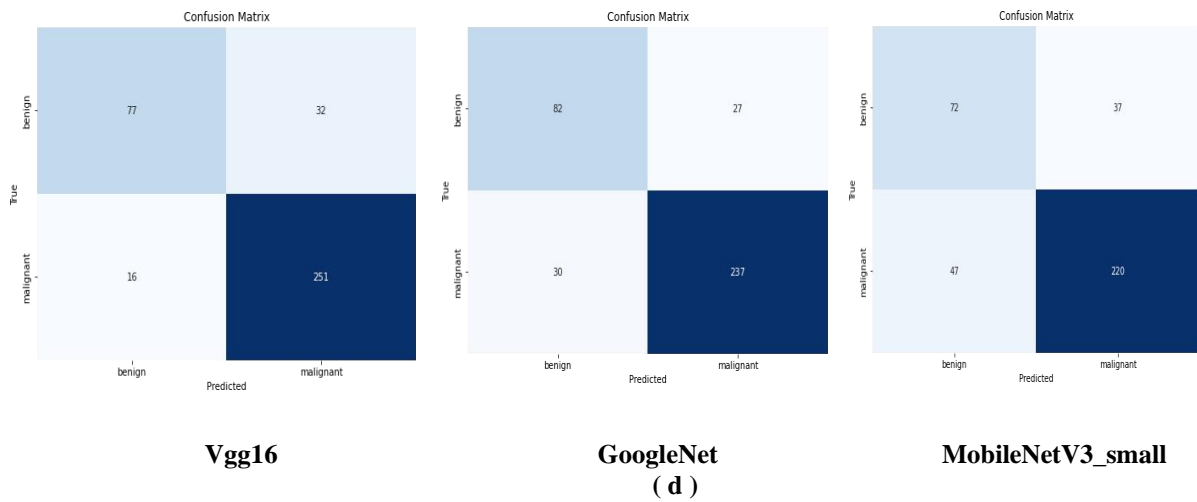


Figure 7.1: Confusion matrices generated by the CNN models on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively

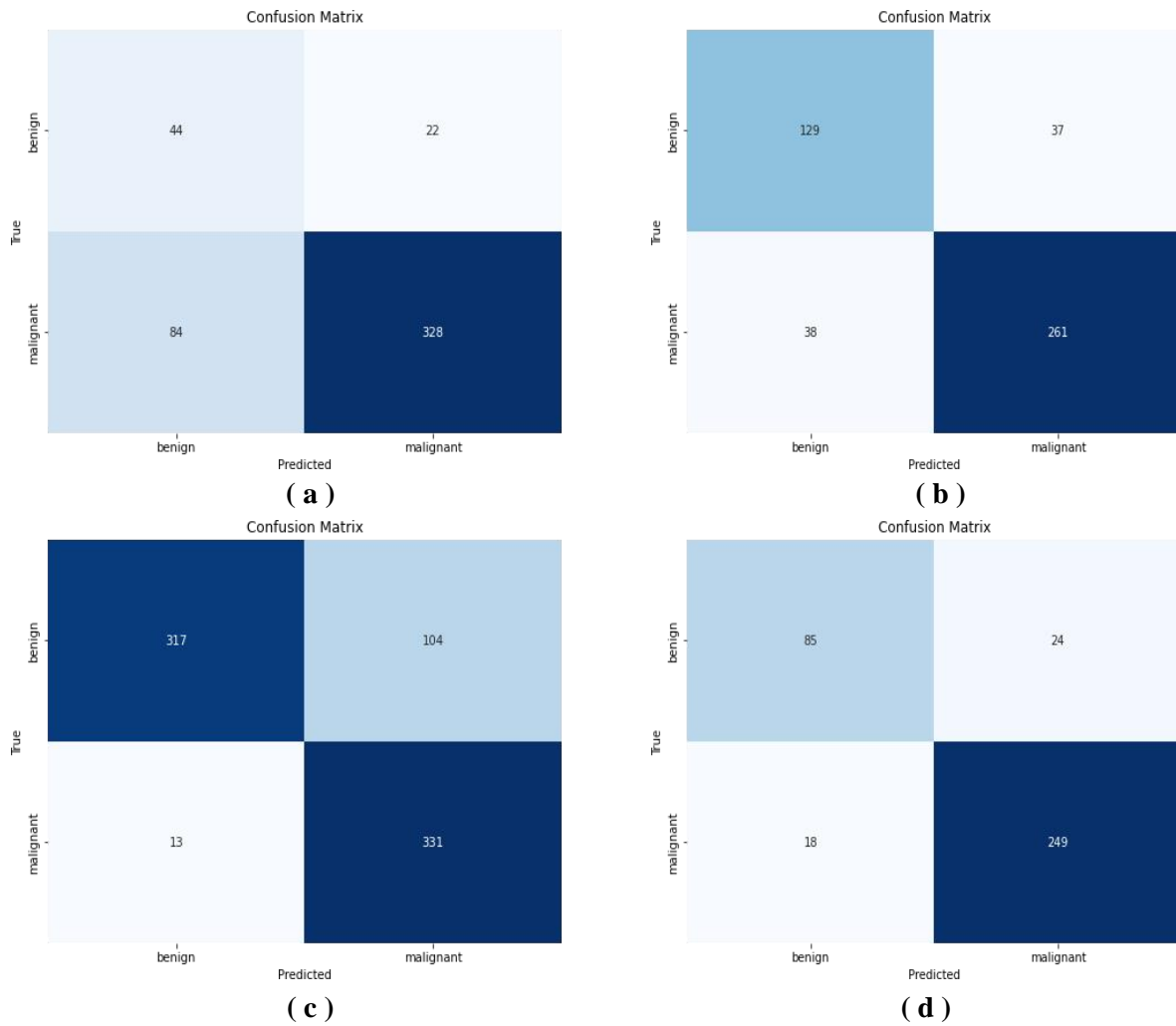


Figure 7.2: Confusion matrices generated by the proposed ensemble model on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively



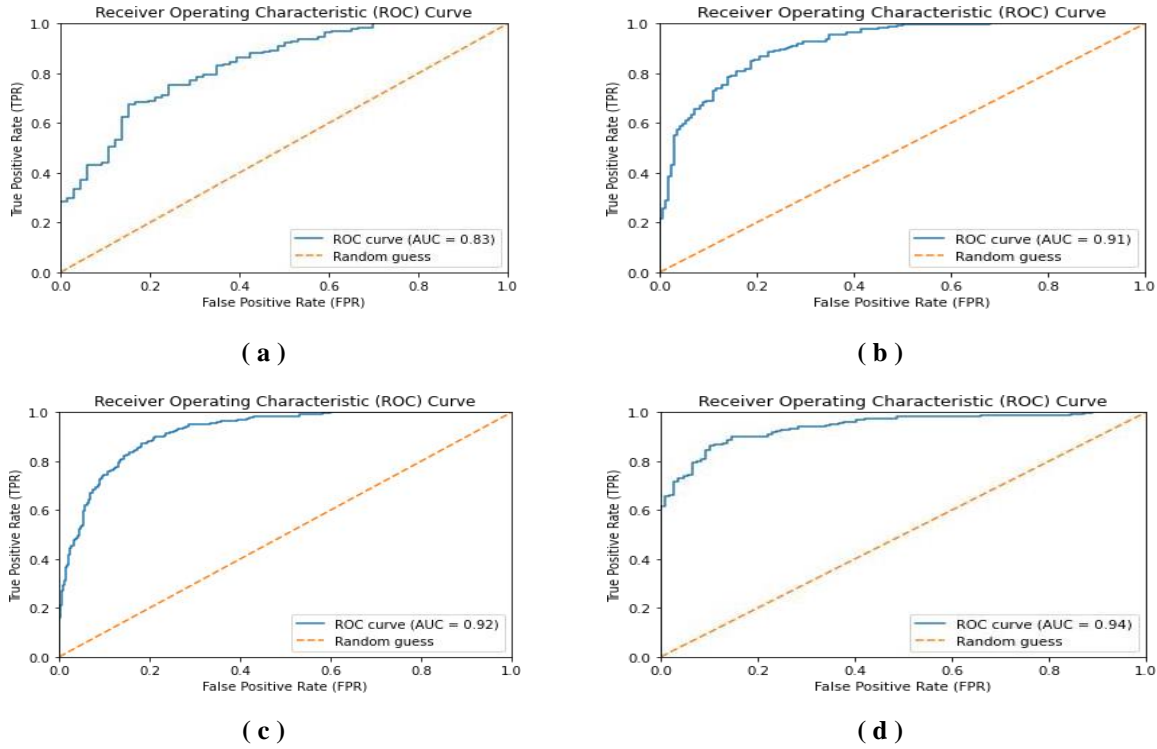


Figure 7.3: ROC curves obtained by the proposed ensemble model from the four sets of magnification i.e., a) 40X, b) 100X, c) 200X and d) 400X

The Receiver Operating Characteristic (ROC) curves obtained by our proposed ensemble model on all the four sets of magnification is shown in Figure 7.3. In all curves, the AUC (Area Under the Curve) value is maintained. Under 400X magnification, the AUC value is 94% which is the highest among all, whereas in 40X magnification the AUC value is 83%. In a ROC curve, higher the value on the X axis indicates the higher value of false positive instances than the number of true negative instances. If the curve is more inclined towards Y axis it indicates the higher number of true positive instances than the number of false negatives. The more ROC curve is inclined towards the top-left corner of the Y axis, greater is the ability of the classifier to distinguish between positive and negative instances. Here in Figure 4 under 400X magnification, the ROC curve is most inclined towards the Y axis, whereas in 40X magnification the curve is least inclined than the other sets of magnification.

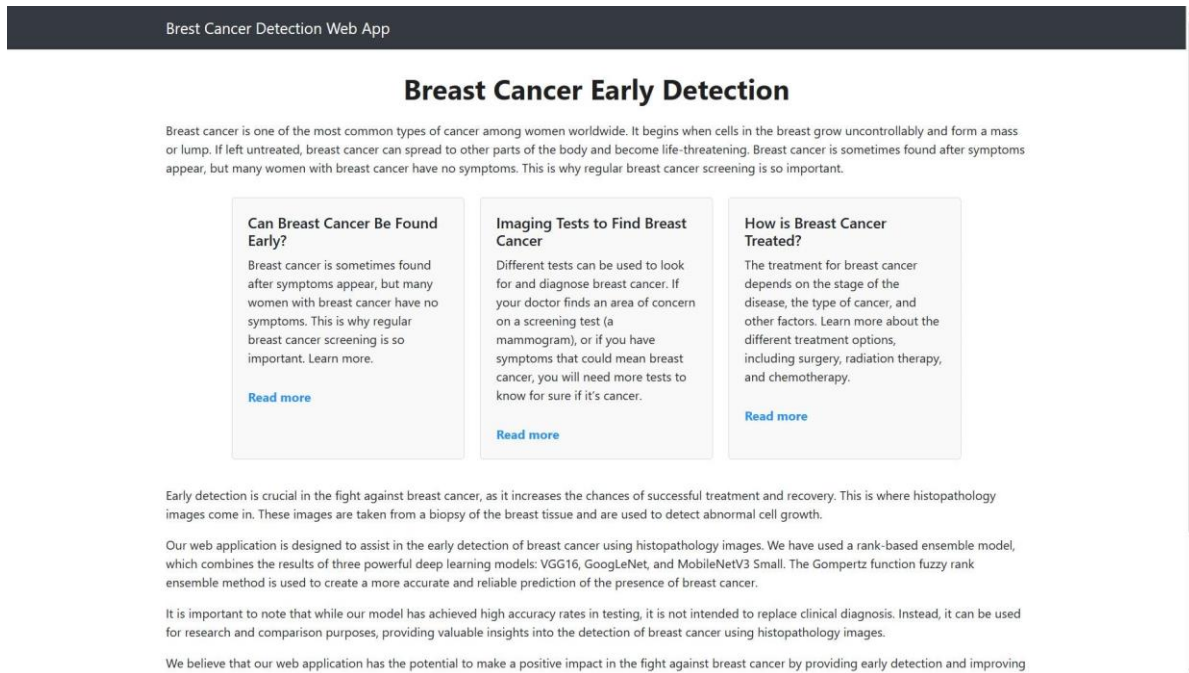


Figure 7.4: Screenshot of Home Page(Part-1)

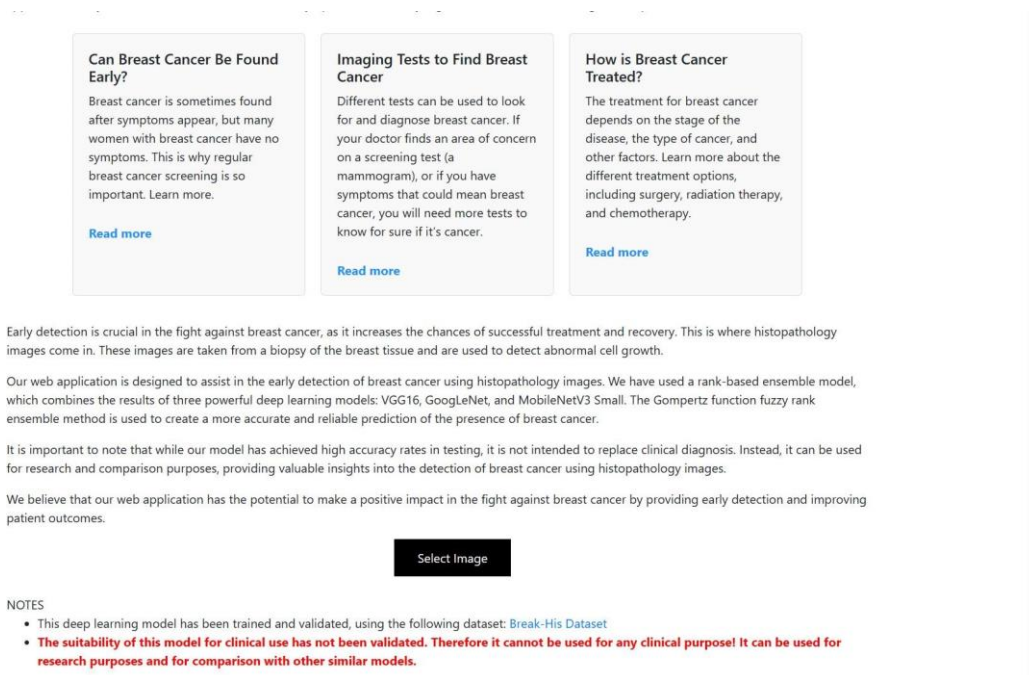


Figure 7.5: Screenshot of Home Page(Part-2)

Figure 7.4 and Figure 7.5 shows the Home page of our project, with a focus on the "Select Image" button. When clicked, this button opens a file dialog, allowing the user to select a histopathology image to be analyzed. This button is prominently displayed on the homepage, making it easy for users to get started with the analysis process.

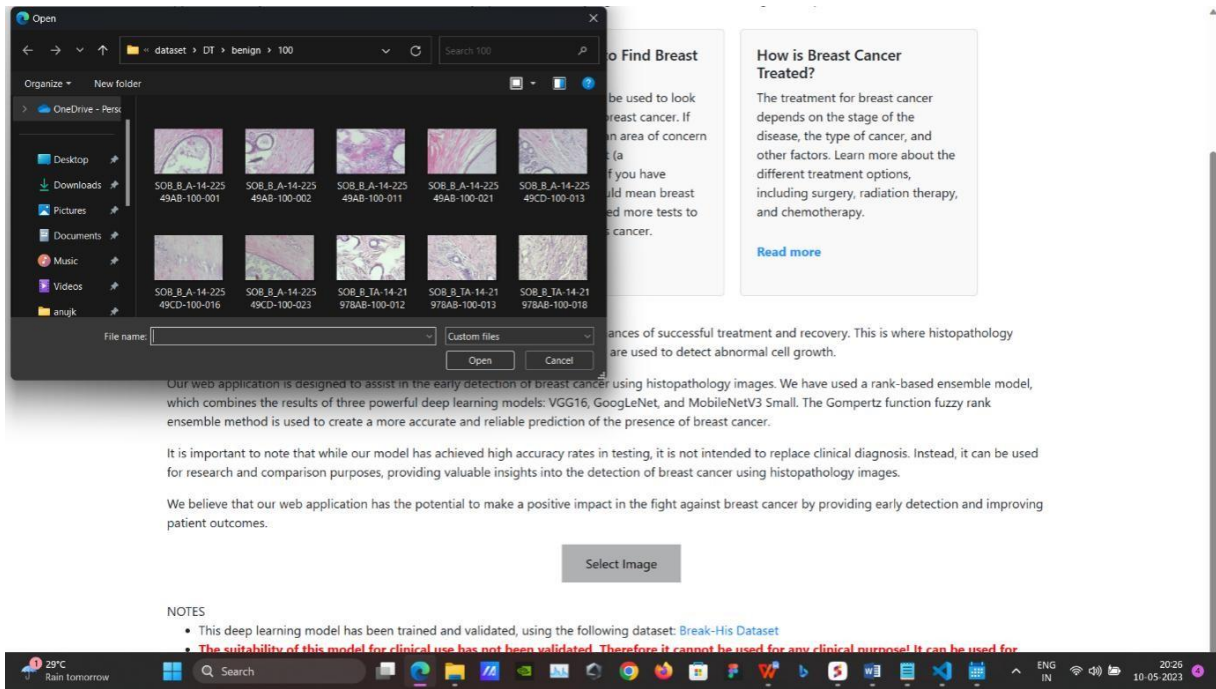


Figure 7.6: Screenshot of before selecting a image

Figure 7.6 shows the file dialog that appears when the "Select Image" button is clicked. The user can navigate through their local file system to find the histopathology image they wish to analyze.

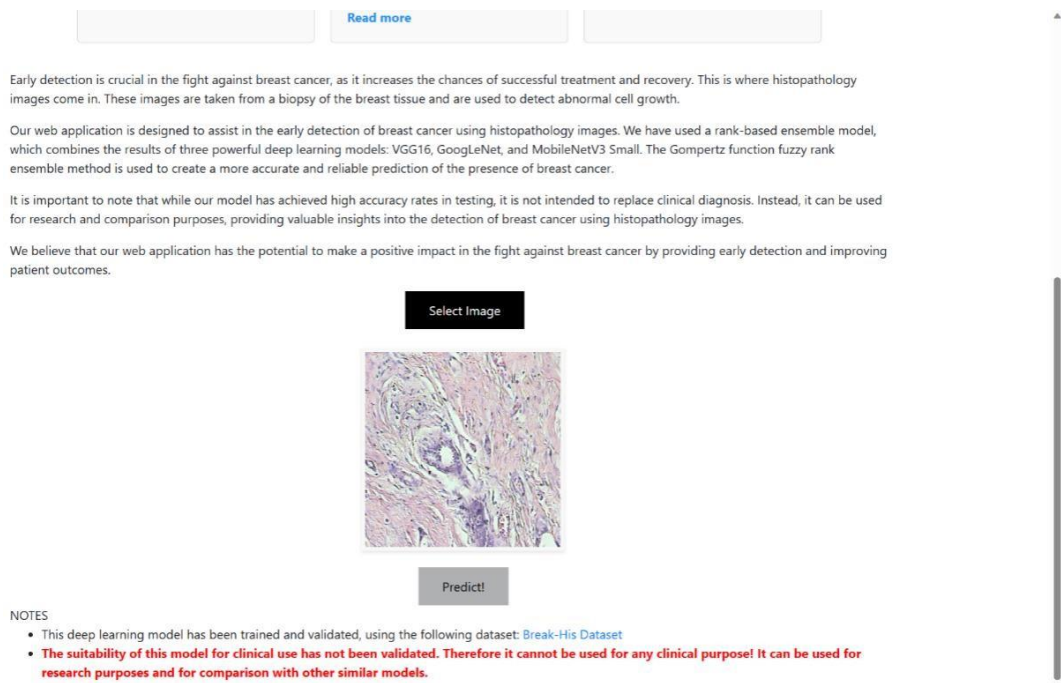


Figure 7.7: Screenshot of after selecting a image

Figure 7.7 shows the screen that appears after an image has been uploaded. The uploaded image is displayed below the “Select Image” button, while the "Predict" button is

prominently displayed below these image. This button triggers the analysis process, which takes a few seconds to complete.

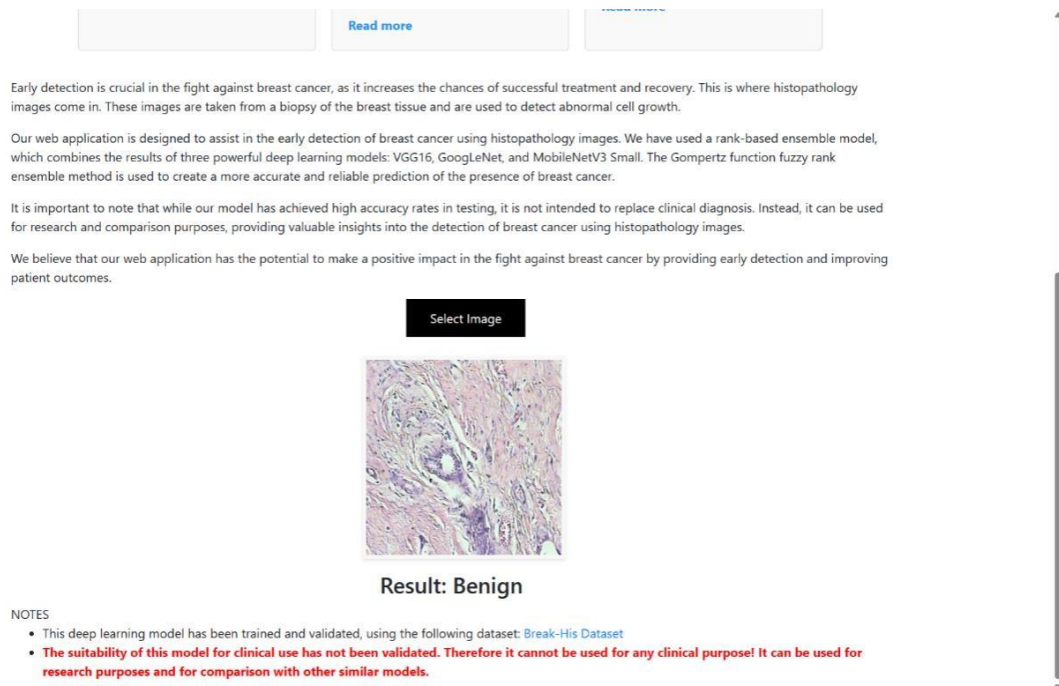


Figure 7.8: Screenshot of result page(Part-1)

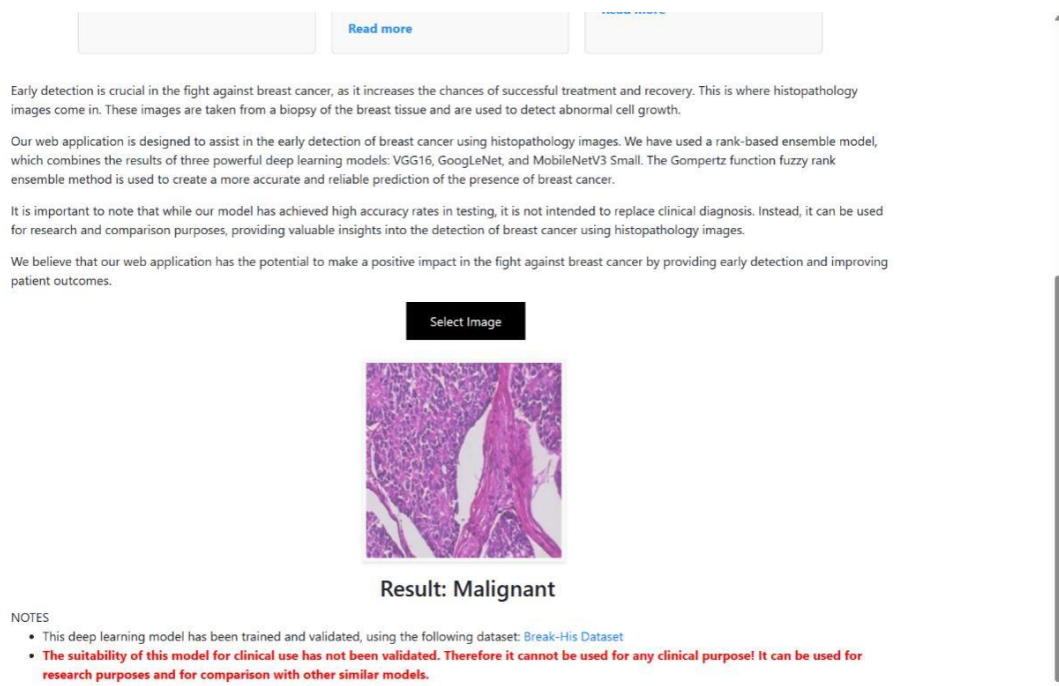


Figure 7.9: Screenshot of result page(Part-2)

Figure 7.8 and Figure 7.9 shows the results of the analysis process. The screen displays a "Results" section.

**Chapter – 8**  
**CONCLUSION**

## CHAPTER 8

### CONCLUSION

Breast cancer is one of the serious concerns among women worldwide. Breast histopathology is a method of microscopic inspection as well as evaluation of a biopsy sample performed by a medical expert to determine cancer growth in the breast. Advancements in digital imaging along with the success of deep learning models in complex image classification tasks motivate many researchers to develop more accurate CAD systems. In this work, we proposed an ensemble model for classifying breast histopathology images by combining the decision scores of three standard CNN models, namely GoogleNet, VGG16, and MobileNetV3 Small. We utilized the Gompertz function for score combination and evaluated our proposed model on the BreakHis dataset, which is a publicly available dataset. Our experimental results demonstrated that the ensemble model outperformed individual CNN models and achieved satisfactory outcomes across various magnification levels of BreakHis images. Therefore, the proposed ensemble model can be a promising approach for accurate and robust classification of breast histopathology images.

#### 8.1 Scope for Future Work

As described below, the project can improve its scope in several ways to further enhance the impact of the models in cancer diagnosis and treatment:

1. **Specialized Models for Multi-Stain Analysis:** Develop dedicated models to accurately analyze multi-stain images, enabling more detailed insights into the presence and progression of cancer.
2. **Expansion to Multiple Organ Histopathology Images:** Include the analysis of histopathology images from different organs that are susceptible to cancer, contributing to early detection and treatment across various types of cancer.
3. **Large-Scale Deployment in Clinical Settings:** Focus on the large-scale deployment of the models in clinical settings by developing efficient algorithms to analyze a substantial volume of images accurately and rapidly.

By incorporating these advancements, the project can provide a more comprehensive and efficient solution for cancer diagnosis and treatment.

# REFERENCES

## REFERENCES

### TextBooks

1. Breast Pathology: A Volume in the Series: Foundations in Diagnostic Pathology" by Frances P. O'Malley and Sarah E. Pinder
2. Diagnostic Pathology: Breast" by Susan C. Lester
3. Breast Cancer: Diagnosis and Pathological Analysis" by Neha S. Shah and Maryam B. Lustberg
4. Breast Cancer: Morphology, Biology, and Prognosis" by Frederick C. Koerner

### Journals

- [1] "DeepBreastNet: A novel and robust approach for automated breast cancer detection from histopathological images," Fatih Demir, Biocybernetics and Biomedical Engineering, Volume 41, Issue 3, 2021, Pages 1123-1139, ISSN 0208-5216, <https://doi.org/10.1016/j.bbe.2021.07.004>.
- [2] "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images," Samriddha Majumdar, Payel Pramanik, Ram Sarkar, Expert Systems with Applications, Volume 213, Part B, 2023, 119022, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2022.119022>.
- [3] "Deep-learning based breast cancer detection for cross-staining histopathology images," Pei-Wen Huang, Hsu Ouyang, Bang-Yi Hsu, Yu-Ruei Chang, Yu-Chieh Lin, Yung-An Chen, Yu-Han Hsieh, Chien-Chung Fu, Chien-Feng Li, Ching-Hung Lin, Yen-Yin Lin, Margaret Dah-Tsyr Chang, Tun-Wen Pai, Heliyon, Volume 9, Issue 2, 2023, e13171, ISSN 2405-8440, <https://doi.org/10.1016/j.heliyon.2023.e13171>.
- [4] "Breast Cancer Histopathology Image Classification Using an Ensemble of Deep Learning Models" , Hameed, Zabit et al. ,Sensors (Basel, Switzerland) vol. 20,16 4373. 5 Aug. 2020, doi:10.3390/s20164373.
- [5] "Breast Cancer Detection From Histopathological Images Using Deep Learning," N. Khuriwal and N. Mishra, 2018 3rd International Conference and Workshops on Recent



Advances and Innovations in Engineering (ICRAIE), Jaipur, India, 2018, pp. 1-4, doi: 10.1109/ICRAIE.2018.8710426.

[6] “Classification of breast cancer from histopathology images using an ensemble of deep multiscale networks,” R. Karthik, R. Menaka, M.V. Siddharth, Biocybernetics and Biomedical Engineering, Volume 42, Issue 3, 2022, Pages 963-976, ISSN 0208-5216, <https://doi.org/10.1016/j.bbe.2022.07.006>.