DAYANANDA SAGAR UNIVERSITY

Devarakaggalahalli, Harohalli
Kanakapura Road, Ramanagara - 562112, Karnataka, India

Bachelor of Technology
in
COMPUTER SCIENCE AND ENGINEERING

Major Project Phase-I Report

HEALTHCARE APPLICATION FOR CANCER DETECTION AND ANALYSIS USING MACHINE LEARNING AND IMAGE PROCESSING

Batch: 96

By

Vivek Belagali	**-**	**ENG20CS0414**
Rahul	**-**	**ENG19CS0242**
Yash C	**-**	**ENG20CS0416**
Rahul B	**-**	**ENG20CS0428**

Under the supervision of
Prof. Pramoda R
Asst. Professor at Department of CSE

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,
SCHOOL OF ENGINEERING
DAYANANDA SAGAR UNIVERSITY,

(2023-2024)

DAYANANDA SAGAR UNIVERSITY

SCHOOL OF ENGINEERING

Department of Computer Science & Engineering
Kudlu Gate, Bangalore–560068 Karnataka, India

CERTIFICATE

This is to certify that the Major Project Stage-I work titled **"HEALTHCARE APPLICATION FOR CANCER DETECTION AND ANALYSIS USING MACHINE LEARNING AND IMAGE PROCESSING"** is carried out by **Vivek Belagali (ENG20CS0414), Rahul (ENG219CS0242), Yash C (ENG0CS0416), Rahul Bhattacharya (ENG20CS428),** bonafide students seventh semester of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2023-2024**.

Prof. Pramoda R	**Dr. Girisha G S**	**Dr. Udaya Kumar Reddy K R**
Assistant/Associate/ Professor Dept. of CS&E, School of Engineering Dayananda Sagar University	Chairman CSE School of Engineering Dayananda Sagar University	Dean School of Engineering Dayananda Sagar University
Date:	Date:	Date:

Name of the Examiner **Signature of Examiner**

1.

2.

DECLARATION

We, **Vivek Belagali (ENG20CS0414), Rahul (ENG19CS0242), Yash C (ENG20CS0416), Rahul Bhattacharya (ENG20CS0428),** are students of eighth semester B. Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the Major Project Stage-I titled **"HEALTHCARE APPLICATION FOR CANCER DETECTION AND ANALYSIS USING MACHINE LEARNING AND IMAGE PROCESSING"** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology in Computer Science and Engineering** during the academic year **2023-2024.**

Student **Signature**

Name1:

USN :

Name2

USN :

Name3:

USN :

Name4:

USN :

Place : Bangalore
Date :

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank ***Dr. Udaya Kumar Reddy K R, Dean, School of Engineering & Technology, Dayananda Sagar University*** *for his constant encouragement and expert advice.*

It is a matter of immense pleasure to express our sincere thanks to ***Dr. Girisha G S, Department Chairman****,* ***Computer Science and Engineering****,* ***Dayananda Sagar University,*** *for providing right academic guidance that made our task possible.*

We would like to thank our guide **……………………..***, Associate /* ***Assistant/ Professor****,* ***Dept. of Computer Science and Engineering****,* ***Dayananda Sagar University****, for sparing his/her valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.*

We would like to thank our ***Project Coordinator Dr. Meenakshi Malhotra*** *and* ***Prof. Mohammed Khurram J*** *as well as all the staff members of Computer Science and Engineering for their support.*

We are also grateful to our family and friends who provided us with every requirement throughout the course.

We would like to thank one and all who directly or indirectly helped us in the Project work.

TABLE OF CONTENTS

Page

NOMENCLATURE USED

AI	Artificial Intelligence
DL	Deep Learning
ML	Machine Learning
CNN	Convolutional Neural Network
IMGP	Image Processing

LIST OF FIGURES

ABSTRACT

The abstract for a cancer detection project using machine learning and image processing would typically highlight key aspects of the research. It may include information about the dataset used, the machine learning algorithms employed, and the results obtained. Here's a generic example:

"This study presents a novel approach to cancer detection leveraging machine learning and image processing techniques. A comprehensive dataset of medical images was utilized, and a convolutional neural network (CNN) was employed to extract relevant features for classification. The model demonstrated promising accuracy in identifying cancerous regions, showcasing the potential for advanced diagnostic tools. The integration of image processing algorithms further enhanced the precision of the detection system. Results indicate a significant advancement in early cancer diagnosis, paving the way for improved patient outcomes and personalized treatment strategies."

.

CHAPTER 1

INTRODUCTION

CHAPTER 1 INTRODUCTION

Cancer is a leading cause of death worldwide, and early detection is crucial for effective treatment and improved survival rates. Medical imaging, such as mammograms for breast cancer and CT scans for lung cancer, plays a vital role in early diagnosis. However, the accurate and timely interpretation of these images can be challenging, often relying on the expertise of radiologists. Machine learning can augment this process by automating the detection and classification of cancerous tumors in medical images.

1.1. SCOPE

The scope of this project for cancer detection using machine learning is to create a comprehensive and user-friendly system that leverages state-of-the-art machine learning models and technologies to aid in the early and accurate identification of cancerous tumors in medical images. This project encompasses the collection and preprocessing of diverse medical image datasets, the development and rigorous evaluation of machine learning models, and the deployment of a secure and compliant system for healthcare professionals. Additionally, the project's scope extends to fostering collaboration with healthcare experts, addressing ethical considerations, ensuring robust data security, and providing a feedback mechanism for continuous improvement. The project aims to contribute significantly to the field of healthcare by enhancing early cancer diagnosis, reducing healthcare costs, and ultimately improving patient outcomes, all while adhering to the highest ethical and regulatory standards.

CHAPTER 2

PROBLEM DEFINITION

CHAPTER 2 PROBLEM DEFINITION

The primary goal of this project is to develop a machine learning system capable of accurately detecting cancer in medical images. This involves the following key aspects:

1. Data Collection: Gather a diverse and representative dataset of medical images (e.g., mammograms, CT scans, MRIs) along with corresponding labels indicating whether the tumor is cancerous or not. The dataset should cover various cancer types and stages.
2. Data Preprocessing: Clean and preprocess the dataset, which may include tasks such as image resizing, normalization, and augmentation. Additionally, handle class imbalance if present.
3. Feature Extraction: Extract relevant features from the medical images that are informative for cancer detection. For instance, texture, shape, and intensity features can be extracted.
4. Model Selection: Choose appropriate machine learning models for cancer detection. This may involve the use of convolutional neural networks (CNNs) for image data and other suitable algorithms for feature-based approaches.
5. Model Training: Train the selected models using the preprocessed dataset. Implement techniques like cross-validation to ensure robustness and avoid overfitting.
6. Evaluation: Evaluate the models' performance using appropriate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Fine-tune the models to achieve the best possible performance.
7. Deployment: Develop a user-friendly interface or API that can accept medical images as input and provide the prediction of cancer presence or absence. Ensure that the deployed model meets regulatory and ethical standards for medical applications.
8. Validation and Testing: Conduct extensive validation and testing of the deployed system using real-world medical data. Collaborate with healthcare professionals to assess the model's clinical utility.
9. Interpretability: Ensure that the model provides interpretable results, allowing healthcare providers to understand the basis for its predictions.

CHAPTER 3

LITERATURE SURVEY

CHAPTER 3 LITERATURE SURVEY

1.**”Lung Cancer Classification and Prediction Using Machine Learning and Image Processing”** by Sharmila Nageswaran, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso

One of the most lethal types of the disease, lung cancer, is
responsible for the passing away of about one million people every year.When doing a CT scan, sophisticated X-ray equipment is
utilized in order to capture images of the human body from
a number of different angles. A dataset of 83 CT images from 70 different patients was
used in the experimental study ½x. Images are preprocessed
using the geometric mean filter. This results in improving
image quality. Then, images are segmented using the K
-means algorithm. This segmentation helps in the identification of the region of interest. Then, machine learning classification techniques are applied.
For performance comparison, three parameters, accuracy, sensitivity, and specificity, are used:
Accuracy =
TP + TN
TP + TN + FP + FN ,
Sensitivity =
TP
TP + FN ,
Specificity = TN
TN + FP ,
ð1Þ
where TP is true positive, TN is true negative, FP is false positive, and FN is false negative.
Results of different machine learning predictors are
shown in Figures 3–5. The accuracy of ANN is better.

2.”**Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis”** by V. Durga Prasad Jasti , Abu Sarwar Zamani, K. Arumugam, Mohd Naved, Harikumar Pallathadka, F. Sammy, Abhishek Raghuvanshi, and Karthikeyan Kaliyaperumal

Normal cells become crowded out as the cancerous growth spreads throughout the body, making it difficult for the body to operate properlyTumors and lumps form as a result of cancer growth. Some anomalies, however, may not be harmful. To determine whether a tumor or lump is cancerous, a little sample is removed by the doctor and examined under a microscope.An X-ray image of the breast is captured by mammography. Computerized mammography has eliminated the need for repeated mammograms in breast screening methods. Computer-based PC programs that warn radiologists to optimum variations in mammography and allow integrated film.

Image preprocessing using geometric mean filter
(ii) Feature extraction using AlexNet
(iii) Feature selection using relief algorithm
(iv) Classification using LS-SVM and other algorithms

In order to get a better and more accurate analysis and interpretation of breast images, it is important to get rid of all the noise [27].
'ree parameters accuracy, sensitivity, and specificity are used in this study to compare the performance of different algorithms.

Accuracy � (TP + TN)

(TP + TN + FP + FN)

,

Sensitivity � TP

(TP + FN)

,

Specificity � TN

(TN + FP)

,

Precision � TP

(TP + FP)

,

Recall � (TP)

(TP + FN)

,

(1)

where TP � True Positive, TN � True Negative, FP�False Positive, FN�False Negative

'e accuracy, sensitivity, and specificity of LS-SVM, KNN, random forest, and Na¨ıve Bayes classifiers for breast cancer disease detection are shown in Figures 2 and 3. Accuracy of LS-SVM is better than the rest of the classifiers. Sensitivity and specificity of the KNN algorithm are better than the rest of the classifiers.

3.“**Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis”** by NOREEN FATIMA 1 , LI LIU 1 , SHA HONG1 , AND HAROON AHMED

Data mining is a process of discovering the useful information from a big dataset, data mining techniques and functions help to discover any kind of disease, data mining techniques such as machine learning, statistics, database, fuzzy set, data warehouse and neural network help in diagnosis and prognosis of different cancer diseases.Artificial Neural Network is a common algorithm for data mining process. Neural network consists of input layer, hidden layer and output layer. This technique is used to extract the pattern that is too complex.

LOGISTICS REGRESSION (LR);It is a supervised learning algorithm that includes more dependent variables. The response of this algorithm is in the binary form.

(KNN)This algorithm is used in pattern recognition. It is a good
approach for breast cancer prediction. In order to recognize
the pattern, each class has given an equal importance .K Nearest Neighbor extract the similar featured data from a
large dataset. On the basis of features similarity we classify a
big dataset.

Features selection and features extraction techniques
on Artificial Neural Network (ANN), Support Vector
Machine (SVM) and Naive Bayes (NB) were applied for the
prediction of breast cancer. Dataset of patients was collected
from Wisconsin Diagnostic Breast Cancer. Feature selection
is a selection of sub features from a huge dataset that helps in
computation process. Authors comparatively analyzed each
technique with different type of features selection such as
coloration based feature selection (CFS), Linear Discriminant
Analysis and Recursive Feature Elimination (RFE). After the
comparative analysis with different features selection methods, authors came to know that the accuracy rate of Artificial
Neural Network was higher than the other algorithms. The
accuracy of Support Vector Machine was 96.4%, Artificial
Neural Network was 97.0% and Naive Bayes accuracy was
91% [81].

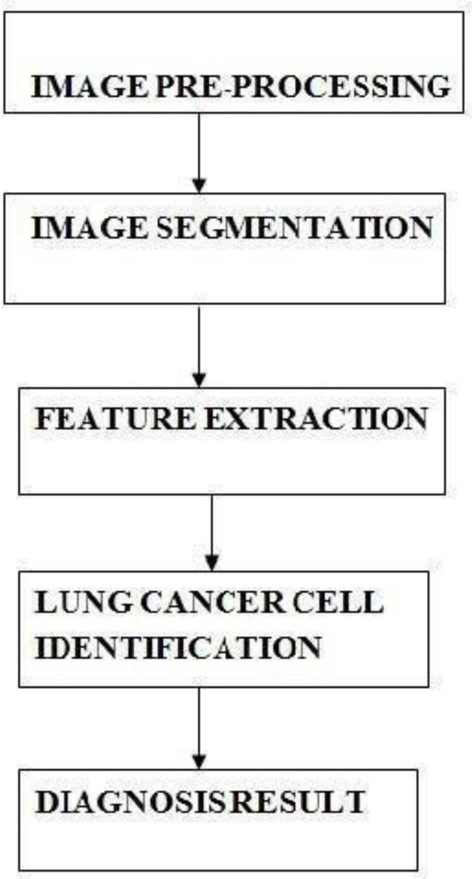
4.”**Bone cancer detection using machine learning techniques”** by Deepshikha Shrivastava1 , Sugata Sanyal2 , Arnab Kumar Maji1 and Debdatta Kanda

Detecting bone cancer through machine learning techniques involves using algorithms to analyze medical images, like X-rays or MRIs, to identify abnormalities or potential signs of cancerous growth in bones. It often involves feature extraction, image segmentation, and classification algorithms to differentiate between healthy and cancerous tissues. Researchers and medical professionals are continually exploring and refining these techniques to improve accuracy and early detection rates. If you're looking for specific articles or studies on this topic, I can certainly help with that.

CHAPTER 4

PROJECT DESCRIPTION

CHAPTER 4 4.1.PROPOSED DESIGN



4.1.Data Flow Diagram

Project Description:

Objective:

- Early Detection: Develop a machine learning system that enables the early detection of cancerous tumors in medical images.
- Accuracy: Achieve a high level of accuracy in cancer detection to improve diagnostic outcomes.
- Data Collection: Gather a comprehensive and diverse dataset of medical images along with accurate labels for model training and testing.
- Data Preprocessing: Implement data preprocessing techniques, including resizing, normalization, and augmentation, to ensure data quality and consistency.

- Model Development: Select and train appropriate machine learning models, such as convolutional neural networks (CNNs), to effectively classify cancerous and non-cancerous tumors.
- Performance Metrics: Evaluate model performance using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to ensure reliable results.
- Interpretability: Implement model explainability tools to provide insights into why specific predictions are made, aiding healthcare professionals' decision-making.
- User-Friendly Interface: Create an intuitive and accessible user interface, allowing healthcare providers to easily upload medical images for analysis.
- Data Security: Ensure robust data security measures to protect patient privacy and comply with data protection regulations.
- Compliance: Ensure compliance with relevant healthcare and data privacy regulations, including HIPAA or GDPR, to maintain ethical standards.
- Deployment: Deploy the system in healthcare settings, either on-premises or in the cloud, to make it readily available to medical professionals.
- Monitoring and Maintenance: Continuously monitor the system's performance and maintain it through regular updates and improvements.
- User Feedback: Establish a feedback mechanism to collect user feedback and utilize it to enhance the model and system.
- Collaboration: Foster collaboration between data scientists and healthcare experts to align the model with clinical needs and standards.
- Ethical Considerations: Develop guidelines and protocols for handling ethical dilemmas, ensuring responsible use of AI in healthcare.
- Documentation: Maintain comprehensive documentation of the project's progress, including model architecture and validation procedures.
- Disaster Recovery: Implement disaster recovery measures to ensure system availability in unforeseen circumstances.
- Cost Management: Manage resource usage and costs effectively throughout the project lifecycle.
- Regulatory Compliance Testing: Conduct rigorous testing to ensure compliance with healthcare regulations before clinical deployment.

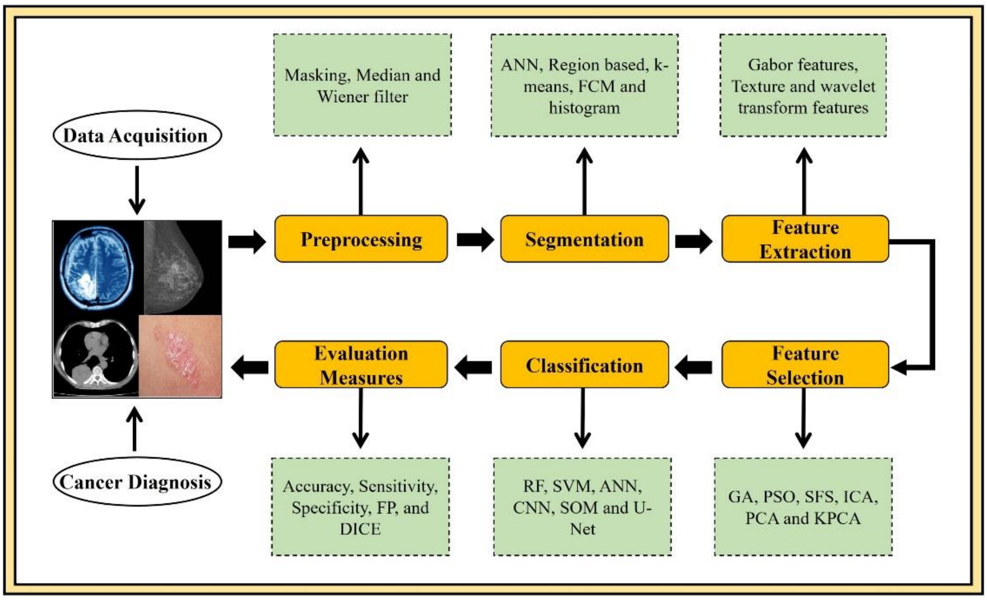
Components:

a. Data Collection:

Gather a diverse dataset of medical images containing cancerous and non-cancerous cases.

b. Preprocessing:

Standardize image sizes, enhance contrast, and remove noise.


Data Acquisition
Masking, Median and Wiener filter
ANN, Region based, k-means, FCM and histogram
Gabor features, Texture and wavelet transform features
Preprocessing
Segmentation
Feature Extraction
Evaluation Measures
Classification
Feature Selection
Cancer Diagnosis
Accuracy, Sensitivity, Specificity, FP, and DICE
RF, SVM, ANN, CNN, SOM and U-Net
GA, PSO, SFS, ICA, PCA and KPCA

4.2.CANCER DETECTION ARCHITECTURE

c. Feature Extraction:

Extract relevant features such as texture, shape, and intensity from the images.

d. Machine Learning:

Train and test various machine learning algorithms (e.g., CNNs, SVM, or decision trees) for classification.

e. Integration:

Develop a user-friendly interface for uploading images and obtaining analysis results.

f. Evaluation:

Assess the system's performance through cross-validation, ROC curves, and other relevant metrics.

CHAPTER 5

REQUIREMENTS

5.1.FUNCTIONAL REQUIREMENTS

Functional requirements specify what a system must do and outline its features and capabilities. Here are the functional requirements for a cancer detection system using machine learning:

1. Data Collection and Ingestion:

   -The system must be able to collect and ingest medical images from various sources, including hospitals, clinics, and research institutions.

   - It should support the collection of diverse and representative datasets covering different cancer types and stages.

2. Data Preprocessing:

   -The system must preprocess the incoming data, including resizing images to a standard format, normalizing pixel values, and handling data augmentation.

   - It should address class imbalance issues if present in the dataset.

3. Model Development:

   - The system must include a module for selecting and developing machine learning models, with an emphasis on deep learning models like CNNs.

   - It should provide options for transfer learning using pre-trained models to improve efficiency.

4. Training Pipeline:

   -The system should establish a robust training pipeline for machine learning models, including data shuffling, batch processing, and early stopping to prevent overfitting.

5. Model Evaluation:

   -The system must evaluate model performance on separate testing datasets, providing metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.

   - It should support cross-validation techniques to assess model generalization.

6. Model Interpretability:

   -The system should offer tools for model interpretability, enabling healthcare professionals to understand and trust the model's predictions.

7. User Interface:

- The system must have an intuitive and user-friendly interface allowing users, primarily healthcare professionals, to upload medical images for analysis.

-It should provide real-time predictions and explanations.

8. Security and Compliance:

-The system must implement stringent data security measures to protect patient privacy and ensure compliance with healthcare regulations (e.g., HIPAA or GDPR).

9. Deployment:

-The system should support deployment strategies both in a cloud environment (e.g., AWS, Azure) and on-premises.

-It must be scalable to handle varying workloads.

10. Monitoring and Maintenance:

-The system should continuously monitor model performance, system health, and resource usage.

-It must support automatic model updates and software maintenance.

11. Feedback and Improvement:

-The system should include a feedback mechanism for collecting user feedback and performance data to iteratively improve the model.

-It must allow for the integration of user suggestions.

12. Collaboration:

-The system should facilitate collaboration between data scientists, healthcare experts, and administrators.

-It must enable efficient communication and data sharing among stakeholders.

13. Ethical Considerations:

-The system must adhere to ethical guidelines for responsible AI use, addressing concerns like bias, fairness, and transparency.

-It should provide mechanisms to handle ethical dilemmas.

Software/System Requirements

Software Requirements

- Operating System : Linux or Windows Server.
- Language : Python, Javascript, Html and Css
- Tool : Scikit-learn, TensorFlow, PyTorch and Open CV

Hardware Requirements

- System : Intel Core i5
- Memory : 8 GB
- Hard Disk : 1 TB.
- Webcam

CHAPTER 6

METHODOLOGY

CHAPTER 6 METHODOLOGY

1. Collect diverse and representative medical image datasets.
2. Preprocess data by resizing, normalizing, and handling class imbalance.
3. Extract relevant features from images.
4. Select appropriate machine learning models (e.g., CNNs).
5. Train models with cross-validation to avoid overfitting.
6. Evaluate performance using accuracy, precision, recall, F1-score, and ROC-AUC.
7. Deploy a user-friendly interface/API for practical use in healthcare.

METHODOLOGY: (CANCER DETECTION)

a. Data Collection:

Gather a diverse dataset of medical images containing cancerous and non-cancerous cases.

b. Preprocessing:

Standardize image sizes, enhance contrast, and remove noise.

c. Feature Extraction:

Extract relevant features such as texture, shape, and intensity from the images.

d. Machine Learning:

Train and test various machine learning algorithms (e.g., CNNs, SVM, or decision trees) for classification.

e. Integration:

Develop a user-friendly interface for uploading images and obtaining analysis results.

f. Evaluation:

Assess the system's performance through cross-validation, ROC curves, and other relevant metrics.

CHAPTER 7

DELIVERABLES

CHAPTER 7 DELIVERABLES

- The project should result in a robust machine learning model for cancer detection, along with a user-friendly interface for practical use in a healthcare setting. The system's performance should meet or exceed the standards set by medical professionals for accuracy and reliability.

The deliverables for a cancer detection project using machine learning and image processing typically include:

1.Dataset Preparation: Gather and preprocess a labeled dataset of medical images for training and testing the machine learning model.

2.Feature Extraction: Extract relevant features from the medical images, focusing on key characteristics indicative of cancer.

3.Model Development: Implement and train a machine learning model, such as a convolutional neural network (CNN), using the prepared dataset to learn patterns and make predictions.

4.Validation and Testing: Evaluate the model's performance on a separate set of validation and test data to ensure generalization and accuracy.

5.User Interface (UI): Develop a user-friendly interface for healthcare professionals to interact with the model and interpret results.

6.Integration with Existing Systems: Ensure seamless integration with existing healthcare systems, allowing for efficient adoption and use.

7.Accuracy Metrics: Provide metrics such as sensitivity, specificity, and accuracy to assess the model's performance and reliability.

8.Documentation: Prepare comprehensive documentation outlining the project details, including methodologies, model architecture, and instructions for use.

9.Ethical Considerations: Address ethical considerations, such as patient privacy and data security, and ensure compliance with relevant regulations.

10.Deployment Plan: Develop a deployment plan for the model, considering scalability, updates, and ongoing maintenance.

11.Training Materials: Create materials for training healthcare professionals on using the system effectively.

12.Continuous Improvement: Establish mechanisms for continuous improvement, incorporating feedback and updates to enhance the model's capabilities over time.

Remember that effective communication with healthcare professionals, patients, and stakeholders is crucial throughout the development and deployment phases of the project.

REFERENCES

- Lung Cancer Classification and Prediction Using Machine Learning and Image Processing by Sharmila Nageswaran, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso https://doi.org/10.1155/2022/1755460

- Computational Technique Based on Machine Learning and Image Processing for Medical Image Analysis of Breast Cancer Diagnosis by V. Durga Prasad Jasti , Abu Sarwar Zamani, K. Arumugam, Mohd Naved, Harikumar Pallathadka, F. Sammy, Abhishek Raghuvanshi, and Karthikeyan Kaliyaperumal https://doi.org/10.1155/2022/1918379

- Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis by NOREEN FATIMA 1 , LI LIU 1 , SHA HONG1 , AND HAROON AHMED https://doi: 10.1007/s41870-020- 00427-7

- Bone cancer detection using machine learning techniques by Deepshikha Shrivastava1 , Sugata Sanyal2 , Arnab Kumar Maji1 and Debdatta Kandar https://www.sciencedirect.com/science/article/pii/B9780128179130000171