Department of Computer Engineering & Technology
CSE
Capstone Project Synopsis

Group No:	Students names	PRN	Email, Mobile number	Panel	Sign of student
P-35	Dharmika Tank	1032210159	1032210159@mitwpu.edu.in, 9511783503	G	
	Jay Mehta	1032210499	1032210499@mitwpu.edu.in, 8488896953	G	
	Devanshu Surana	1032210755	1032210755@mitwpu.edu.in, 8999300859	G	
	Tejas Redkar	1032210937	1032210937@mitwpu.edu.in, 9175012103	G	

Name of Guide : Dr. Pratvina Talele

Project Title: DeepTumorNet

Project Domain: Deep Learning

Problem Identification after Literature survey in Tabular form

Sr. No.	Title of Paper	Conference /Journal	Paper Objectives	Gaps
1	Lung Cancer Detection and Classification Using Deep CNN	Journal	1. Early Cancer Detection using CNN on chest CT scans. 2. Automated Diagnosis with deep learning for benign vs. malignant classification. 3. End-to-End AI Pipeline including preprocessing, segmentation, and classification. 4. Performance Evaluation with metrics like accuracy (96%), precision, recall, etc.	1. Hyperparameter Tuning Needed – No AutoML/advanced optimizations. 2. No Cloud Deployment – Misses AWS/GCP integration for real-time AI diagnosis. 3. Limited Explainability – No Grad-CAM/SHAP for AI interpretability.
2	Deep Learning for Accurate Chest Disease Classification: A CNN-Based Approach for Lung Cancer Subtypes and Normal Cells.	Conference	1. Lung Cancer Subtype Classification – Identify Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and Normal Cells. 2. AI-Powered Diagnosis – Implement CNNs for accurate chest disease classification. 3. High-Performance Model – Optimize accuracy, precision, recall, and AUC-ROC. 4. Clinical Validation – Ensure model aligns with real-world	1. Limited Interpretability – Grad-CAM used, but lacks detailed explainability techniques like SHAP/LIME. 2. Lack of Transfer Learning – No use of pre-trained models for enhanced feature learning.

			medical diagnostics. 5. Explainability – Use Grad-CAM for visualization of decision-making areas.	3. No 3D Image Analysis – Model is 2D-based, missing depth details in lung scans.
3	Lung Cancer Detection in Radiological Imaging using Deep Learning: A Review	Conference	1. Review state-of-the-art deep learning models for lung cancer detection. 2. Assess performance on different imaging modalities like CT scans and X-rays. 3. Analyze advancements, trends, and challenges in AI-driven lung cancer detection. 4. Explore clinical applications for improved early detection and patient outcomes. 5. Emphasize the need for diverse datasets and model interpretability. 6. Investigate ethical concerns and deployment challenges in AI-based diagnostics.	1. Limited dataset diversity leading to biases and reduced generalizability. 2. Lack of standardized evaluation metrics for model comparison. 3. Model interpretability issues making AI decisions less transparent. 4. Integration challenges preventing seamless adoption in clinical workflows. 5. High computational costs limiting real-time clinical applicability. 6. Ethical and legal concerns regarding data privacy, accountability, and transparency.
4	Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images	IEEE Journal	1. Automated Cancer Segmentation – Develop deep learning methods to accurately segment lung cancer tissues in whole-slide histopathology images (WSI). 2. Challenge Benchmarking – The ACDC@LungHP Challenge 2019 aimed to evaluate different AI-based segmentation methods on lung cancer WSI. 3. Model Evaluation – Assess CNN-based models using metrics like Dice Coefficient (DC), Precision, Sensitivity, and Specificity. 4. Multi-Model vs. Single Model Performance – Compare the effectiveness of multi-model vs. single-model architectures. 5. Enhance Pathologists' Workflow – Use AI to assist in cancer tissue detection, reducing pathologists' workload.	1. Lack of Clinical Deployment – Model not tested in real-world pathology labs. 2. Limited Performance on Challenging Cases – Models struggled with complex tumor structures, reducing accuracy. 3. No Integration with Cloud AI – The study does not explore scalable cloud-based implementations for remote pathology.
5	Lung Cancer Detection and Classification using	Journal	1. To propose a novel approach for lung cancer detection and classification using CT scan	1. Existing methods for lung cancer detection using CT

	CT Scan Image Processing		image processing. 2. To implement image preprocessing techniques such as smoothing and enhancement for better image quality. 3. To apply thresholding and edge detection methods for segmentation of the lung tumor region from the lung mask. 4. To extract geometrical features like area, perimeter, eccentricity, compactness, and circularity of the segmented tumor regions. 5. To classify the extracted features into benign or malignant tumors using a Support Vector Machine (SVM) classifier. 6. To evaluate the performance of the proposed system in terms of accuracy and detection of lung cancer nodules.	scans often require prior information, which limits the accuracy and applicability in real-world scenarios. 2. Many current approaches for lung cancer classification focus on only benign or malignant detection but lack clear distinction in the severity stages. 3. Image preprocessing methods, such as filtering and enhancement, have room for optimization to improve segmentation and feature extraction processes. 4. There is a need for more reliable and generalized classification models that can work efficiently across different datasets without manual adjustments. 5. Current methods lack robustness in handling noise and artifacts in CT images, which can lead to incorrect tumor identification and classification.
6	Lung Cancer Detection and Classification from Chest CT Scans Using Machine Learning Techniques.	Conference	1. Machine Learning-Based Classification – Using Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) for accurate classification. 2. Clinical Impact – Aims to improve cancer detection and assist radiologists in diagnosis. 3. Early Lung Cancer Detection – Identify Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma from CT scan.	1. Lack of Clinical Validation – Model is not tested in real-world hospitals or integrated into radiology workflows. 2. Annotation Challenges – Potential noise in labeled datasets affecting accuracy.
7	Real-time Detection of Lung Cancer Using CNN.	Conference	1. Preprocessing & Model Optimization – Implements image normalization, resizing, and contrast enhancement for better accuracy.	1. No Clinical Trials – Model is not tested in hospitals or integrated into radiology workflows.

			2. CT Scan-Based Classification – Uses Convolutional Neural Networks (CNNs) to classify malignant vs. non-cancerous cases. 3. Real-time Lung Cancer Detection – Develop a CNN-based model for rapid lung cancer diagnosis.	2. No Multi-Modal Data Integration – Lacks fusion with additional clinical data like blood reports or symptoms. 3. Limited Explainability – No Grad-CAM or SHAP-based interpretability to aid radiologists.
8	Lung Cancer Detection with Machine Learning and Deep Learning: A Narrative Review.	Conference	1. AI for Lung Cancer Detection – Explore the role of Machine Learning (ML) and Deep Learning (DL) in improving lung cancer diagnosis. 2. Automated Image Processing – Use AI-based Computer-Aided Detection (CAD) systems for lung nodule detection.	1. Feature Extraction Needs Improvement – Requires better hybrid AI models (ML + DL) for enhanced performance. 2. Dataset Limitations – Uses small datasets, limiting generalization. 3. No Real-World Clinical Testing – Not validated in hospitals for real time diagnostics.
9	Lung Cancer Prediction Model Using Ensemble Learning Techniques and a Systematic Review Analysis	Conference	1. Develop an effective machine learning-based prediction model for lung cancer diagnosis. 2. Applied and compared different ensemble learning techniques such as XGBoost, LightGBM, Bagging, and AdaBoost. 3. Evaluating the models using accuracy, precision, recall, F1-score, and AUC. 4. Use SMOTE (Synthetic Minority Oversampling Technique) to balance the dataset and improve model reliability.	1. Most past studies focused on single classifiers like SVM, Random Forest, or Naïve Bayes, without leveraging ensemble learning techniques. 2. Many previous models used imbalanced datasets, leading to biased predictions. The study addressed this using SMOTE. 3. Existing models often relied on specific datasets that were not representative. The study highlights the need for larger and more diverse datasets for better generalization.
10	Enhancing Predictive Accuracy in Lung Disease Diagnosis Through Hybrid ResNet and Transfer Learning	Conference	Improve Lung Disease Diagnosis Accuracy: 1. Develop a hybrid model combining ResNet and Transfer Learning to enhance predictive performance in	Lack of Explainability & Robustness: 1. Deep learning models in healthcare

	Models		lung disease classification. Utilize Deep Learning for Medical Imaging: 2. Compare multiple deep learning models (CNN, ResNet, Transfer Learning, and a Hybrid ResNet-Transfer Learning Model) for lung disease classification. Benchmark Model Performance: 3. Evaluate model accuracy, sensitivity, and F1-score, demonstrating the superiority of the hybrid approach.	often struggle with interpretability and robustness when diagnosing complex diseases. 2. The study emphasizes feature extraction and fusion techniques to enhance the model's understanding of lung disease patterns. Need for Generalized Predictive Models: 3. Prior research often focused on specific lung diseases, limiting applicability. 4. The study extends deep learning across multiple lung conditions, making it more applicable for broad clinical use.

Abstract:

The End-to-End Chest Cancer Classification project provides a comprehensive framework for detecting and classifying chest cancer using deep learning techniques. Leveraging a Convolutional Neural Network (CNN), the system processes medical images to identify malignancies with high accuracy. The project integrates MLflow for experiment tracking, DVC for data and model versioning, and Flask for deployment, ensuring a seamless and reproducible machine learning pipeline. It offers automated preprocessing, model training, evaluation, making it adaptable for research and real-world implementation. The primary goal is to create a scalable and efficient system that aids early cancer detection, reducing diagnostic delays and improving patient outcomes.

Scope:

1) Medical Image Classification

Automates the detection and classification of chest cancer using deep learning.
Enhances accuracy and efficiency in medical diagnostics.

2) Experiment Tracking & Model Versioning

Uses MLFlow to log experiments, compare models, and track performance.
Implements DVC for data and model version control, ensuring reproducibility.

3) Scalable & Reproducible Machine Learning Pipeline

Provides an end-to-end ML workflow from data preprocessing to deployment.
Designed for scalability, allowing easy integration of additional features.

4) Deployment-Ready System

Flask-based API enables real-time model inference.
Supports Dockerization, making the model portable and production-ready.

5) Healthcare Applications

Can be extended to classify other types of cancer and medical conditions.
Potential integration with cloud-based AI healthcare systems for remote diagnostics.

6) Performance Evaluation & Improvement

Logs key evaluation metrics (accuracy, precision, recall, F1-score).
Allows continuous model refinement through iterative improvements.

Project Objectives for Team:

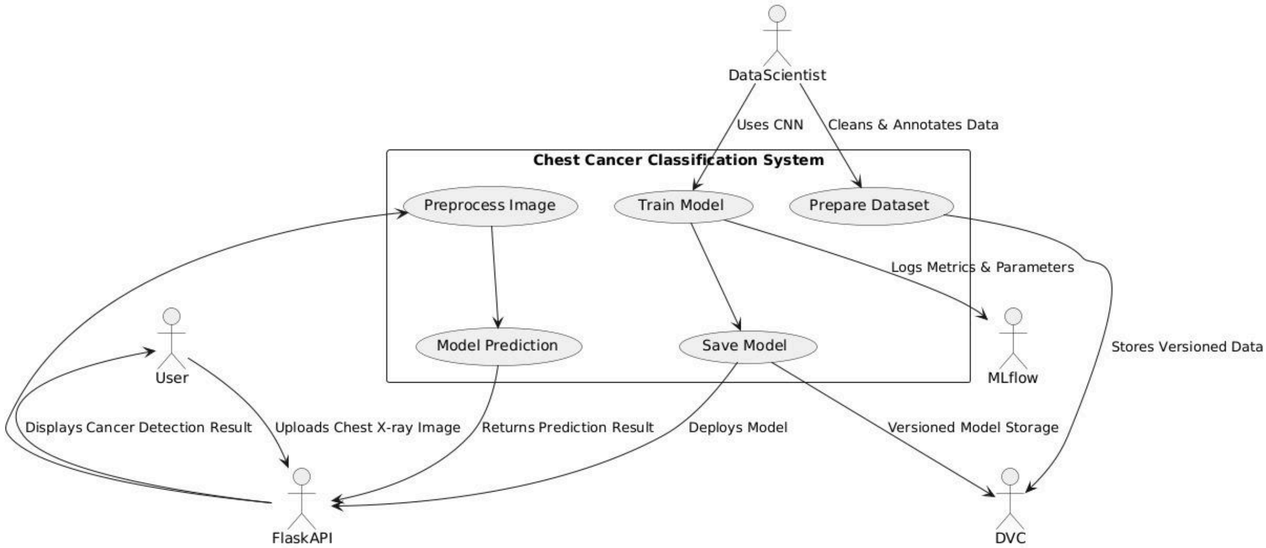
1) Develop a CNN-based Model – Train a deep learning model for classifying chest cancer from medical images.
2) Automate Preprocessing – Implement data cleaning, augmentation, and normalization for better accuracy.
3) Integrate MLflow for Experiment Tracking – Enable real-time tracking of metrics, parameters, and artifacts.
4) Use DVC for Data & Model Versioning – Ensure reproducibility and efficient management of datasets and models.
5) Optimize Model Performance – Improve accuracy, precision, and recall through hyperparameter tuning.
6) Deploy a Web-Based Prediction System – Use Flask to create a user-friendly web interface.
7) Enable Scalability for Future Enhancements – Design a modular architecture to allow integration of new AI techniques.

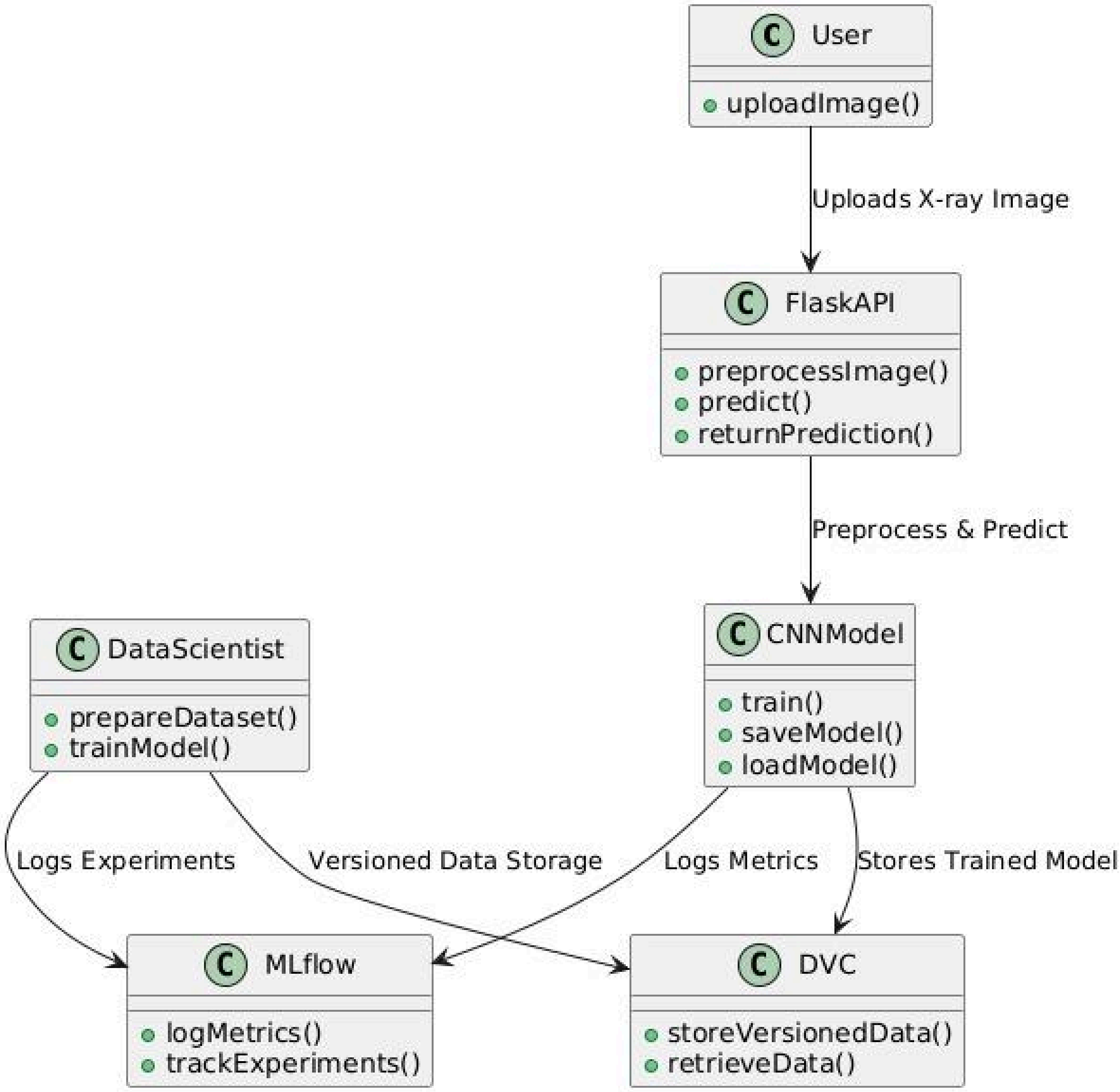
Individual Project Objectives Identification per Team member:

1. Jay Metha, Tejas Redkar - Designed the model
2. Dharmika Tank - Performance Metrics
3. Devanshu Surana - Orchestration tools

H/w & S/w Requirements: <u>Software</u> - **MLFlow, DVC**
<u>Hardware</u> - **Mac Mini M4, Macbook M3**
16 GB RAM, 24 GB RAM

High Level Design [Architecture diagram/block diagram] and its working:





Name & Sign of the Guide