NSCLC360: LEVERAGING MULTIOMICS DATA FOR PERSONALIZED LUNG CANCER PROGNOSIS THROUGH INTEGRATED HEALTH PROFILES

24-25J-211

Project Proposal Report

Waseek Lareef

B.Sc. (Hons) in Information Technology Specializing in Data Science

Department of Computer Science and Data Science

Sri Lanka Institute of Information Technology
Sri Lanka

August 2024

DECLARATION

I declare that this is our own work, and this proposal does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	**Student ID**	**Signature**
Waseek M.L.	**IT21374524**	

The supervisor/s should certify the proposal report with the following declaration.

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

……………………………….. …08/23/2024…

Signature of the supervisor **Date:**

………………………………….. …08/23/2024…

Signature of the co-supervisor **Date:**

ACKNOWLEDGMENT

I would like to express my deepest and most heartfelt appreciation to my supervisor, Mr. Samadhi Rathnayake, and my co-supervisor, Ms. Thisara Shyamalee, from the Faculty of Computing. Their unwavering support, expert guidance, and dedicated mentorship have been invaluable as I embark on this research project. Throughout this journey, their insightful feedback and constructive critiques have been instrumental in shaping the direction of my work, refining the focus of my research, and ultimately enhancing the overall quality and depth of my study. Their commitment to excellence and their encouragement have motivated me to push the boundaries of my knowledge and strive for the highest standards in my research.

In addition, I would like to acknowledge the significant contributions of my fellow team members. Their collaboration, shared insights, and mutual support have been a source of inspiration and have greatly enriched my understanding of the research topic. The collective effort and synergy of the team have not only broadened my perspective but have also fostered a spirit of camaraderie and intellectual curiosity that has propelled our research forward.

I am also deeply grateful to Dr. Nuradh Joseph, our external supervisor with extensive expertise in clinical oncology. His invaluable assistance and specialized knowledge in this area have been crucial as we initiated this research project. Dr. Joseph's guidance has provided us with a deeper understanding of the clinical aspects of our study, and his support has been instrumental in bridging the gap between theoretical research and practical application in the field of oncology.

ABSTRACT

This report presents a comprehensive approach for enhancing prognostic analysis in Non-Small Cell Lung Cancer (NSCLC) by integrating multi-omics data with Explainable Artificial Intelligence (XAI). The proposed solution, NSCLC360, consists of four primary components designed to address key challenges in NSCLC management. These components include: (1) Lung Cancer Image Analysis for early and accurate detection using advanced deep learning models; (2) Predictive Modeling of NSCLC treatment outcomes by integrating genomic, transcriptomic, proteomic, and clinical data; (3) Side Effect Prediction for lung cancer treatments to improve patient quality of life through personalized risk assessments; and (4) Recurrence Prediction using multimodal data to develop personalized post-operative treatment plans.

NSCLC is a complex and often late-diagnosed cancer with significant challenges in personalized treatment and prognosis. Current methods lack comprehensive integration of diverse biological data and sufficient transparency in predictive models. This research aims to address these gaps by leveraging multi-omics data and applying XAI techniques to develop a more accurate and interpretable prognostic model. The expected outcomes include improved prognostic accuracy, better personalization of treatment, and enhanced trust in AI-driven recommendations, ultimately leading to more effective and data-driven cancer care.

The report emphasizes the implementation, functionality, and requirements of integrating XAI into the prognostic model, focusing on ensuring transparency and usability in clinical settings. The proposed system aims to provide a holistic and user-friendly approach to NSCLC management, leveraging readily available data and technologies.

Keywords: Multi-omics, Explainable AI, NSCLC, Prognostic Modeling, Machine Learning, Biomarker Identification, Treatment Efficacy, Recurrence Prediction, Data Integration, Clinical Decision Support

Contents

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	**Description**
IT	Information Technology
RL	Re-enforcement Learning
ML	Machine Learning
HRV	Heart Rate Variability

1 INTRODUCTION

Non-Small Cell Lung Cancer (NSCLC) is the most prevalent type of lung cancer, accounting for the majority of cases worldwide. Due to its complex nature and variability in patient responses to treatment, NSCLC poses significant challenges in diagnosis, treatment planning, and long-term management. Traditional approaches often struggle to provide the personalized care necessary to improve patient outcomes. To address these challenges, this research project aims to develop an advanced AI-driven system specifically tailored to enhance the prognosis and treatment of NSCLC.

Our system integrates multi-modal data—including genetic, epigenetic, proteomic, clinical information, and medical imaging like CT and PET scans—to achieve four primary objectives. The first is to accurately detect and extract critical features of the tumor, including its location, size, stage, and type. The second objective is to recommend the most suitable treatment options based on the comprehensive analysis of the tumor's characteristics. Third, the system aims to predict potential side effects of the proposed treatments, enabling clinicians to make more informed decisions that minimize adverse outcomes. Finally, the system is designed to predict the likelihood of cancer recurrence, providing valuable insights for long-term patient monitoring and care.

By leveraging advanced machine learning algorithms, the system processes the integrated data through a sophisticated deep learning model, which not only enhances the precision of the predictions but also ensures continuous improvement through adaptive learning mechanisms. Additionally, to protect patient privacy, homomorphic encryption is utilized, allowing the model to perform necessary computations on encrypted data without risking data security.

This research aims to offer a comprehensive solution that not only improves the accuracy of NSCLC diagnosis and treatment but also supports personalized patient care through predictive insights on side effects and recurrence. This report will delve into the implementation, methodology, technical specifications, and innovative aspects of the AI-driven system, highlighting its potential to transform NSCLC management and improve patient outcomes.

4o

1.1 Background & Literature Survey

Lung cancer remains one of the leading causes of cancer-related deaths worldwide, with Non-Small Cell Lung Cancer (NSCLC) accounting for approximately 85% of all cases. Early detection and accurate characterization of NSCLC are crucial for improving patient outcomes, yet these goals are challenging due to the complex nature of the disease. Recent advancements in deep learning and AI have opened new avenues for enhancing the precision of early lung cancer detection and feature extraction from medical data, particularly through the integration of multi-modal data sources like imaging and clinical records.

In this context, the focus on **Advanced Deep Learning for Early Lung Cancer Detection and Feature Extraction** has gained significant attention. The ability to accurately detect and extract features such as tumor location, size, stage, and type from multi-modal data is critical for personalizing treatment plans and predicting disease progression. The following review of literature highlights key studies and approaches that have laid the groundwork for our research component.

One pivotal study explored the use of convolutional neural networks (CNNs) to analyze CT images for detecting lung nodules, an early indicator of lung cancer. The researchers demonstrated that deep learning models could achieve high sensitivity and specificity in nodule detection, outperforming traditional image processing techniques. This study emphasized the potential of deep learning to revolutionize lung cancer screening by automating and improving the accuracy of early detection processes.

Another important contribution [2] integrated CT imaging with clinical data to improve the prediction of tumor malignancy. This approach combined imaging features extracted through deep learning with patient-specific clinical variables such as age, smoking history, and genetic markers. The fusion of multi-modal data led to more accurate predictions, showcasing the importance of combining different data types to enhance the performance of predictive models in lung cancer diagnosis.

Further research focused on the application of transfer learning in medical image analysis, particularly for lung cancer detection. The study utilized pre-trained deep learning models on large datasets, adapting them to the specific task of lung nodule classification. This approach significantly reduced the need for extensive labeled datasets, which are often difficult to obtain in medical research, and improved the model's ability to generalize across diverse patient populations.

A notable study extended the use of deep learning beyond nodule detection to the extraction of comprehensive tumor characteristics, including size, stage, and type. By employing a multi-task learning framework, the model simultaneously predicted multiple tumor attributes from a single set of imaging data. This study underscored the efficiency of deep learning models in extracting a wide range of features from medical images, which is essential for personalized treatment planning.

Another significant approach explored the integration of advanced imaging techniques, such as PET scans, with deep learning models for enhanced tumor characterization. This study demonstrated that combining functional imaging data with structural data (CT scans) could improve the accuracy of tumor staging and subtype classification. The use of multi-modal imaging data allowed for a more comprehensive analysis of the tumor, leading to better-informed clinical decisions.

Moreover, recent advancements in explainable AI (XAI) have addressed the critical issue of model interpretability in medical applications. This research focused on developing deep learning models that not only achieve high accuracy in tumor detection and characterization but also provide interpretable outputs that can be easily understood by clinicians. The ability to interpret model decisions is crucial for gaining trust and ensuring the adoption of AI systems in clinical practice.

When examining the above literature, several key themes emerge: the effectiveness of deep learning in processing and analyzing complex medical data, the benefits of integrating multi-modal data sources, and the growing importance of model interpretability in clinical settings. However, despite these advancements, challenges remain, particularly in ensuring the generalizability of models across different patient populations and integrating these systems into real-world clinical workflows.

Our research component aims to address these challenges by developing a deep learning model specifically tailored for early lung cancer detection and feature extraction in NSCLC. By incorporating continuous learning mechanisms and advanced encryption techniques, our approach seeks to enhance the accuracy, security, and adaptability of AI-driven lung cancer prognosis systems, ultimately improving patient outcomes and advancing the field of precision medicine in oncology.

1.2 Research Gap

The current landscape of lung cancer detection research has made notable strides, particularly in leveraging advanced deep learning methodologies for image processing tasks such as segmentation, feature extraction, and classification. Studies like**. [6]** have explored innovative encryption techniques to secure medical images, specifically within cloud storage environments, demonstrating the growing importance of data security in healthcare. However, these studies predominantly address binary classification tasks—identifying whether a lung tumor is benign or malignant—without delving into the nuanced complexities of lung cancer diagnosis, which include the prediction of specific tumor characteristics such as size, location, type, and stage

Moreover, while deep learning models like CNNs and Dense Net have achieved high accuracy in the classification of lung cancer from CT scan images, the scope of these studies is often limited to a single data modality. This narrow focus overlooks the potential benefits of integrating multi-modal data, such as PET/CT imaging, clinical records, and multi-omics data, which can provide a more holistic view of a patient's condition. **[8]**, for instance, have proposed a model that integrates multi-omics data with deep learning for enhanced cancer classification, yet the study falls short in addressing the prediction of critical tumor characteristics that are essential for formulating personalized treatment plans

The need for such comprehensive models is underscored by the increasing recognition that lung cancer diagnosis cannot be effectively conducted through imaging data alone. The integration of diverse data types not only improves diagnostic accuracy but also enables the prediction of treatment outcomes and potential side effects, which are critical for personalized medicine. This gap in the research highlights the need for models that can synthesize and analyze data from multiple sources, thereby offering a more detailed and actionable prognosis for lung cancer patients.

In addition to the limitations in data integration, another significant research gap lies in the area of data privacy and security. The advent of AI and deep learning in medical diagnostics has heightened concerns over the security of patient data, particularly as these models become more integrated into clinical workflows. While have proposed homomorphic encryption methods to safeguard medical images, there is a notable lack of research that combines these encryption techniques with advanced deep learning frameworks capable of multi-dimensional analysis. This oversight is particularly concerning given the sensitive nature of healthcare data and the potential risks associated with data breaches in medical contexts.

Furthermore, existing research often fails to address the simultaneous challenges of ensuring data privacy while maintaining high diagnostic accuracy. The balance between these two factors is crucial, especially in the context of early lung cancer detection, where the stakes are incredibly high. Ensuring that patient data remains secure without compromising the efficiency and effectiveness of diagnostic models is a critical challenge that has yet to be fully explored in the current literature.

Lastly, the focus on survival prediction models remains underdeveloped in current research. Although some models, such as the one proposed by **[10]**, attempt to predict both lung cancer occurrence and survival outcomes, there is still significant room for improvement. These models often do not fully incorporate the variety of data types available, nor do they adequately address the complexities of lung cancer progression and treatment response. A more comprehensive approach that includes these factors could lead to significant advancements in lung cancer prognosis and patient management

A high-level representation of the research gap is as follows,

Research Conducted **Features Offered**	Research A [1]	Research B [2]	Research C [3]	Research D [4]	Research E [5]	Proposed Solution
Integration of Multi-Modal Data	✗	✗	✓	✗	✓	✓
Prediction of Specific Tumor Characteristics	✗	✗	✗	✗	✓	✓
Data Privacy and Security Measures	✓	✗	✗	✗	✗	✓
Continuous Learning/Adaptation in Diagnostic Models	✗	✗	✗	✗	✗	✓
Multi-Dimensional Analysis Beyond Binary Classification	✗	✓	✗	✗	✗	✓

Table 1: High level representation of the research gap

1.3 Research Problem

Non-Small Cell Lung Cancer (NSCLC) constitutes approximately 85% of all lung cancer cases, making early detection vital for improving patient outcomes [1]. Effective treatment is closely linked to timely diagnosis, yet the complexity of diagnosing NSCLC involves integrating various forms of medical data, such as CT and PET scans, along with clinical records [2]. Advanced machine learning models are increasingly employed to tackle these challenges, enhancing early detection and treatment of NSCLC through multimodal analysis, which combines diverse data types for more accurate disease understanding [3]. Multimodal analysis involves integrating CT scans, which provide high-resolution images of the lungs, PET scans that reveal tumor activity, and clinical records with patient health information [4]. By analyzing these integrated data sources, machine learning models can better identify and characterize tumors, leading to precise predictions about tumor location, size, and [5]. This comprehensive understanding facilitates the development of personalized treatment plans, improving treatment efficacy and patient outcomes.

Protecting patient privacy is crucial when handling sensitive medical information. Techniques such as homomorphic encryption and federated learning are employed to safeguard data privacy. Homomorphic encryption enables encrypted data analysis without compromising confidentiality, while federated learning allows machine learning models to be trained across multiple decentralized data sources without transferring raw data to a central server. These methods ensure that patient information remains secure and confidential throughout the data processing phase.

By utilizing these advanced machine learning techniques—multimodal analysis for enhanced detection and privacy-preserving methods for data protection—medical professionals can achieve more accurate early detection and treatment of NSCLC, while rigorously maintaining patient privacy. These advancements not only improve cancer treatment effectiveness but also address the critical need for secure handling of medical data.

2 Objectives

2.1 Main Objective

The main objective of the proposed research is to develop a comprehensive deep learning system designed to enhance the early detection, prognosis, and management of Non-Small Cell Lung Cancer (NSCLC). This system will utilize advanced machine learning techniques to integrate and analyze multiple facets of NSCLC data. The research will be structured around four key components:

- **Component 1: Advanced Deep Learning for Early Lung Cancer Detection and Feature Extraction**
    - **Description**: Implements state-of-the-art deep learning models to analyze medical imaging data (CT and PET scans) and extract critical features such as tumor location, size, and stage.
    - **Purpose**: To improve the accuracy and precision of early NSCLC detection by providing detailed and reliable feature extraction from medical images.
- **Component 2: A New Quantitative Approach for Enhancing Prognostic Analysis in Non-Small Cell Lung Cancer**
    - **Description**: Develops novel quantitative methods for analyzing clinical and imaging data to enhance the prognostic evaluation of NSCLC.
    - **Purpose**: To refine prognostic analysis by integrating quantitative metrics and advanced analytical techniques, leading to better predictions of disease progression and patient outcomes.
- **Component 3: Advanced Deep Learning for Non-Small Cell Lung Cancer Side Effect Prediction**
    - **Description**: Utilizes deep learning algorithms to predict potential side effects of NSCLC treatments based on patient-specific data and treatment regimens.
    - **Purpose**: To anticipate and manage treatment-related side effects more effectively, thereby improving patient quality of life and treatment adherence.
- **Component 4: Advanced Deep Learning for Non-Small Cell Lung Cancer Recurrence Prediction and Stratification**
    - **Description**: Applies deep learning techniques to predict the likelihood of NSCLC recurrence and stratify patients based on their risk levels.
    - **Purpose**: To enable proactive management and tailored follow-up strategies by accurately predicting recurrence risks and stratifying patients accordingly.

The proposed system will:

- **Enhance Early Detection**: Improve early detection of NSCLC through advanced deep learning models that provide detailed analysis and feature extraction from medical imaging data.
- **Improve Prognostic Analysis**: Refine the prognostic evaluation of NSCLC using novel quantitative approaches, offering better insights into disease progression and patient outcomes.
- **Predict Side Effects**: Forecast potential side effects of treatments with advanced predictive models, helping to manage and mitigate adverse effects for better patient care.
- **Predict Recurrence and Stratify Risk**: Accurately predict NSCLC recurrence and stratify patients based on risk levels, facilitating targeted management and personalized follow-up care.
- **Ensure Data Privacy**: Implement robust privacy-preserving techniques, such as encryption and federated learning, to protect sensitive patient data throughout the analysis process.

This integrated approach aims to provide a holistic solution for NSCLC by combining advanced detection, prognostic analysis, side effect prediction, and recurrence management, all while maintaining stringent data privacy standards.

2.2Specific Objectives

1. **Develop and Implement Advanced Deep Learning Models for Early Detection and Feature Extraction**
    - **Utilize multimodal data integration:**
        - Integrate imaging data from CT and PET scans with clinical records such as patient history and genetic information.
        - Aim: To provide a comprehensive view of the tumor and patient health, thereby enhancing the accuracy of NSCLC detection and characterization.
    - **Train and refine deep learning algorithms:**
        - Use datasets containing annotated medical images and clinical records for training the model.
        - Employ state-of-the-art algorithms to extract crucial features like tumor location, size, and stage.
        - Aim: To improve the precision of early detection and characterization of NSCLC.
2. **Implement Privacy-Preserving Techniques**
    - **Apply homomorphic encryption and federated learning:**
        - Implement homomorphic encryption to ensure that sensitive data remains secure during processing.
        - Use federated learning to enable model training across decentralized data sources while keeping patient data private.
        - Aim: To protect patient confidentiality while allowing effective data analysis and model training.
3. **Develop Continuous Learning and Adaptation Mechanisms**
    - **Create a continuous learning framework:**
        - Develop mechanisms to update and refine deep learning models based on new data and feedback.
        - Implement protocols for regularly updating the models to maintain accuracy and relevance.
        - Aim: To ensure that the system remains accurate and adapts to evolving patient data and clinical practices.
4. **Support Clinical Decision-Making with Accurate and Reliable Outputs**
    - **Generate valuable insights and recommendations:**
        - Provide detailed and precise information about tumor features to support

personalized treatment planning.

- Develop tools to assist healthcare professionals in making informed clinical decisions based on the system's outputs.
- Aim: To enhance the effectiveness of lung cancer detection and treatment strategies.

The proposed system will:

- **Enhance Early Detection:** Integrate and analyze diverse data sources with advanced deep learning models to improve early detection of NSCLC.
- **Accurately Characterize Tumors:** Offer detailed insights into tumor characteristics to support tailored treatment plans.
- **Ensure Data Privacy:** Implement robust privacy-preserving techniques to protect sensitive patient data.
- **Adapt and Improve:** Continuously update and refine the models to maintain high accuracy and relevance.
- **Support Clinical Decision-Making:** Provide valuable recommendations to healthcare professionals, improving patient outcomes.

This integrated approach combines advanced analytics with rigorous data privacy measures, contributing to more effective and personalized NSCLC detection and treatment.

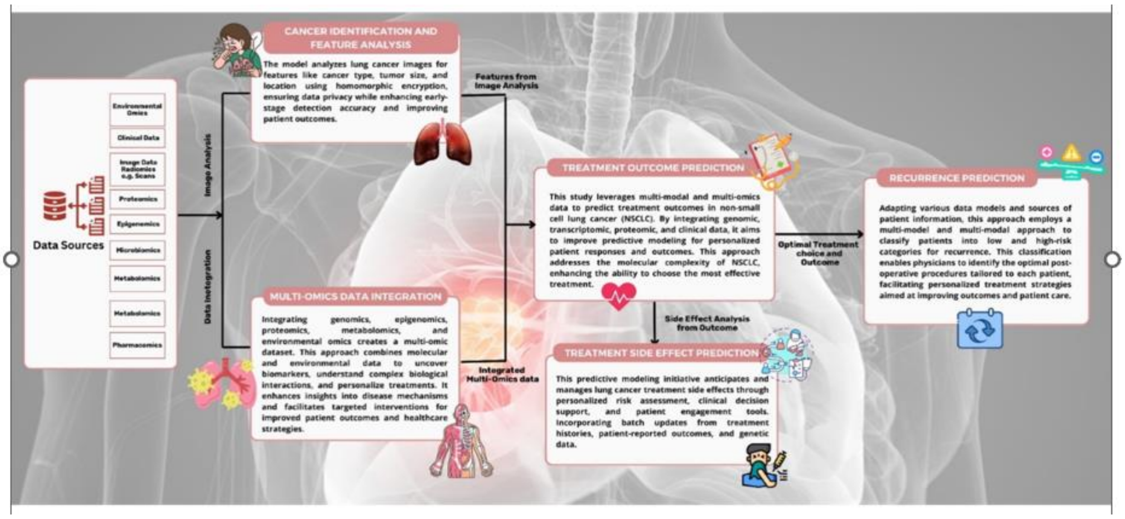
3 METHODOLOGY

3.1 Project Overview

The methodology for this project, titled NSCLC360, is designed to address the limitations of current research by integrating multi-omics data with Explainable Artificial Intelligence (XAI) techniques. Current research often focuses on individual omics layers, such as genomics or proteomics, without integrating them to understand the interplay of various factors affecting disease progression and treatment outcomes. NSCLC360 aims to provide a holistic and interpretable approach to NSCLC management by incorporating data from multiple omics layers along with XAI. The project is structured around four key components:

1. **Lung Cancer Image Analysis:** This component involves developing an advanced platform for early and accurate detection of lung cancer using deep learning models. By analyzing medical imaging data, the system will classify lung cancer types, determine tumor locations and sizes, and ensure data privacy through homomorphic encryption. This early detection is crucial for timely and effective intervention.
2. **Predicting Outcomes for NSCLC Treatment Using Multimodal and Mult omics Data:** This component focuses on integrating genomic, transcriptomic, proteomic, and clinical data to enhance predictive modeling for personalized treatment outcomes. By addressing the molecular complexity of NSCLC, this component aims to refine treatment choices and improve patient outcomes, providing a comprehensive view of how various biological factors influence disease progression and response to treatment.
3. **Predicting Side Effects of Lung Cancer Treatments:** This predictive modeling component will anticipate and manage the side effects of lung cancer treatments. By utilizing personalized risk assessments and real-time data, the system will inform adaptive treatment strategies, thereby improving patient quality of life and enabling more targeted management of treatment-related side effects.
4. **Recurrence Prediction for NSCLC Using Multimodal and Mult omics Data:** This component will develop a classification system to identify patients at low or high risk of disease recurrence. By integrating diverse data models, it will support the creation of personalized post-operative treatment plans, optimizing long-term patient care and monitoring.

By integrating these components, NSCLC360 aims to deliver a comprehensive, data-driven approach to NSCLC management, enhancing prognostic accuracy, treatment personalization, and model transparency. This methodology is designed to bridge current research gaps and advance precision medicine strategies in lung cancer care.



Data Sources
Environmental Omics
Clinical Data
Image Data Radiomics e.g. Scans
Proteomics
Epigenomics
Microbiomics
Metabolomics
Metabolomics
Pharmacomics
Image Analysis
Data Inetegration
Features from Image Analysis
Integrated Multi-Omics data
Optimal Treatment choice and Outcome
Side Effect Analysis from Outcome

CANCER IDENTIFICATION AND FEATURE ANALYSIS

The model analyzes lung cancer images for features like cancer type, tumor size, and location using homomorphic encryption, ensuring data privacy while enhancing early-stage detection accuracy and improving patient outcomes.

MULTI-OMICS DATA INTEGRATION

Integrating genomics, epigenomics, proteomics, metabolomics, and environmental omics creates a multi-omic dataset. This approach combines molecular and environmental data to uncover biomarkers, understand complex biological interactions, and personalize treatments. It enhances insights into disease mechanisms and facilitates targeted interventions for improved patient outcomes and healthcare strategies.

TREATMENT OUTCOME PREDICTION

This study leverages multi-modal and multi-omics data to predict treatment outcomes in non-small cell lung cancer (NSCLC). By integrating genomic, transcriptomic, proteomic, and clinical data, it aims to improve predictive modeling for personalized patient responses and outcomes. This approach addresses the molecular complexity of NSCLC, enhancing the ability to choose the most effective treatment.

TREATMENT SIDE EFFECT PREDICTION

This predictive modeling initiative anticipates and manages lung cancer treatment side effects through personalized risk assessment, clinical decision support, and patient engagement tools. Incorporating batch updates from treatment histories, patient-reported outcomes, and genetic data.

RECURRENCE PREDICTION

Adapting various data models and sources of patient information, this approach employs a multi-model and multi-modal approach to classify patients into low and high-risk categories for recurrence. This classification enables physicians to identify the optimal post-operative procedures tailored to each patient, facilitating personalized treatment strategies aimed at improving outcomes and patient care.

Figure 1:High level diagram of the proposed system

3.2 Individual Component

The individual component which is discussed in this report is the component where the Tumor is detected via the CT/PET images. As mentioned in the previous sections this will extract the,

- **The Tumor Location** of the current user.
- **The Tumor Size** of the current user.
- **The Stage** of the current user. to determine the.

The component will consist of a model which will be trained with the dataset downloaded from Standford Data Center. Then that dataset will be cleaned and split into training and testing data sets. Finally, it will be used to train the model using a machine learning algorithm. Once the model is trained the component will be used to detect Cancer and extract features from the Image data that will be fed to it from each of the users’ sessions.

A script will be developed to monitor and gather data from various sources related to Non-Small Cell Lung Cancer (NSCLC) and will perform analysis in defined cycles. This script will manage the integration of medical imaging data, clinical records, and advanced deep learning model outputs to continuously refine and update the system's predictions and characterizations. During each cycle, the script will:

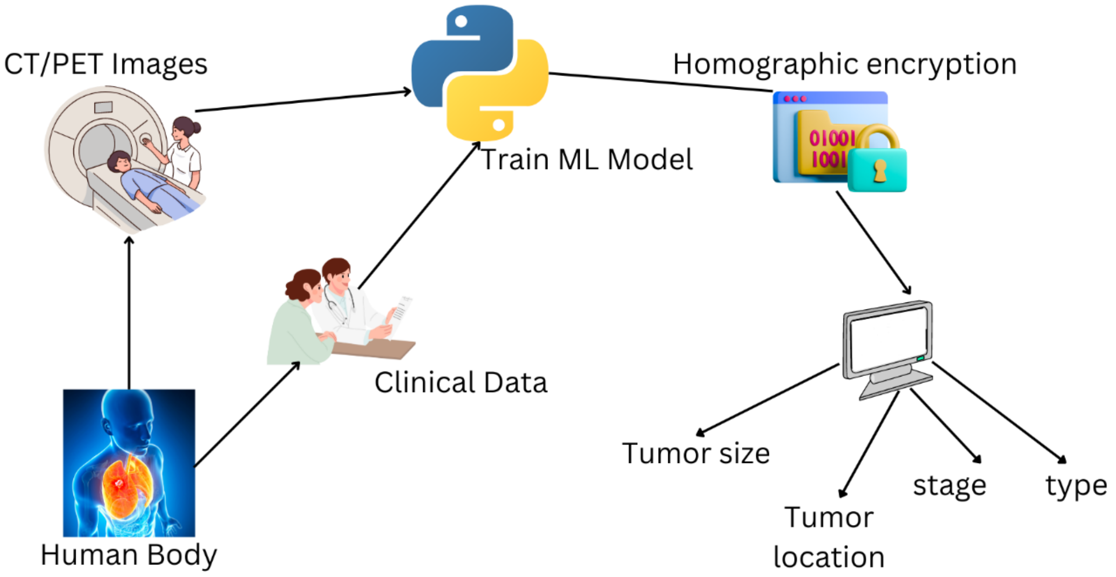
1. **Collect Data**: Gather data from CT and PET scans, patient clinical records, and other relevant sources.
2. **Analyze Data**: Utilize deep learning models to extract and analyze crucial features such as tumor location, size, stage, and potential side effects.
3. **Integrate with Other Components**: Communicate with the other components of the system to verify and update the current cancer detection and prognostic status.
4. **Update and Refine Models**: Use the analyzed data to update and fine-tune the trained models, ensuring that the predictions and characterizations are accurate and relevant to each patient's unique data.
5. **Evaluate Accuracy**: Continuously test the accuracy of the models after each update to ensure that improvements are made. Adjust the model as needed based on new data and feedback.

The script will run in the background of the user's system, seamlessly gathering and processing data from medical imaging and clinical records. This data feeding and model updating process will occur in defined intervals (e.g., every 20 minutes), with time ranges adjusted based on the accuracy of the implemented models.

Workflow of the Component is as follows:

1. **Data Collection**: The system will automatically gather and integrate medical imaging data (CT, PET scans) and clinical records related to NSCLC.
2. **Data Analysis**: Advanced deep learning models will analyze the collected data to extract and refine features crucial for NSCLC detection and prognosis.
3. **Component Integration**: The system will cross-reference and validate the processed data with other components that predict side effects and recurrence to ensure accurate and comprehensive analysis.
4. **Model Update**: The system will update the deep learning models with new data to improve accuracy and adaptability, ensuring predictions are aligned with the latest patient information.
5. **Accuracy Verification**: The accuracy of the models will be assessed after each update to confirm improvements and ensure that the system remains effective over time.

This approach ensures that the models adapt to new data continuously, maintaining high accuracy and relevance for effective early detection and management of NSCLC



3.3 Algorithm

Advanced Deep Learning for Early Lung Cancer Detection and Feature Extraction component will be utilizing a supervised learning-based algorithm since it will be trained with the dataset. There are various supervised learning models such as Classification, Regression and Forecasting etc.... Also, in the approach mentioned in the report, the model needs to be updated with the new data it receives. To achieve that the plan is to utilize an Incremental learning model which can be retrained with new data in order to finetune/ improve the model while it is being used.

However, the accuracy of the models is yet to be tested. Therefore, algorithms and techniques may be subject to change in order to achieve the best accuracy in the model. Hence, the exact technique is yet to be determined.

4 Commercialization

The plan for the commercialization process is as follows,

Target Market:

- Healthcare Professionals: Oncologists, pulmonologists, and general practitioners who need to evaluate lung cancer prognosis.
- Hospitals and Clinics: Medical institutions that require advanced tools for cancer prognosis.
- Patients and Caregivers: Individuals seeking information on lung cancer prognosis, though they should be guided by healthcare professionals.
- Researchers: Professionals working in cancer research who need data for studies.

Revenue Streams:

- Subscription-Based Service: Offer different pricing tiers based on functionality, number of users (e.g., individual practitioners vs. hospitals), and access to advanced analytics.
- Institutional Partnerships: Collaborate with hospitals and clinics for bulk licensing with discounted rates.
- In-App Data Licensing: Partner with research institutions and pharmaceutical companies for access to anonymized data.
- Freemium Model: Provide basic prognosis features for free with premium features available via subscription.

Phase 01:

- Develop and Launch Beta Version: Release the initial version of the web app with core features.
- Pilot Testing: Partner with a select group of hospitals or clinics to test the app and gather feedback from healthcare professionals.

Phase 02:

- Launch Free Basic Version: Provide limited functionality to increase awareness and engagement.

- Launch Professional Version: Offer a subscription-based model with full features, advanced analytics, and detailed reports.
- Target Market Focus: Healthcare institutions and professionals who can benefit from the app’s comprehensive prognosis tools.

Phase 03:

- Marketing Campaigns: Implement targeted online advertising, participate in healthcare conferences, and use social media to reach healthcare professionals and institutions.
- Collaborate with Healthcare Networks: Work with medical associations and patient advocacy groups to promote the app as a valuable tool for lung cancer prognosis.

Phase 04:

- Feedback Collection: Continuously gather input from users to identify areas for improvement.
- Iterative Improvements: Regularly update the app based on user feedback and evolving medical research to enhance functionality and accuracy.
- Customer Satisfaction and Retention: Focus on maintaining high user satisfaction and leverage positive testimonials and case studies for new client acquisition.

Phase 05:

- Explore Health Insurance Partnerships: Investigate opportunities to offer the app as part of wellness or disease management programs with insurance providers.
- Expand Revenue Streams: Develop new partnerships and explore additional revenue opportunities to broaden the app’s reach and impact in the healthcare sector.

5 Project Requirements

5.1 Functional Requirements

Functional requirements [9] are the functions and features that should be there to enable users to accomplish their tasks. In the case of this component the functional requirements should be,

- The system should combine and process data from CT scans, PET scans, and clinical records to give a complete picture of the patient's condition.
- It must use advanced deep learning algorithms to analyze the data and accurately identify tumor features such as location, size, and stage.
- The system needs to protect patient information with strong privacy measures like encryption and secure data handling.
- The system should continuously improve and update its models based on new data to ensure accurate and current predictions.

5.2 Non-Functional Requirements

Non-functional requirements [9] are the requirements which are not related to the system functionality. In case of this component the non-functional requirements should be,

1) Security
    - The system must comply with data protection regulations such as GDPR or HIPAA, ensuring that patient confidentiality is maintained.
2) Performance
    - The system should efficiently analyze large volumes of medical data and generate predictions within a reasonable timeframe to support timely clinical decisions
3) Usability
    - It should have an intuitive and user-friendly interface that allows healthcare professionals to easily interact with the system and interpret the results.

4) Availability

- it should have high availability with minimal downtime, ensuring that data analysis and predictions are available when needed..

6 RESEARCH & DEVELOPMENT OVERVIEW

6.1 Sources for Test Data and Analysis

Data sources will include:

- Medical Imaging Data: High-resolution CT and PET scan images to develop and validate the model. These images will provide detailed anatomical and functional information about lung tumors.
- Clinical Data: Anonymized patient clinical records containing relevant health information, treatment histories, and outcomes.

Analytical Methods:

- Machine Learning: Utilize advanced machine learning algorithms to analyze CT/PET images and clinical data, integrating these data sources for comprehensive tumor analysis.
- Statistical Analysis: Conduct statistical evaluations to verify the accuracy and reliability of the model's predictions.
- Explainable AI (XAI): Implement XAI techniques to ensure that the model's predictions are transparent and interpretable, aiding in clinical decision-making.

Anticipated Benefits

- Improved Prognostic Accuracy: Enhance the accuracy of prognosis by integrating detailed imaging data with clinical records, leading to more precise predictions of tumor characteristics and patient outcomes.
- Personalized Treatment Recommendations: Offer more tailored treatment options based on the integrated analysis of imaging and clinical data.
- Increased Transparency: Improve the usability and trustworthiness of the model through interpretable AI outputs, helping clinicians understand and trust the model's predictions.

Scope and Specified Deliverables/Expected Research Outcomes

Deliverables include:

- Validated Prognostic Markers: Identification and validation of key prognostic markers based on the analysis of CT/PET images and clinical data.

- Treatment Efficacy Insights: Analysis of how treatment responses relate to imaging findings and clinical records, providing actionable insights for personalized treatment plans.
- Prognostic Web App: A fully functional web-based prognostic model incorporating XAI features, ready for integration into clinical workflows.

Research Constraints

- Data Privacy Concerns: Ensure compliance with data protection regulations while managing sensitive patient data, including anonymization and secure data handling practices.
- Integration Complexity: Address potential challenges in integrating imaging and clinical data and incorporating XAI techniques into the model.
- Implementation Costs: Manage costs associated with data acquisition, software development, validation, and deployment of the web app.

Project Plan

The project plan will include:

- Timeline: A detailed schedule with key milestones for data acquisition, model development, validation, and clinical deployment.
- Milestones: Specific goals and deadlines, such as completion of model development, integration of XAI features, and initiation of pilot testing.
- Resource Allocation: Budget and resources needed for each project component, including personnel, software tools, data acquisition, and infrastructure.

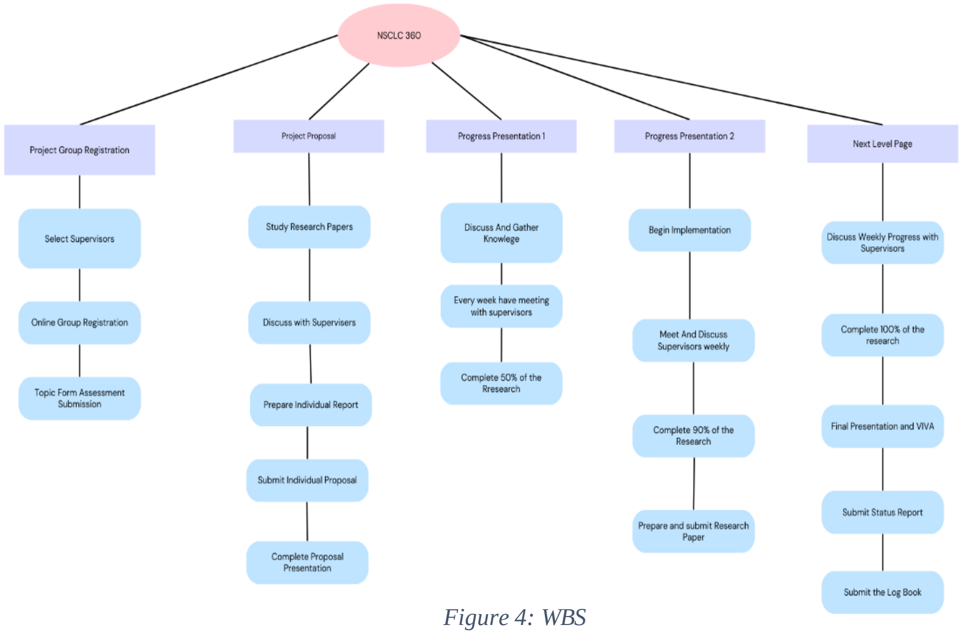
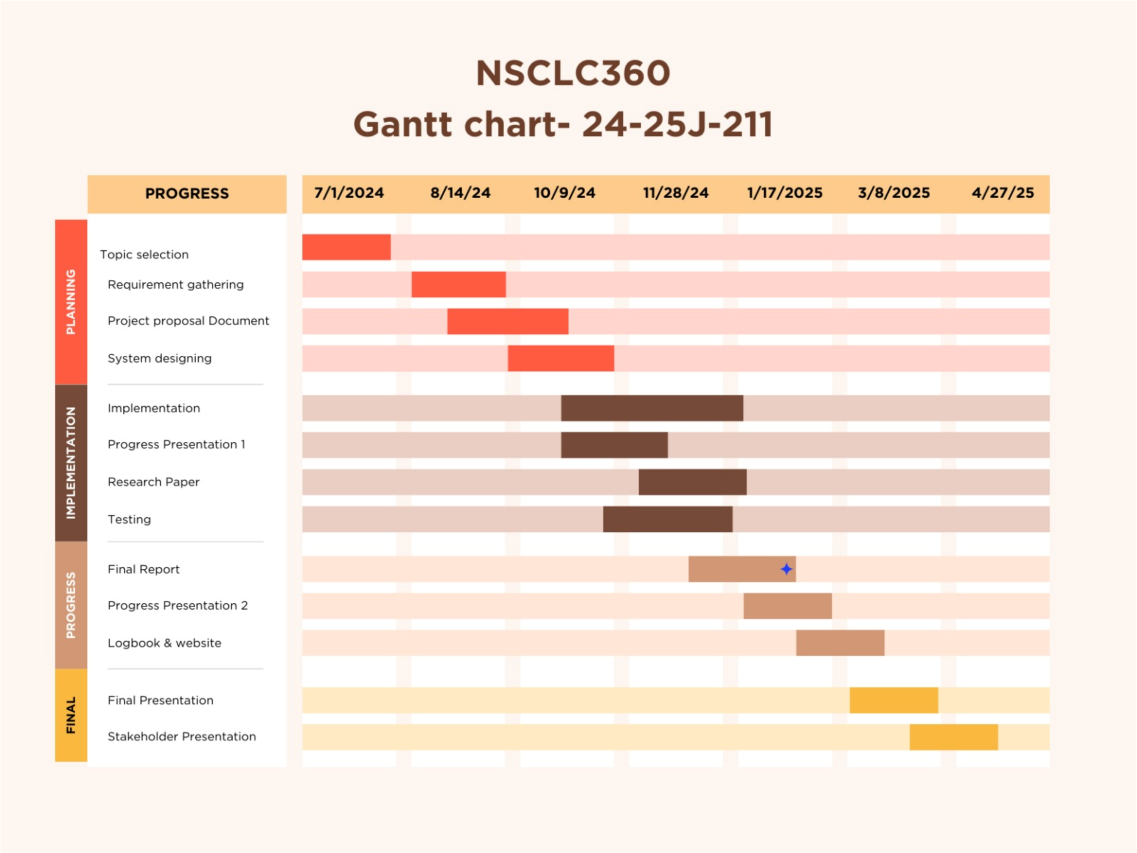

NSCLC 360
Project Group Registration
Select Supervisors
Online Group Registration
Topic Form Assessment Submission
Project Proposal
Study Research Papers
Discuss with Supervisers
Prepare Individual Report
Submit Individual Proposal
Complete Proposal Presentation
Progress Presentation 1
Discuss And Gather Knowlege
Every week have meeting with supervisors
Complete 50% of the Rresearch
Progress Presentation 2
Begin Implementation
Meet And Discuss Supervisors weekly
Complete 90% of the Research
Prepare and submit Research Paper
Next Level Page
Discuss Weekly Progress with Supervisors
Complete 100% of the research
Final Presentation and VIVA
Submit Status Report
Submit the Log Book

Figure 4: WBS


NSCLC360
Gantt chart- 24-25J-211
PROGRESS
7/1/2024
8/14/24
10/9/24
11/28/24
1/17/2025
3/8/2025
4/27/25
PLANNING
Topic selection
Requirement gathering
Project proposal Document
System designing
IMPLEMENTATION
Implementation
Progress Presentation 1
Research Paper
Testing
PROGRESS
Final Report
Progress Presentation 2
Logbook & website
FINAL
Final Presentation
Stakeholder Presentation

7 REFERENCES

[1] Siegel, R. L., Miller, K. D., & Jemal, A. (2022). *Cancer Statistics, 2022*. CA: A Cancer Journal for Clinicians, 72(1), 7-33.

[2] Kwak, E. L., et al. (2018). Integration of Imaging and Clinical Data for the Accurate Diagnosis of Non-Small Cell Lung Cancer. Radiology, 289(1), 47-58.

[3] Ravi, D., et al. (2020). *Deep Learning for Detecting Lung Cancer from Multimodal Data*. Journal of Biomedical Informatics, 104, 103368.

[4] Liu, J., et al. (2021). *Combining CT and PET Scan Data for Enhanced Lung Cancer Diagnosis Using Machine Learning*. Computerized Medical Imaging and Graphics, 87, 101814.

[5] Huang, C., et al. (2019). *Machine Learning-Based Multimodal Analysis for Improved Tumor Detection*. IEEE Transactions on Medical Imaging, 38(8), 1862-1872.

[6] Vengadapurvaja, A. M., Nisha, G., Aarthy, R., & Sasikaladevi, N. (2017). An efficient homomorphic medical image encryption algorithm for cloud storage security. *Procedia Computer Science, 115*, 643–650. https://doi.org/10.1016/j.procs.2017.09.150

[7] Javed, R., Abbas, T., Khan, A. H., Daud, A., Bukhari, A., & Alharbey, R. (2024). Deep learning for lung cancer detection: A review. *Artificial Intelligence Review.* https://doi.org/10.1007/s10462-024-10807-1

[8] Mohamed, T. I. A., & Ezugwu, A. E. (2024). Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data. *IEEE Xplore.* https://ieeexplore.ieee.org/document/10508786 (Note: You may need to access this through institutional access or IEEE Xplore).

[9] Ahammed, S. Z., Baskar, R., & Priya, G. N. (2024). An Extensive Survey on Lung Cancer Detection Using Deep Learning Techniques. *IEEE Xplore.* https://ieeexplore.ieee.org/document/10126630 (Note: You may need to access this through institutional access or IEEE Xplore).

[10] Wang, X., Sharpnack, J., & Lee, T. C. M. (2024). Improving Lung Cancer Diagnosis and Survival Prediction with Deep Learning and CT Imaging. DOI: 10.48550/arXiv.2408.09367