PROJECT REPORT

LYMPHORAPHY CLASSIFICATION USING MACHINE LEARNING

TEAM ID: 593213

TEAM MEMBERS:

A. JEBA PEARLLIN (21BDS0320)

Sudarsanaa C R (21BDS0143)

Shruthi Gangam (21BCB0063)

Jasti Pavan (21BDS0309)

INTRODUCTION

1.1 Project Overview

Lymphography, a diagnostic imaging technique that involves the visualization of the lymphatic system, plays a crucial role in the diagnosis and treatment of various medical conditions, including lymphedema and lymphatic disorders. In recent years, machine learning has demonstrated its potential to enhance the accuracy and efficiency of lymphography interpretation. This project aims to develop a comprehensive solution for lymphography classification using machine learning, leveraging cutting-edge technologies and techniques to provide valuable insights for healthcare professionals.
Project Objectives:
1. Accurate Classification: The primary objective of this project is to develop a robust machine learning model capable of accurately classifying lymphography images into relevant categories. This includes distinguishing between normal and abnormal lymphatic patterns, as well as identifying specific disorders and conditions.

2. Real-time Inference: The project will focus on enabling real-time image classification, providing rapid results to medical professionals for timely decision-making. This will be achieved through the deployment of a user-friendly web application with a responsive interface for image upload and result display.

3.Scalability: As the demand for lymphography classification services grows, the system will be designed for scalability, allowing it to handle an increasing volume of incoming image data without compromising performance or availability. Load balancing and efficient resource management will be key considerations.

4.Data Security and Privacy: Given the sensitive nature of medical data, robust security and privacy measures will be implemented to ensure the protection of patient information. Compliance with healthcare regulations, such as HIPAA and GDPR, will be a top priority.

5. Monitoring and Continuous Improvement: The project will include comprehensive monitoring and logging mechanisms to track system health and performance. This data will be used for system maintenance and optimization. Additionally, a feedback loop will be established to gather user input and continuously improve the system's accuracy and usability.

1.2 Purpose

The purpose of the project focused on lymphography classification using machine learning is multifaceted, driven by the need for improved medical diagnostics and healthcare efficiency. The primary objectives and purposes of this endeavour are as follows:

1. Enhancing Diagnostic Accuracy:

One of the central purposes of this project is to improve the accuracy of lymphography image interpretation. Human interpretation of medical images can be subject to variability and error. Machine learning, through the use of advanced algorithms and deep learning models, aims to provide more consistent and precise results in classifying lymphography images. By reducing diagnostic errors, the project contributes to better patient outcomes and more reliable clinical decisions.

2. Timely Medical Decision-Making:

The project seeks to facilitate rapid and real-time lymphography image classification. Timely access to classification results is critical in healthcare, enabling medical professionals to make informed decisions swiftly. By reducing the time required for image analysis and interpretation, the project supports timely medical interventions and treatment planning, which can be particularly crucial in cases of lymphatic disorders and related conditions.

3. Scalability and Efficiency:

Scalability is another significant purpose of this project. As medical facilities and healthcare providers generate an increasing volume of medical imaging data, the system will be designed to efficiently process and classify a large number of lymphography images. Scalability ensures that the system remains responsive and accessible, even in the face of growing data demands, ultimately benefiting both medical professionals and patients.

4. Data Security and Privacy:

Ensuring data security and privacy is a core purpose. Medical data, especially images containing sensitive patient information, must be safeguarded against unauthorized access or breaches. The project is dedicated to implementing state-of-the-art security measures and compliance with healthcare data protection regulations, such as HIPAA and GDPR, to protect patient privacy and maintain data integrity.

2. LITERATURE SURVEY

2.1 Existing problem

Lymphography classification using machine learning has emerged as a promising approach to improve the accuracy and efficiency of lymphatic system diagnosis. However, the field faces several challenges and problems that require attention and innovative solutions. This literature survey aims to highlight the existing problems in the domain of lymphography classification to inform further research and development.

1. Limited Data Availability: The scarcity of labelled lymphography datasets is a significant hurdle. Machine learning models, particularly deep learning algorithms, require substantial amounts of data

to achieve high accuracy. With lymphography data being relatively limited, the challenge lies in obtaining diverse and comprehensive datasets for model training.

2.Interoperability and Data Standardization: Lymphography images are often stored in various formats, and the lack of standardized data representation poses a challenge for machine learning applications. Ensuring data interoperability and consistency across different healthcare systems and institutions is crucial for creating robust classification models.

3.Imbalanced Data: In lymphography, certain conditions or abnormalities may be rare, leading to imbalanced datasets. Imbalanced data can result in biased models that perform well on majority classes but struggle with minority classes. Developing effective strategies to handle imbalanced data is a problem that requires attention.

4.Model Interpretability: Deep learning models, such as Convolutional Neural Networks (CNNs), are known for their black-box nature. Interpretability of model predictions is a critical concern in the medical domain, where clinicians and radiologists need to understand the reasoning behind a classification. Ensuring transparency and interpretability of machine learning models in lymphography classification is a complex problem.

5. Data Privacy and Security: The integration of machine learning into healthcare requires a robust framework for data privacy and security. Patient data in lymphography images is sensitive and must be protected in compliance with healthcare regulations. Balancing data utility with privacy is a challenging problem in the development of machine learning systems for lymphography.

In conclusion, lymphography classification using machine learning holds great promise, but it also presents several existing problems and challenges that need to be addressed. Solutions to these problems will pave the way for more accurate, efficient, and secure lymphatic system diagnosis, benefiting both healthcare providers and patients. Future research and development efforts must focus on these challenges to advance the field.

2.2 References

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

2. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

3. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500-510.

4. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

5. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR).

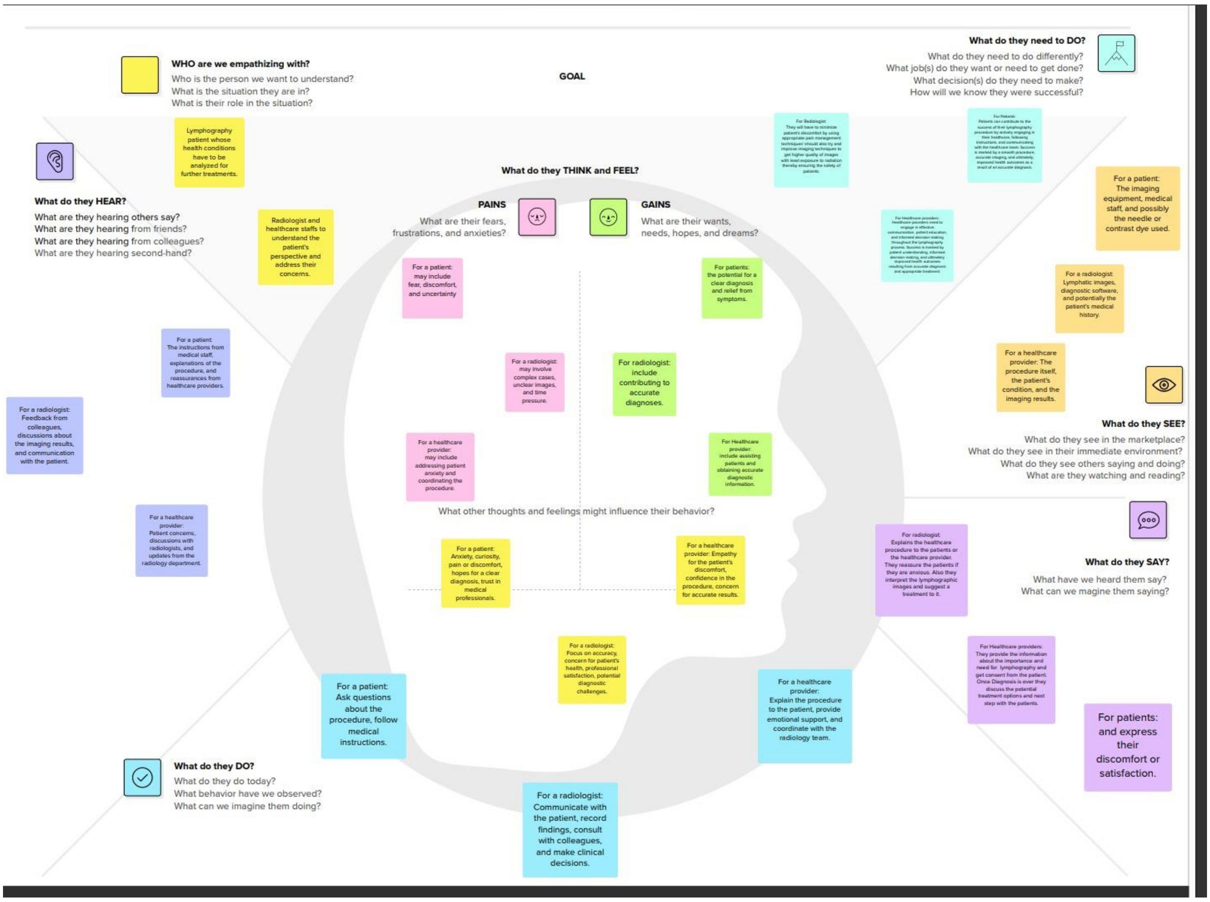
6. Su, H., Zheng, Y., Lu, F., Zhu, L., Li, S., & Lu, Z. (2020). A deep-learning-based computer-aided diagnosis system for lymphoma classification on histopathological images. Journal of healthcare engineering, 2020.

2.3 Problem Statement Definition

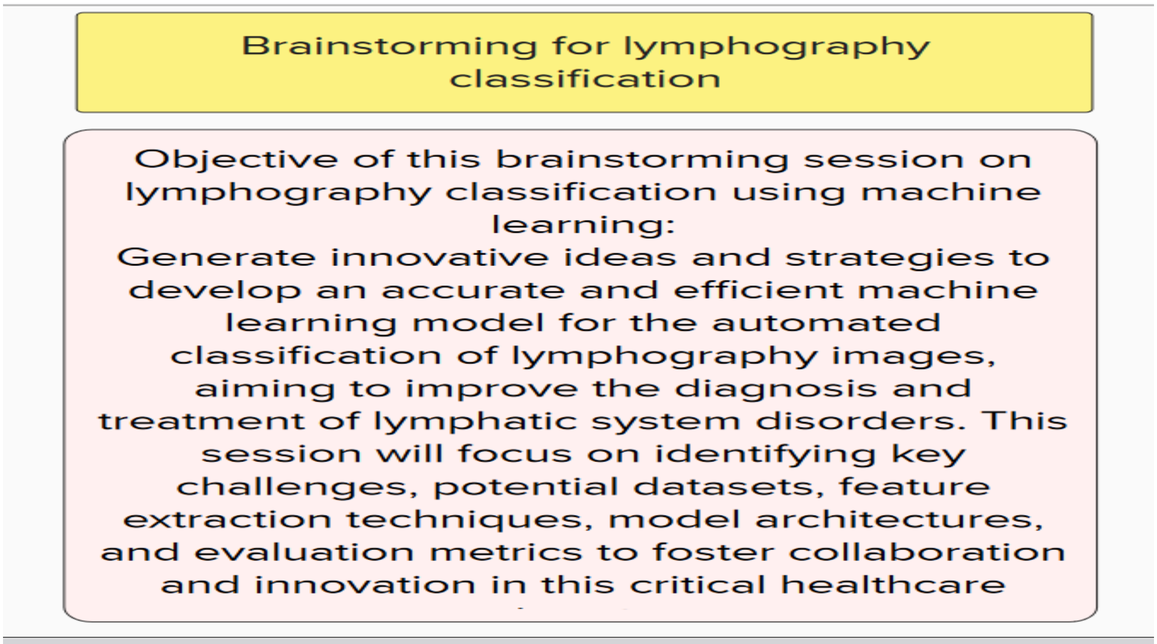
The problem at the core of lymphography classification using machine learning is the accurate and efficient categorization of lymphography images, which is pivotal for the timely diagnosis and treatment of various lymphatic disorders. This challenge encompasses the need to differentiate between normal and abnormal lymphatic patterns, identify specific conditions, and provide invaluable support to healthcare professionals. However, this endeavour faces several significant hurdles. These include the scarcity of comprehensive and diverse labelled lymphography datasets required for robust model training, the imbalanced distribution of classes in these datasets, and the lack of standardized data formats. Achieving real-time inference capabilities, while ensuring data privacy, model interpretability, and regulatory compliance, also poses complex problems. Moreover, the challenge extends to building machine learning models that can generalize across diverse patient populations and healthcare settings. Addressing these multifaceted challenges is fundamental to the development of accurate, reliable, and secure lymphography classification systems, which have the potential to revolutionize the field of medical diagnostics, benefiting both healthcare providers and patients.

3. IDEATION & PROPOSED SOLUTION

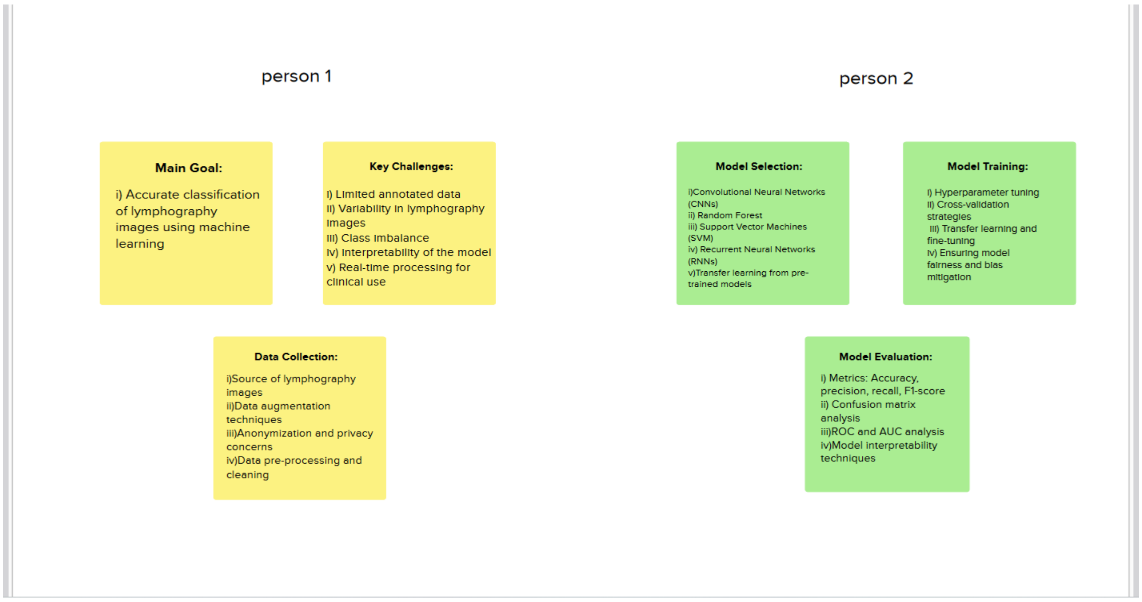
3.1. Empathy Map Canvas:

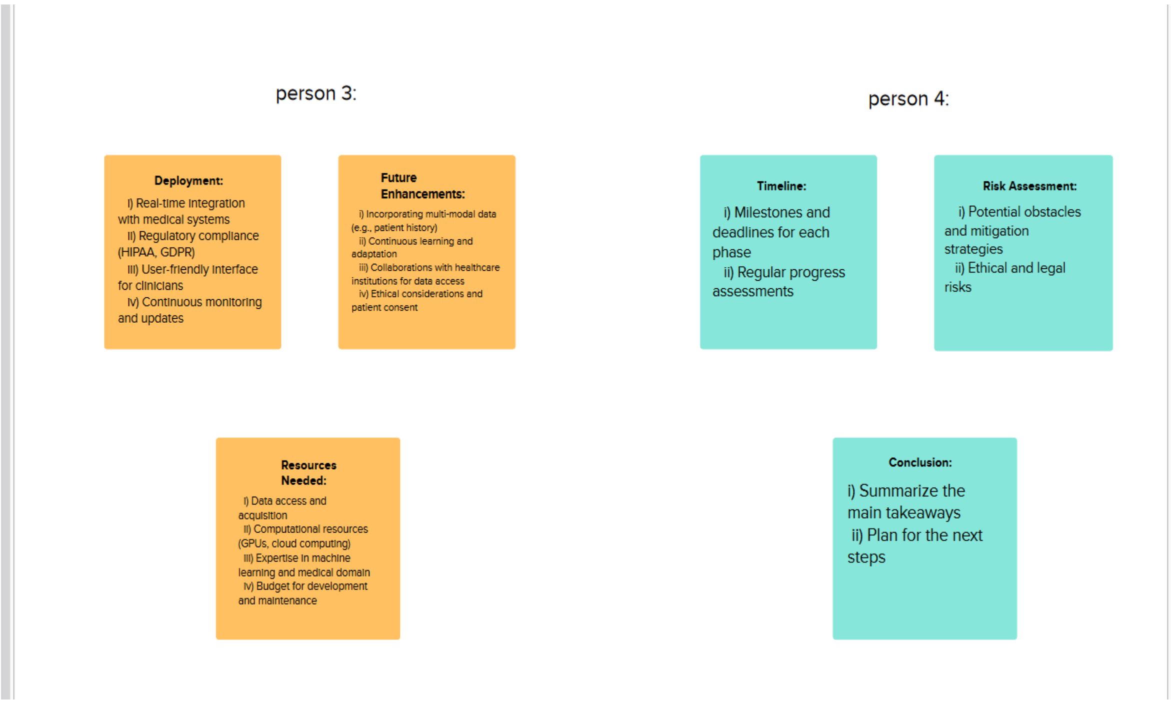


3.2.Ideation & Brainstorming:

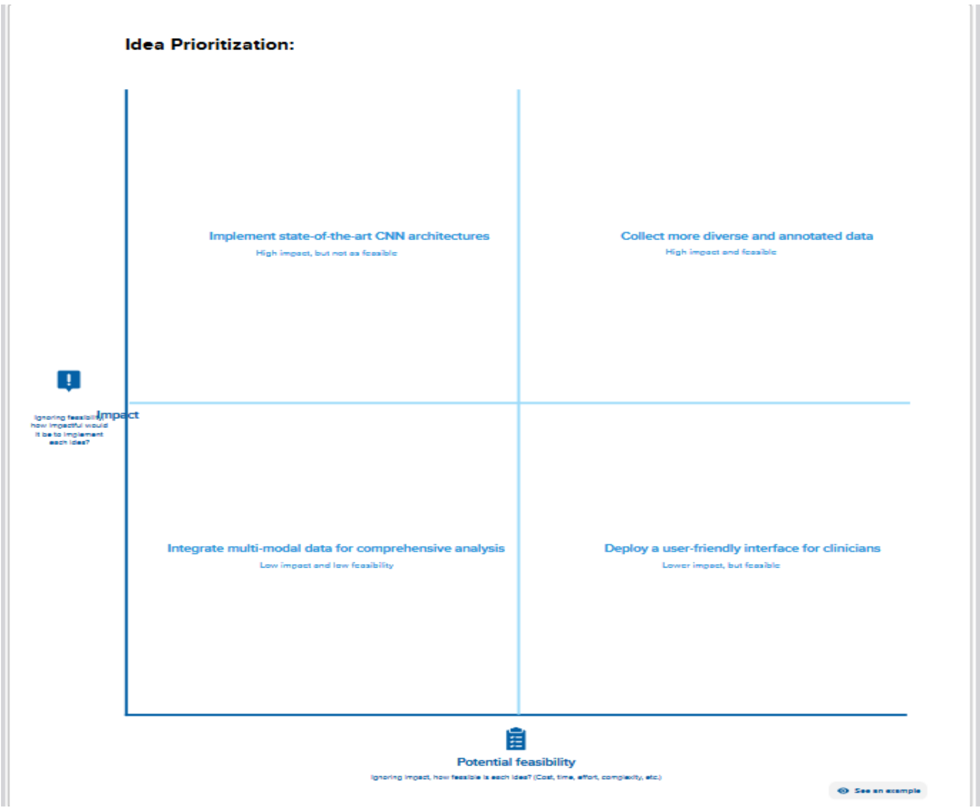


<u>Brainstorm solo session:</u>





Idea prioritization:



4.REQUIREMENT ANALYSIS

4.1.Functional requirement:

Certainly, let's divide the provided requirements into functional and non-functional categories:

1. Idea / Solution Description:

- Develop a system that can accurately classify lymphography data using machine learning techniques.

- Implement various components that work together to achieve efficient classification.

2. Social Impact / Customer Satisfaction:

- Improve healthcare delivery and patient outcomes through accurate diagnosis of lymphatic disorders.

- Ensure the solution positively impacts individuals affected by these disorders.

3. Business Model (Revenue Model):

- Focus on advancing and licensing technology to support the project's goals.

- Prioritize enhanced health and well-being as part of the business strategy.

4.2. Non-Functional requirements

1. Novelty / Uniqueness:

- Stand out in the field of lymphatic disorder diagnosis by integrating cutting-edge machine learning and advanced data processing techniques.

- Contribute to the field with innovative features.

2. Scalability of the Solution:

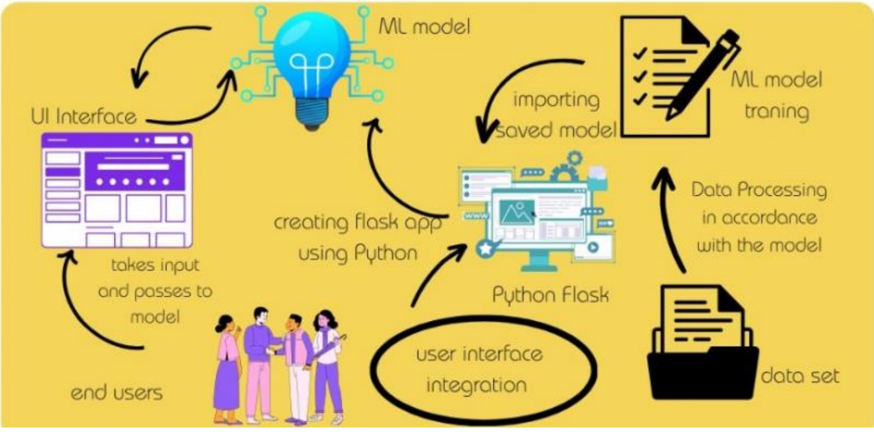
- Ability to handle larger datasets effectively as data requirements grow.

- Seamless integration with current and future technologies to ensure long-term efficacy.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories:

Data Flow Diagrams: A Data Flow Diagram (DFD) is a conventional graphical representation illustrating the flow of information within a system. A well-organized and lucid DFD can visually portray the precise system requirements. It illustrates the entry and exit points of data in the system, tracks alterations in information, and identifies data.



User Stories:

User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Patient	Lymphography Workflow Optimization	USN-1	As a user, I want a secure patient portal that allows me to access my lymphography data, reports, and post-procedure care instructions online.	I can trust my data with the portal	Medium	Sprint 1
Patient		USN-2	As a user, I want to be able to request prescription refills online through the patient portal for a streamlined medication management process.	I can make request through online portal	Medium	Sprint 1
Radiologist		USN-3	As a user, I want a digital system for securely storing and sharing lymphography data and reports, ensuring quick access for diagnosis and consultation.	I can trust my data's security with the portal	Low	Sprint 2
Radiologist		USN-4	As a user, I want an automated scheduling system that optimizes the allocation of lymphography appointments to available slots, reducing wait times and improving resource utilization.	I can quickly update and automate data.	Low	Sprint 2
Healthcare provider	Enhancing the patient experience	USN-5	As a user, I want to offer patients the option to provide feedback on their experiences and suggest improvements, helping us continuously enhance our services.	I can improve through people's review	Medium	Sprint 1
Patient		USN-6	As a user, I want to receive clear and timely communication about my upcoming appointments and any necessary preparations, so I can be well prepared and reduce anxiety.	I can be informed with my schedule	High	Sprint 1

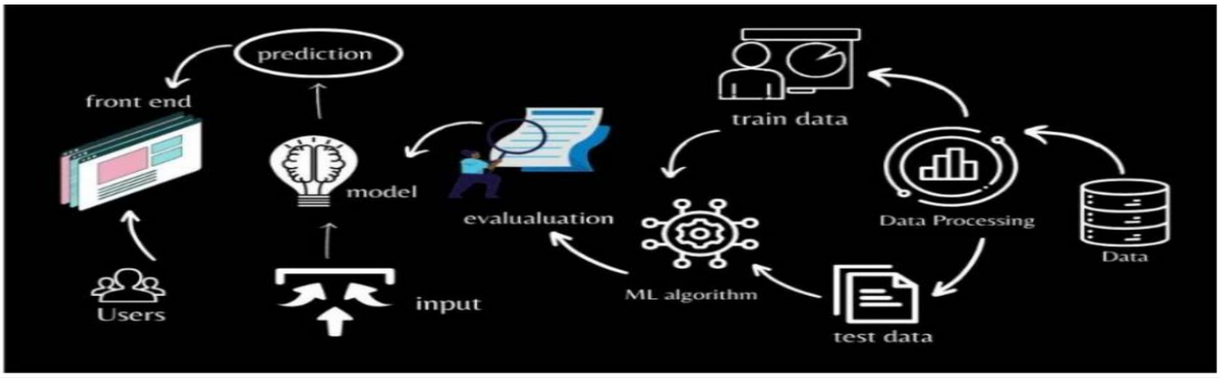
Patient		USN-7	As a user, I want to receive personalized post procedure care instructions and reminders through email or a mobile app, aiding my recovery and follow-up process.	I can be reminded of my appointments	Medium	Sprint 1
Healthcare Administrator		USN-8	As a user, I want to offer educational materials within our facilities and online resources on various health topics to empower patients to take control of their wellbeing.	I can share educational resources	Medium	Sprint 1
Healthcare Provider		USN-9	As a user, I want to easily access a patient's medical history, including allergies and prior diagnoses, during clinical visits to provide informed and personalized care.	I can be informed with patient's past medical record	High	Sprint 2
Patient		USN-10	As a user, I want to securely access and share my medical records with other healthcare providers, ensuring seamless continuity of care during referrals and consultations.	I can share my past medical record	High	Sprint 3
Healthcare Manager		USN-11	As a user, I want to offer e-learning modules on new healthcare regulations and best practices, ensuring that staff can easily access and complete their training.	I can share information with the staff	Low	Sprint 3

5.2 Solution Architecture:

The solution architecture for Lymphography classification is a sophisticated framework designed to automate the diagnosis of lymphatic disorders using advanced machine learning techniques. At its core, the architecture comprises a user-friendly interface that allows for the seamless upload of lymphography data. This data then undergoes a rigorous preprocessing stage, enabling us to train the model in an effective way. It offers a powerful tool to healthcare professionals for more precise and timely assessments.

Our solution leverages ML models to address the Lymphography classification problem effectively

- Data Collection: This component is responsible for collecting lymphography data and their corresponding labels. The data is stored in a data lake, while the labels are stored in a database.
- Data Processing: This component is responsible for processing the lymphography data to prepare them for training. This includes tasks such as outlier removal, SMOTE, etc.
- Model Training: This component is responsible for training the machine learning model using the pre-processed lymphography data. The trained model is stored in a model registry.
- Model Deployment: This component is responsible for deploying the trained model to a production environment. The deployed model is used to classify new lymphography data.
- User Interface: This component provides a user interface for interacting with the system. Users can input data to be classified, and view the results of classifications.



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

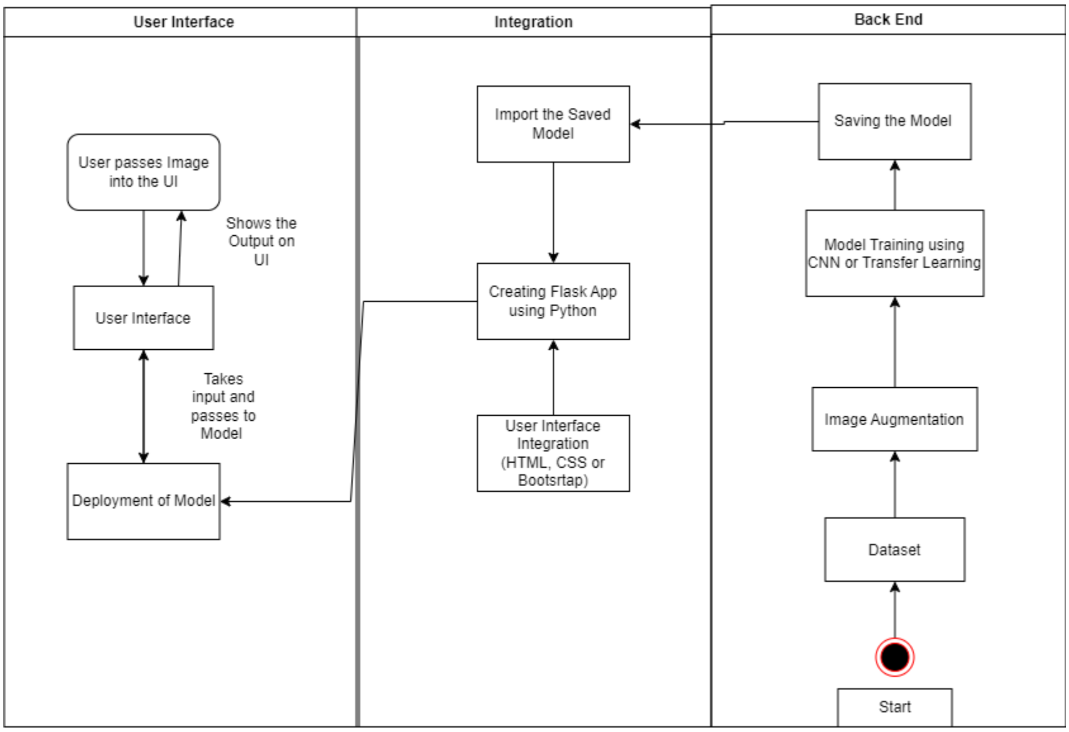


Table-1 : Components & Technologies:

Component	Description	Technology
1.Data Collection	Collecting lymphography images for training and testing.	• Medical imaging equipment (e.g., MRI, CT) • DICOM (Digital Imaging and Communications in Medicine) format for storage and retrieval of medical images • Data augmentation tools (e.g., OpenCV)
2.Data Preprocessing	Preparing data for model training and testing, including normalization, resizing, and data augmentation.	• Python (NumPy, Pandas, scikit-learn) • Image processing libraries (e.g., OpenCV)
3.Data Storage	Storing raw and preprocessed lymphography images for easy access and retrieval.	• Relational databases (e.g., PostgreSQL) • Distributed storage (e.g., Hadoop HDFS)
4.Machine Learning Model	Building and training a model to classify lymphography images.	• Convolutional Neural Networks (CNNs) • Deep learning frameworks (e.g., TensorFlow, PyTorch)
5.Model Evaluation	Evaluating model performance using validation data and metrics such as accuracy, precision, recall, and F1-score.	• Python (scikit-learn, TensorFlow, or PyTorch)
6.Model Deployment	Deploying the trained model to make predictions on new data.	• Web frameworks (e.g., Flask, Django) • Cloud services (e.g., AWS, Azure) for scalability
7.User Interface	Providing an interface for users to interact with the model, upload images, and view classification results.	• HTML, CSS, JavaScript for web applications • Front-end frameworks (e.g., React, Angular)
8.Security and Privacy	Implementing security measures to ensure data privacy and protect against unauthorized access.	• Authentication protocols (e.g., OAuth, JWT) • SSL/TLS for data encryption • Compliance with data protection regulations

9.Monitoring and Logging	Monitoring system health and logging activities, errors, and performance.	• Logging frameworks(e.g.Log4j, Logstash) • Monitoring tools
10.Scalability	Ensuring the system can handle increased load by distributing requests.	• Load balancers • Container orchestration (e.g., Kubernetes)

Table-2: Application Characteristics:

Characteristics	Description	Technology
1.Open-Source Frameworks	List the open-source frameworks used	Python's Flask
2.Security Implementations	List all the security / access controls implemented,use of firewalls etc.	e.g.SHA-256, Encryptions, IAM Controls, OWASP etc.
3.Scalable Architecture	Justify the scalability of architecture (3 – tier,Micro-services)	Technology used
4.Availability	Justify the availability of application (e.g.use of load balancers, distributed servers etc.)	Technology used
5.Performance	Design consideration for the performance of the application (number of requests per sec, use of Cache, use of CDN's) etc.	Technology used

6.2 Project Planning Phase: Sprint Planning and Estimation in Lymphography

Introduction:

In the dynamic landscape of healthcare and medical imaging, precision and efficiency are paramount. The field of lymphography, which involves imaging and diagnosing conditions related to the lymphatic system, is no exception. The effective utilization of lymphography requires a meticulous and organized approach to ensure that healthcare professionals can make accurate diagnoses, plan treatments, and enhance patient outcomes.

Purpose:

This document is dedicated to "Sprint Planning and Estimation in Lymphography," a pivotal component of managing projects and workflow in the realm of medical imaging, particularly lymphography. By applying the principles of sprint planning and estimation to lymphography, we can ensure that resources are optimized, timelines are adhered to, and patient care is enhanced.

Importance:

As the medical community continuously advances in its quest to understand and treat lymphatic system disorders, lymphography is a versatile tool that has a crucial role to play. By adopting sprint planning and estimation practices, we aim to streamline the process of diagnosing and treating lymphatic conditions, while maintaining a keen focus on patient care, scientific research, and medical innovation.

Sections Overview:

In the following sections, we will delve into the intricacies of sprint planning, the specifics of estimation techniques, and their application in the field of lymphography. We will provide insights into how agile principles can be adapted to enhance the delivery of healthcare services, with a particular focus on lymphatic system imaging and diagnostics.

Sprint Details in Lymphography

The "Sprint Details" section focuses on breaking down the intricate process of lymphography into digestible components, each represented by a sprint. These sprints encapsulate a set of objectives, tasks, and deliverables that collectively drive our journey towards more precise diagnostics and treatment planning in the realm of lymphatic system disorders.

Content of Sprint Details:

We will outline the specific details of each sprint, including its duration, scope, and key milestones.

Planned Deliveries:

In the same row as the sprint details, indicate the planned deliveries for that sprint. This may include features, user stories, or other work items that were expected to be completed during the sprint.

Progress Indicators in Lymphography Sprints

Measuring progress is an essential aspect of effective sprint planning and execution, especially in the dynamic field of lymphography. To provide stakeholders with a clear and concise understanding of our sprint performance, we employ progress indicators and completion percentages.

Purpose of Progress Indicators:

These metrics serve as the compass guiding our journey through each sprint, offering a quantitative perspective on our achievements and remaining tasks.

Importance of Completion Percentages:

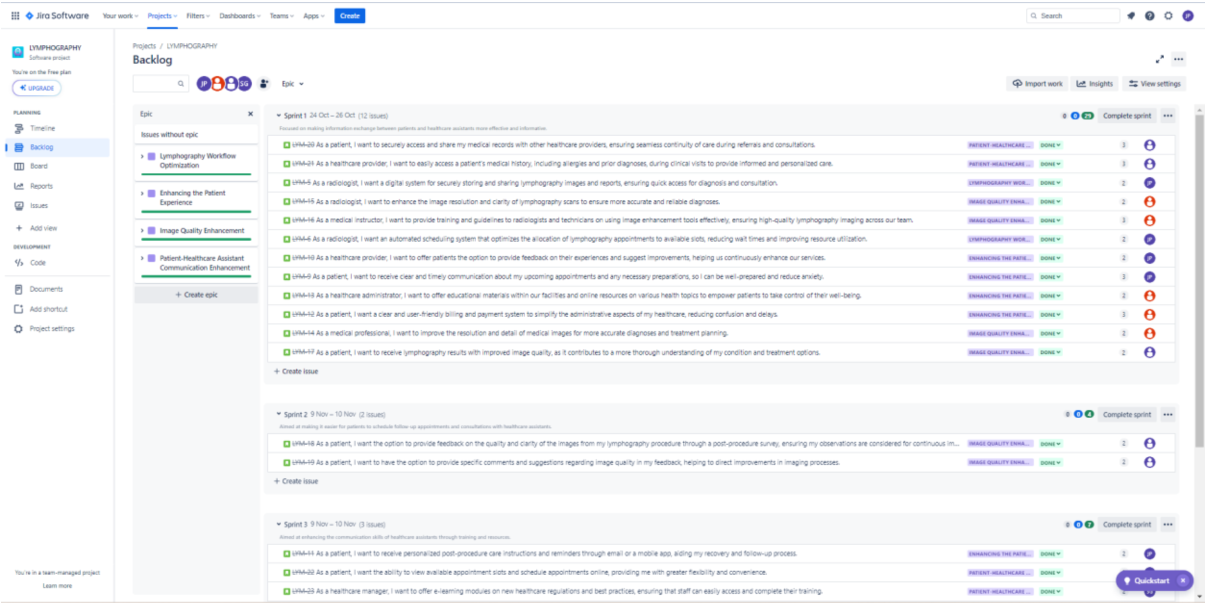
Progress indicators, often presented in the form of completion percentages, allow us to gauge the extent to which the planned work for a given sprint has been successfully executed. They provide an instant snapshot of our performance, enabling us to celebrate our accomplishments, identify areas for improvement, and make data-driven decisions for the future.

Inclusion in Sprint Details:

The progress indicators alongside the details of each sprint. By encapsulating our achievements in numerical form, they offer a tangible representation of our commitment to advancing lymphography.

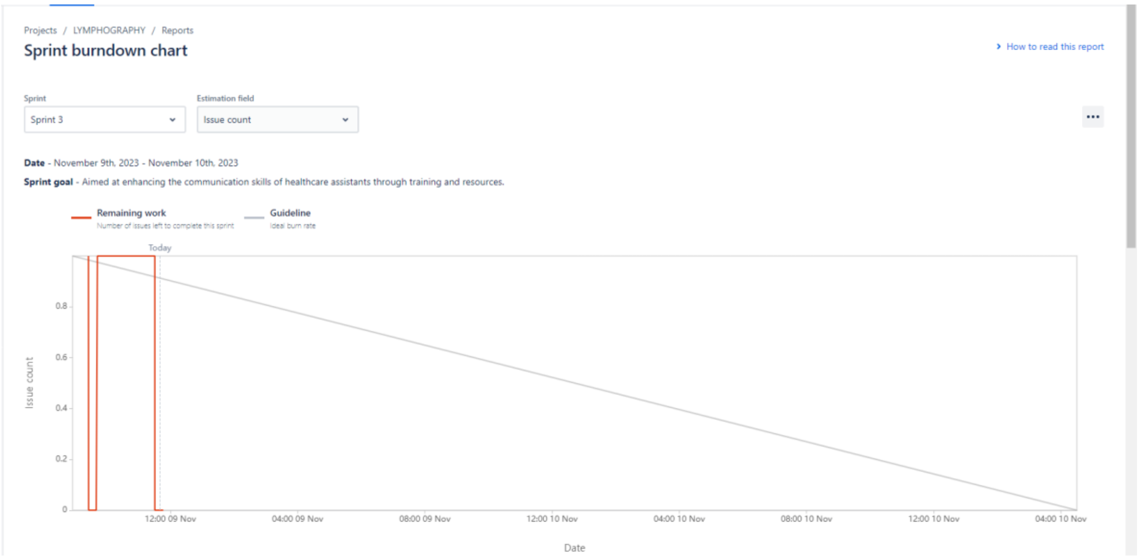
Whether it's the exploration of new diagnostic techniques, enhancements in image quality, or innovations in patient care, these indicators are a testament to our dedication to excellence.

By presenting the information in this structured manner, the document becomes more organized and easier to follow for stakeholders and readers interested in the project planning and execution for lymphography.

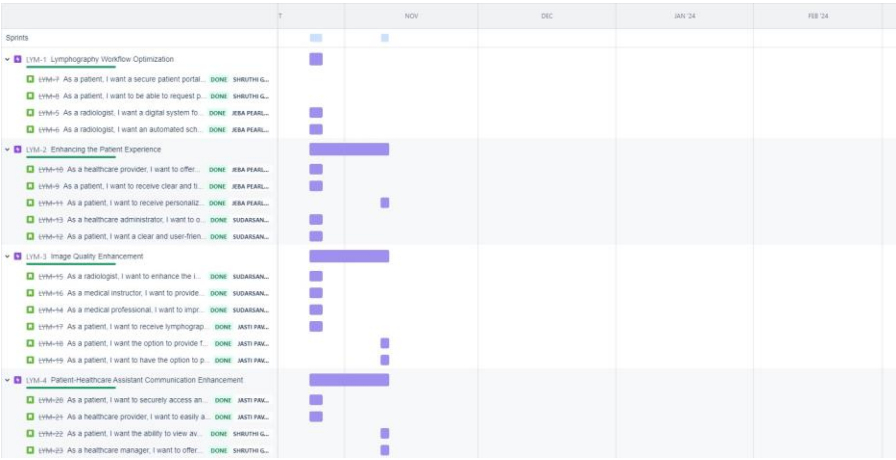

Jira Software
Your work
Projects
Filters
Dashboards
Teams
Apps
Create
Search
LYMPHOGRAPHY
Software project
Projects / LYMPHOGRAPHY
Backlog
Timeline
Backlog
Board
Reports
Issues
Add view
Code
Documents
Add shortcut
Project settings
Epic
Issues without epic
Lymphography Workflow Optimization
Enhancing the Patient Experience
Image Quality Enhancement
Patient-Healthcare Assistant Communication Enhancement
+ Create epic
Import work
Insights
View settings
Complete sprint
+ Create issue
Quickstart


Sprint
Sprint 1
Estimation field
Issue count
Date - October 24th, 2023 - October 26th, 2023
Sprint goal - Focused on making information exchange between patients and healthcare assistants more effective and informative.
Remaining work
Number of issues left to complete this sprint
Guideline
Ideal burn rate
Planned end
Issue count
12
9
6
3
0
Oct 25
Oct 27
Oct 29
Oct 31
Nov 01
Nov 03
Nov 05
Nov 07
Nov 09
Date

Sprint burndown chart
How to read this report
Sprint
Sprint 2
Estimation field
Issue count
Date - November 9th, 2023 - November 10th, 2023
Sprint goal - Aimed at making it easier for patients to schedule follow-up appointments and consultations with healthcare assistants.
Remaining work
Number of issues left to complete this sprint
Guideline
Ideal burn rate
Today
Issue count
2.5
2
1.5
1
0.5
0
12:00 09 Nov
04:00 09 Nov
08:00 09 Nov
12:00 10 Nov
04:00 10 Nov
08:00 10 Nov
12:00 10 Nov
04:00 10 Nov



Projects / LYMPHOGRAPHY / Reports
Sprint burndown chart
How to read this report
Sprint
Sprint 3
Estimation field
Issue count
Date - November 9th, 2023 - November 10th, 2023
Sprint goal - Aimed at enhancing the communication skills of healthcare assistants through training and resources.
Remaining work
Number of issues left to complete this sprint
Guideline
Ideal burn rate
Today
Issue count
0.8
0.6
0.4
0.2
0
12:00 09 Nov
04:00 09 Nov
08:00 09 Nov
12:00 10 Nov
04:00 10 Nov
08:00 10 Nov
12:00 10 Nov
04:00 10 Nov
Date



NOV
DEC
JAN '24
FEB '24
Sprints
LYM-1 Lymphography Workflow Optimization
LYM-7 As a patient, I want a secure patient portal... DONE SHRUTHI G...
LYM-8 As a patient, I want to be able to request p... DONE SHRUTHI G...
LYM-5 As a radiologist, I want a digital system fo... DONE JEBA PEARL...
LYM-6 As a radiologist, I want an automated sch... DONE JEBA PEARL...
LYM-2 Enhancing the Patient Experience
LYM-10 As a healthcare provider, I want to offer... DONE JEBA PEARL...
LYM-9 As a patient, I want to receive clear and ti... DONE JEBA PEARL...
LYM-11 As a patient, I want to receive personaliz... DONE JEBA PEARL...
LYM-13 As a healthcare administrator, I want to o... DONE SUDARSAN...
LYM-12 As a patient, I want a clear and user-frien... DONE SUDARSAN...
LYM-3 Image Quality Enhancement
LYM-15 As a radiologist, I want to enhance the i... DONE SUDARSAN...
LYM-16 As a medical instructor, I want to provide... DONE SUDARSAN...
LYM-14 As a medical professional, I want to impr... DONE SUDARSAN...
LYM-17 As a patient, I want to receive lymphograp... DONE JASTI PAV...
LYM-18 As a patient, I want the option to provide f... DONE JASTI PAV...
LYM-19 As a patient, I want to have the option to p... DONE JASTI PAV...
LYM-4 Patient-Healthcare Assistant Communication Enhancement
LYM-20 As a patient, I want to securely access an... DONE JASTI PAV...
LYM-21 As a healthcare provider, I want to easily a... DONE JASTI PAV...
LYM-22 As a patient, I want the ability to view av... DONE SHRUTHI G...
LYM-23 As a healthcare manager, I want to offer... DONE SHRUTHI G...

6.3 Sprint Delivery Schedule

Lymphography in Healthcare:

Lymphography is a diagnostic imaging technique used to visualize the lymphatic system, a vital but often understated component of the human body. The lymphatic system is responsible for maintaining fluid balance, filtering harmful substances, and supporting the immune response. It plays a crucial role in diagnosing and treating conditions related to the lymphatic system, such as lymphedema and lymphatic tumors.

Significance of Lymphography:

Lymphography has gained prominence as a valuable tool in modern healthcare. It enables healthcare professionals to observe the lymphatic system's structure, identify potential issues, and formulate accurate diagnoses. The information obtained through lymphography contributes to the development of effective treatment plans, ultimately leading to improved patient outcomes.

Agile Project Management in Healthcare:

The principles of agile project management, commonly associated with software development, have found applicability in healthcare and diagnostic imaging, particularly in projects like lymphography. Agile emphasizes iterative, adaptive, and collaborative approaches to project planning, execution, and review. By applying these principles, healthcare professionals can optimize resource allocation, adhere to timelines, and enhance patient care.

Sprint Planning and Estimation:

Sprint planning is a central component of agile project management, allowing teams to work in defined, time-bound phases called sprints. During each sprint, specific objectives are set, and tasks are executed collaboratively. Estimation techniques, like completion percentages, assist in tracking progress and making data-driven decisions.

Purpose of this Report:

This Sprint Delivery Schedule report is an endeavor to document the application of agile project management principles, specifically sprint planning and estimation, to the domain of lymphography. It seeks to illustrate the adaptability of these principles in enhancing project management, resource optimization, and ultimately, patient care within the context of lymphatic system imaging and diagnostics.

By combining the power of lymphography with agile project management techniques, we aim to provide a comprehensive view of how these methodologies can be used to streamline the diagnosis and treatment of lymphatic conditions, furthering scientific research and medical innovation.

Sprint	Functional Requirement(Epic)	User Story Number	User Story/Task	Story Points	Priority	Team Members
Sprint 1	Lymphography Workflow Optimization	USN-1	As a patient, I want a secure patient portal that allows me to access my lymphography images, reports, and post-procedure care instructions online.	2	medium	Jeba pearllin
Sprint 1		USN-2	As a patient, I want to be able to request prescription refills online through the patient portal for a streamlined medication management process.	3	medium	Jeba pearllin
Sprint 2		USN-3	As a radiologist, I want a digital system for securely storing and sharing lymphography images and reports, ensuring quick access for diagnosis and consultation.	2	low	Jeba pearllin
Sprint 2		USN-4	As a radiologist, I want an automated scheduling system that optimizes the allocation of lymphography appointments to available slots, reducing wait times and improving resource utilization.	2	low	Jeba pearllin
Sprint 1	Enhancing the Patient Experience	USN-5	As a healthcare provider, I want to	3	medium	Jeba pearllin

Project Tracker, Velocity and Burndown chart: (4 marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date(planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	28	4 days	19 oct 2022	22 oct 2022	28	22 oct 2022
Sprint-2	11	3 days	23 oct 2022	25 oct 2022	11	25 oct 2022
Sprint-3	8	3 days	26 oct 2022	28 oct 2022	8	28 oct 2022

AV = sprint duration/velocity

= 28+11+8/3

=15.67

7.Coding and Solutioning:

Importing the data from the data set:

```python
col_names = ['class', 'lymphatics', 'block of affere', 'bl. of lymph. c', 'bl. of lymph. s', 'by pass',
 'extravasates', 'regeneration of', 'early uptake in', 'lym.nodes dimin', 'lym.nodes enlar',
'changes in lym.', 'defect in node', 'changes in node', 'changes in stru', 'special forms',
'dislocation of', 'exclusion of no', 'no. of nodes in']
```

```python
df = pd.read_csv("lymphography.data", names=col_names)
df.head()
```

	class	lymphatics	block of affere	bl. of lymph. c	bl. of lymph. s	by pass	extravasates	regeneration of	early uptake in	lym.nodes dimin	lym.nodes enlar	changes in lym.	defect in node	changes in node	changes in stru	special forms	dislocation of	exclusion of no
0	3	4	2	1	1	1	1	1	2	1	2	2	2	4	8	1	1	2
1	2	3	2	1	1	2	2	1	2	1	3	3	2	3	4	2	2	2
2	3	3	2	2	2	2	2	2	2	1	4	3	3	4	8	3	2	2
3	3	3	1	1	1	1	2	1	2	1	3	3	4	4	4	3	1	2
4	2	3	1	1	1	1	1	1	1	1	2	2	4	3	5	1	2	2

Describing the data frame:

```python
df.describe()
```

	class	lymphatics	block of affere	bl. of lymph. c	bl. of lymph. s	by pass	extravasates	regeneration of	early uptake in	lym.nodes dimin	lym.nodes enlar	changes in lym.	defect in node	changes in node
count	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000	148.000000
mean	2.452703	2.743243	1.554054	1.175676	1.047297	1.243243	1.506757	1.067568	1.702703	1.060811	2.472973	2.398649	2.966216	2.804054
std	0.575396	0.817509	0.498757	0.381836	0.212995	0.430498	0.501652	0.251855	0.458621	0.313557	0.836627	0.568323	0.868305	0.761834
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
25%	2.000000	2.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000	2.000000	2.000000	2.000000
50%	2.000000	3.000000	2.000000	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000	1.000000	2.000000	2.000000	3.000000	3.000000
75%	3.000000	3.000000	2.000000	1.000000	1.000000	1.000000	2.000000	1.000000	2.000000	1.000000	3.000000	3.000000	4.000000	3.000000
max	4.000000	4.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000	3.000000	4.000000	3.000000	4.000000	4.000000

Checking information of data frame:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 148 entries, 0 to 147
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   class            148 non-null    int64
 1   lymphatics       148 non-null    int64
 2   block of affere  148 non-null    int64
 3   bl. of lymph. c  148 non-null    int64
 4   bl. of lymph. s  148 non-null    int64
 5   by pass          148 non-null    int64
 6   extravasates     148 non-null    int64
 7   regeneration of  148 non-null    int64
 8   early uptake in  148 non-null    int64
 9   lym.nodes dimin  148 non-null    int64
 10  lym.nodes enlar  148 non-null    int64
 11  changes in lym.  148 non-null    int64
 12  defect in node   148 non-null    int64
 13  changes in node  148 non-null    int64
 14  changes in stru  148 non-null    int64
 15  special forms    148 non-null    int64
 16  dislocation of   148 non-null    int64
 17  exclusion of no  148 non-null    int64
 18  no. of nodes in  148 non-null    int64
dtypes: int64(19)
memory usage: 22.1 KB
```

Checking for the null values:

```python
df.isnull().sum()
```

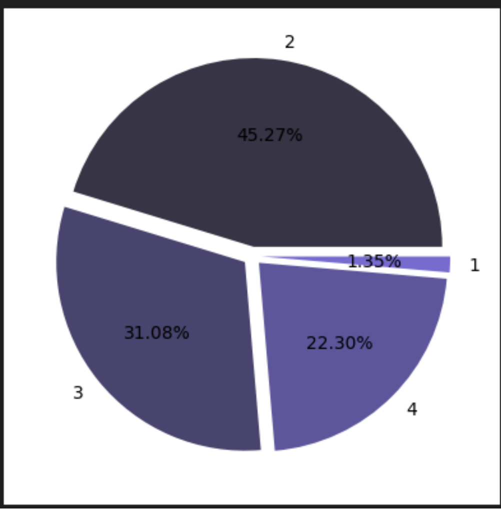
```
class                0
lymphatics           0
block of affere      0
bl. of lymph. c      0
bl. of lymph. s      0
by pass              0
extravasates         0
regeneration of      0
early uptake in      0
lym.nodes dimin      0
lym.nodes enlar      0
changes in lym.      0
defect in node       0
changes in node      0
changes in stru      0
special forms        0
dislocation of       0
exclusion of no      0
no. of nodes in      0
dtype: int64
```

Few of the visualizations(rest of the visualizations are in the ipynb file):

Uni-variant Analysis:

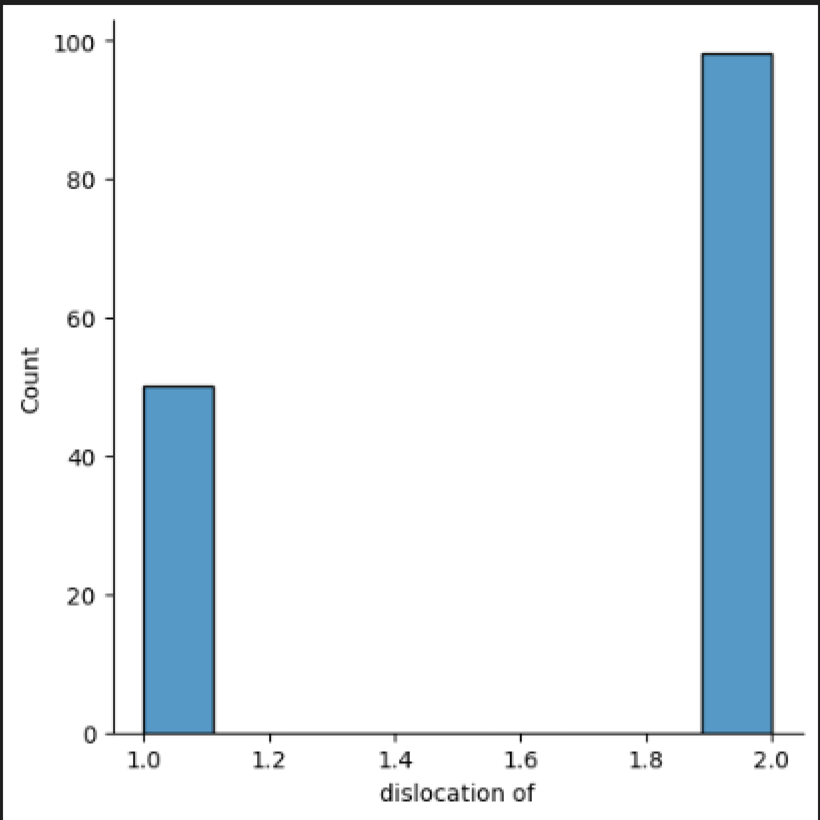
```python
plt.pie(df['lymphatics'].value_counts(),autopct='%0.2f%%',colors=["#363445", "#48446e", "#5e569b", "#776bcd"],
labels=df['lymphatics'].value_counts().keys(),explode=[0.05,0.05,0.05,0.05])
plt.show()
```

Python



```
plt.figure(figsize=(3,3))
sns.displot(df['dislocation of'])
```

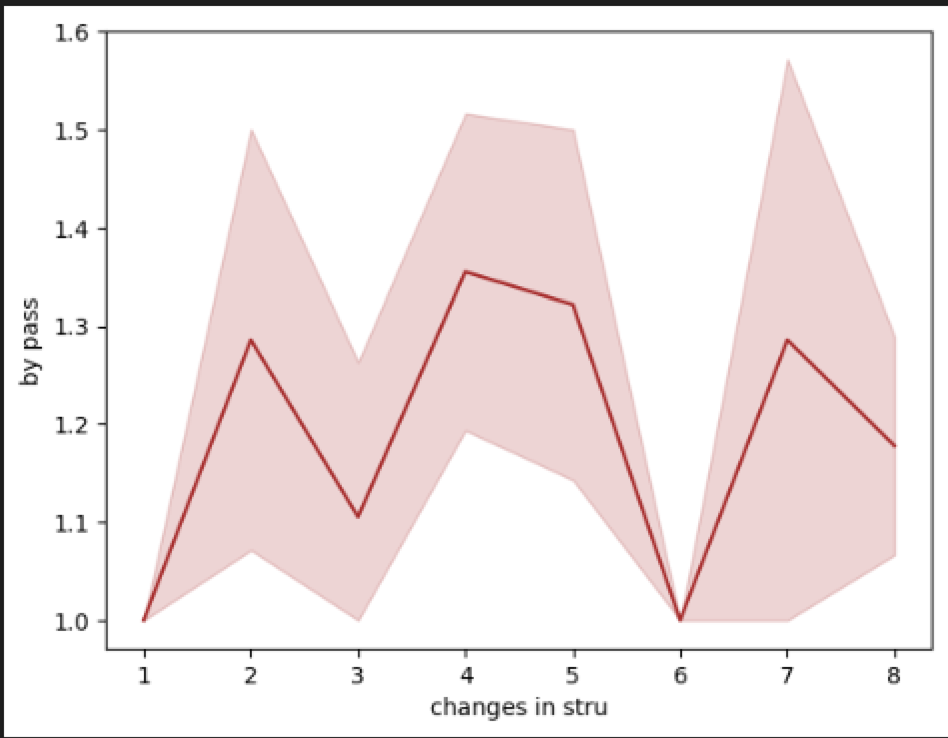
```
<seaborn.axisgrid.FacetGrid at 0x7b55dfe2a980>

<Figure size 300x300 with 0 Axes>
```



Bi-variant Analysis:

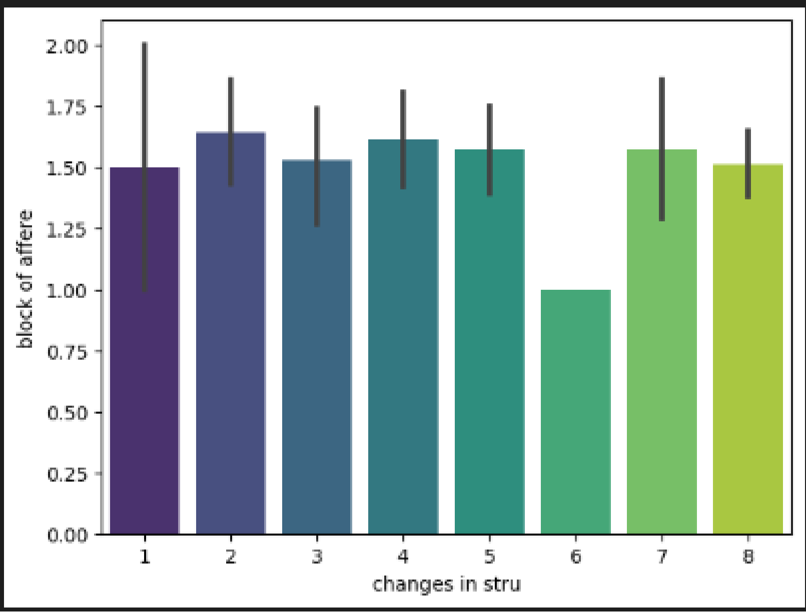
```
sns.lineplot(x='changes in stru',y='by pass',data=df,color='brown')
```

```
<Axes: xlabel='changes in stru', ylabel='by pass'>
```




```
sns.barplot(y='block of affere',x='changes in stru',data=df,palette='viridis')
```

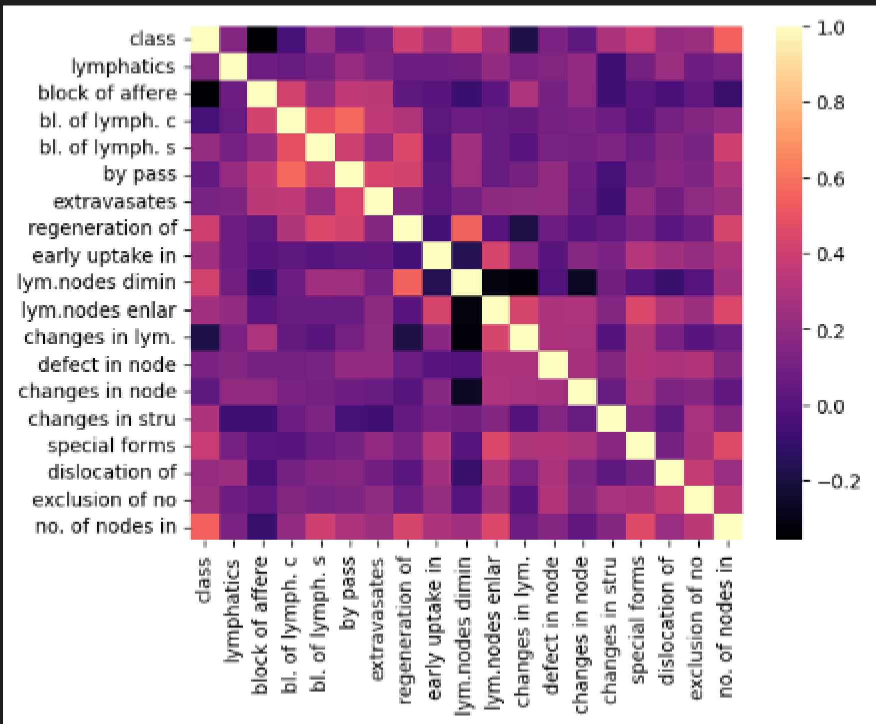
```
<Axes: xlabel='changes in stru', ylabel='block of affere'>
```



Multi-variant Analysis:

```
sns.heatmap(df.corr(),cmap ="magma")
```

```
<Axes: >
```



Splitting data into testing and training data:

Splitting testing and training data

```
from sklearn.model_selection import train_test_split
x = df.drop('class', axis=1)
y = df['class']
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=2)
```

[18]

```
y_train.value_counts()
```

[19]

```
2    64
3    50
4     3
1     1
Name: class, dtype: int64
```

```
y_test.value_counts()
```

[20]

```
2    17
3    11
4     1
1     1
Name: class, dtype: int64
```

Models used:

a)KNN model

```python
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report,accuracy_score
knn=KNeighborsClassifier()
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(x_train, y_train)
p_pre1p=knn.predict(x_test)
p_pre1t=knn.predict(x_train)
print("classification_report")
print(classification_report(y_true=y_test, y_pred=p_pre1p))
```

```
classification_report
              precision    recall  f1-score   support

           1       0.00      0.00      0.00         1
           2       0.82      0.82      0.82        17
           3       0.75      0.82      0.78        11
           4       1.00      1.00      1.00         1

    accuracy                           0.80        30
   macro avg       0.64      0.66      0.65        30
weighted avg       0.78      0.80      0.79        30
```

b)Decision Tree Classifier

```python
from sklearn.tree import DecisionTreeClassifier
model1 = DecisionTreeClassifier(max_depth=6,splitter='best',
criterion='entropy')
model1.fit(x_train,y_train)
```

Python

```
DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy', max_depth=6)
```

```python
p_pre1p=model1.predict(x_test)
p_pre1t=model1.predict(x_train)
```

Python

c)Random Forest Classifier

Random Forest Classifer

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report,accuracy_score
model2 =RandomForestClassifier(n_estimators=5,criterion='entropy')
model2.fit(x_train,y_train)
```

Python

```
RandomForestClassifier
RandomForestClassifier(criterion='entropy', n_estimators=5)
```

```
p_pre1p=model2.predict(x_test)
p_pre1t=model2.predict(x_train)
```

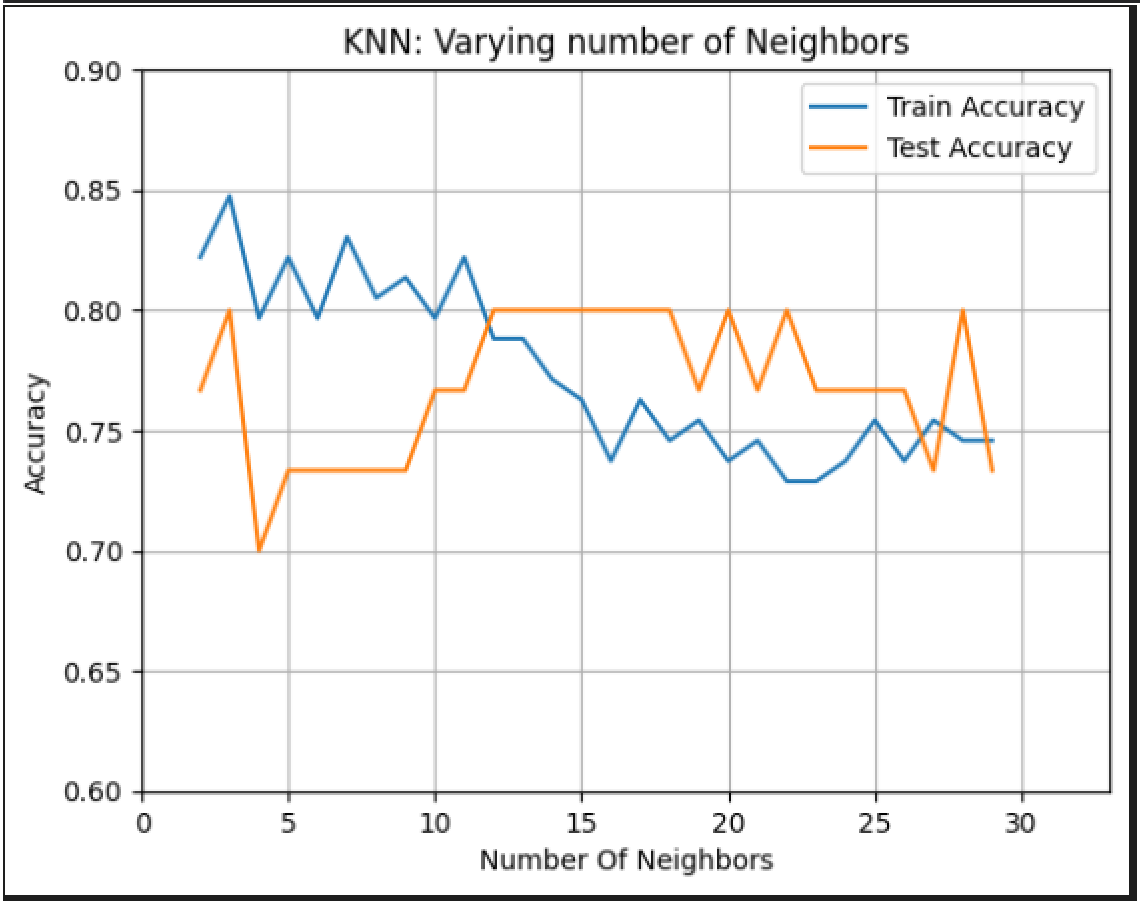
Python

8.Performance Testing:

a)KNN model

Hyper-tuning for KNN model

```
from sklearn.model_selection import GridSearchCV
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier()
train_score = {}
test_score = {}
n_neighbors = np.arange(2, 30, 1)
for neighbor in n_neighbors:
    knn = KNeighborsClassifier(n_neighbors=neighbor)
    knn.fit(x_train, y_train)
    train_score[neighbor]=knn.score(x_train, y_train)
    test_score[neighbor]=knn.score(x_test, y_test)
plt.plot(n_neighbors, train_score.values(), label="Train Accuracy")
plt.plot(n_neighbors, test_score.values(), label="Test Accuracy")
plt.xlabel("Number Of Neighbors")
plt.ylabel("Accuracy")
plt.title("KNN: Varying number of Neighbors")
plt.legend()
plt.xlim(0, 33)
plt.ylim(0.60, 0.90)
plt.grid()
plt.show()
```



Classification Report of test data

```python
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report,accuracy_score
knn=KNeighborsClassifier()
knn = KNeighborsClassifier(n_neighbors=3)
knn.fit(x_train, y_train)
p_pre1p=knn.predict(x_test)
p_pre1t=knn.predict(x_train)
print("classification_report")
print(classification_report(y_true=y_test, y_pred=p_pre1p))
```

```
classification_report
              precision    recall  f1-score   support

           1       0.00      0.00      0.00         1
           2       0.82      0.82      0.82        17
           3       0.75      0.82      0.78        11
           4       1.00      1.00      1.00         1

    accuracy                           0.80        30
   macro avg       0.64      0.66      0.65        30
weighted avg       0.78      0.80      0.79        30
```

Accuracy Score and Confusion matrix

```python
from sklearn.metrics import accuracy_score,
confusion_matrix
accuracy_score(y_test,p_pre1p)
```

```
0.8
```

+ Code + Markdown

```python
confusion_matrix(y_test,p_pre1p)
```

```
array([[ 0,  1,  0,  0],
       [ 0, 14,  3,  0],
       [ 0,  2,  9,  0],
       [ 0,  0,  0,  1]])
```

b)Decision Tree Classifier

Classification Report of test data

```python
print("classification_report for test data")
print(classification_report(y_true=y_test, y_pred=p_pre1p))
```

Python

```
classification_report for test data
              precision    recall  f1-score   support

           1       0.00      0.00      0.00         1
           2       0.84      0.94      0.89        17
           3       0.89      0.73      0.80        11
           4       0.50      1.00      0.67         1

    accuracy                           0.83        30
   macro avg       0.56      0.67      0.59        30
weighted avg       0.82      0.83      0.82        30
```

Classification Report of train data

```python
print("classification_report for train data")
print(classification_report(y_true=y_train, y_pred=p_pre1t)
```

Python

```
classification_report for train data
              precision    recall  f1-score   support

           1       1.00      1.00      1.00         1
           2       0.90      1.00      0.95        64
           3       1.00      0.86      0.92        50
           4       1.00      1.00      1.00         3

    accuracy                           0.94       118
   macro avg       0.98      0.96      0.97       118
weighted avg       0.95      0.94      0.94       118
```

Accuracy Score and Confusion matrix

```python
accuracy_score(y_test,p_pre1p)
```

```
0.8333333333333334
```

```python
confusion_matrix(y_test,p_pre1p)
```

```
array([[ 0,  1,  0,  0],
       [ 0, 16,  1,  0],
       [ 0,  2,  8,  1],
       [ 0,  0,  0,  1]])
```

c)Random Forest Classifier

Classification Report of test data

```python
print("classification_report of test data")
print(classification_report(y_true=y_test, y_pred=p_pre1p))
```

Python

```
classification_report of test data
              precision    recall  f1-score   support

           1       1.00      1.00      1.00         1
           2       0.79      0.88      0.83        17
           3       0.78      0.64      0.70        11
           4       1.00      1.00      1.00         1

    accuracy                           0.80        30
   macro avg       0.89      0.88      0.88        30
weighted avg       0.80      0.80      0.80        30
```

Classification Report of train data

```python
print("classification_report for train data")
print(classification_report(y_true=y_train, y_pred=p_pre1t))
```

Python

```
classification_report for train data
              precision    recall  f1-score   support

           1       1.00      1.00      1.00         1
           2       0.95      0.98      0.97        64
           3       0.98      0.94      0.96        50
           4       1.00      1.00      1.00         3

    accuracy                           0.97       118
   macro avg       0.98      0.98      0.98       118
weighted avg       0.97      0.97      0.97       118
```

Accuracy Score and Confusion matrix

```python
pd.crosstab(y_test,p_pre1p)
```

col_0	1	2	3	4
class				
1	1	0	0	0
2	0	15	2	0
3	0	4	7	0
4	0	0	0	1

```python
accuracy_score(y_test,p_pre1p)
```

```
0.8
```

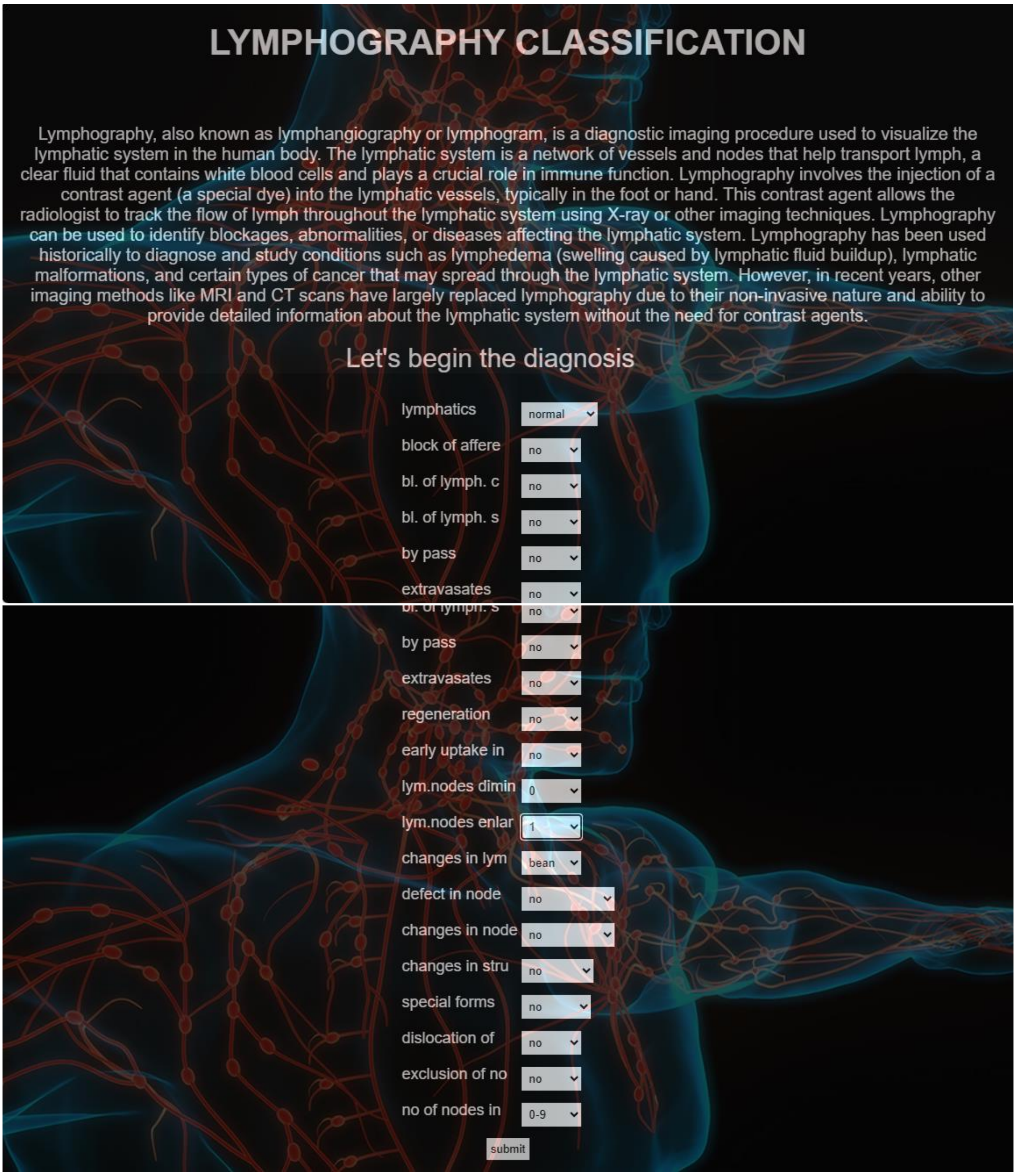
```python
confusion_matrix(y_test,p_pre1p)
```

```
array([[ 1,  0,  0,  0],
       [ 0, 15,  2,  0],
       [ 0,  4,  7,  0],
       [ 0,  0,  0,  1]])
```

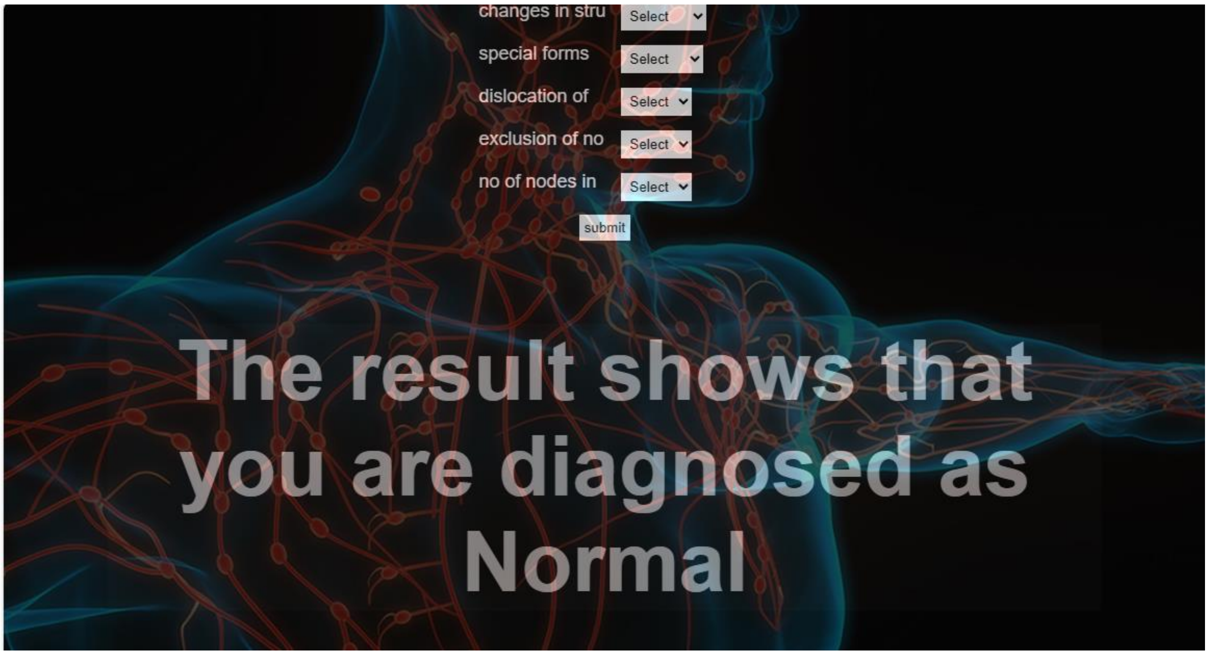
9.Results:

For class “Normal”:

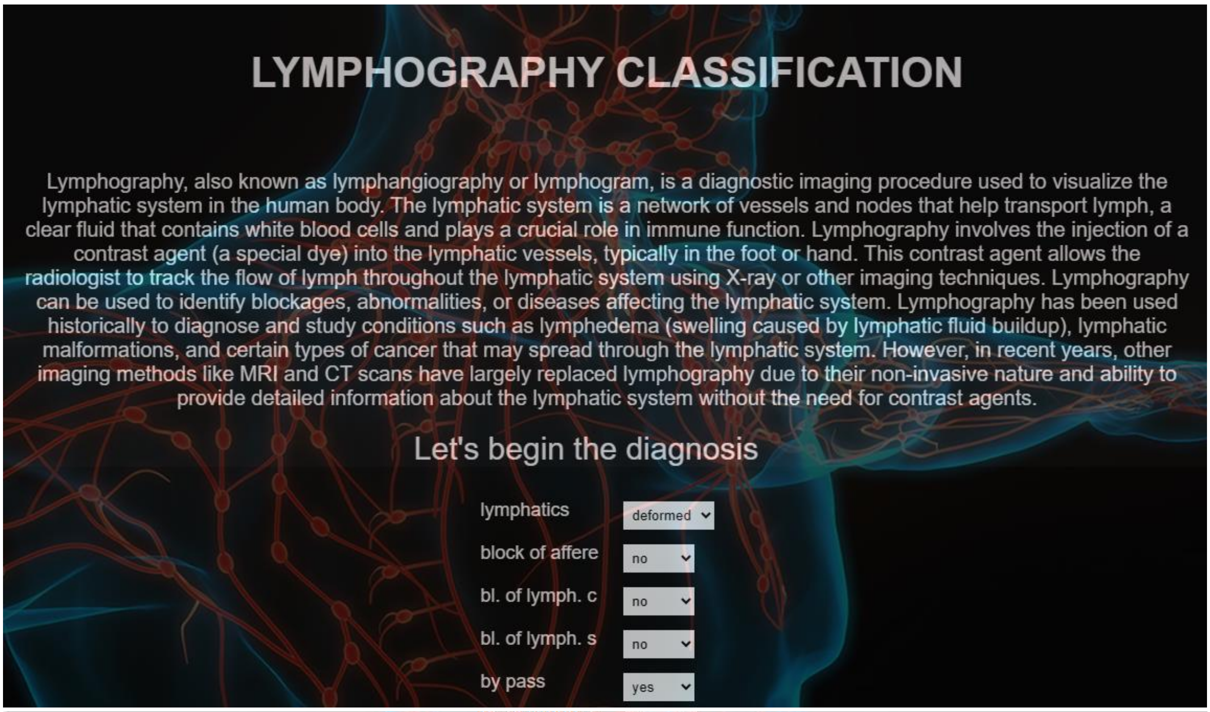
Before clicking on submit and after giving data:

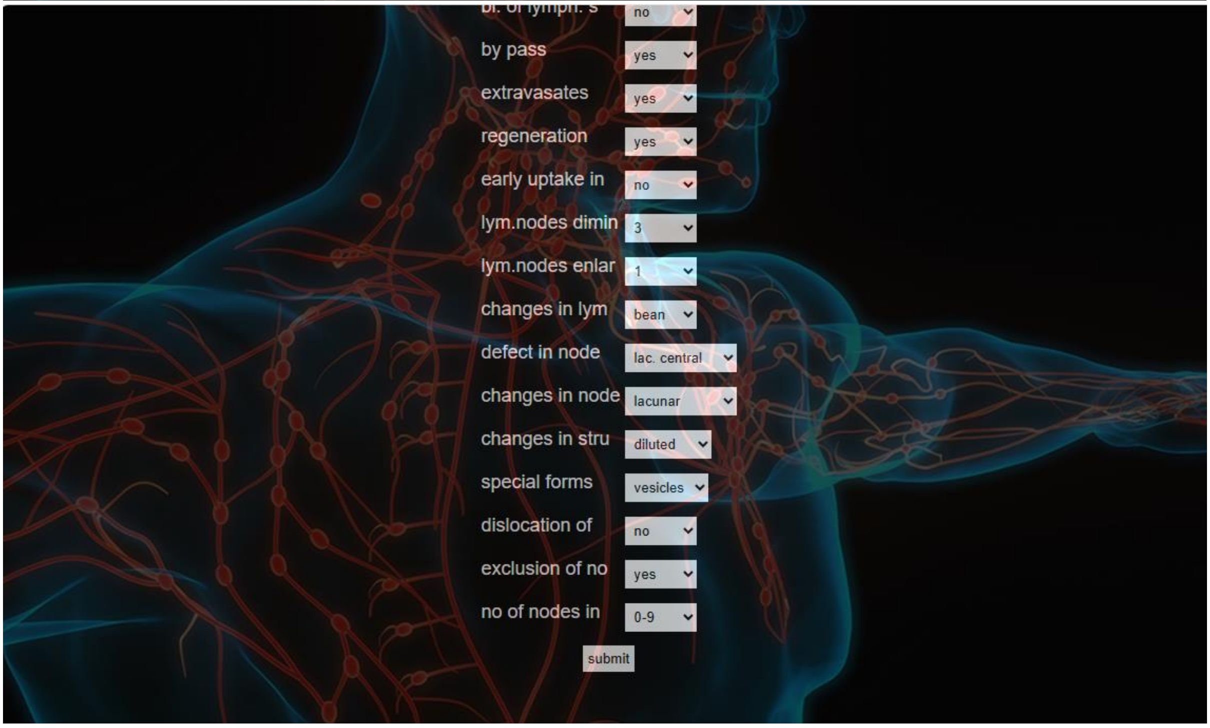


After clicking on submit:

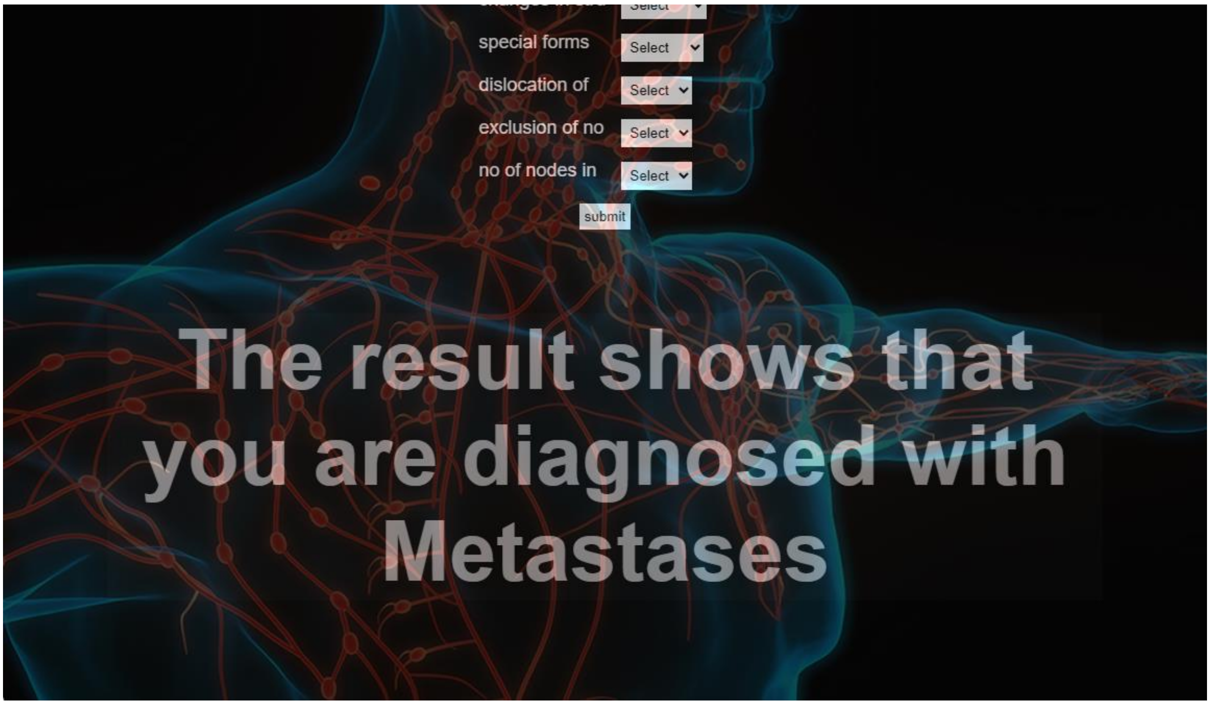


For class “Metastases”:
Before clicking on submit and after giving data:

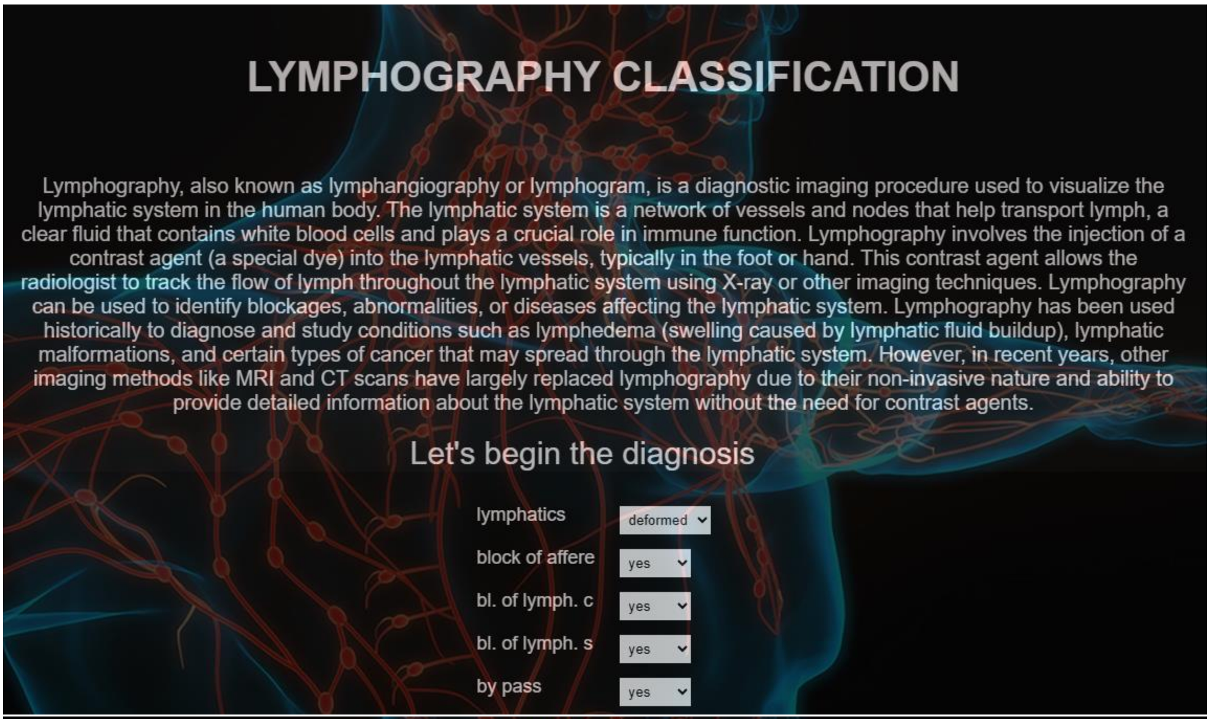


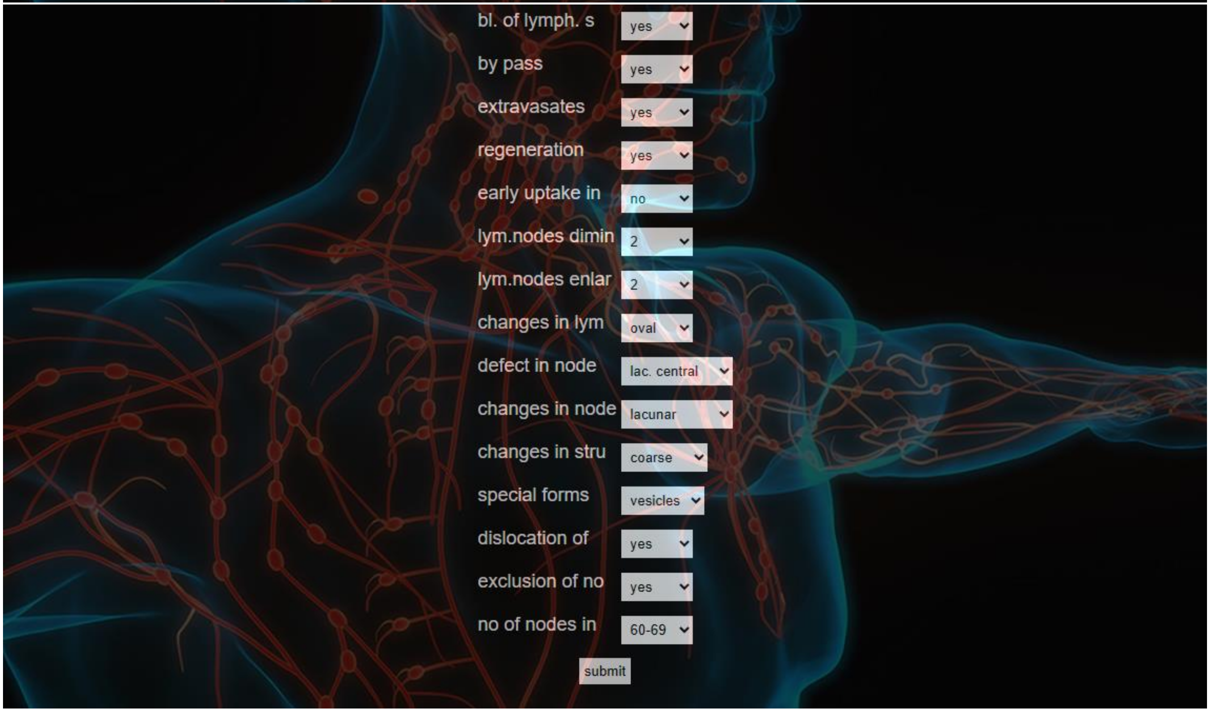


After clicking on submit:



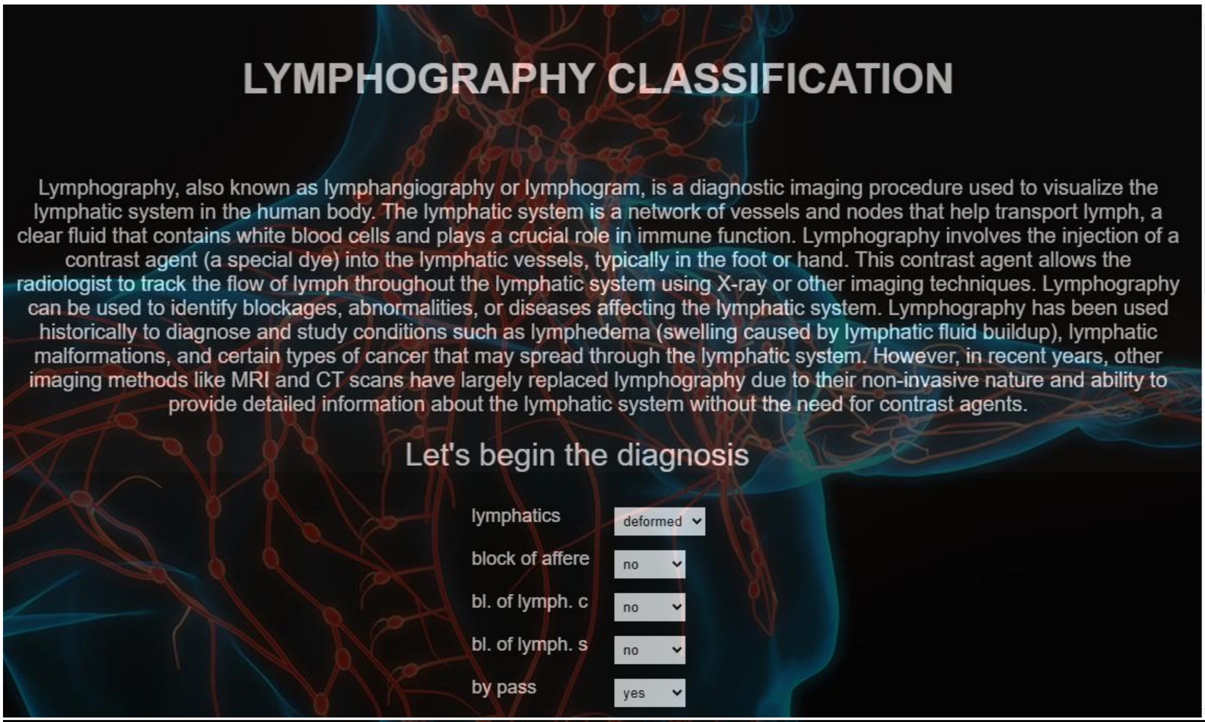
For class "Malign Lymph":
Before clicking on submit and after giving data:

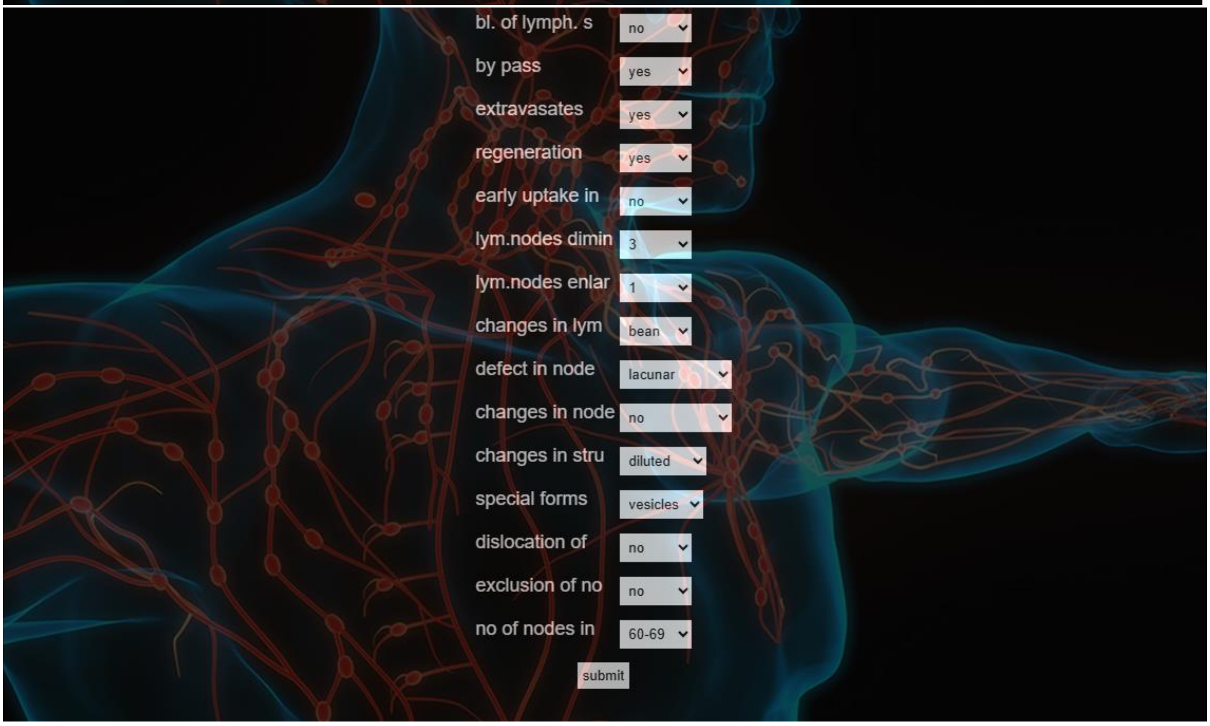




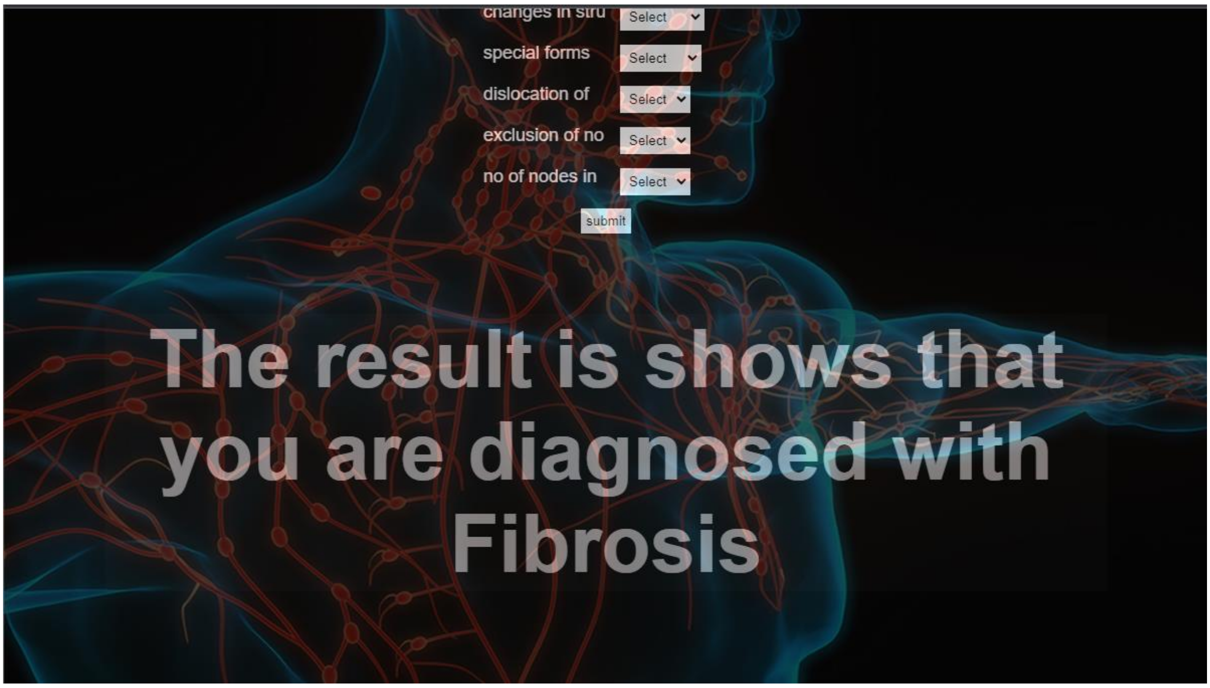
After clicking on submit:

For class “Fibrosis”:
Before clicking on submit and after giving data:





After clicking on submit:



10. ADVANTAGES & DISADVANTAGES

Advantages of Lymphography Classification Using Machine Learning:*

1. Enhanced Diagnostic Accuracy: Machine learning models can analyse lymphography images with a high degree of accuracy, reducing the likelihood of human error and increasing the reliability of diagnoses.

2. Efficient Diagnosis: Machine learning algorithms can quickly process and classify lymphography images, providing rapid results to medical professionals. This efficiency can be particularly crucial for time-sensitive medical decisions.

3. Scalability: Machine learning-based systems can handle a large volume of medical images and data, making them suitable for healthcare institutions with a high patient load.

4.Continuous Learning: Machine learning models can improve over time through continuous learning and adaptation to new data, leading to increased accuracy and reliability.

5. Improved Access to Healthcare: By automating parts of the diagnostic process, machine learning can enhance access to healthcare services, especially in regions with a shortage of specialized medical professionals.

6. Objective Decision Support: Machine learning models provide objective decision support, reducing subjectivity in medical image interpretation.

Disadvantages of Lymphography Classification Using Machine Learning:

1. Data Dependency: Machine learning models rely on large, high-quality datasets for training. In medical imaging, obtaining such datasets can be challenging, leading to potential biases and limitations in model performance.

2. Interpretability: Deep learning models are often considered black boxes, making it difficult for medical professionals to understand the rationale behind specific classifications. Model interpretability is a significant challenge.

3. Regulatory Compliance: Compliance with healthcare regulations, such as HIPAA and GDPR, is essential when dealing with patient data. Ensuring that machine learning systems adhere to these regulations can be complex and resource-intensive.

4. Data Privacy: Protecting patient privacy is critical. The storage and analysis of sensitive medical images require robust data privacy and security measures to prevent breaches and unauthorized access.

5. Complexity and Resource Intensiveness: Developing and maintaining machine learning systems can be resource-intensive and require expertise in data science and AI, which may not be readily available in all healthcare settings.

6. Initial Development Cost: Building and training machine learning models can be costly in terms of time and resources, which may pose a barrier to implementation in some healthcare facilities.

11. CONCLUSION

In this project, we successfully developed a Machine Learning model for classifying lymphography images. We have used Random Forest Classifier approach to analyse and classify lymph node data into distinct categories, ultimately aiding in the diagnosis process.

The results obtained were promising, demonstrating the effectiveness of our model in accurately distinguishing lymph nodes. The performance metrics, including accuracy, showcased the robustness of our approach.

Furthermore, this project exemplifies the potential of Machine Learning in the medical field. By automating the classification process, we can enhance diagnostic efficiency and provide valuable decision support to healthcare professionals. This is particularly crucial in situations where time plays a critical role in patient outcomes.

In conclusion, this project represents a meaningful step towards leveraging Machine Learning for lymphography classification. The potential impact on healthcare, particularly in the realm of diagnostic radiology, is substantial. We hope that this work serves as a foundation for further research and development in this important area.

12. FUTURE SCOPE

We are expecting the following to be done in the future:

Fine-tuning and Transfer Learning:

Exploring techniques like fine-tuning pre-trained models or using transfer learning can potentially boost the model's performance. Leveraging knowledge from related tasks or domains could lead to even more accurate classifications.

Real-time Application and Deployment:

Adapting the model for real-time applications within clinical settings can significantly impact patient care. Streamlining the inference process to operate in real-time scenarios would be a crucial next step.

Continual Learning and Adaptation:

Implementing strategies for the model to adapt and learn from new data in an incremental manner will ensure its continued relevance and accuracy over time.

Clinical Validation and Regulatory Compliance:

Collaborating with medical professionals for extensive validation and compliance with regulatory standards is essential before deploying the model in clinical practice. This includes thorough testing on diverse patient populations.

Explain ability and Interpretability:

Developing methods to provide interpretable explanations for the model's decisions will be critical for gaining trust from healthcare practitioners and ensuring transparency in the diagnostic process.

13. APPENDIX:

DEMO VIDEO LINK:
https://drive.google.com/file/d/1slKZIGo3g6vLnZS4aLkQ4O3V7r4R9v7L/view?usp=drive_link

GIT REPO LINK:

https://github.com/smartinternz02/SI-GuidedProject-601058-1698591106