People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Mohamed Khider University - Biskra
Faculty of Exact Science
Computer Science Department

Revolutionizing Lung Cancer Diagnosis with LungPathAI: An AI-Powered Digital Pathology Platform

Presented by:

- GUESBAYA Islem
- CHERIFI Kacem
- OUAMANE Takieddine

Supervised by:

- BELAALA Abir

Academic year : 2024 – 2025

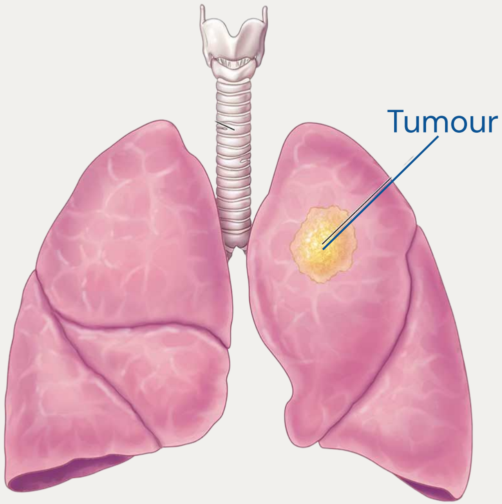
Table of contents

01

Introduction

Introduction

Lung cancer is a major health concern in Algeria.
In 2022, it ranked among the most diagnosed cancers, with **5,040 new cases**.
Among men, it had the **highest cancer-related death rate**, with an incidence of **19.5** and mortality of **17.9 per 100,000** (GLOBOCAN, 2022).



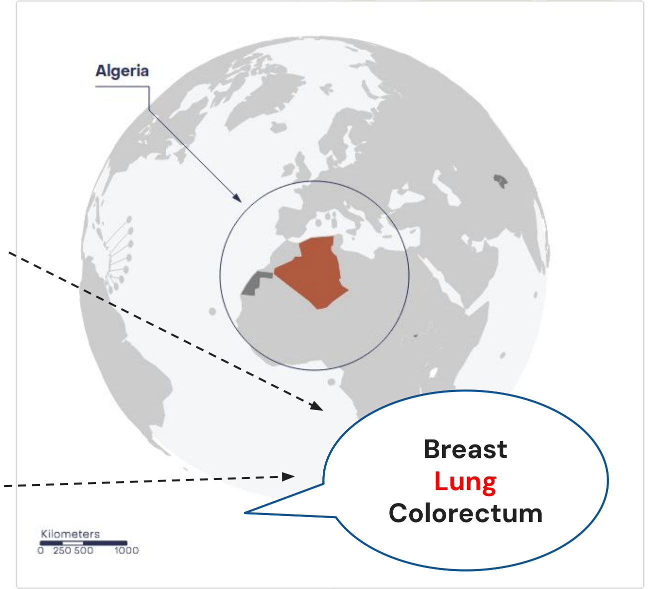
Cancer Statistics in Algeria (2022)

World Health Organization

Number of new cases	Number of deaths	Number of prevalent cases (5-year)
64 713	35 778	177 718

Statistics at a glance, 2022

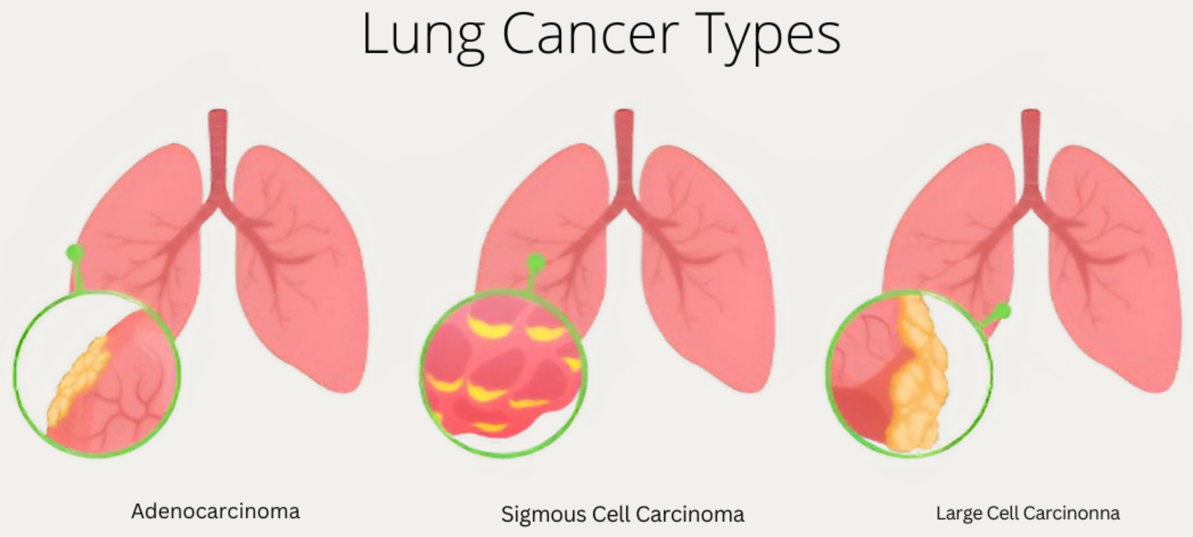
	Males	Females	Both sexes
Population	22 912 687	22 437 454	45 350 141
Incidence*			
Number of new cancer cases	29 387	35 326	64 713
Age-standardized incidence rate	130.6	152.2	141.2
Risk of developing cancer before the age of 75 years (cum. risk %)	14.1	15.1	14.6
Top 3 leading cancers (ranked by cases)**	Lung Colorectum Prostate	Breast Colorectum Thyroid	Breast Colorectum Lung
Mortality*			
Number of cancer deaths	18 809	16 969	35 778
Age-standardized mortality rate	82.7	73.1	77.7
Risk of dying from cancer before the age of 75 years (cum. risk %)	8.7	7.8	8.2
Top 3 leading cancers (ranked by deaths)**	Lung Colorectum Bladder	Breast Colorectum Cervix uteri	Breast Lung Colorectum
Prevalence*			
5-year prevalent cases	72 408	105 310	177 718



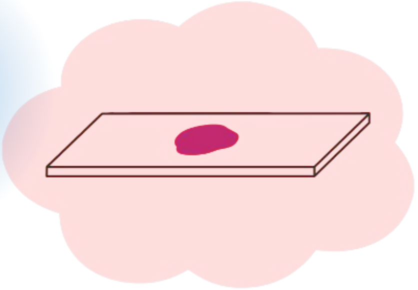
Lung
Colorectum
Bladder

Breast
Lung
Colorectum

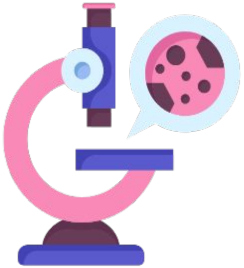
Introduction | Lung Cancer Types



Introduction | TRADITIONAL SLIDE DIAGNOSIS



slide preparation



slides scanning

Whole slide image

Manual Whole slide Image Analysis

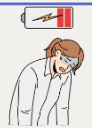
Limitations of Traditional Lung Cancer Diagnosis

Time-intensive Process

Manual examination of tissue samples creates significant diagnostic delays.

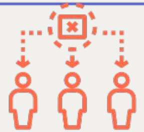
Diagnostic Variability

Substantial inter-observer differences among pathologists affect consistency.



Volume Challenges

Rising cancer incidence is overwhelming existing diagnostic capacity.



Limited Standardization

Inconsistent sample preparation and interpretation practices across institutions

02

Problem Statement

AI Solutions in Literature: Limitations

Limitations

01
Limited Diagnostic Scope

Focus only on cancer subtypes and fail to detect normal tissue

02
Explainability Gap

Operate as "black boxes," limiting clinical trust and adoption

03
Dataset Biases

Non-diverse training data leads to poor performance across varied demographic groups

04
Workflow Integration

Integration failing with existing clinical systems

Proposed Solution

01

Interpretable Results
Provides transparent reasoning

02

Clinical Integration
Integration with current pathology workflows

03

Efficiency Improvement
reduce diagnostic time and improve accuracy

04

Decision Support
Augment pathologist expertise

Solutions

03

Related Work

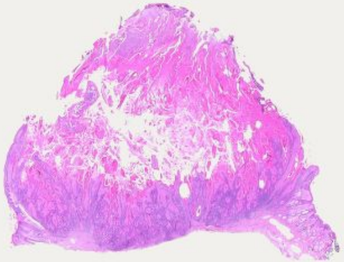
Overview of Leading MIL Methods

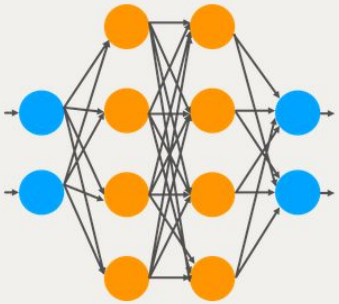
Approach	Year	Learning Paradigm	Training/Inference Speed	Clinical Interpretability Quality
CLAM	2021	Weakly Supervised + Clustering	Moderate/Fast	Good
TransMIL	2021	Correlated instance learning	Slow/Moderate	Normal
DSMIL	2021	Self-supervised + supervised	Fast/Fast	Normal
CAMIL	2025	Channel-aware feature learning	Moderate/Fast	Normal

Our work

1. Adopt CLAM architecture to classify lung cancer subtypes.
2. Train and Evaluate the model using the public CPTAC dataset.
3. Generate heatmaps highlighting regions of interest (ROIs) using attention scores.
4. Integrate the model into a web application.
5. Validate heatmaps quality by comparing model outputs with pathologist-provided annotations.

Why Not Traditional CNN ?





WSI

Convolutional Neural Network

04

Methodology

Operational steps

1- Data Acquisition

Whole Slide Images (WSIs) collected from public dataset (TCGA)

2- Preprocessing

- Tissue Segmentation
- Patching
- Feature Extraction

3- Deep Learning Model

Train a model that classifies each slide into LUAD, LSCC or Normal with high accuracy

4- Explainability (XAI)

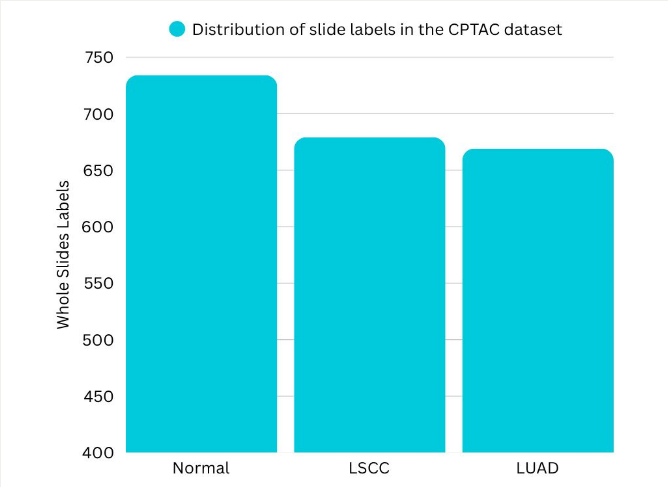
- Attention Scores
- Generate Heatmap over WSI

5- Model Validation

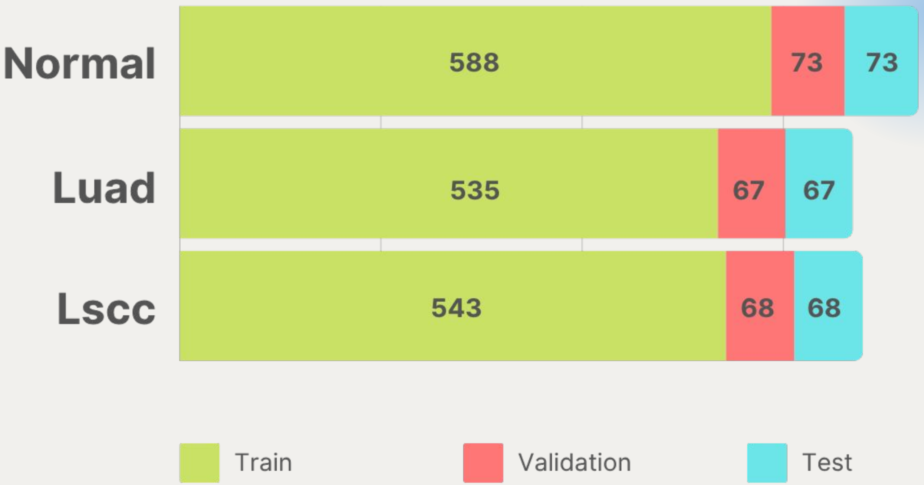
Validate the model with unseen data

1- Data Acquisition

Data-set Description



Data-set Split



2- Preprocessing

1- Segmentation & Background removing

- Isolate tissue regions from WSI.
- Remove background.

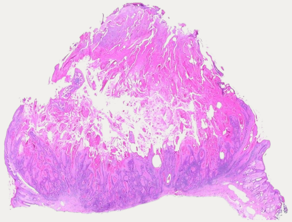
2- Patching

- Divide tissue regions.
- Ensure consistent magnification across all patches.

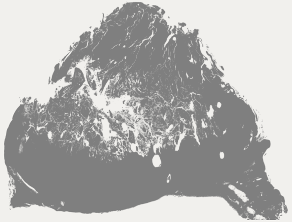
3- Feature Extraction

- Convert image tiles into feature vectors.
- Using a pre-trained CNN.

Step 1 : Removing background and Patching

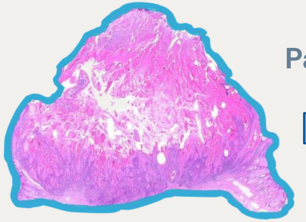


Whole slide image : a high-resolution digital pathology image (in svs format).



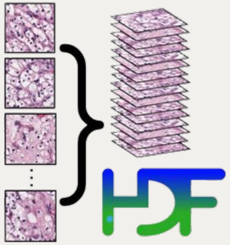
Grayscale image

Segmentation process
(removing background)



segmented tissue

Patching process



Output : 256*256 pixels extracted patches saved in **HDF5** format

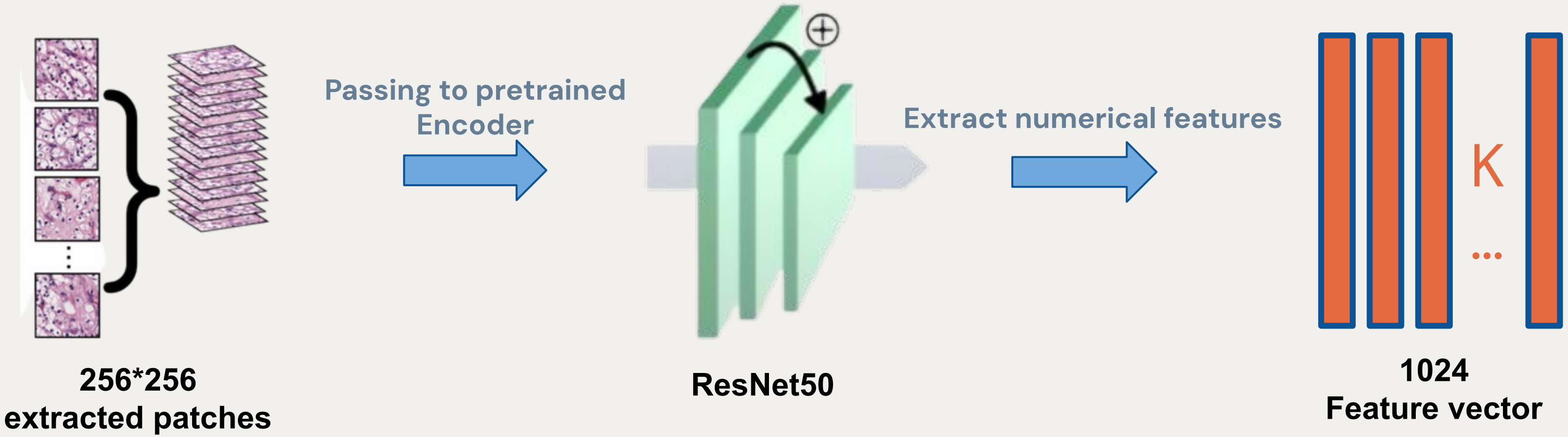
Segmentation Params

Parameter	Value	Description
Segmentation Level	-1	Resolution level used for segmentation -1 indicates the lowest available resolution level in the WSI
Saturation Threshold	8	determine if a pixel belongs to tissue based on its color saturation
Close	4	Size of the morphological closing operation to fill small holes and gaps in tissue regions
Area Threshold	100	Minimum area threshold for keeping a tissue region. Regions smaller than this are removed
Area holes	16	Minimum hole area to be considered for removal from tissue mask
Max_n_holes	8	Maximum number of holes allowed per tissue region

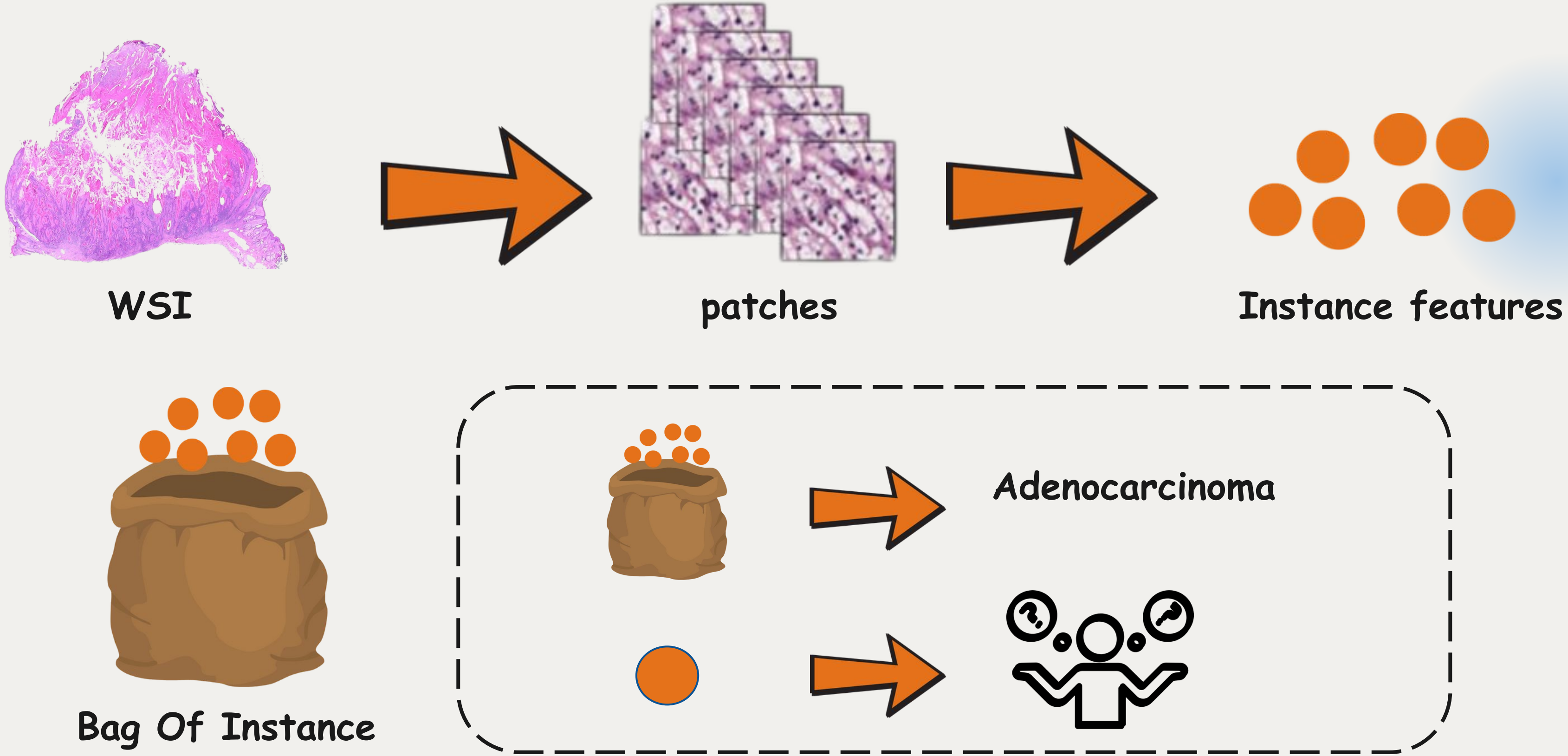
Patching Params

Parameter	Value	Description
Patch Level	0	The resolution level of the WSI from which patches are extracted (0 is is for the best details)
Patch Size	256	Size of each patch
Step Size	256	This control how far to move when extracting the next patch it determines the overlap between adjacent patches

Step 2 : Feature Extractions



What is Next ?



WSI
patches
Instance features
Bag Of Instance
Adenocarcinoma

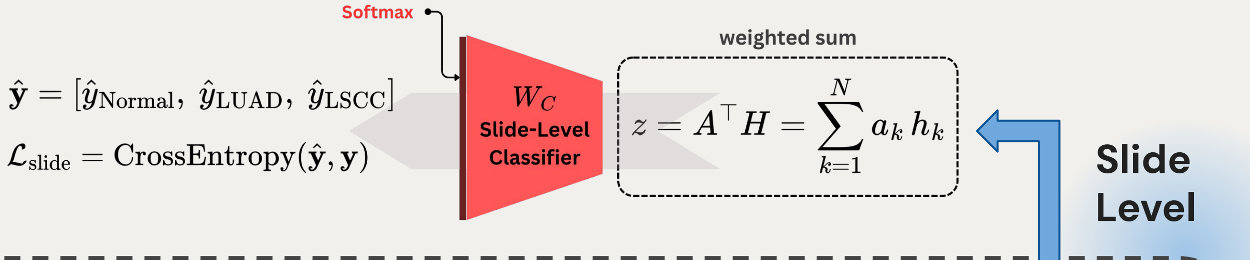
3- Deep Learning Model

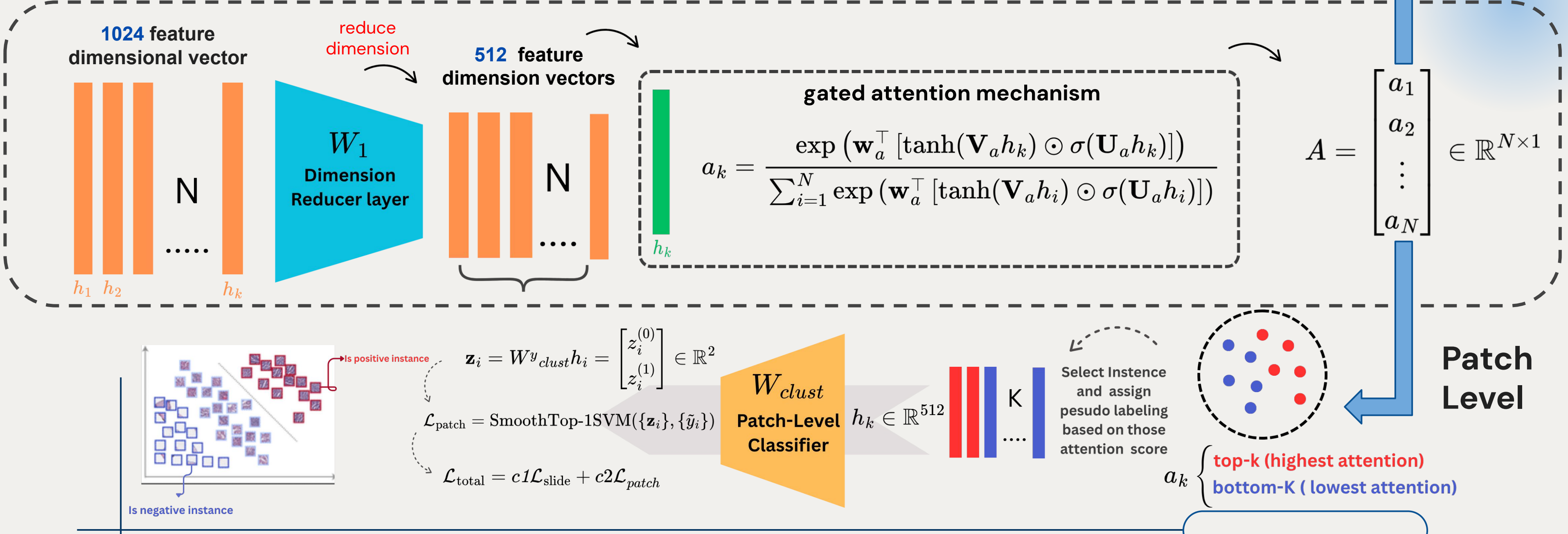
Single-Branch VS **Multi-Branch**

Components	Single Branch	Multi Branch
Attention Branches	Single Attention Pathway (one attention score per patch)	Multiple Attention Pathways (class-specific attention score per patch)
Slide Representation	A single aggregated slide-level representation	3 class-specific slide-level representations (Normal , Luad , Lscc)
Slide-Level Classifier	A single classifier acts on the aggregated slide-level representation	3 parallel classifiers act on 3 class-specific representations
Patch-Level Classifier (Clustering)	A single patch-level objective based on general attention.	N parallel patch-level objectives based on class-specific attention.

3- Deep Learning Model

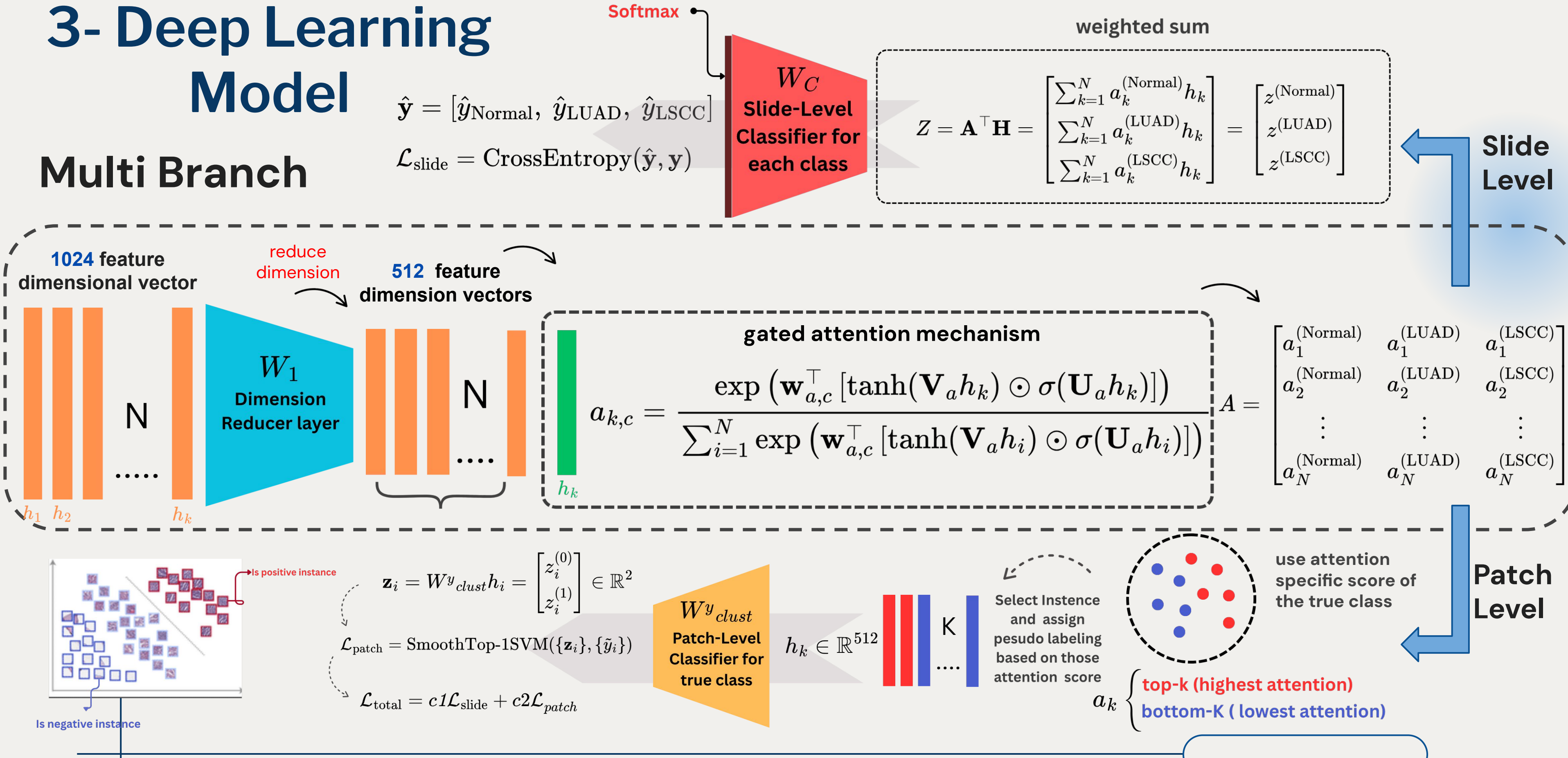
Single Branch





3- Deep Learning Model

Multi Branch



Training HyperParameters

Parameter	Value
Maximum epochs	200
Learning rate	0.0002
Optimizer	Adam
Bag loss	Cross-entropy (ce)
Instance loss	Smooth Top-1 SVM (svm)
Bag weight	0.7
Dropout rate	0.25
Early stopping	True

04

Experimental Results

Results | Evaluation Metrics:

Single Branch

Accuracy = 89%

	auc	accuracy	precision	recall	f1_score	specificity
Normal	0.9962455606	0.971153846	0.9466666667	0.9726027397	0.9594594595	0.9703703704
Luad	0.9611516884	0.913461538	0.9152542373	0.8059701493	0.8571428571	0.9645390071
Lscc	0.9712184874	0.913461538	0.8378378378	0.9117647059	0.8732394366	0.9142857143

Multi Branch

Accuracy = 88%

	auc	accuracy	precision	recall	f1_score	specificity
Normal	0.998782344	0.971153846	0.971830985	0.9452054795	0.9583333333	0.9851851852
Luad	0.96305705	0.894230769	0.816901408	0.8656716418	0.8405797101	0.9078014184
Lscc	0.970378151	0.903846153	0.86363636	0.8382352941	0.8507462687	0.9357142857

Results |

Single Branch Model

Multi Branch Model

Confusion Matrix

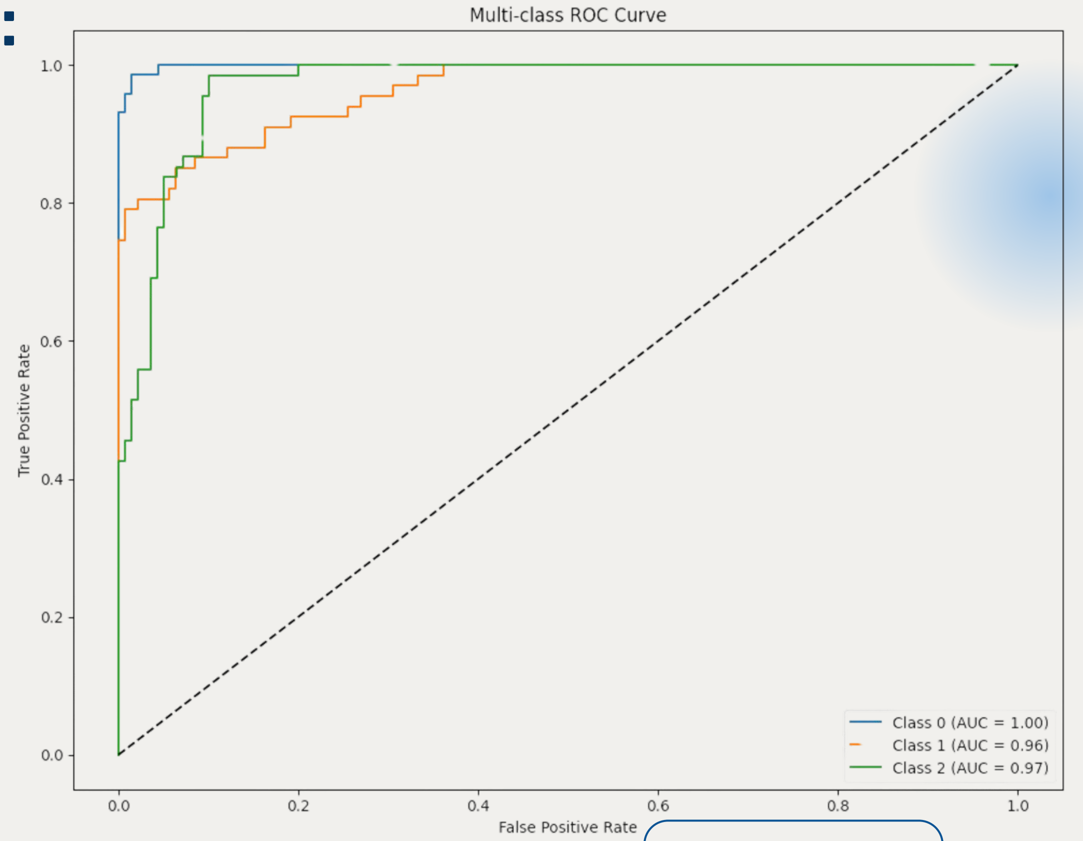
	Normal	Luad	Lscc
Normal	69	3	1
Luad	1	58	8
Lscc	1	10	57

Results | ROC Curve:

Class 0 (Normal): AUC = 1.00

Class 1 (LUAD): AUC = 0.96

Class 2 (LSCC): AUC = 0.97



Pathologist business card

Pathologist validation

Pathologist Validation Report

AI-Generated HeatMap Analysis for Lung Pathology (LungPathAI)

Validating Pathologist: Dr. Khelifi.N Ep karasad
Institution/Affiliation: Ex assistante at EPH biskra / CHU at beni messous
Board Certifications: Anatomic Pathology
Date of Validation: 20/05/2025
AI System Under Review: LungPathAI Heat Map
Developers: Ouamane Takieddine, Cherifi Kacem, Guesbaya Islem
Supervisor: ABIR belaala

Cases Reviewed: 10 lung pathology specimens

- Lung Adenocarcinoma (LUAD)
- Lung Squamous Cell Carcinoma (LSCC)
- Normal Lung Tissue

Validation Method: Independent pathologist review comparing AI-generated heat map outputs against standard histopathological assessment.

Based on my review as a board-certified pathologist with expertise in lung pathology, I hereby validate that:
The LungPathAI heat map generation system produces clinically relevant outputs that accurately identify pathologically significant areas in lung tissue specimens.
The heat maps demonstrate sufficient accuracy and clinical utility to serve as an effective diagnostic aid when used under appropriate pathologist supervision.

Recommendation: APPROVED for clinical use as a supplementary diagnostic tool in lung pathology practice.

Limitations: This validation is based on the specific case cohort reviewed. The AI system should be used as an adjunct to, not replacement for, standard pathological assessment. Continued monitoring and periodic revalidation are recommended.

EHP ERRAZI -2- LABORATOIRE D'ANATOMIE PATHOLOGIE

Professional Signature:

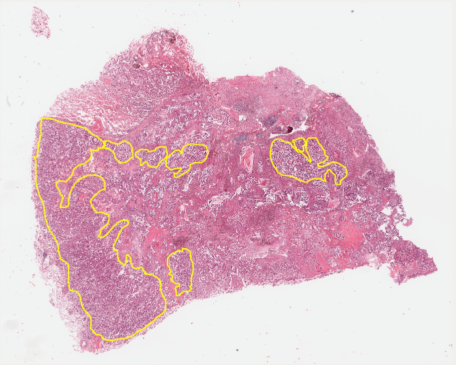
Dr. KHELIFI N Ep KARASAD
Date: 15/06/2025

Dr KHELIFI.N
MEDECIN SPECIALISTE EN ANATOMIE PATHOLOGIQUE

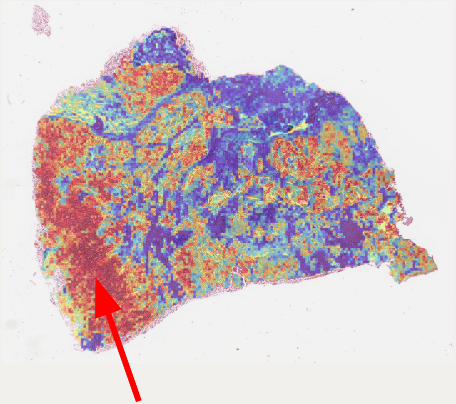
1/1

5- Explainability (XAI)

Annotated WSI by expert

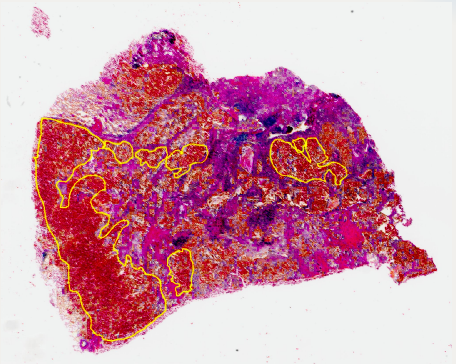


Generated Heatmap



Region
of Interest
(High attention)

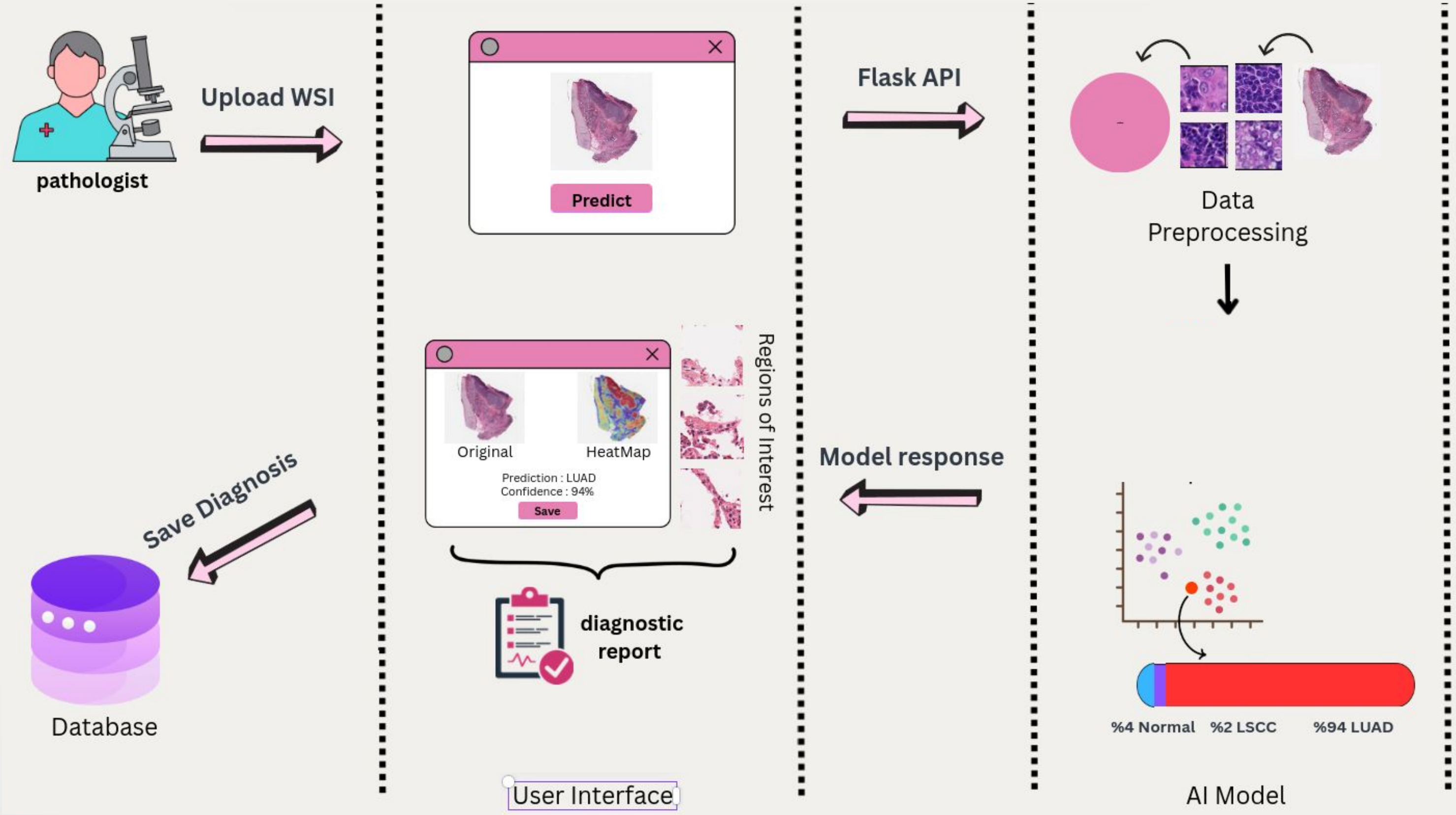
Overlay



06

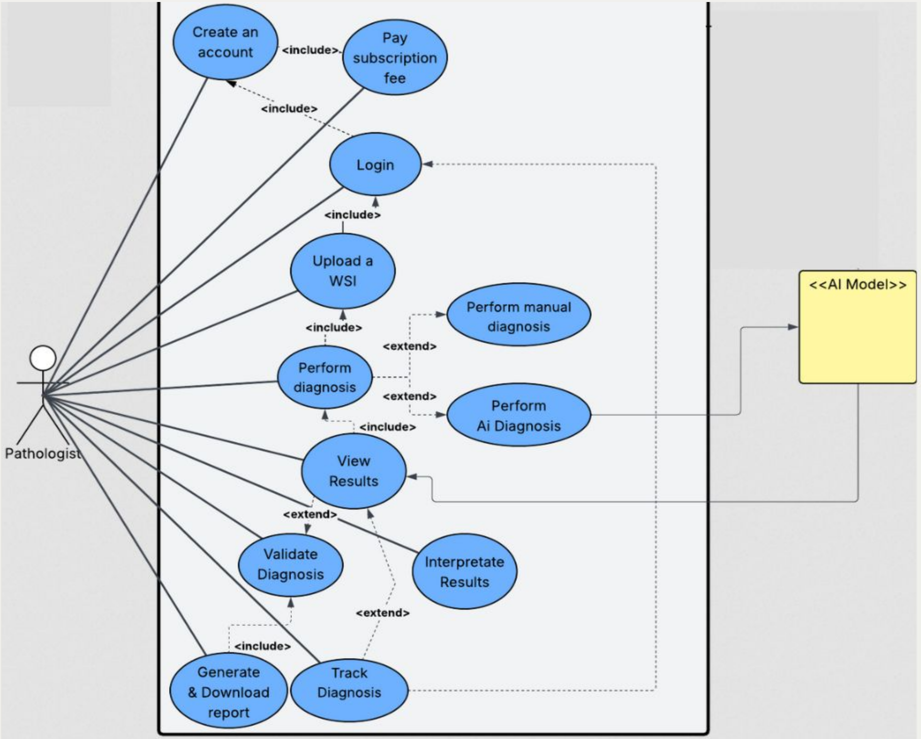
Platform Development

Platform Development | Design

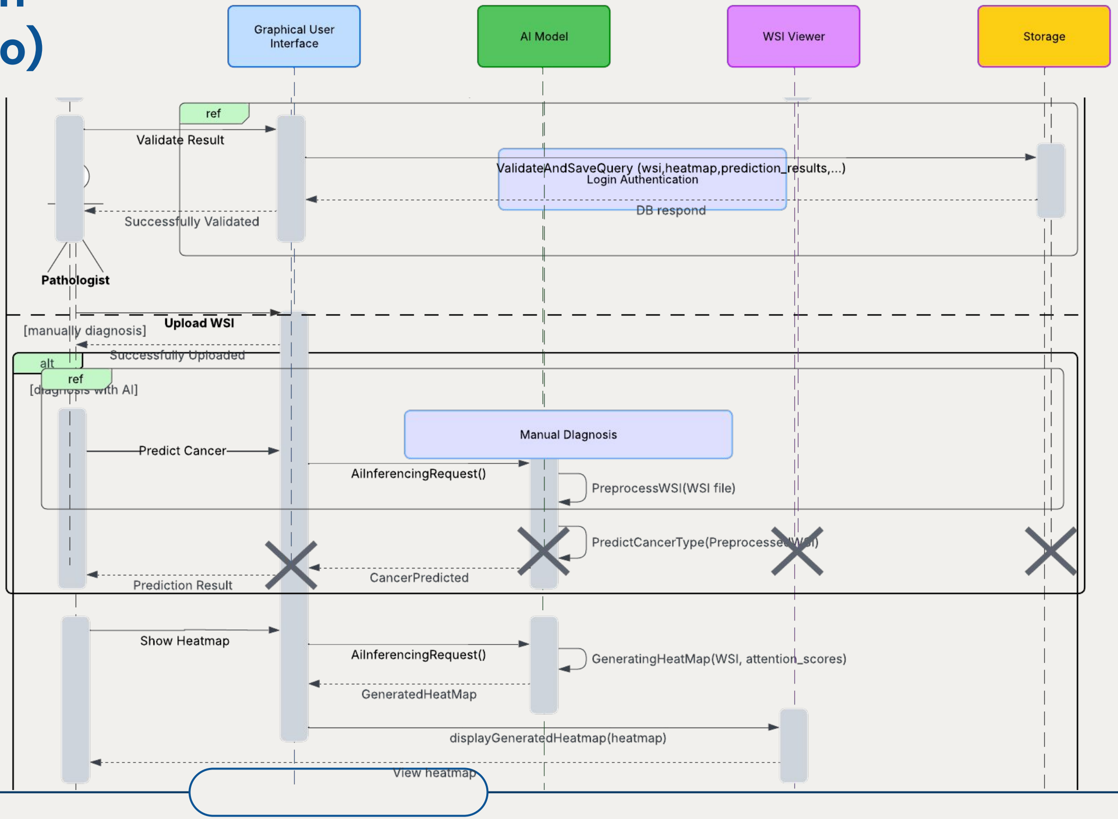


Platform Development | Design

1 – Use Case Diagram



2 – Sequence Diagram (AI Diagnosis Scenario)



Platform Development | Tools

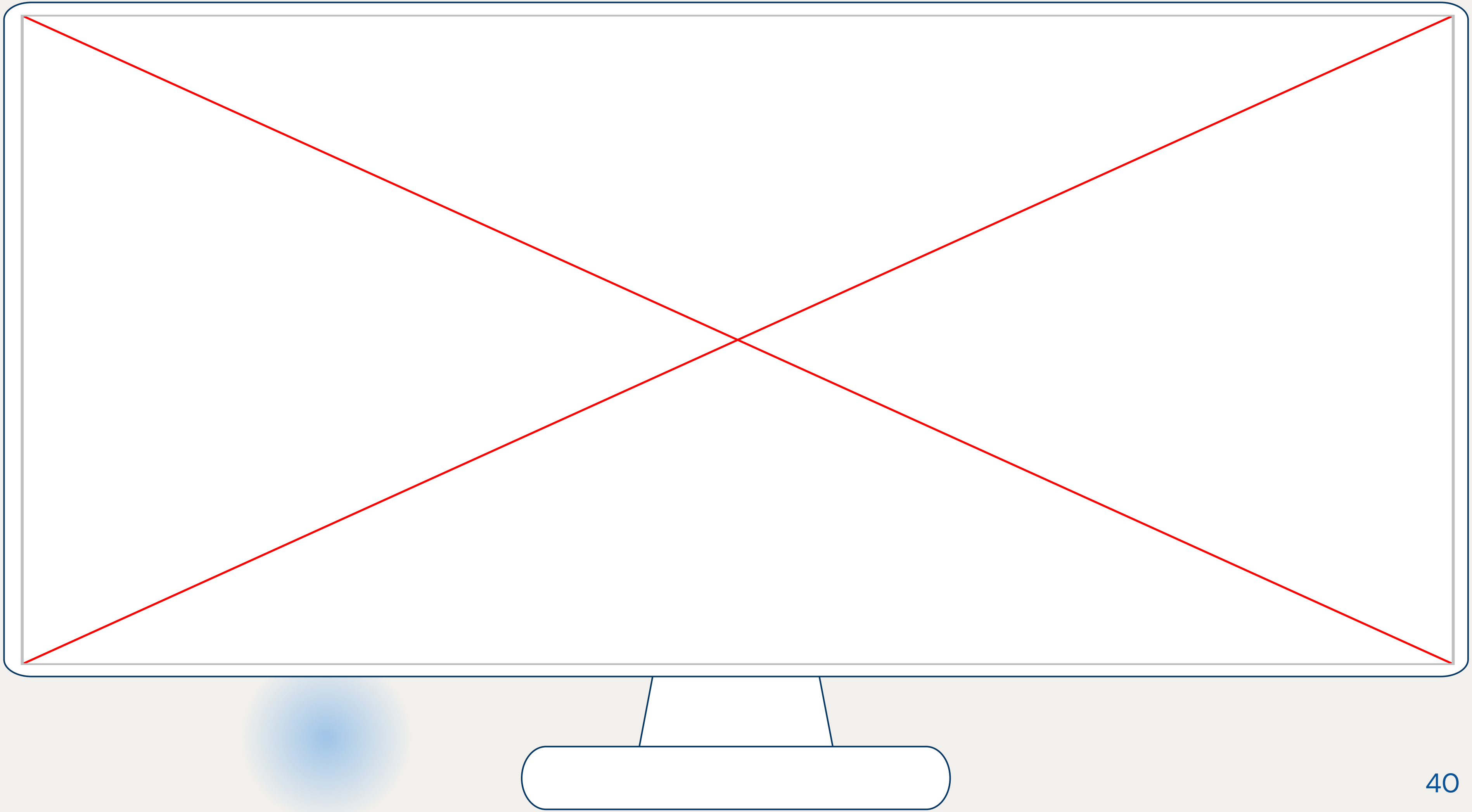
CONDA®

NumPy

scikit learn

OpenSlide

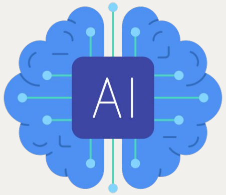
PC



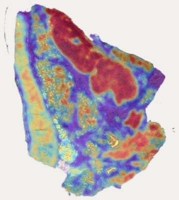
07

Conclusion

Conclusion



Leveraging deep learning techniques to enhance Histopathology diagnosis accuracy by training a model on WSIs for lung cancer subtype classification



Provides clinical explainability through attention-based heatmaps

Integrate the validated model into a web application

Future Perspectives

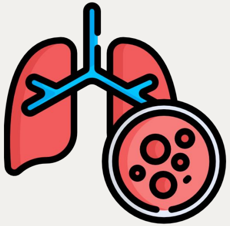
Expanding the dataset from diverse institutions and regions to improve model generalizability

Include multi modal approach by Combining histopathological image features with clinical data

Implementing comprehensive multi-user collaboration features

Thank you

فكرة المشروع



سرطان الرئة مرض منتشر و خطير عالميا ومحليا ، حيث أن نسبة الوفيات من المرض عالية

تجاريا ، يقدم المشروع على شكل خدمة طبية تسوق للمخابر و المستشفيات بشكل أساسي .

منصة على الويب توفر خدمة تشخيص السرطان عن طريق صور الشرائح الكاملة **(WSIs)**

إدماج خبرات الأطباء

تحليل SWOT

نقاط القوة
S

- استخدام الذكاء الاصطناعي.
- انشاء خرائط حرارية.
- منصة متكاملة سهلة الاستخدام
- عدم وجود منافسين مباشرين
- دعم القرار الطبي.

نقاط الضعف
W

- محدودية الموارد المالية و البشرية
- نقص الوعي المحلي.
- تحديات في الحصول على تراخيص رسمية
- الحاجة إلى بيانات طبية.

الفرص
O

- الشراكة مع المؤسسات الصحية
- قلة المنافسة في السوق المحلية
- ازدياد الاهتمام بالتحول الرقمي.
- دعم المشاريع المبتكرة والتكنولوجيا من قبل الحكومة
- إمكانية التوسع

التهديدات
T

- دخول شركات أجنبية
- التردد في اعتماد تقنيات الذكاء الاصطناعي
- القيود القانونية المرتبطة باستخدام الذكاء الاصطناعي

نموذج العمل التجاري BMC

شرائح العملاء	العلاقات مع الزبائن	القيمة المقدمة	الأنشطة الرئيسية	الشركاء
- أخصائي علم الأمراض (Pathologists) - مخابر التحاليل والتشخيص - المصحات - طالب الطب والباحثون	- الشبكات الإجتماعية لمشاركة التحديثات وتلقي المقترحات - الموقع الإلكتروني الخاص بالخدمة - دعم فني متخصص	- تحديد نوع سرطان الرئة - الكشف عن المناطق المتضررة بالسرطان - استخدام خوارزميات التعلم العميق - واجهة مستخدم سهلة الاستعمال - تقليل وقت التشخيص بشكل ملحوظ. - أمان بيانات المرضى	- البحث والتطوير - تصميم وتطوير نماذج الذكاء الإصطناعي - تقييم نتائج النموذج من خلال المخابر - تطوير تطبيق الويب	- جامعة بسكرة - المصحات والمستشفيات - مراكز البحث - جمعيات و منظمات الأمراض والأورام
	القنوات		**الموارد الرئيسية**	
	- المؤتمرات الطبية والمعارض - التسويق بالمحتوى - اشتراكات عبر الإنترنت - شراكات مع المخابر		- اشتراكات الخدمات السحابية - أدوات وبرامج تطوير البرمجيات - فريق التطوير	

مصادر الإيرادات	التكالـــــيف
- الإشتراكات - نموذج Freemium - عقود الخدمة و الصيانة - تحديث النماذج	- نفقات البحث والتطوير - التسويق - تكاليف الخدمات السحابية