

A Hybrid Deep Learning Approach For Breast Tumor Analysis and Phytochemical Recommendation

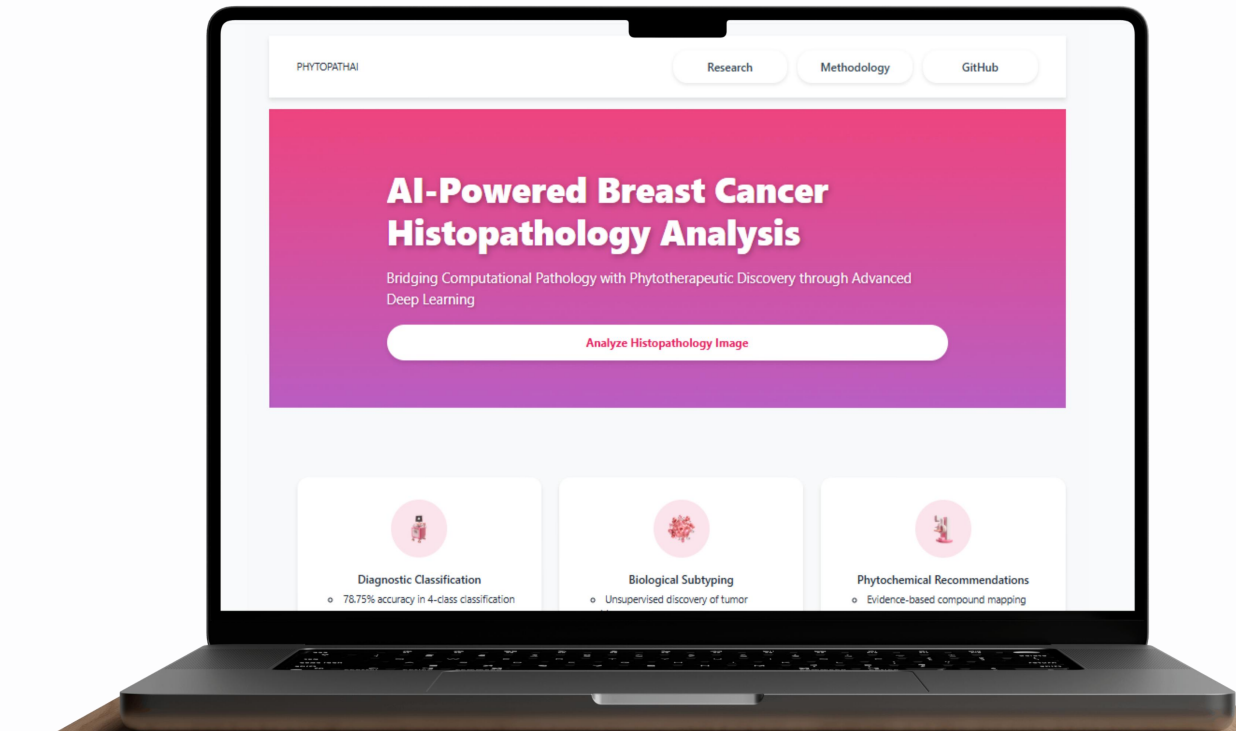
PhytopathAI

Name: DOMINION AKINROTIMI

Matric Number: CSC/20/031

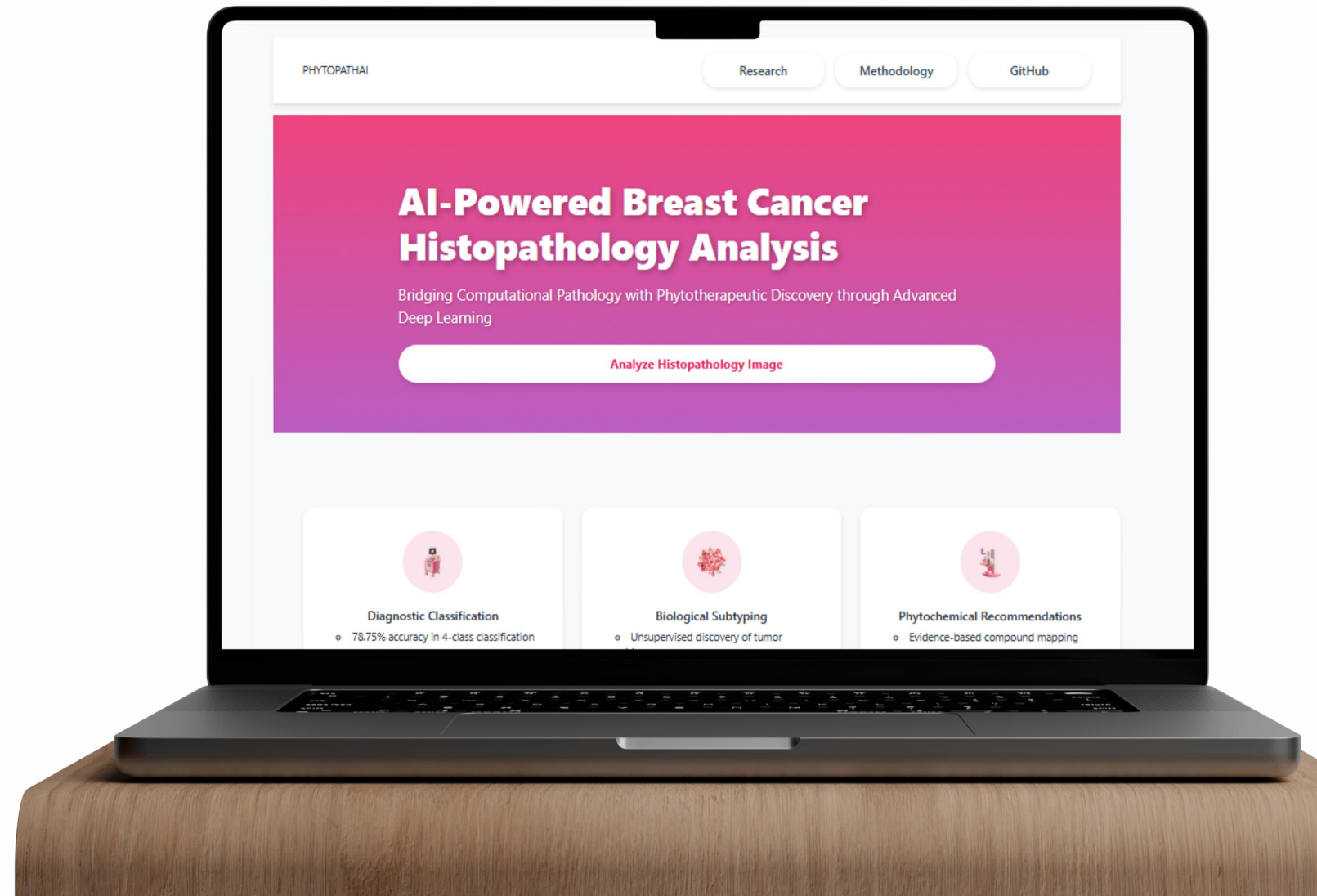
Supervisor's Name: PROF. D. T. AKOMOLAFE

Department: Computer Science



Presentation Outline

1. Introduction
2. Motivation
3. Statement of Problem
4. Aim and Objectives
5. Significance of the Study
6. Literature Review
7. Methodology
8. Implementation & Workflow
9. Contribution to Knowledge
10. Summary and Conclusion
11. References



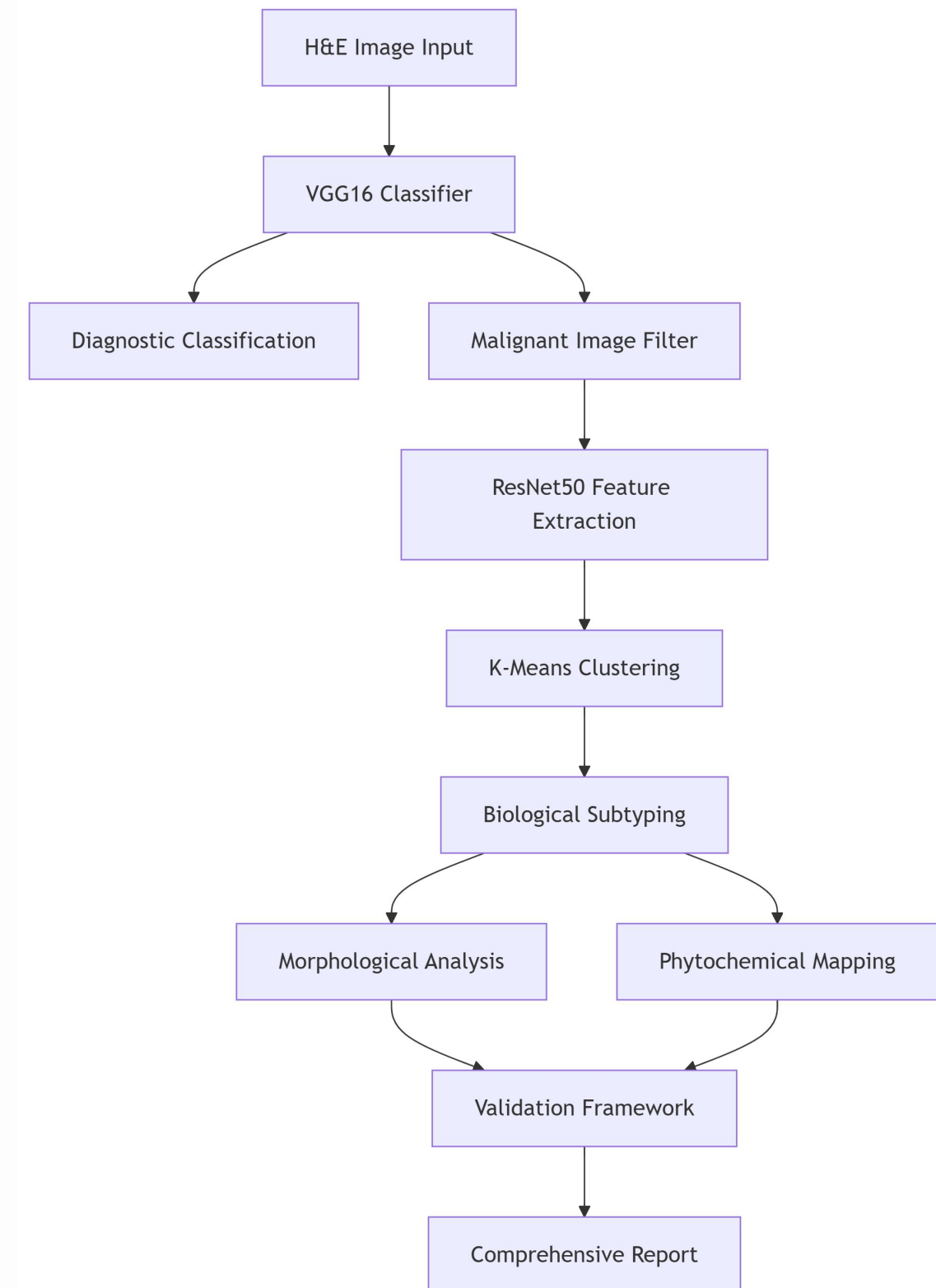
Introduction

Most AI systems in medicine are **single-task models** (e.g., just classification).

This project builds an **integrated pipeline** that does more:

1. **Diagnostic Classification** (What is it?)
2. **Unsupervised Morphological Subtyping** (What patterns exist within the diagnosis?)
3. **Actionable Output Generation** (What does that pattern imply?)

Core Technology: Deep Learning (VGG16, ResNet50), Unsupervised ML (K-Means), and Web Framework (Gradio).



Motivation

Beyond Automation

To push computational pathology
beyond **automation** and towards
discovery.

Pattern Discovery

To demonstrate that AI can be used to **uncover hidden patterns** in complex data, not just predict predefined labels.

Complete Workflow

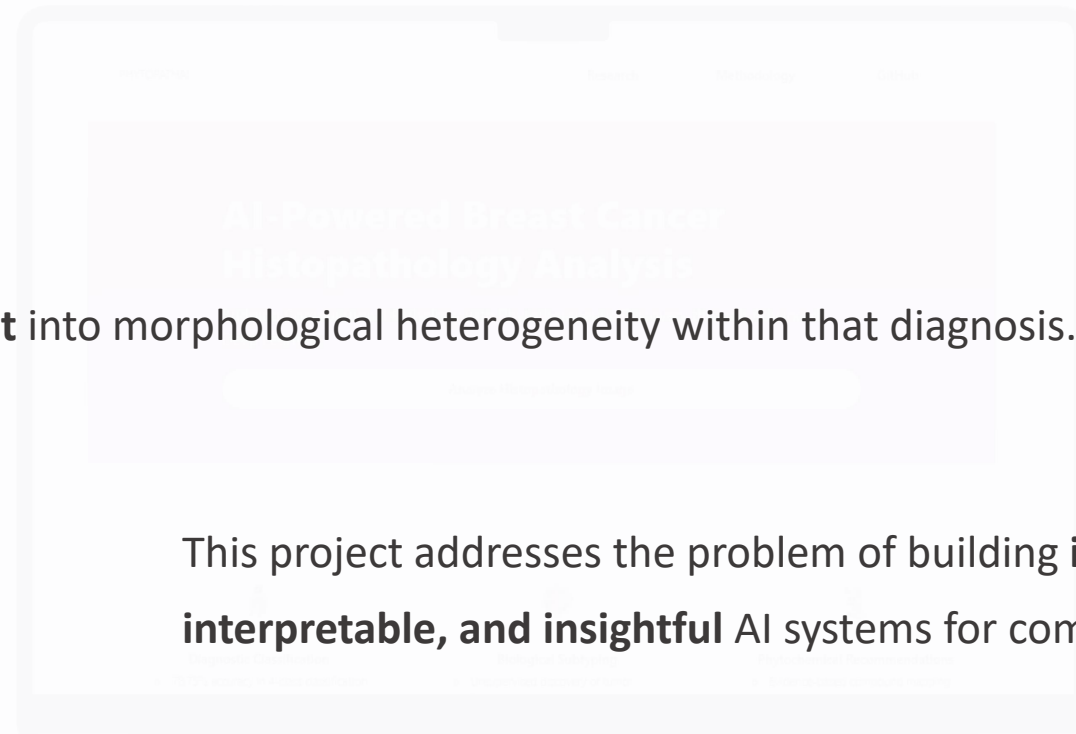
To build a **fully functional application** that showcases a complete AI workflow, from raw data to user-friendly insights.

Statement of Problem

Existing models provide a diagnosis but offer **no deeper insight** into morphological heterogeneity within that diagnosis.

A gap exists between **pattern recognition** (what the AI sees) and **actionable knowledge** (what a user can do with it).

This project addresses the problem of building **integrated, interpretable, and insightful** AI systems for complex data.



Aim and Objectives of the Study

Aim:

To develop an integrated deep learning system for comprehensive histopathology image analysis.

Objectives:



To preprocess and extract deep features from unannotated histology images using convolutional neural networks.



Apply supervised classification and unsupervised clustering to analyze tumor microenvironments.



To construct a rule-based engine mapping inferred biological features to relevant phytochemicals.



To develop an output format and user interface that presents tumor features alongside phytochemical and botanical recommendations.

Significance of the Study

Technical Significance

Demonstrates a novel **hybrid AI architecture** combining supervised, unsupervised, and knowledge-based modules.

Practical Significance

Provides a **blueprint for building more insightful AI tools** that can discover patterns and suggest actions, applicable beyond healthcare.

Academic Significance

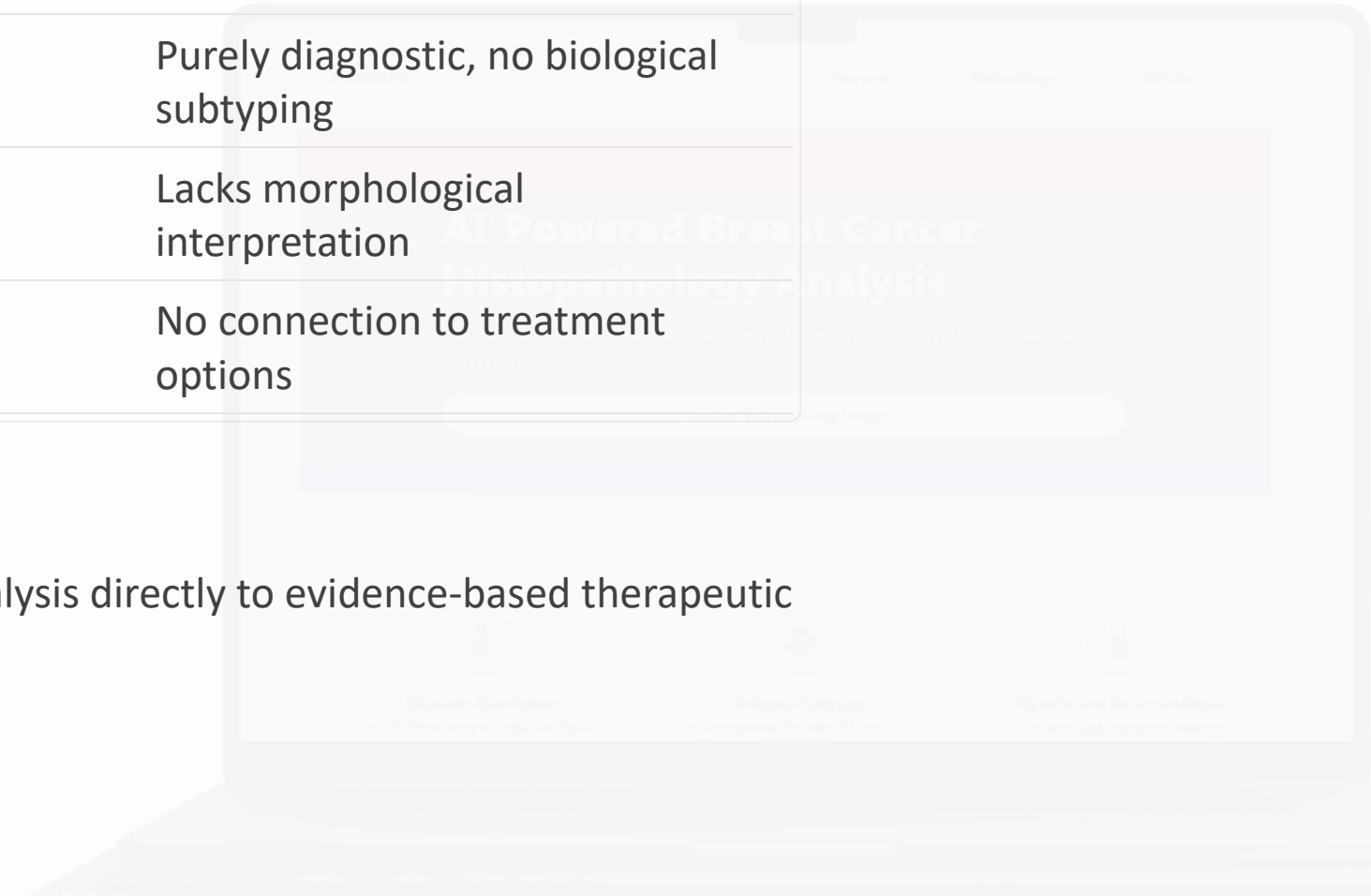
Serves as a robust case study in the **complete machine learning lifecycle**, from data preprocessing to model deployment.

Literature Review

Author (Year)	Focus	Key Limitation / Gap
Ciresan et al. (2013)	Mitosis detection	Task-specific, no therapeutic link
Spanhol et al. (2016)	CNN classification	Purely diagnostic, no biological subtyping
Araújo et al. (2017)	Histopathology analysis	Lacks morphological interpretation
Couture et al. (2018)	Tissue classification	No connection to treatment options

Core Novelty:

The first framework to connect AI-based histopathological analysis directly to evidence-based therapeutic discovery.



Methodology

A Three-Stage Technical Methodology:

Supervised Learning Phase

- Model: Fine-tuned VGG16.
- Task: 4-Class Image Classification.

Result: **78.75% Validation Accuracy.**

Unsupervised Learning

- Model: ResNet50 (for features) + K-Means (for clustering).
- Task: Discover morphological subtypes.

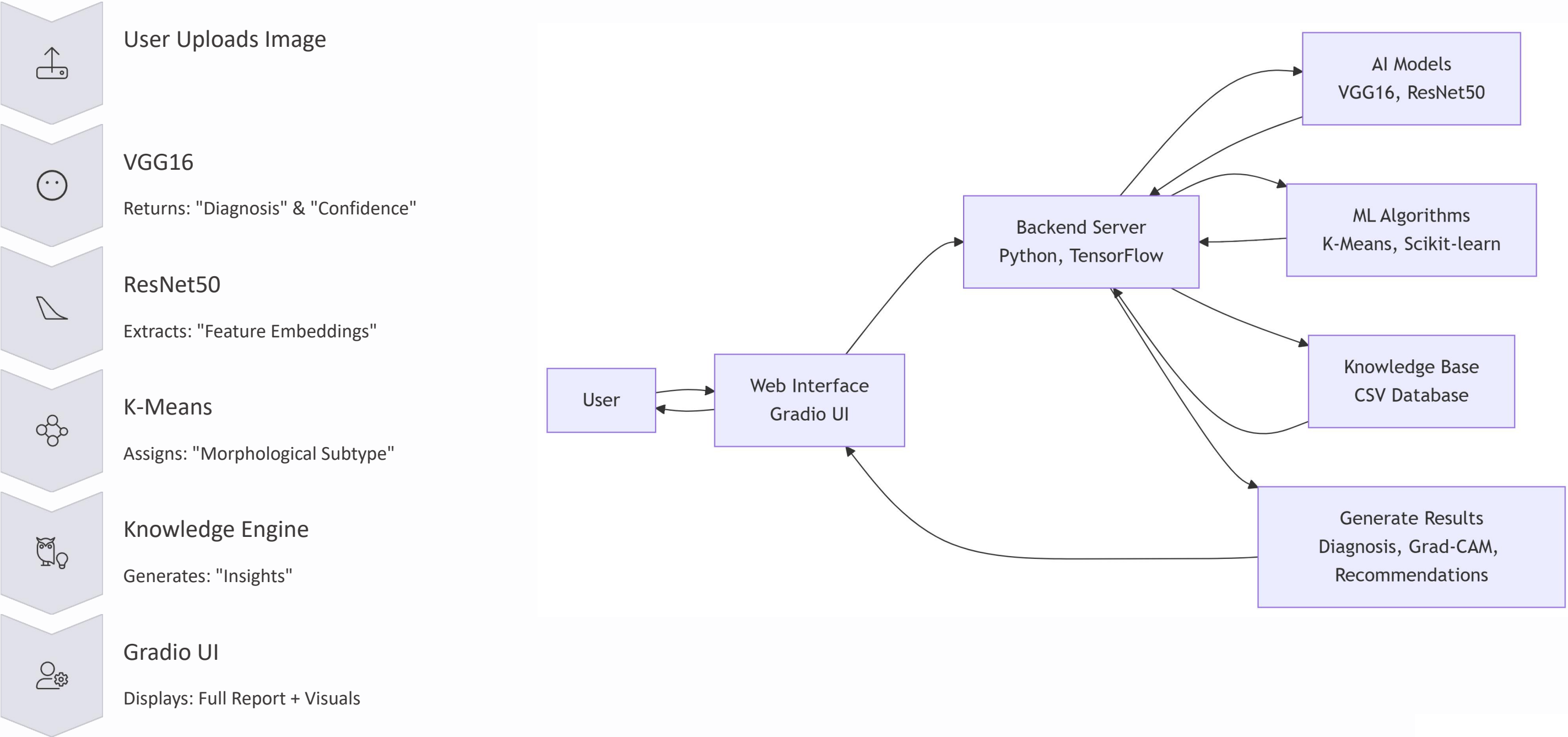
Result: **2 significant clusters found.**

Integration & Deployment

- Framework: Gradio.
- Task: Build a web application.

Result: A single, user-friendly interface for the entire pipeline.

Implementation and Workflow



Results and Findings

	precision	recall	f1-score	support
Normal	0.9444	0.85	0.8947	20
Benign	0.6818	0.75	0.7143	20
InSitu	0.7391	0.85	0.7907	20
Invasive	0.8235	0.7	0.7568	20
accuracy	0.7875	0.7875	0.7875	0.7875
macro avg	0.7972	0.7875	0.7891	80
weighted avg	0.7972	0.7875	0.7891	80

- 1. 78.75% Accuracy in 4-class cancer classification
- 2. 2 biological subtypes identified with unique profiles
- 3. AI-powered histopathology with visual explanations

Innovations

Dual-model architecture (VGG16 + ResNet50)

Unsupervised subtyping with clinical relevance

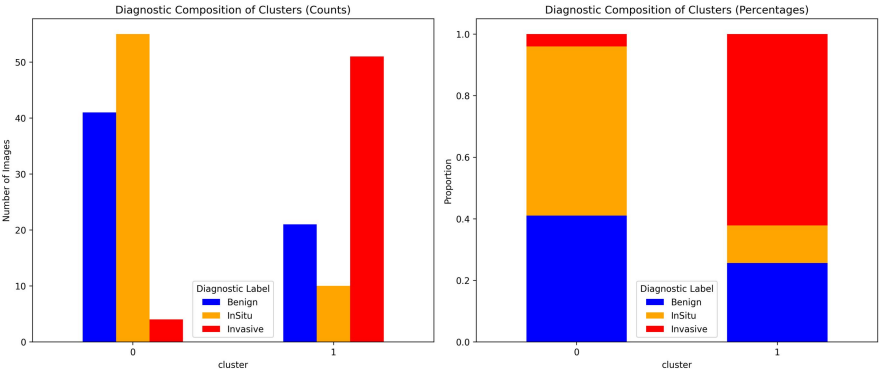
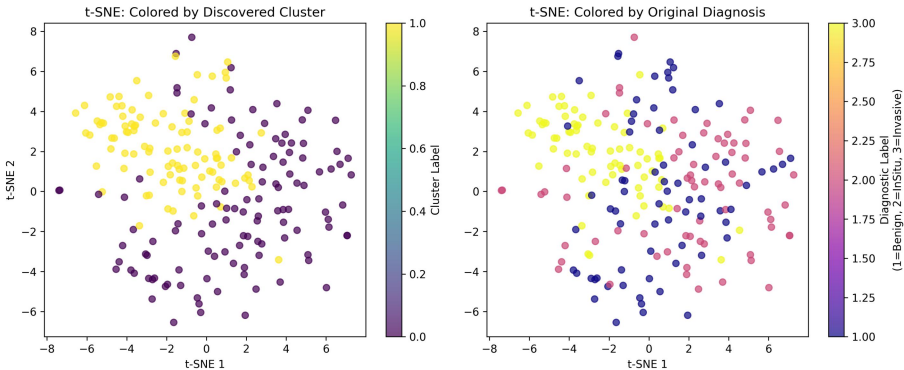
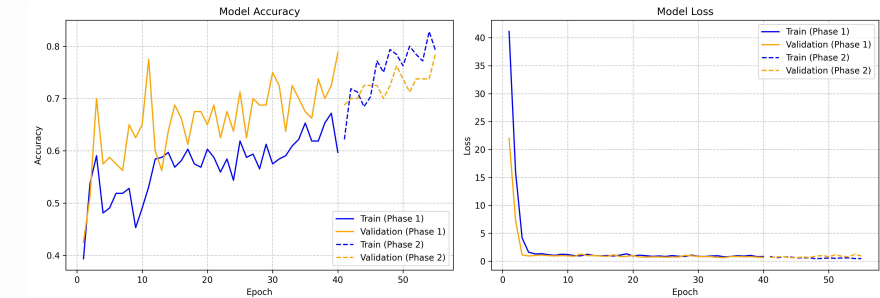
20+ phytochemical recommendations mapped

Impact

Automated diagnostic support for pathologists

Subtype-specific treatment insights

Natural compound discovery platform



Morphological Feature	Cluster 0 (Median ± SD)	Cluster 1 (Median ± SD)	p-value
Nuclear Size (px)	24.25 ± 20.92	23.50 ± 12.92	0.5792 (ns)
Nuclear Pleomorphism	26.23 ± 16.95	22.07 ± 16.98	0.0370 (*)
Glandular Score	0.61 ± 0.19	0.79 ± 0.20	< 0.0001 (**)

The test **Cluster 1** (Invasive-enriched) exhibited significantly higher **nuclear pleomorphism** (p=0.0370) and a profoundly higher **glandular disorganization score** (p<0.0001), known hallmarks of malignancy (Elston et al., 1991). Nuclear size was not significantly different (p=0.5792)

Contribution to Knowledge



A Novel Architecture

Demonstrates the design and efficacy of a hybrid AI pipeline for layered data analysis.



Validation of Unsupervised Discovery

Provides evidence that clustering deep feature embeddings can reveal biologically meaningful and statistically significant patterns.



A Complete Deployment Blueprint

Offers a practical guide for integrating complex ML models into a single, deployable web application.

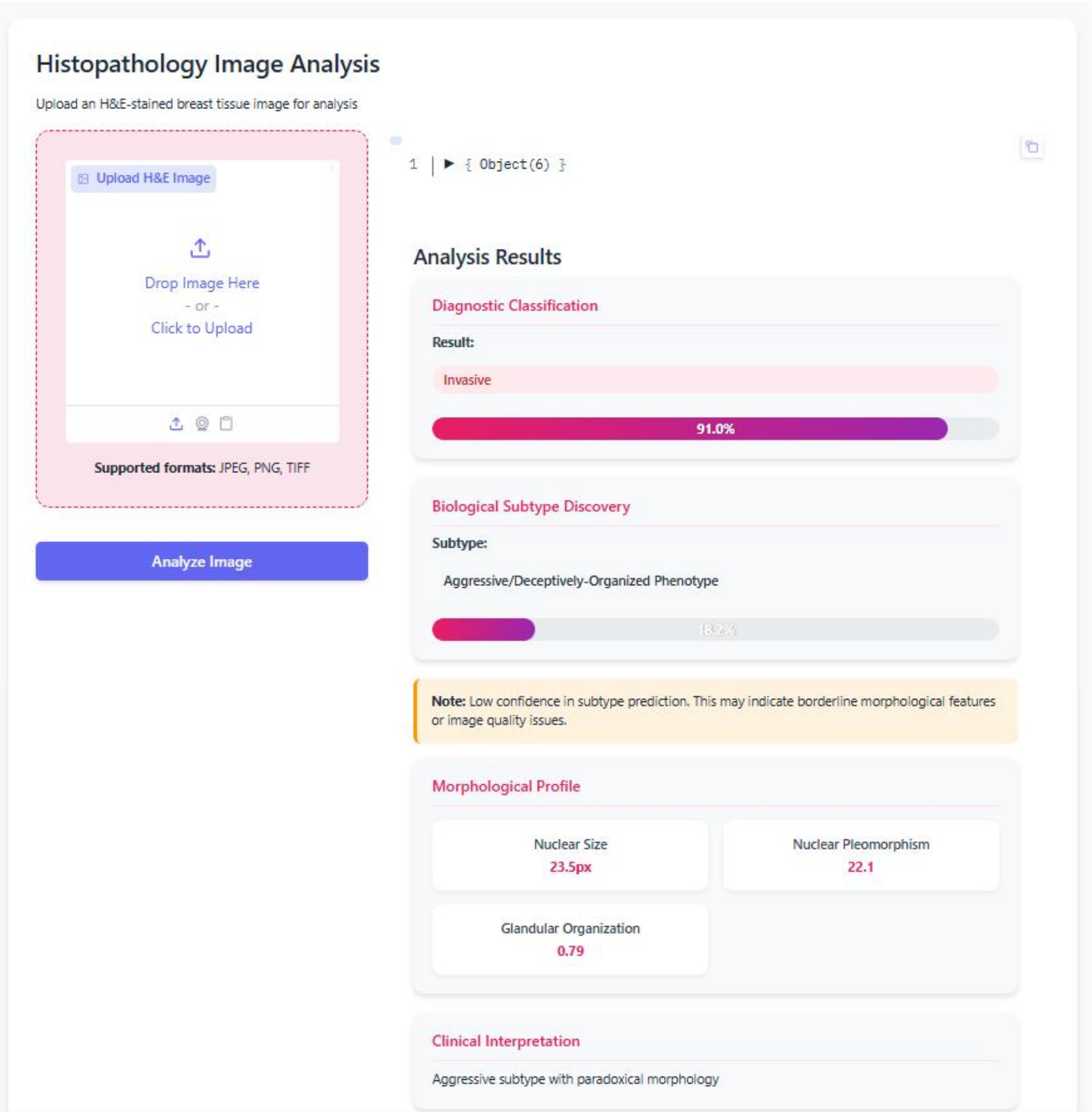
Summary and Conclusion

Successfully developed an integrated system that performs diagnostic classification and unsupervised discovery.

Key Achievement:

The system doesn't just classify; it finds hidden patterns and provides tailored insights.

The model achieved **78.75% accuracy** and discovered morphologically distinct clusters validated by statistical analysis ($p < 0.05$).



References

- Ciresan et al. (2013). Mitosis detection in breast cancer histology using deep neural networks. MICCAI, pp. 411–418.
- Spanhol et al. (2016). Breast cancer histopathological image dataset. IEEE Trans. Biomed. Eng., 63(7), 1455–1462.
- Araújo et al. (2017). CNN-based classification of breast cancer histology images. PLoS One, 12(6), e0177544.
- Couture et al. (2018). Deep learning for breast cancer grade and subtype prediction. NPJ Breast Cancer, 4(1), 1–8.
- Elston, C. W., & Ellis, I. O. (1991). Pathological prognostic factors in breast cancer, 19(5), 403–410.

Thank you for listening

