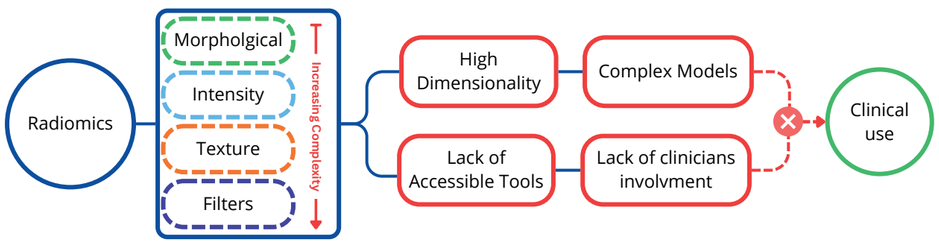
SIMPLER RADIOMICS CAN BE ENOUGH: IDENTIFYING THE MINIMAL FEATURE COMPLEXITY REQUIRED FOR ACCURATE PREDICTION

www.mediml.app

By demonstrating that simple and readily interpretable radiomic features can provide strong predictive performance across multiple clinical problems, MEDiml provides a practical, accessible framework for reducing the clinical translation gap in radiomics.

INTRODUCTION

Radiomics, defined as the extraction of quantitative data from medical images, has gained a lot of interest since its introduction in 2012. However, very few, if any, radiomics studies are translated to clinics. The high dimensionality of feature sets and the lack of accessible tools for clinicians are amongst the reasons for this.



We are introducing **MEDiml**, a new radiomic tool that focuses on simpler models and removes technical barriers for radiomics research.

RESEARCH OBJECTIVES

- Develop a methodology that, for a given clinical problem, determines the smallest and simplest feature set required for accurate predictive modeling.
- Validate our methodology across distinct imaging cohorts of cancer patients with distinct clinical problems.
- Remove the technical barriers to radiomics research by offering users a new collaborative radiomics tool.

RESEARCH BENEFITS

1. Show that simpler, more interpretable models can often match or outperform complex ones.
2. Make radiomics research accessible to clinicians and provide a practical path toward the integration of radiomic biomarkers into clinical workflows.
3. Help multidisciplinary teams work better together through radiomics pipeline sharing.

METHODOLOGY

Data Processing

MEDiml incorporates all processing steps required for radiomic feature extraction in accordance with international standards defined by the IBSI.[1]

Feature selection

Based on the False Discovery Avoidance[2] approach, where only features with the highest correlation to the outcome and the least inter-correlation are kept for model training.

Model Development

Models are trained using the XGBoost algorithm on 10 cross-validation splits.

Testing & Interpreting

Feature importance, computed during training, is used to interpret the model's decisions.

DATASETS

Cohorts	Non-Small Cell Lung Cancer (NSCLC)	Low Grade Glioma (LGG)	Meningioma	Renal Cell Carcinoma (RCC)	RCC
Imaging Modality	CT	MRI-T2F	MRI-T1CE	MRI-T2WI	CECT
Clinical Endpoint	Histological subtype: • Adenocarcinoma (n=240) • Other (n=266)	IDH1 mutation: • Yes (n=239) • No (n=90)	Pathological tumor grade: • Grade 1 (n=197) • Grade 2 & 3 (n=147)	Subtype discrimination: • Papillary (n=158) • Clear Cell (n=441)	Subtype discrimination • Non-Clear Cell (n=79) • Clear Cell (n=247)

RESULTS AND DISCUSSION

AUC-based Model Performance

LGG 0.68

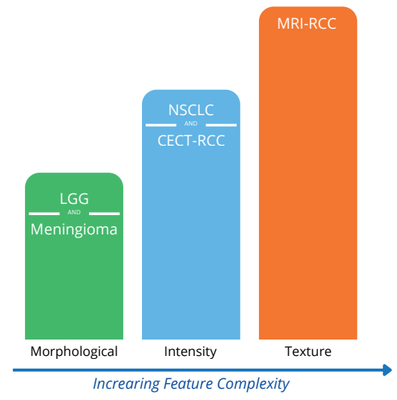
Meningioma 0.65

NSCLC 0.77

CECT-RCC 0.82

MRI-RCC 0.72

Complexity-based Performance



Results Interpretation

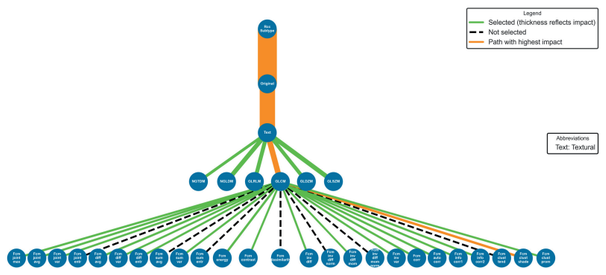


Figure 1 - Importance tree for the Papillary versus Clear Cell subtype classification in RCC MRI images. The orange line draws the path leading to the most predictive feature in the set. Green lines represent other features that participated in the model's decision-making.

In-Depth Analysis - *CECT-RCC*

For clear versus non-clear RCC classification, the intensity-based level was selected as optimal, which allowed us to refine feature-extraction settings, particularly the re-segmentation range, which directly affects the intensities within the ROI. **This refinement led to a 0.04 improvement in AUC from 0.82 to 0.86**

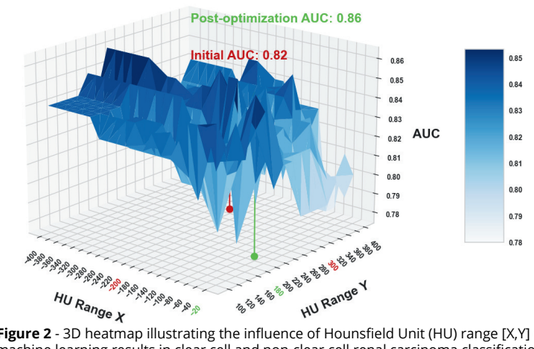


Figure 2 - 3D heatmap illustrating the influence of Hounsfield Unit (HU) range [X,Y] on machine learning results in clear cell and non-clear cell renal carcinoma classification.

SYSTEM SPECIFICATIONS

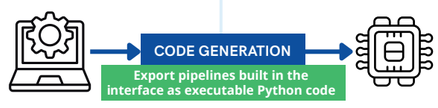
User Interface

- Installable Desktop application
- No coding required
- Fully customizable radiomics pipelines
- Convert graphical pipelines into Python code.

Python Package

- Installable via a single command
- Modular architecture
- Over 500 PyPI downloads.

```
>_ pip install mediml
```



CONCLUSIONS

Our proposed methodology and software can estimate the optimal radiomics complexity level for specific medical outcomes, potentially simplifying the use of radiomics in predictive modeling across various contexts.

MEDiml lowers the barrier to entry for clinicians and provides a scalable pathway for the clinical integration of radiomic biomarkers.

REFEFENCES

1. Zwanenburg A, Vallières M, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. Radiology. 2020;295(2):328-338. doi:10.1148/radiol.2020191145
2. Chatterjee A, Vallières M, et al. An Empirical Approach for Avoiding False Discoveries When Applying High-Dimensional Radiomics to Small Datasets. IEEE Trans Radiat Plasma Med Sci. 2019;3(2):201-209. doi:10.1109/TRPMS.2018.2880617

FUTURE WORK

- Investigate the impact of radiomic feature extraction parameters on predictive modeling.
- Introduce deep radiomics features to this comparative analysis.
- Address medical data scarcity and privacy concerns by implementing a federated learning approach.

AUTHORS:

Mahdi Ait Lhaj Loutfi[1], Martin Vallières[2]

[1] Medical Physic Unit, Department of Biomedical Engineering, McGill University, Montreal, Canada

[2] Medical Physic Unit, Department of Oncology, McGill University, Montreal, Canada

Contact: **mahdi.aitlhajloutfi@mail.mcgill.ca**