Beyond Single dimensions with XAI: Investigating the Challenges Hindering eXplainable Artificial Intelligence Implementation in the Multi-modal Medical Domain

Presenter: Amandeep Kaur, MMath

Project presentation for CS886: Trust Modeling, Explainability and Online Social Networks

TABLE OF CONTENTS

01
Introduction

Describes the problem statement.

02
Research questions

Discusses the 3 RQs proposed

03
Methodology

Discusses the methods for answering the 3 RQs

04
Progress

Describing the work done so far

05
SWOT analysis

Discusses the strength, weaknesses, future work

06
Conclusion

Finally the summary of project

Introduction

- With advancements in computation and medical data availability, the field of **AI has huge potential for healthcare** industry.

- Why hasn't the healthcare industry embraced active deployment of AI?

1. **Trust and Reliability concern**
2. AI's Black box effect
3. **Lack of transparency**
4. **Multimodal data Integration Challenges**
5. Resistance to Change
6. Lack of accountability

Research Questions (RQs)

Probing Challenges in Medical XAI

What are the challenges impeding the effective implementation of eXplainable Artificial Intelligence (XAI) within the medical field?

Enhancing Insights: Integrating Diverse Medical Data

Evaluation of the effectiveness of integrating multiple datasets in comparison to using standalone medical data sources

Integration of XAI: Initial vs post-hoc

Contrasting the integration of eXplainable AI (XAI) during the initial stages of algorithm design vs the post-hoc implementation. Proposing design of WatNeT

Methodology: RQ1

Probing Challenges in medical XAI

1. **Medical datasets** have **different modalities**, and have to be interpreted cautiously.
2. The traditional **black box models** are **not transparent**.
3. **XAI algorithms** were trained on real-world datasets like **ImageNet**.
4. **Answering RQ1:** Conducted case studies on 9 XAI algorithms on medical image and text data.
5. In the slides: Highlighting the fallbacks of 4 algorithms - LASSO vs Shapley(Text data), and LIME vs GradCAM(Image data)

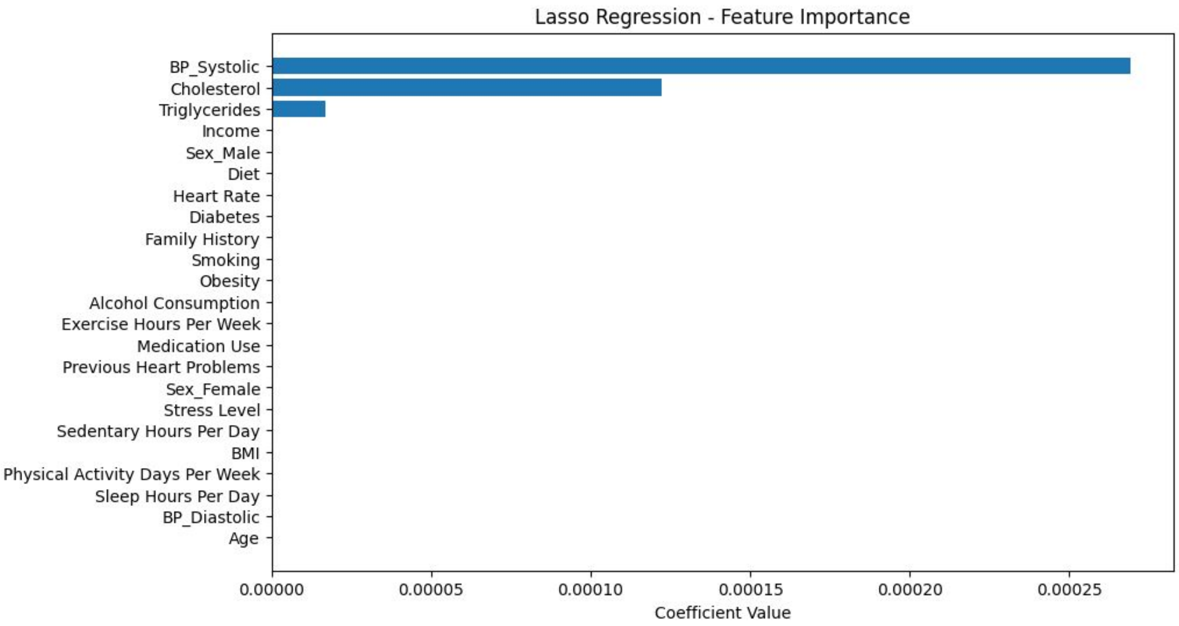
XAI Algorithms categorization

- XAI via **Dimension reduction** (LASSO, Cluster analysis)
- XAI via **feature importance** (Shap, DeepLift)
- XAI via **attention mechanism** (GradCam, super pixel maps)
- XAI via **surrogate representation** (LIME, LRP)
- XAI via **knowledge distillation and rule extraction** (Bayesian)

1. LASSO

LASSO assumes linear relationships with the target variable, chooses 1 of the highly correlated features.

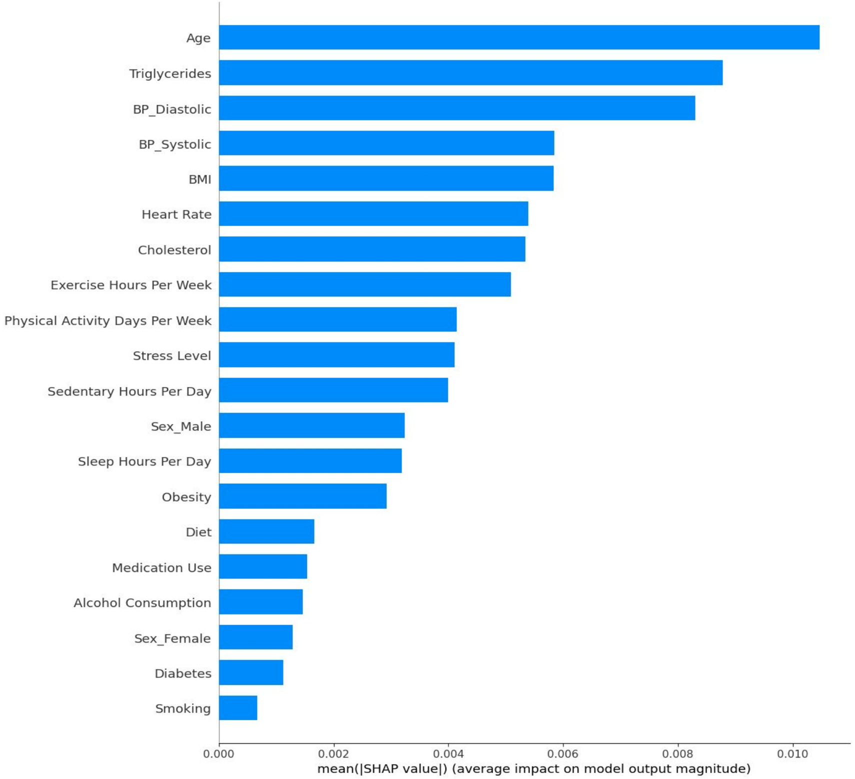
Fallback: The possibility of losing crucial features, which may still be relevant for clinical predictions on a case-by-case basis, can be common and these important features may be neglected unintentionally by the dimensional reduced models.



2. SHAPley

Shapley values are calculated in the context of cooperative game theory.

Fallback: Values are model-specific, might not be transferable. Interpreting the working of the algorithm might be a challenge for non-experts.

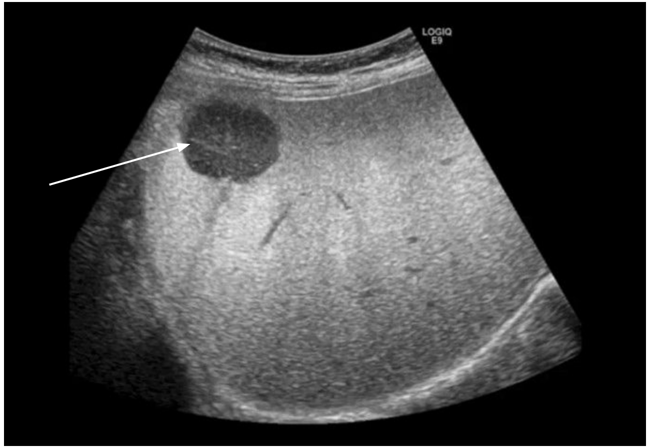


3. LIME

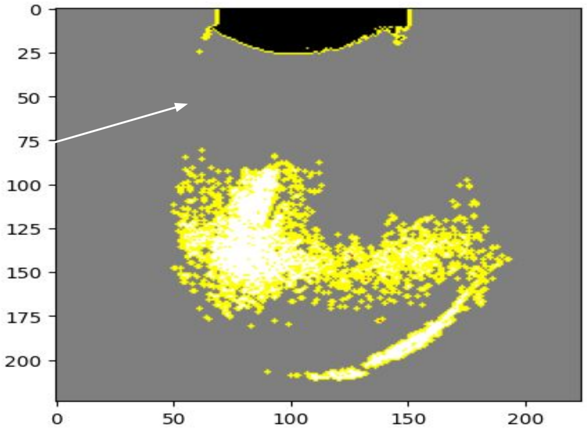
LIME approximates the behaviour of black-box model with local interpretable model.

Fallback: Limited to input features, might lack clinical or medical context.

```
Top predictions: [('n02895154', 'breastplate',
0.7047977), ('n03146219', 'cuirass', 0.08550827),
('n03950228', 'pitcher', 0.03341381),
('n03041632', 'cleaver', 0.023403496),
('n03498962', 'hatchet', 0.020561848)]
```



Original image with tumor



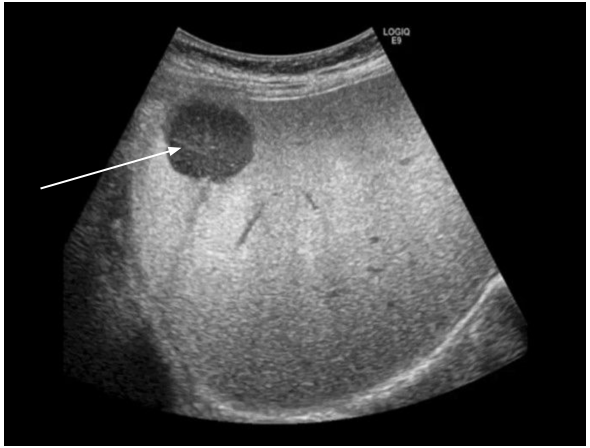
LIME explanations

https://link.springer.com/article/10.1007/s10396-020-01012-y (Metastatic liver cancer (25mm)

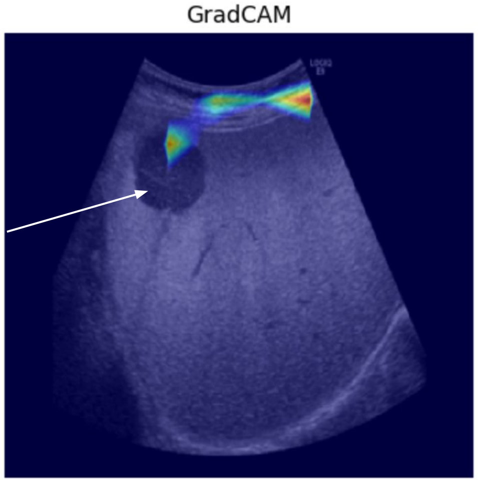
4. GradCAM

GradCAM highlights the important regions in image that led to the model’s final prediction.

Fallbacks: These algorithms do not advice on what makes the region important to a clinical user? Can be frustrating for the end user especially when they do not know the rationale behind the region highlighted.



```
Model prediction:
        backpack         (414)    with probability 0.970
        ping-pong_ball   (722)    with probability 0.027
        lipstick         (629)    with probability 0.002
        lens_cap         (622)    with probability 0.000
        groenendael      (224)    with probability 0.000
Explanation for 'black_swan'
```



https://link.springer.com/article/10.1007/s10396-020-01012-y (Metastatic liver cancer (25mm)

Methodology: RQ2

Enhancing Insights: Integrating Diverse Medical Data

1. The expectation in machine learning is that data fusion efforts will result in an improvement in predictive power.
2. Multimodality fusion models generally led to increased accuracy (**1.2–27.7%**) and AUROC (**0.02–0.16**) over traditional single modality models for the same task. [Huang et al.]
3. **Answering RQ2:** Discussion on 10 multimodal medical studies
4. In the slides: Highlight **2** such studies.

Huang et al. https://www.nature.com/articles/s41746-020-00341-z

Case study 1

Citation: Kharazmi, P., Kalia, S., Lui, H., Wang, Z. J. & Lee, T. K. A feature fusion system for basal cell carcinoma detection through data-driven feature learning and patient profile. *Skin Res. Technol.* 24, 256–264 (2018)

Clinical domain	Outcome	Fusion details	Input: Medical images	Input: Text records	Number of samples	Model performance
Dermatology	Basal cell carcinoma detection	CNN extracted features	Dermoscopic images	Patient data (age, sex, elevation, lesion location, lesion size)	1191	Fusion: 91.1% Accuracy Dermoscopic Images: 84.7% Accuracy Patient Profile: 75.6% Accuracy **(6.4% and 15.5% increase)**

Case study 2

Citation: Nie, D. et al. Multi-channel 3D deep feature learning for survival time prediction of brain tumor patients using multi-modal neuroimages. *Sci. Rep.* 9, 1103 (2019).

Clinical domain	Outcome	Fusion details	Input: Medical images	Input: Text records	Number of samples	Model performance
Radiology/Oncology	Prediction of survival time for brain tumor patients	CNN extracted features	MRI	Patient data (age, tumor size, histological type)	93	Fusion: 90.66% Accuracy MRI: 81.04% Accuracy Demographics and tumor features: 62.96% Accuracy **(9.62% and 27.7% increase)**

Methodology: RQ3

Integration of XAI in multi-modal dataset: Initial vs post-hoc

1. RQ1 proposes the need to improve and deploy XAI algorithms, while RQ2 discusses the efficiency increase by using multimodal dataset.
2. The need is to combine RQ1 and RQ2 and propose an architecture that is both explainable and can handle multimodal dataset as well.
3. Answering RQ3- **WatNetFuse**: A multimodal explainable AI architecture for medical domain
4. In the slides - The architectural design for WatNetFuse.
   (Implementation would be a future task, given the time constraints)

Multimodal data fusion methods

Early fusion

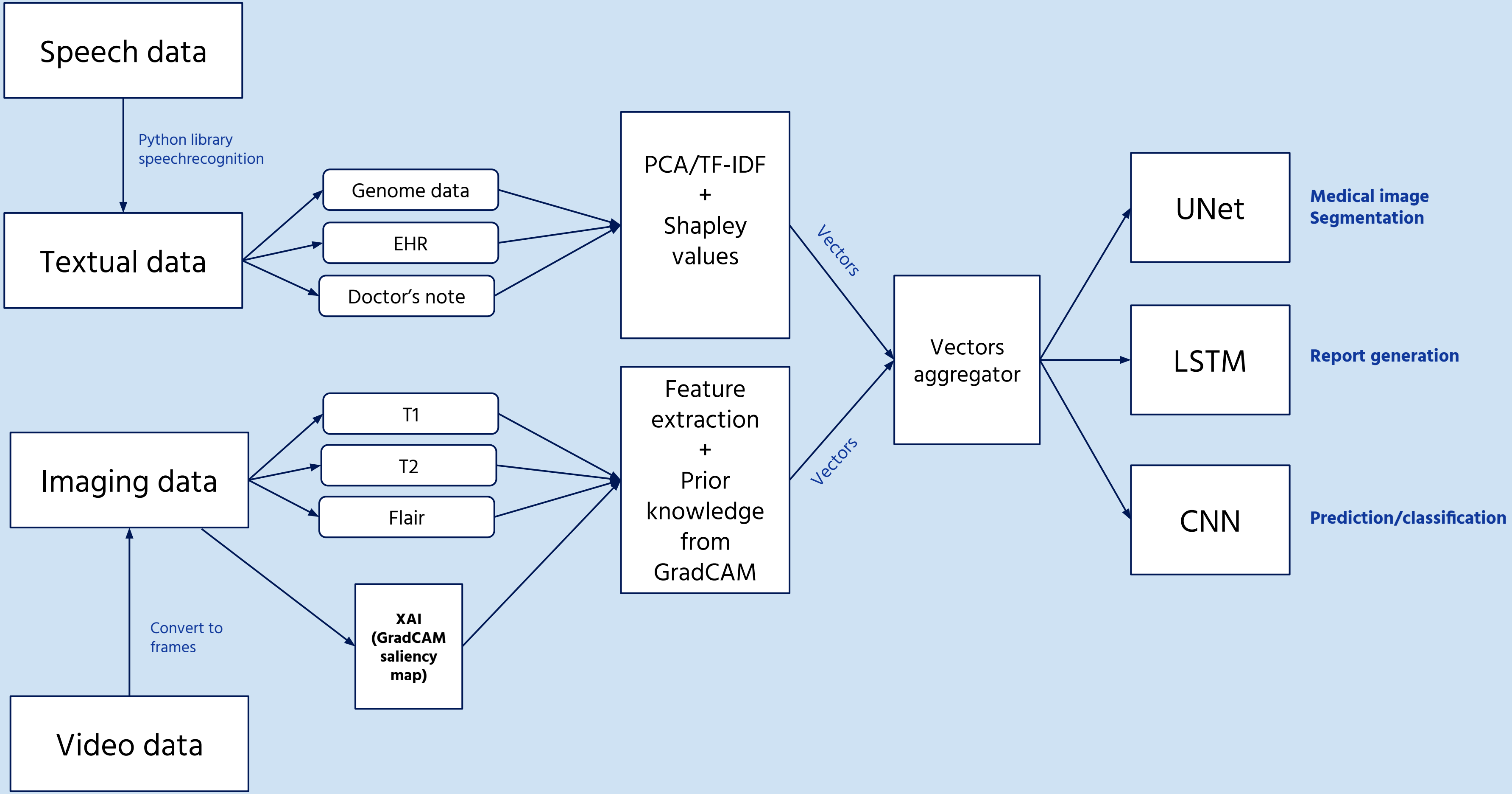
- Feature level fusion
- Joining multiple input modalities into single feature vector

Joint fusion

- Joining learned feature representation from intermediate layers of neural networks with features from other modalities as input to final model (loss is propagated)

Late fusion

- Leveraging predictions from multiple models to make a final prediction (averaging/majority voting)



WatNetFuse architecture

Progress so far

1. Case studies on RQ1 and RQ2 completed
2. Conceptualised the architecture for WatNetFuse.
3. Report writing is in progress

SWOT Analysis

Strengths	**W**eaknesses	**O**pportunities	**T**hreats
Comprehensive study done on existing (most popular) XAI algorithms in the context of medical domain	Peer-review and domain expert analysis is missing	Huge potential for innovating XAI algorithms specific to medical domain	Time constraints for implementation
Fusion of RQ1 and RQ2 conceptualised as RQ3	Case studies done cannot be generalised	Implementation of WatNetFuse with real-world medical data	Implementation in real-world as a deployable architecture is questionable
Proposed a new architecture for fusion of multimodal dataset with XAI		A deployable UI application for the model	

Conclusion

1. XAI for healthcare is the need of the hour.
2. Present XAI algorithms lack transferability to medical domain.
3. The introduction of multimodal dataset for medicine can be utilised with AI.
4. An approach discussed for fusing XAI and multimodal medical data.
5. Future prospects : To deploy the architecture

THANKS!

Do you have any questions?
a3kaur@uwaterloo.ca

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution